



Cisco IP Telephony Solution Reference Network Design (SRND)

Cisco CallManager Release 4.1
April 2005

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

Text Part Number: OL-7447-01



THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

CCSP, CCVP, the Cisco Square Bridge logo, Follow Me Browsing, and StackWise are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, and iQuick Study are service marks of Cisco Systems, Inc.; and Access Registrar, Aironet, ASIST, BPX, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Empowering the Internet Generation, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Fast Step, FormShare, GigaDrive, GigaStack, HomeLink, Internet Quotient, IOS, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, LightStream, Linksys, MeetingPlace, MGX, the Networkers logo, Networking Academy, Network Registrar, *Packet*, PIX, Post-Routing, Pre-Routing, ProConnect, RateMUX, ScriptShare, SlideCast, SMARTnet, StrataView Plus, TeleRouter, The Fastest Way to Increase Your Internet Quotient, and TransPath are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0502R)

Cisco IP Telephony SRND

Copyright © 2005 Cisco Systems, Inc. All rights reserved.



Preface **xv**

New or Changed Information for This Release	xv
Revision History	xvi
Obtaining Documentation	xvi
Cisco.com	xvi
Ordering Documentation	xvi
Documentation Feedback	xvii
Obtaining Technical Assistance	xvii
Cisco Technical Support Website	xvii
Submitting a Service Request	xviii
Definitions of Service Request Severity	xviii
Obtaining Additional Publications and Information	xviii

CHAPTER 1

Introduction **1-1**

Overview of Cisco IP Communications Solutions	1-1
Overview of the Cisco IP Telephony Solution	1-2
Cisco IP Network Infrastructure	1-4
Quality of Service	1-4
Call Processing Agent	1-4
Communication Endpoints	1-5
Applications	1-6
Security	1-7
Network Management Tools	1-8

CHAPTER 2

IP Telephony Deployment Models **2-1**

Single Site	2-2
Benefits of the Single-Site Model	2-3
Best Practices for the Single-Site Model	2-4
Multi-Site WAN with Centralized Call Processing	2-4
Best Practices for the Centralized Call Processing Model	2-6
Remote Site Survivability	2-6
Voice Over the PSTN as a Variant of Centralized Call Processing	2-9
VoPSTN Using AAR	2-11
VoPSTN Using Dial Plan	2-12

Multi-Site WAN with Distributed Call Processing	2-13
Benefits of the Distributed Call Processing Model	2-14
Best Practices for the Distributed Call Processing Model	2-14
Call Processing Agents for the Distributed Call Processing Model	2-15
Clustering Over the IP WAN	2-16
WAN Considerations	2-17
Intra-Cluster Communications	2-17
Failover Between Subscriber Servers	2-18
Cisco CallManager Publisher	2-18
Call Detail Records (CDR)	2-19
Delay Testing	2-19
Error Rate	2-20
Troubleshooting	2-20
Local Failover Deployment Model	2-21
Cisco CallManager Provisioning for Local Failover	2-22
Gateways for Local Failover	2-23
Voicemail for Local Failover	2-23
Music on Hold and Media Resources for Local Failover	2-23
Remote Failover Deployment Model	2-24
Design Considerations for Section 508 Conformance	2-25

CHAPTER 3

Network Infrastructure 3-1

LAN Infrastructure	3-4
LAN Design for High Availability	3-4
Campus Access Layer	3-4
Campus Distribution Layer	3-7
Campus Core Layer	3-9
Network Services	3-10
Power over Ethernet (PoE)	3-19
Category 3 Cabling	3-20
IBM Type 1A and 2A Cabling	3-20
LAN Quality of Service (QoS)	3-21
Traffic Classification	3-22
Interface Queuing	3-23
Bandwidth Provisioning	3-23
Impairments to IP Communications if QoS is Not Employed	3-23
WAN Infrastructure	3-24
WAN Design and Configuration Best Practices	3-24
Deployment Considerations	3-24

Guaranteed Bandwidth	3-25
Best-Effort Bandwidth	3-26
WAN Quality of Service (QoS)	3-26
Bandwidth Provisioning	3-28
Traffic Prioritization	3-34
Link Efficiency Techniques	3-36
Traffic Shaping	3-38
Wireless LAN Infrastructure	3-40
WLAN Design and Configuration	3-40
Wireless Infrastructure Considerations	3-41
Wireless AP Configuration and Design	3-44
Wireless Security	3-45
WLAN Quality of Service (QoS)	3-47
Traffic Classification	3-47
Interface Queuing	3-47
Bandwidth Provisioning	3-48

CHAPTER 4

Voice Gateways 4-1

Understanding Cisco Gateways	4-1
Cisco Access Analog Gateways	4-1
Cisco Access Digital Trunk Gateways	4-2
Gateway Selection	4-2
Core Feature Requirements	4-2
Gateway Protocols	4-3
Gateway Protocols and Core Feature Requirements	4-5
DTMF Relay	4-5
Supplementary Services	4-6
Cisco CallManager Redundancy	4-8
Site-Specific Gateway Requirements	4-10
QSIG Support	4-17
Fax and Modem Support	4-18
Gateway Support for Fax Pass-Through and Cisco Fax Relay	4-18
Gateway Support for Modem Pass-Through	4-19
Supported Platforms and Features	4-20
Platform Protocol Support	4-21
Gateway Combinations and Interoperability of Features	4-22
Feature Support Between Similar Gateways	4-23
Gateway Configuration Examples	4-23
Cisco IOS Gateway Configuration	4-23

Cisco VG248 Configuration	4-24
Cisco CallManager Configuration for Cisco IOS Gateways	4-25
Clock Sourcing for Fax and Modem Pass-Through	4-27
T.38 Fax Relay	4-27
Loose Gateway Controlled with Network Services Engine (NSE)	4-27
Gateway Controlled with Capability Exchange Through H.245 or Session Definition Protocol (SDP)	4-28
Call-Agent-Controlled T.38 with H.323 Annex D and MGCP	4-29

CHAPTER 5

Cisco CallManager Trunks 5-1

H.323 Trunks	5-2
Intercluster Trunk (Non-Gatekeeper Controlled)	5-2
Intercluster Trunk (Gatekeeper Controlled)	5-2
H.225 Trunk (Gatekeeper Controlled)	5-3
Gatekeeper Trunk Redundancy, Resilience, and Load Balancing	5-3
H.323 Trunks with Media Termination Points	5-7
SIP Trunks with Media Termination Points	5-8
H.323 Operation in Cisco CallManager	5-8

CHAPTER 6

Media Resources 6-1

Voice Termination Resources	6-1
Medium and High Complexity Mode	6-2
Flex Mode	6-2
DSP Resources for Voice Termination	6-3
Conferencing, Transcoding, and MTP Resources	6-7
Conferencing	6-8
Transcoding	6-8
Media Termination Point (MTP)	6-9
Hardware Resources for MTP, Conferencing, and Transcoding	6-10
Allocating Resources to Voice Termination or a DSP Farm	6-12
Calculating DSP Requirements for NM-HDV	6-13
Platform Support of DSP Resources	6-13
Software Conferencing	6-14
Annunciator	6-15
Cisco IP Voice Media Streaming Application	6-16
Conferencing Guidelines and Application Scenarios	6-17
Conferencing Guidelines for All Deployment Models	6-17
Conferencing Guidelines for Single-Site Deployments	6-17
Conferencing Guidelines for Multi-Site WAN Deployments with Centralized Call Processing	6-18

Media Resource Groups and Lists	6-20
Media Resource Allocation in Cisco CallManager	6-20
Conferencing Guidelines for Multi-Site WAN Deployments with Distributed Call Processing	6-21
Software MTP Resources	6-23
Transcoding Guidelines and Application Scenarios	6-23
Single-Site Deployments	6-23
Multi-Site WAN Deployments with Centralized Call Processing	6-23
Multi-Site WAN Deployments with Distributed Call Processing	6-25
IP PSTN Access	6-26

CHAPTER 7

Music on Hold 7-1

Deployment Basics of MoH	7-2
Unicast and Multicast MoH	7-2
Coresident and Standalone MoH Servers	7-3
Fixed and Audio File MoH Sources	7-3
MoH Server as Part of the Cisco CallManager Cluster	7-4
Basic MoH and MoH Call Flows	7-5
Basic MoH	7-5
User and Network Hold	7-6
Unicast and Multicast MoH Call Flows	7-8
MoH Configuration Considerations and Best Practices	7-8
Codec Selection	7-8
Multicast Addressing	7-8
MoH Audio Sources	7-9
Using Multiple Fixed or Live Audio Sources	7-9
Unicast and Multicast in the Same Cisco CallManager Cluster	7-10
Redundancy	7-11
Quality of Service (QoS)	7-12
Hardware and Capacity Planning for MoH Resources	7-12
Server Platform Limits	7-12
Resource Provisioning and Capacity Planning	7-13
Implications for MoH With Regard to IP Telephony Deployment Models	7-14
Single-Site Campus (Relevant to All Deployments)	7-14
Centralized Multi-Site Deployments	7-14
Call Admission Control and MoH	7-15
Multicast MoH from Branch Router Flash	7-16
Distributed Multi-Site Deployments	7-19
Clustering Over the WAN	7-19
Detailed Unicast and Multicast MoH Call Flows	7-19

CHAPTER 8

Call Processing 8-1

- Cisco CallManager Cluster Guidelines 8-1
 - Hardware Platforms 8-1
 - Cisco CallManager Cluster Services 8-3
 - Publisher 8-4
 - Call Processing Subscriber 8-5
 - TFTP Server 8-8
 - CTI Manager 8-9
 - Media Resources 8-10
 - Voice Activity Detection 8-10
 - Extension Mobility 8-10
- Cisco CallManager Platform Capacity Planning 8-11
 - Call Processing with Cisco CallManager Releases 3.1 and 3.2 8-12
 - Call Processing with Cisco CallManager Release 3.3 and Later 8-12
 - Capacity Calculations 8-12
 - Cisco CallManager Capacity Tool 8-13
- Gatekeeper Considerations 8-16
 - Hardware Platform Selection 8-16
 - Gatekeeper Redundancy 8-16
 - Hot Standby Router Protocol (HSRP) 8-17
 - Gatekeeper Clustering (Alternate Gatekeeper) 8-19
 - Directory Gatekeeper Redundancy 8-22
- Interoperability of Cisco CallManager and CallManager Express 8-25
 - Multi-Site IP Telephony Deployments Using Cisco CallManager and CME 8-26
 - Multi-Site IP Telephony Deployments Using Cisco CallManager, CME, and an H.450 Tandem Gateway 8-28

CHAPTER 9

Call Admission Control 9-1

- Call Admission Control Elements 9-2
 - Cisco CallManager Locations 9-2
 - Gatekeeper 9-6
 - RSVP 9-7
 - RSVP Principles 9-8
 - RSVP Operation Models 9-10
 - Design Best Practices 9-13
 - IP-to-IP Gateway 9-14
 - Via-Zone Gatekeeper 9-15
 - Design Best Practices 9-16
 - Redundancy 9-17

Configuration Guidelines	9-17
Call Admission Control Design	9-20
Simple Hub-and-Spoke Topologies	9-21
Purely Centralized Deployments	9-22
Purely Distributed Deployments	9-22
Hybrid Centralized/Distributed Deployments	9-25
Two-Tier Hub-and-Spoke Topologies	9-26
Purely Centralized Deployments	9-27
Purely Distributed Deployments	9-31
Hybrid Centralized/Distributed Deployments	9-33
MPLS-Based Topologies	9-34
Purely Centralized Deployments	9-36
Purely Distributed Deployments	9-39
Hybrid Centralized/Distributed Deployments	9-39
Complex Topologies	9-40
Case Study	9-41

CHAPTER 10

Dial Plan 10-1

Planning Considerations	10-2
On-Net versus Off-Net Dialing	10-2
Abbreviated Dialing	10-3
Avoiding Overlap of Extension Dialing	10-3
Dialing String Length	10-3
Uniform On-Net Dial Plan	10-4
Variable Length On-Net Dial Plan	10-5
On-Net and Off-Net Access Codes	10-6
Plan Ahead	10-7
Dial Plan Elements	10-7
Call Routing in Cisco CallManager	10-8
Route Patterns	10-10
Route Lists	10-13
Route Groups	10-13
Route Group Devices	10-13
Calling Privileges in Cisco CallManager	10-14
Partitions	10-15
Calling Search Spaces	10-16
Digit Manipulation in Cisco CallManager	10-18
Automated Alternate Routing	10-19
Establish the PSTN Number of the Destination	10-19

Prefix the Required Access Codes	10-20
Select the Proper Dial Plan and Route	10-20
Special Considerations for Sites Located Within the Same Local Dialing Area	10-21
Extension Mobility	10-22
Hunt Lists and Line Groups	10-23
Hunt Pilot	10-24
Hunt List	10-25
Line Group	10-25
Line Group Devices	10-26
Time-of-Day Routing	10-26
Call Routing in Cisco IOS with H.323 Dial Peers	10-27
Call Routing in Cisco IOS with a Gatekeeper	10-30
Centralized Gatekeeper Configuration	10-34
Distributed Gatekeeper Configuration	10-36
Distributed Gatekeeper Configuration with Directory Gatekeeper	10-37
Calling Privileges in Cisco IOS with H.323 Dial Peers	10-39
Digit Manipulation in Cisco IOS with H.323 Dial Peers	10-41
Design Considerations	10-43
Design Guidelines for Multi-Site Deployments	10-44
Choosing a Dial Plan Approach	10-46
Deploying Uniform On-Net Dial Plans	10-48
Inter-Site Calls Within a Cluster	10-49
Outgoing PSTN and IP WAN Calls	10-49
Incoming Calls	10-49
Voicemail Calls	10-49
Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing	10-50
Inter-Site Calls Within a Cluster	10-52
Outgoing PSTN and IP WAN Calls	10-53
Incoming Calls	10-54
Voicemail Calls	10-54
Deploying Variable-Length On-Net Dial Plans with Flat Addressing	10-54
Inter-Site Calls Within a Cluster	10-57
Outgoing PSTN and IP WAN Calls	10-58
Incoming Calls	10-61
Voicemail Calls	10-61
Special Considerations for Deployments Without Site Codes	10-61
Building Classes of Service for Cisco CallManager with the Traditional Approach	10-63
Building Classes of Service for Cisco CallManager with the Line/Device Approach	10-67
Guidelines for the Line/Device Approach	10-71
Extension Mobility Considerations for the Line/Device Approach	10-72

Building Classes of Service in Cisco IOS with H.323	10-75
Deploying Call Coverage	10-78
Deploying Call Coverage in a Multi-Site Centralized Call Processing Model	10-79
Deploying Call Coverage in a Distributed Call Processing Model	10-80
Hunt Pilot Scalability	10-81

CHAPTER 11

Emergency Services 11-1

Planning for 911 Functionality	11-2
Public Safety Answering Point (PSAP)	11-2
911 Network Service Provider	11-2
Interface Points into the Appropriate 911 Networks	11-3
Interface Type	11-4
Dynamic ANI (Trunk Connection)	11-5
Static ANI (Line Connection)	11-6
Emergency Response Location Mapping	11-6
Emergency Location Identification Number Mapping	11-7
Nomadic Phone Considerations	11-9
Cisco Emergency Responder	11-9
Emergency Call String	11-11
Gateway Considerations	11-11
Gateway Placement	11-11
Gateway Blocking	11-12
Answer Supervision	11-12
Cisco Emergency Responder Considerations	11-13
Device Mobility Across Call Admission Control Locations	11-13
Default Emergency Response Location	11-13
Soft Clients	11-14
Test Calls	11-14
PSAP Callback to Shared Directory Numbers	11-14
Multi-Cluster Considerations	11-15
Single Cisco ER Group	11-15
Multiple Cisco ER Groups	11-16
Emergency Call Routing within a Cisco ER Cluster	11-18
Scalability Considerations for Cisco ER Clustering	11-19
ALI Formats	11-19

CHAPTER 12

Voicemail Design 12-1

Cisco Unity	12-1
Cisco Unity Express	12-1

Deployment Models for Cisco Unity Express	12-2
Best Practices for Deploying Cisco Unity Express	12-6
Third-Party Voicemail Systems	12-7
SMDI	12-7
Cisco Messaging Interface	12-8
Cisco VG248	12-8
Considerations When Using FXS Ports	12-9
Digital Set Emulation	12-9
Dual PBX Integration	12-11
Centralized Voicemail	12-12
Positive Disconnect Supervision	12-14
Summary of Third-Party Voicemail Integration	12-15

CHAPTER 13

Cisco Unity 13-1

Messaging Deployment Models	13-2
Single-Site Messaging	13-2
Centralized Messaging	13-2
Distributed Messaging	13-2
Messaging Failover	13-3
Messaging System Infrastructure Components	13-4
Managing Bandwidth	13-4
Cisco Unity Native Transcoding Operation	13-6
Voice Port Integration with a Cisco CallManager Cluster	13-6
Voice Port Integration with Dedicated Cisco CallManager Backup Servers	13-12
Centralized Messaging and Centralized Call Processing	13-13
Distributed Messaging with Centralized Call Processing	13-15
Combined Messaging Deployment Models	13-17
Centralized Messaging with Clustering Over the WAN	13-18
Distributed Messaging with Clustering Over the WAN	13-20
Cisco Unity Messaging Failover	13-21
Cisco Unity Failover and Clustering Over the WAN	13-22
Centralized Messaging with Multiple Cisco CallManager Servers	13-23

CHAPTER 14

Directory Access and Integration 14-1

Directory Access Versus Directory Integration	14-2
Directory Access for Cisco IP Telephony Endpoints	14-3
Directory Integration with Cisco CallManager	14-6

The Cisco Customer Directory Configuration Plugin	14-7
Security Considerations	14-8
Adding Cisco CallManager Servers to a Domain	14-9
Best Practices for Directory Integration	14-9
Planning the Directory Integration	14-10
Preparing the Directory for Integration	14-11
Integrating Cisco CallManager with the Directory	14-15
Maintaining the Directory Integration	14-18

CHAPTER 15
IP Telephony Migration Options 15-1

Phased Migration	15-1
Flash Cutover, or Big-Bang Migration	15-3
The Need for QSIG in Multiple-Site Enterprises	15-3
Summary	15-4

CHAPTER 16
IP Telephony Endpoints 16-1

Analog Gateway	16-1
Analog Network Module	16-1
Low-Density Voice/Fax Network Module	16-2
High-Density Voice/Fax Network Module	16-2
Cisco IP Communications Voice/Fax Network Module	16-2
Supported Platforms and Cisco IOS Requirements for Analog Network Modules	16-3
Cisco Communication Media Module (CMM)	16-3
WS-X6624-FXS Analog Interface Module	16-4
Cisco VG224 Gateway	16-4
Cisco VG248 Gateway	16-4
Cisco ATA 186 and 188	16-4
Cisco Desktop IP Phone	16-5
Low-End Cisco Desktop IP Phone	16-5
Cisco IP Phone 7902G	16-5
Cisco IP Phone 7905G	16-5
Cisco IP Phone 7910G and 7910G+SW	16-5
Cisco IP Phone 7912G	16-6
Mid-Range Cisco Desktop IP Phone	16-6
Barge	16-6
cBarge	16-7
Security	16-8
High-End Cisco Desktop IP Phone	16-9
Software-Based Endpoint	16-10

Cisco IP SoftPhone	16-10
Maximum Cisco IP SoftPhone Configuration Limits	16-11
Codec Selection	16-11
Call Admission Control	16-12
Cisco IP Communicator	16-13
Maximum IP Communicator Configuration Limits	16-13
Codec Selection	16-13
Call Admission Control	16-14
Wireless Endpoint	16-14
Site Survey	16-14
Authentication	16-15
Capacity	16-15
Phone Configuration	16-16
Roaming	16-16
AP Call Admission Control	16-17
Device Mobility and Cisco CallManager	16-18
Cisco IP Conference Station	16-19
QoS Recommendations	16-19
Cisco VG224 and VG248	16-19
Cisco ATA 186 and Conference Station	16-20
Cisco ATA 188 and IP Phones	16-20
Software-Based Endpoints	16-24
Cisco Wireless IP Phone 7920G	16-27
Endpoint Features Summary	16-30

APPENDIX A

Recommended Hardware and Software Combinations A-1

GLOSSARY

INDEX



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 builds upon ideas and concepts presented in previous versions of the *Cisco IP Telephony Solution Reference Network Design* (SRND), which are available online at

<http://cisco.com/go/srnd>

This document assumes that you are already familiar with the basic terms and concepts presented in previous versions of the *Cisco IP Telephony SRND*. 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 4.1. [Table 1](#) lists the topics that are new in the current release of this document or that have changed significantly from previous releases of Cisco CallManager.

Table 1 *New or Changed Information for Cisco CallManager Release 4.1*

Topic	Described in:
Call admission control	Call Admission Control, page 9-1
Call processing	Call Processing, page 8-1
Centralized voicemail	Centralized Voicemail, page 12-12
Cisco Unity	Cisco Unity, page 13-1
IP-to-IP gateway	IP-to-IP Gateway, page 9-14
Media resources	Media Resources, page 6-1
Music-on-hold sessions	Hardware and Capacity Planning for MoH Resources, page 7-12
QoS recommendations for endpoints	QoS Recommendations, page 16-19
Resource Reservation Protocol (RSVP)	RSVP, page 9-7
Security considerations for directory integration	Security Considerations, page 14-8
Security considerations for endpoints	Security, page 16-8
Trunks	Cisco CallManager Trunks, page 5-1

Revision History

This document may be updated at any time without notice. You can obtain the latest version of this document online at

<http://cisco.com/go/srnd>

Visit this Cisco.com website periodically and check for documentation updates by comparing the revision date (on the front title page) of your copy with the revision date of the online document.

The following table lists the revision history for this document.

Revision Date	Comments
April, 2005	Initial version of this document for Cisco CallManager Release 4.1.

Obtaining Documentation

Cisco documentation and additional literature are available on Cisco.com. Cisco also provides several ways to obtain 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 at this URL:

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

You can access the Cisco website at this URL:

<http://www.cisco.com>

You can access international Cisco websites at this URL:

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

Ordering Documentation

You can find instructions for ordering documentation at this URL:

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

You can order Cisco documentation in these ways:

- Registered Cisco.com users (Cisco direct customers) can order Cisco product documentation from the Ordering tool:
<http://www.cisco.com/en/US/partner/ordering/index.shtml>
- Nonregistered Cisco.com users can order documentation through a local account representative by calling Cisco Systems Corporate Headquarters (California, USA) at 408 526-7208 or, elsewhere in North America, by calling 1 800 553-NETS (6387).

Documentation Feedback

You can send comments about technical documentation to bug-doc@cisco.com.

You can submit comments by using the response card (if present) 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

For all customers, partners, resellers, and distributors who hold valid Cisco service contracts, Cisco Technical Support provides 24-hour-a-day, award-winning technical assistance. The Cisco Technical Support Website on Cisco.com features extensive online support resources. In addition, Cisco Technical Assistance Center (TAC) engineers provide telephone support. If you do not hold a valid Cisco service contract, contact your reseller.

Cisco Technical Support Website

The Cisco Technical Support Website provides online documents and tools for troubleshooting and resolving technical issues with Cisco products and technologies. The website is available 24 hours a day, 365 days a year, at this URL:

<http://www.cisco.com/techsupport>

Access to all tools on the Cisco Technical Support Website requires a Cisco.com user ID and password. If you have a valid service contract but do not have a user ID or password, you can register at this URL:

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



Note

Use the Cisco Product Identification (CPI) tool to locate your product serial number before submitting a web or phone request for service. You can access the CPI tool from the Cisco Technical Support Website by clicking the **Tools & Resources** link under Documentation & Tools. Choose **Cisco Product Identification Tool** from the Alphabetical Index drop-down list, or click the **Cisco Product Identification Tool** link under Alerts & RMAs. The CPI tool offers three search options: by product ID or model name; by tree view; or for certain products, by copying and pasting **show** command output. Search results show an illustration of your product with the serial number label location highlighted. Locate the serial number label on your product and record the information before placing a service call.

Submitting a Service Request

Using the online TAC Service Request Tool is the fastest way to open S3 and S4 service requests. (S3 and S4 service requests are those in which your network is minimally impaired or for which you require product information.) After you describe your situation, the TAC Service Request Tool provides recommended solutions. If your issue is not resolved using the recommended resources, your service request is assigned to a Cisco TAC engineer. The TAC Service Request Tool is located at this URL:

<http://www.cisco.com/techsupport/servicerequest>

For S1 or S2 service requests or if you do not have Internet access, contact the Cisco TAC by telephone. (S1 or S2 service requests are those in which your production network is down or severely degraded.) Cisco TAC engineers are assigned immediately to S1 and S2 service requests to help keep your business operations running smoothly.

To open a service request by telephone, use one of the following numbers:

Asia-Pacific: +61 2 8446 7411 (Australia: 1 800 805 227)

EMEA: +32 2 704 55 55

USA: 1 800 553-2447

For a complete list of Cisco TAC contacts, go to this URL:

<http://www.cisco.com/techsupport/contacts>

Definitions of Service Request Severity

To ensure that all service requests are reported in a standard format, Cisco has established severity definitions.

Severity 1 (S1)—Your network is “down,” or there is a critical impact to your business operations. You and Cisco will commit all necessary resources around the clock to resolve the situation.

Severity 2 (S2)—Operation of an existing network is severely degraded, or significant aspects of your business operation are negatively affected by inadequate performance of Cisco products. You and Cisco will commit full-time resources during normal business hours to resolve the situation.

Severity 3 (S3)—Operational performance of your network is impaired, but most business operations remain functional. You and Cisco will commit resources during normal business hours to restore service to satisfactory levels.

Severity 4 (S4)—You require information or assistance with Cisco product capabilities, installation, or configuration. There is little or no effect on your business operations.

Obtaining Additional Publications and Information

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

- Cisco Marketplace provides a variety of Cisco books, reference guides, and logo merchandise. Visit Cisco Marketplace, the company store, at this URL:

<http://www.cisco.com/go/marketplace/>

- 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://cisco.com/univercd/cc/td/doc/pcat/>

- *Cisco Press* publishes a wide range of general networking, training and certification titles. Both new and experienced users will benefit from these publications. For current Cisco Press titles and other information, go to Cisco Press at this URL:

<http://www.ciscopress.com>

- *Packet* magazine is the Cisco Systems technical user magazine for maximizing Internet and networking investments. Each quarter, Packet delivers coverage of the latest industry trends, technology breakthroughs, and Cisco products and solutions, as well as network deployment and troubleshooting tips, configuration examples, customer case studies, certification and training information, and links to scores of in-depth online resources. You can access Packet magazine at this URL:

<http://www.cisco.com/packet>

- *iQ Magazine* is the quarterly publication from Cisco Systems designed to help growing companies learn how they can use technology to increase revenue, streamline their business, and expand services. The publication identifies the challenges facing these companies and the technologies to help solve them, using real-world case studies and business strategies to help readers make sound technology investment decisions. You can access iQ Magazine at this URL:

<http://www.cisco.com/go/iqmagazine>

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

<http://www.cisco.com/ipj>

- World-class networking training is available from Cisco. You can view current offerings at this URL:

<http://www.cisco.com/en/US/learning/index.html>



Introduction

The communications industry, and industry analysts in general, now widely accept and acknowledge that IP has become the universal transport medium for enterprise networks. The rapid adoption and migration of vendors to IP as a transport for data, voice, and video applications further endorses this transition to a converged networking paradigm. This migration even includes those vendors who have historically used time-division multiplexing (TDM) infrastructures and relied upon traditional practices. The message is clear: the move toward IP is happening now.

Cisco provides end-to-end IP network solutions based on open standards, with an established portfolio of applications and an ecosystem of partners to support your transition to a converged IP network. Cisco IP Telephony forms the core of this ever-expanding solution set known as Cisco IP Communications.

Overview of Cisco IP Communications Solutions

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

Cisco IP Communications encompasses the following solutions:

- **IP Telephony**

IP telephony refers to technology that transmits voice communications over a network using IP standards. The Cisco IP Telephony solution includes a wide array of hardware and software products such as call processing agents, IP phones, video devices, and special applications.

- **Unified Communications**

Cisco Unified Communications solutions deliver powerful unified messaging (email, voice, and fax messages managed from a single inbox) and intelligent voice messaging (full-featured voicemail providing advanced capabilities) to improve communications, boost productivity, and enhance customer service capabilities across an organization. Cisco Unified Communications solutions also enable users to streamline communication processes through the use of features such as rules-based call routing, simplified contact management, and speech recognition.

- Rich-Media Conferencing

Cisco Rich-Media Conferencing solutions enhance the virtual meeting environment with a integrated set of IP-based tools for voice, video, and Web conferencing.

- Video Telephony

The Cisco Video Telephony solution enables real-time video communications and collaboration using the same IP network and call processing agent as the Cisco IP Telephony solution. With Cisco Video Telephony, making a video call is now as easy as dialing a phone number.

- Customer Contact

Cisco Customer Contact solutions are a combination of strategy and architecture that promote efficient and effective customer communications across a globally capable network by enabling organizations to draw from a broader range of resources to service customers, including access to an unlimited pool of agents and multiple channels of communication as well as customer self-help tools.

- Third-Part Applications

Cisco works with leading-edge companies to provide the broadest selection of innovative third-party IP telephony applications and products focused on critical business needs such messaging, customer care, and workforce optimization.

The remainder of this document focuses on the Cisco IP Telephony solution. For information about the other Cisco IP Communications solutions, refer to the documentation available online at

<http://www.cisco.com>

You can also find additional design guides for these solutions at

<http://cisco.com/go/smd>

Overview of the Cisco IP Telephony Solution

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

Figure 1-1 illustrates a typical IP Telephony solution employing the Cisco IP network infrastructure, with Cisco CallManager as the call processing agent.

The diagram illustrates a Cisco Unified Communications Manager (CallManager) network architecture. It shows three main locations: Headquarters, Branch (with Call Processing and Applications), and Branch (with SRST and Applications). The Headquarters location contains a Cisco CallManager cluster (four CallManager nodes) and Applications (Voice mail, IVR, ICD, ...). The Branch locations contain IP phones and a Cisco CallManager cluster (three CallManager nodes). The network is connected via a central cloud labeled PSTN and IP WAN. The Rest of world is also connected to the central cloud. The diagram shows the flow of traffic between these components, including the use of a Cisco CallManager cluster and the integration of various applications.

- Cisco IP Network Infrastructure, page 1-4
- Quality of Service, page 1-4
- Call Processing Agent, page 1-4
- Communication Endpoints, page 1-5
- Applications, page 1-6
- Security, page 1-7
- Network Management Tools, page 1-8

Cisco IP Network Infrastructure

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

Quality of Service

Voice, as a class of IP network traffic, has strict requirements concerning packet loss, delay, and delay variation (also known as jitter). To meet these requirements for voice traffic, the Cisco IP Telephony solution includes Quality of Service (QoS) features such as classification, queuing, traffic shaping, compressed Real-Time Transport Protocol (cRTP), and Transmission Control Protocol (TCP) header compression.

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

- Traffic marking
- Enhanced queuing services
- Link fragmentation and interleaving (LFI)
- Compressed RTP (cRTP)
- Low-latency queuing (LLQ)
- Link efficiency
- Traffic shaping
- Call admission control

Call Processing Agent

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

You can deploy the call processing capabilities of Cisco CallManager according to one of the following models, depending on the size, geographical distribution, and functional requirements of your enterprise:

- Single-site call processing model

In the single-site model, each site or campus has its own Cisco CallManager or cluster of Cisco CallManagers to perform call processing functions. No voice traffic travels over the IP WAN; instead, external calls or calls to remote sites use the public switched telephone network (PSTN).

- Multi-site WAN model with centralized call processing

In the multi-site WAN model with centralized call processing, the Cisco CallManager cluster resides at the main (or central) campus, and communication with remote branch offices normally takes place over the IP WAN. If either the central site or the IP WAN is down, the remote sites can continue to have service through a feature called Survivable Remote Site Telephony (SRST), which runs on Cisco IOS gateways. The remote sites can also place calls over the PSTN if the IP WAN is temporarily oversubscribed, and you can interconnect a number of central sites through intercluster trunks.

- Multi-site WAN model with distributed call processing

In the multi-site WAN model with distributed call processing, each site has its own Cisco CallManager cluster for call processing. Communication between sites normally takes place over the IP WAN, with the PSTN serving as a backup voice path. With this model, you can interconnect any number of sites across the IP WAN.

- 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. To provide call processing redundancy, you can deploy backup servers either locally at each site or at a remote site across the WAN. Clustering over the WAN is well suited as a disaster recovery plan for business continuance sites or as a single solution for small or medium sites.

The remaining chapters of this guide explain how to apply these deployment models in designing your Cisco IP Telephony network.

Communication Endpoints

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

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

- Software-based IP phones

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

- Video telephony endpoints

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

- Wireless IP phones

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

Applications

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

- Extension mobility

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

- Cisco MeetingPlace

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

- Unified messaging

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

- Web services for Cisco IP Phones

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

- Cisco IP Contact Center (IPCC) Express

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

- Cisco IP Contact Center (IPCC) Enterprise Edition

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

Security

The Cisco IP Telephony solution addresses security in the following main areas, among others:

- Physical security for restricting physical access to important application servers and network components
- Network access security to prevent hostile logins or attacks
- Security measures for Cisco CallManager, endpoint devices, and various directories and databases
- Mechanisms for defining calling privileges for various classes of users
- Careful network design and management to enhance security

Network Management Tools

The Cisco IP network infrastructure offers a number of network management, QoS, and security management tools that support the IP Telephony solution. Cisco CallManager offers enhanced software and configuration management tools that leverage the strength and flexibility of IP networks. The Cisco CallManager user interface simplifies the most common subscriber and telephony configuration tasks by building upon legacy telephony administration systems and adding software and web-based applications.

In addition, CiscoWorks2000 includes a number of network management tools to manage the operations, administration, and maintenance of IP Telephony networks. In particular, CiscoWorks IP Telephony Environment Monitor (ITEM) provides a suite of applications and tools that facilitate effective management of both small and large IP Telephony installations based on the Cisco Architecture for Voice, Video, and Integrated Data (AVVID) as well as Cisco IOS software. CiscoWorks ITEM provides the following major features:

- Problem-focused fault analysis — Provides timely information about the health of IP Telephony environments.
- Confidence testing and monitoring — Use synthetic testing to emulate normal day-to-day operations and to validate operational readiness of the IP infrastructure and the Cisco IP Telephony deployment.
- Intelligent integration with existing management infrastructures — Generates intelligent traps that can be forwarded to other event-management systems installed in the network, sent to email or pager gateways, or displayed on the Alerts and Activities Display (AAD).
- Evaluation and correlation capabilities — Evaluate the general health of the IP Telephony environment in the monitored network environment.
- Alerts and Activities Display (AAD) — Provides a proactive, web-based operations screen for real-time status and alerting of actual and suspected problems in the underlying IP network as well as in the Cisco IP Telephony implementation.
- ITEM Multi-View — Enables large enterprise customers and managed service providers to partition specific user communities and manage all of them from a single ITEM implementation.



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 2-2](#)

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 2-4](#)

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 2-13](#)

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 2-16](#)

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, the section on [Design Considerations for Section 508 Conformance, page 2-25](#), presents guidelines for designing your IP telephony network to provide accessibility to users with disabilities, in conformance with U.S. Section 508.

**Note**

For the most recent information on recommended hardware platforms and software releases, frequently refer to the documentation at http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/ccmcomp.htm

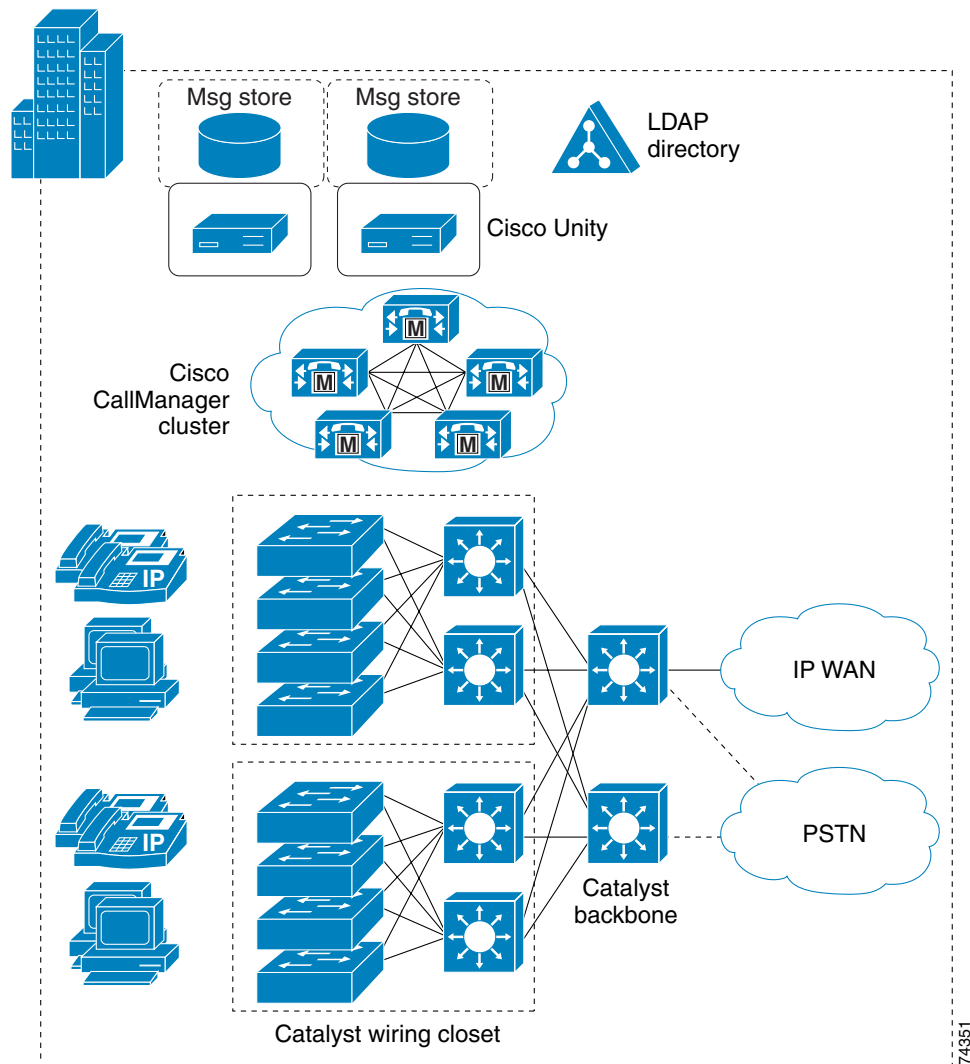
Single Site

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 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)
- Voicemail 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 voicemail systems

[Figure 2-1](#) illustrates the model for an IP telephony network within a single campus or site.

Figure 2-1 Single-Site Model

Benefits of the Single-Site Model

A single infrastructure for a converged network solution provides significant cost benefits and enables IP telephony to take advantage of the many IP-based applications in the enterprise. Single-site deployment also allows each site to be completely self-contained. There is no dependency for service in the event of an IP WAN failure or insufficient bandwidth, and there is no loss of call processing service or functionality.

In summary, the main benefits of the single-site model are:

- Ease of deployment
- A common infrastructure for a converged solution
- Simplified dial plan
- No transcoding resources required, due to the use of only G.711 codecs

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 3-1](#).)
- Follow the provisioning recommendations listed in the chapter on [Call Processing, page 8-1](#).

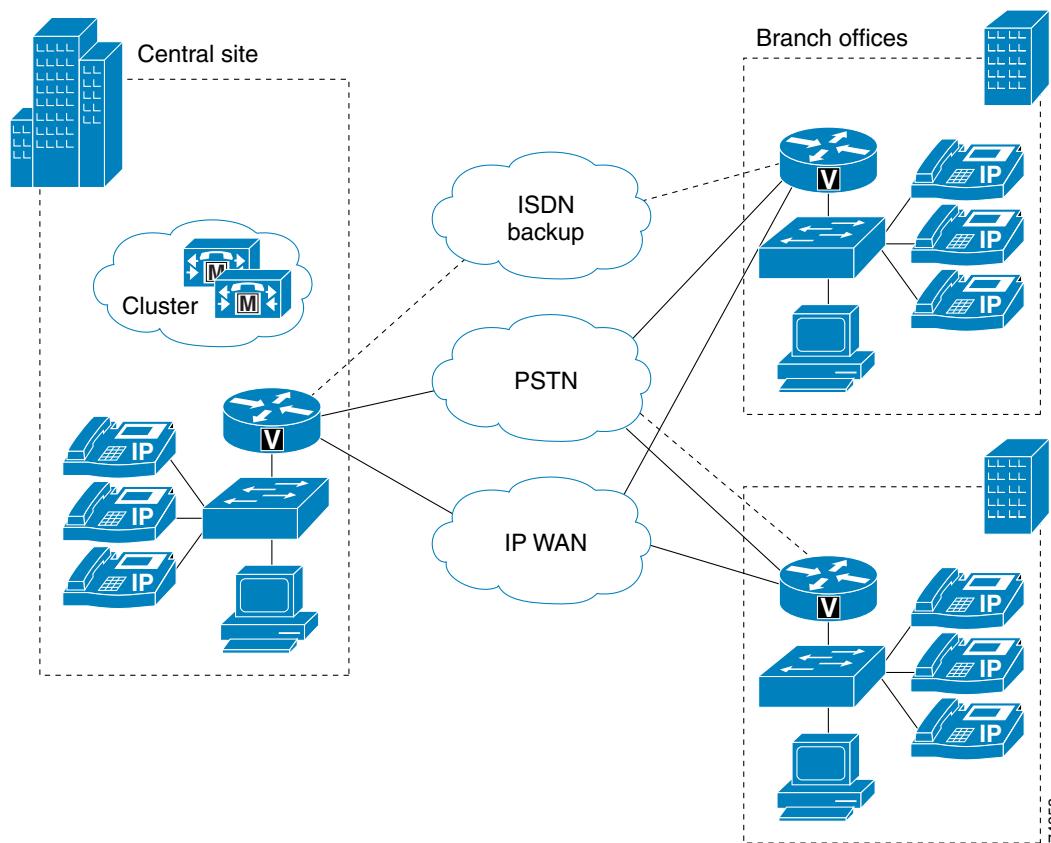
Multi-Site WAN with Centralized Call Processing

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 2-2](#) 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 voicemail 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 2-2 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 [Cisco CallManager Locations, page 9-2](#), 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.

Best Practices for the Centralized Call Processing Model

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).
- Use the locations mechanism in Cisco CallManager to provide call admission control into and out of remote branches. See the chapter on [Call Admission Control, page 9-1](#), for details on how to apply this mechanism to the various WAN topologies.
- 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 Cisco CallManager Express on the branch router.

Remote Site Survivability

When deploying IP Telephony across a WAN with the centralized call processing model, you should take additional steps to ensure that data and voice services at the remote sites are highly available. [Table 2-1](#) summarizes the different strategies for providing high availability at the remote sites. The choice of one of these strategies may depend on several factors, such as specific business or application requirements, the priorities associated with highly available data and voice services, and cost considerations.

Table 2-1 Strategies for High Availability at the Remote Sites

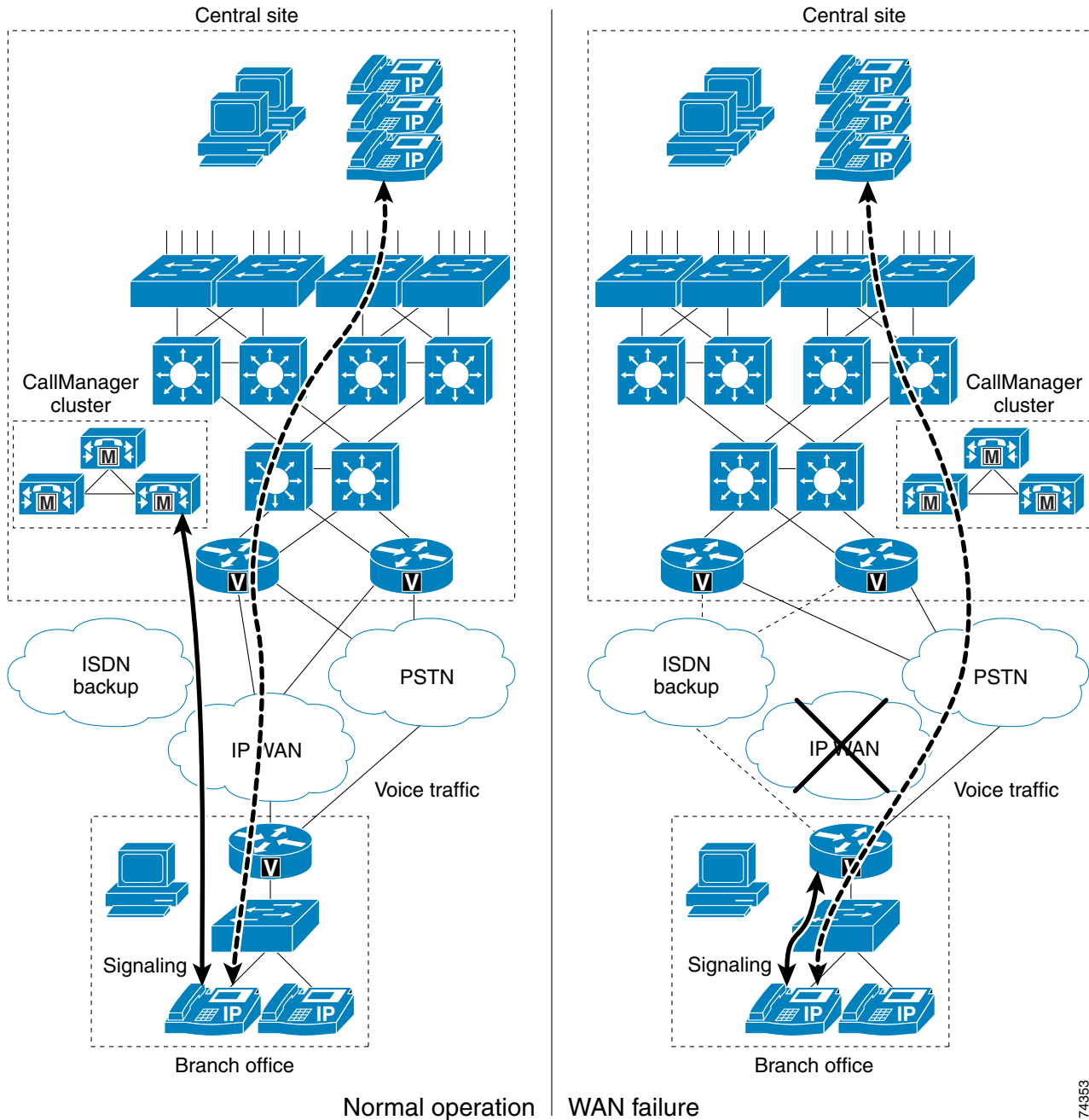
Strategy	High Availability for Data Services?	High Availability for Voice Services?
Redundant IP WAN links in branch router	Yes	Yes
Redundant branch router platforms + Redundant IP WAN links	Yes	Yes
Survivable Remote Site Telephony (SRST) only	No	Yes

Table 2-1 Strategies for High Availability at the Remote Sites (continued)

Strategy	High Availability for Data Services?	High Availability for Voice Services?
Data-only ISDN backup + SRST	Yes	Yes
Data and voice ISDN backup	Yes	Yes (see rules below)

The first two solutions listed in [Table 2-1](#) provide high availability at the network infrastructure layer by adding redundancy to the IP WAN access points, thus maintaining IP connectivity between the remote IP phones and the centralized Cisco CallManager at all times. These solutions apply to both data and voice services, and are entirely transparent to the call processing layer. The options range from adding a redundant IP WAN link at the branch router to adding a second branch router platform with a redundant IP WAN link.

The third solution listed in [Table 2-1](#), Survivable Remote Site Telephony (SRST), provides high availability for voice services only, by providing a subset of the call processing capabilities within the remote office router and enhancing the IP phones with the ability to “re-home” to the call processing functions in the local router if a WAN failure is detected. [Figure 2-3](#) illustrates a typical call scenario with SRST.

Figure 2-3 Survivable Remote Site Telephony (SRST) Application Scenario

74353

Under normal operations shown in the left part of [Figure 2-3](#), the branch office connects to the central site via an IP WAN, which carries data traffic, voice traffic, and call signaling. The IP phones at the branch office exchange call signaling information with the Cisco CallManager cluster at the central site and place their calls across the IP WAN. The branch router or gateway forwards both types of traffic (call signaling and voice) transparently and has no knowledge of the IP phones.

If the WAN link to the branch office fails, or if some other event causes loss of connectivity to the Cisco CallManager cluster, the branch IP phones re-register with the branch router. The branch router queries the IP phones for their configuration and uses this information to build its own configuration

automatically. The branch IP phones can then make and receive calls either internally or through the PSTN. The phone displays the message “CM fallback mode,” and some advanced Cisco CallManager features are unavailable and are grayed out on the phone display.

When WAN connectivity to the central site is reestablished, the branch IP phones automatically re-register with the Cisco CallManager cluster and resume normal operation. The branch router deletes its information about the IP phones and reverts to its standard routing or gateway configuration.

The last two solutions in [Table 2-1](#) use an ISDN backup link to provide survivability during WAN failures. The two deployment options for ISDN backup are:

- Data-only ISDN backup

With this option, ISDN is used for data survivability only, while SRST is used for voice survivability. Note that you should configure an access control list on the branch router to prevent Skinny Client Control Protocol (SCCP) traffic from entering the ISDN interface, so that signaling from the IP phones does not reach the Cisco CallManager at the central site.

- Data and voice ISDN backup

With this option, ISDN is used for both data and voice survivability. In this case, SRST is not used because the IP phones maintain IP connectivity to the Cisco CallManager cluster at all times. However, Cisco recommends that you use ISDN to transport data and voice traffic only if all of the following conditions are true:

- The bandwidth allocated to voice traffic on the ISDN link is the same as the bandwidth allocated to voice traffic on the IP WAN link.
- The ISDN link bandwidth is fixed.
- All the required QoS features have been deployed on the router's ISDN interfaces. Refer to the chapter on [Network Infrastructure, page 3-1](#), for more details on QoS.

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 this Voice over the PSTN (VoPSTN) model variation still requires a QoS-enabled WAN with appropriate bandwidth configured for the signaling traffic.

You can implement VoPSTN in one of the following ways:

- Using the automated alternate routing (AAR) feature. (For more information on AAR, see the section on [Automated Alternate Routing, page 10-19](#).)
- Using a combination of dial plan constructs in both Cisco CallManager and the PSTN gateways.

VoPSTN can be an attractive option in deployments where IP WAN bandwidth is either scarce or expensive with respect to PSTN charges, or where IP WAN bandwidth upgrades are planned for a later date but the IP telephony system is already being deployed.

**Note**

Because of its very nature, the VoPSTN deployment model variation offers basic voice functionality that is a reduced subset of the Cisco CallManager feature set.

In particular, regardless of the implementation choice, the system designer should address the following issues, among others:

- Centralized voicemail requires:
 - 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 voicemail carry the redirecting DN, to ensure proper voicemail box selection.
 - If the voicemail system is accessed through an MGCP gateway, the voicemail pilot number must be a fully qualified E.164 number.
- The Extension Mobility feature is limited to IP phones contained within a single branch site.
- 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 destination phones require a fully qualified Direct Inward Dial (DID) PSTN number that can be called directly. Non-DID DNs cannot be reached directly from a different branch site.
- With VoPSTN, music on hold (MoH) is limited to cases where the holding party is co-located with the MoH resource. If MoH servers are deployed at the central site, then only calls placed on hold 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 voicemail system located outside the branch.
- Conferencing resources must be co-located with the phone initiating the conference.
- 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
- Use of the Attendant Console outside of the central site can require a considerable amount of bandwidth if the remote sites must access large user account directories without caching them.
- Because all inter-branch media (including transfers) are sent through the PSTN, the gateway trunk group must be sized to accommodate all inter-branch traffic, transfers, and centralized voicemail access.
- Cisco recommends that you do not deploy shared lines across branches, such that the devices sharing the line are in different branches.
- 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 these cases, Cisco recommends requiring the user to enter the fully qualified PSTN number of the destination.

In addition to these general considerations, the following sections present recommendations and issues specific to each of the following implementation methods:

- [VoPSTN Using AAR, page 2-11](#)
- [VoPSTN Using Dial Plan, page 2-12](#)

VoPSTN Using AAR

This method consists of configuring the Cisco CallManager dial plan as in a traditional centralized call processing deployment, with the automated alternate routing (AAR) feature also properly configured. AAR provides transparent re-routing over the PSTN of inter-site calls when the locations mechanism for call admission control determines that there is not enough available WAN bandwidth to accept an additional call.

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 be 1 kbps, thus preventing *all* calls from traversing the WAN. With this configuration, all inter-site calls trigger the AAR functionality, which automatically re-routes the calls over the PSTN.

The AAR implementation method for VoPSTN offers the following benefits:

- An easy migration path to a complete IP telephony deployment. When bandwidth becomes available to support voice media over the WAN, the dial plan can be maintained intact, and the only change needed is to update the location bandwidth value for each site.
- Support for some supplementary features, such as callback on busy.

In addition to the general considerations listed for VoPSTN, the following design guidelines apply to the AAR implementation method:

- AAR functionality must be configured properly.
- As a general rule, supported call initiation devices include IP phones, gateways, and line-side gateway-driven analog phones.
- Inter-branch calls can use AAR only if the destination devices are IP phones or Cisco Unity ports.
- Inter-branch calls to other endpoints must use a fully qualified E.164 number.
- All on-net, inter-branch calls will display the message, "Network congestion, rerouting."
- If destination phones become unregistered (for example, due to WAN connectivity interruption), AAR functionality will not be invoked and abbreviated dialing will not be possible. If the destination phone has registered with an SRST router, then it can be reached by directly dialing its PSTN DID number.
- If originating phones become unregistered (for example, due to WAN connectivity interruption), they will go into SRST mode. To preserve abbreviated dialing functionality under these conditions, configure the SRST router with an appropriate set of translation rules.
- 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.

VoPSTN Using Dial Plan

This method relies on a specific dial plan configuration within Cisco CallManager and the PSTN gateways to route all inter-site calls over the PSTN. The dial plan must place IP phone DN's at each site into a different partition, and their calling search space must provide access only to the site's internal partition and a set of route patterns that point to the local PSTN gateway.

Abbreviated inter-site dialing can still be provided via a set of translations at each branch site, one for each of the other branch sites. These translations are best accomplished with H.323 gateways and translation rules within Cisco IOS.

The dial plan method for implementing VoPSTN offers the following benefits:

- Easier configuration because AAR is not needed.
- Abbreviated dialing automatically works even under WAN failure conditions on either the originating or destination side, because the Cisco IOS translation rules within the H.323 gateway are effective in SRST mode.

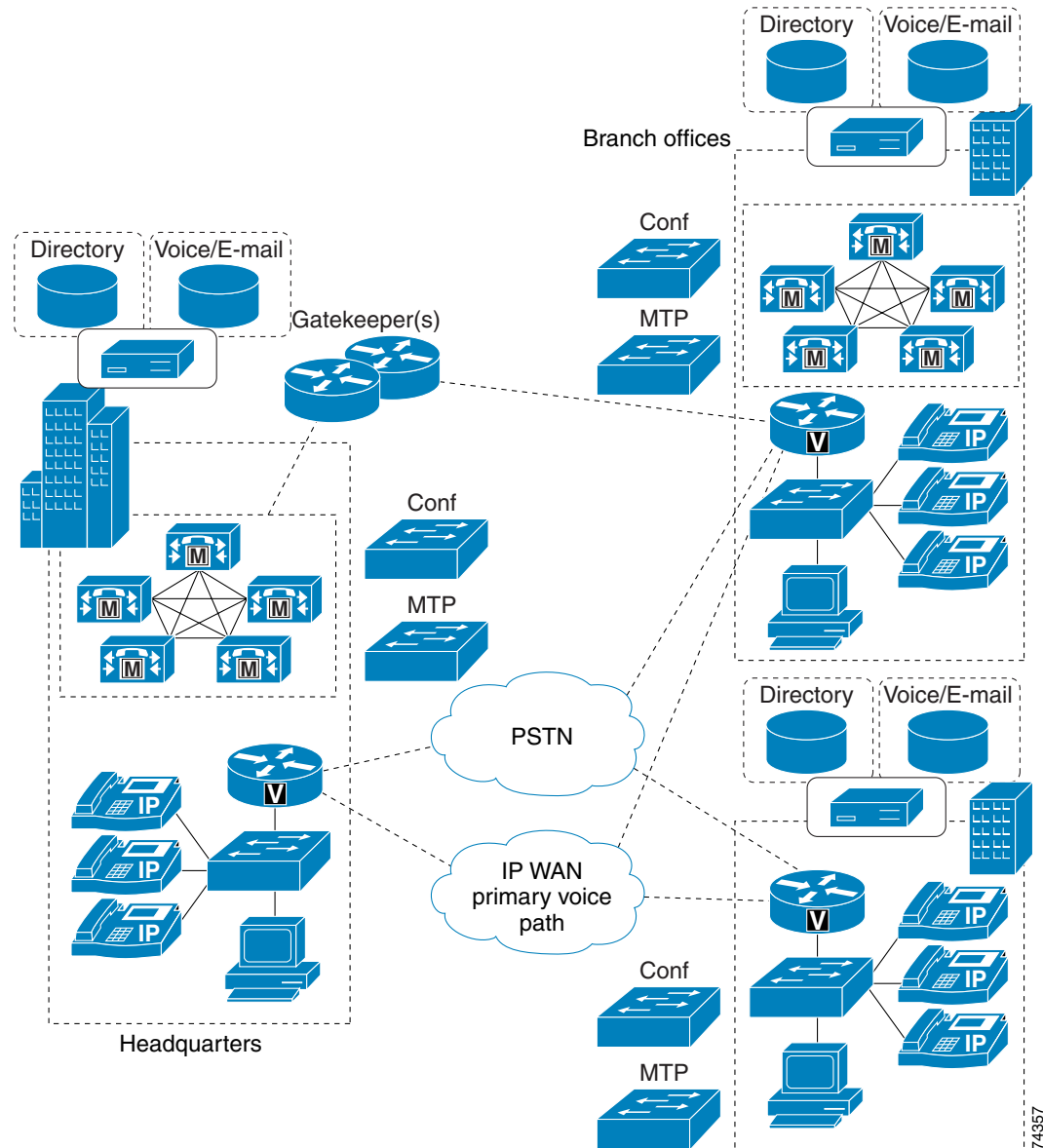
In addition to the general considerations listed for VoPSTN, the following design guidelines apply to the dial plan implementation method:

- There is no support for supplementary features such as callback on busy.
- Some CTI-based applications do not support overlapping extensions (that is, two or more phones configured with the same DN, although in different partitions).
- There is no easy migration to a complete IP telephony deployment because the dial plan needs to be redesigned.

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. [Figure 2-4](#) illustrates a typical distributed call processing deployment.

Figure 2-4 A Distributed Call Processing Deployment



74357

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 CallManager Express
 - 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 2-2](#).)

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)

Benefits of the Distributed Call Processing Model

The multi-site WAN model with distributed call processing provides the following benefits:

- PSTN call cost savings when using the IP WAN for calls between sites.
- Use of the IP WAN to bypass toll charges by routing calls through remote site gateways, closer to the PSTN number dialed. This practice is known as tail-end hop-off (TEHO).
- Maximum utilization of available bandwidth by allowing voice traffic to share the IP WAN with other types of traffic.
- No loss of functionality during IP WAN failure because there is a call processing agent at each site.
- Scalability to hundreds of sites.

Best Practices for the Distributed Call Processing Model

A multi-site WAN deployment with distributed call processing has many of the same requirements as a single site or a multi-site WAN deployment 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 2-2](#), and [Multi-Site WAN with Centralized Call Processing, page 2-4](#).)

Gatekeeper or Session Initiation Protocol (SIP) proxy servers are among the key elements in the multi-site WAN model with distributed call processing. They each provide dial plan resolution, with the gatekeeper also providing call admission control. 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 Cisco IOS gatekeeper to provide call admission control into and out of each site.
- To provide high availability of the gatekeeper, use Hot Standby Router Protocol (HSRP) gatekeeper pairs, gatekeeper clustering, and alternate gatekeeper support. In addition, use multiple gatekeepers to provide redundancy within the network.
- Size the platforms appropriately to ensure that performance and capacity requirements can be met.
- 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 WAN topology.

For more information on the various functions performed by gatekeepers, refer to the following sections:

- For gatekeeper call admission control, see [Call Admission Control, page 9-1](#).
- For gatekeeper scalability and redundancy, see [Call Processing, page 8-1](#).
- For gatekeeper dial plan resolution, see [Dial Plan, page 10-1](#).

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

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

- Provide adequate redundancy for the SIP proxies.
- Ensure that the SIP proxies have the capacity for the call rate and number of calls required in the network.
- Planning for call admission control is outside the scope of this document.

Call Processing Agents for the Distributed Call Processing Model

Your choice of call processing agent will vary, based on many factors. The main factors, for the purpose of design, are the size of the site and the functionality required.

For a distributed call processing deployment, each site has its own call processing agent. The design of each site varies with the call processing agent, the functionality required, and the fault tolerance required. For example, in a site with 500 phones, a Cisco CallManager cluster containing two servers can provide one-to-one redundancy, with the backup server being used as a publisher and TFTP (Trivial File Transfer Protocol) server.

The requirement for IP-based applications also greatly affects the choice of call processing agent because only Cisco CallManager provides the required support for many Cisco IP applications.

[Table 2-2](#) lists recommended call processing agents.

Table 2-2 Recommended Call Processing Agents

Call Processing Agent	Recommended Size	Comments
Cisco CallManager Express (CME)	Up to 120 phones	- For small remote sites - Capacity depends on Cisco IOS platform
Cisco CallManager	50 to 30,000 phones	- Small to large sites, depending on the size of the Cisco CallManager cluster - Supports centralized or distributed call processing
Legacy PBX with VoIP gateway	Depends on PBX	Number of IP WAN calls and functionality depend on the PBX-to-VoIP gateway protocol and the gateway platform

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 chapter on [Call Processing, page 8-1](#).

Clustering over the WAN can support two types of deployments:

- [Local Failover Deployment Model, page 2-21](#)

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

- [Remote Failover Deployment Model, page 2-24](#)

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

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

The key advantages of clustering over the WAN are:

- Single point of administration for users for all sites within the cluster
- Feature transparency
- Shared line appearances
- Extension mobility within the cluster
- Unified dial plan

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

WAN Considerations

For clustering over the WAN to be successful, you must carefully plan, design, and implement various characteristics of the WAN itself. The Intra-Cluster Communication Signaling (ICCS) between Cisco CallManager servers consists of many traffic types. The ICCS traffic types are classified as either priority or best-effort. Priority ICCS traffic is marked with IP Precedence 3 (DSCP 26 or PHB AF31 in Cisco CallManager releases prior to 4.0 and DSCP 24 or PHB CS3 for release 4.0 and later). Best-effort ICCS traffic is marked with IP Precedence 0 (DSCP 0 or PHB BE). The various types of ICCS traffic are described in [Intra-Cluster Communications, page 2-17](#), which also provides further guidelines for provisioning. The following design guidelines apply to the indicated WAN characteristics:

- Delay

The maximum one-way delay between any Cisco CallManager server for all priority ICCS traffic should not exceed 20 ms, or 40 ms round-trip time (RTT). Delay for other ICCS traffic should be kept reasonable to provide timely database and directory access. Measuring the delay is covered in [Delay Testing, page 2-19](#). Propagation delay between two sites introduces 6 microseconds per kilometer without any other network delays being considered. This equates to a theoretical maximum distance of approximately 3000 km for 20 ms delay or approximately 1860 miles. These distances are provided only as relative guidelines and in reality will be shorter due to other delay incurred within the network.

- Jitter

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

- Packet loss and errors

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

- Bandwidth

Provision the correct amount of bandwidth between each server for the expected call volume, type of devices, and number of devices. This bandwidth is in addition to any other bandwidth for other applications sharing the network, including voice and video traffic between the sites. The bandwidth provisioned must have QoS enabled to provide the prioritization and scheduling for the different classes of traffic. The general rule of thumb for bandwidth is to over-provision and under-subscribe.

- Quality of Service

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

Intra-Cluster Communications

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

The intra-cluster traffic between the servers consists of the following:

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

Failover Between Subscriber Servers

With Cisco CallManager Release 3.1 and 3.2, failover behavior is dependant on the reachability of the publisher and the delay between the subscriber and the publisher. If the publisher is reachable, the subscriber will request the relevant device configuration records directly from the publisher during device registration. The round-trip delay and the available bandwidth for the SQL database traffic will affect the speed of registrations. The effect of this is that failover for devices at remote locations to the publisher may experience delays of approximately 20 minutes before all devices on a full server complete the failover process. If the publisher is unreachable during failover, the subscriber will use its own most recent copy of the database for the configuration information. Because there is no incurred delay for the subscriber to access its own database, the failover time in this case is approximately 5 minutes for a full server.

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

Cisco CallManager Publisher

The publisher replicates a read-only copy of the master database to all other servers in the cluster. If changes are made in the publisher's master database during a period when another server in the cluster is unreachable, the publisher will replicate the updated database when communications are re-established. During any period when the publisher is unreachable or offline, no changes can be made to the configuration database. All subscriber databases are read-only and may not be modified. Most normal operations of the cluster will not be affected during this period, including the following:

- Call processing
- Failover

- Installation registration of previously configured devices

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

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

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

Call Detail Records (CDR)

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

Delay Testing

The maximum round-trip time (RTT) between any two servers must not exceed 40 ms at any time. This time limit must include all delays in the transmission path between the two servers. Verifying the round trip delay using the **ping** utility on the Cisco CallManager server will not provide an accurate result. The ping is sent as a best-effort tagged packet and is not transported using the same QoS-enabled path as the ICCS traffic. Therefore, Cisco recommends that you verify the delay by using the closest network device to the Cisco CallManager servers, ideally the access switch to which the server is attached. Cisco IOS provides an extended ping capable of setting the Layer 3 type of service (ToS) bits to make sure the ping packet is sent on the same QoS-enabled path that the ICCS traffic will traverse. The time recorded by the extended ping is the round-trip time (RTT), or the time it takes to traverse the communications path and return. The maximum RTT between any two Cisco CallManager servers is 40 ms, and therefore the maximum one-way delay would be 20 ms. This delay is critical to the performance of the call processing capabilities of the Cisco CallManager cluster and must be strictly enforced.

The following example shows a Cisco IOS extended ping with the ToS bit (IP Precedence) set to 3:

```
Access_SW#ping
Protocol [ip]:
Target IP address: 10.10.10.10
Repeat count [5]:
Datagram size [100]:
Timeout in seconds [2]:
Extended commands [n]: y
Source address or interface:
Type of service [0]: 3
Set DF bit in IP header? [no]:
Validate reply data? [no]:
Data pattern [0xABCD]:
```

```

Loose, Strict, Record, Timestamp, Verbose[none]:
Sweep range of sizes [n]:
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 10.10.10.10, timeout is 2 seconds:
!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/2/4 ms

```

Error Rate

The expected error rate should be zero. Any errors, dropped packets, or other impairments to the IP network can have an impact to the call processing performance of the cluster. This may be noticeable by delay in dial tone, slow key or display response on the IP phone, or delay from off-hook to connection of the voice path. Although Cisco CallManager will tolerate random errors, they should be avoided to avoid impairing the performance of the cluster.

Troubleshooting

If the Cisco CallManager subscribers in a cluster are experiencing impairment of the ICCS communication due to higher than expected delay, errors, or dropped packets, some of the following symptoms might occur:

- IP phones, gateways, or other devices on a remote Cisco CallManager server within the cluster might temporarily be unreachable.
- Calls might be disconnected or might fail during call setup.
- Users might experience longer than expected delays before hearing dial tone.
- Busy hour call completions (BHCC) might be low.
- The ICCS (SDL session) might be reset or disconnected. The following example shows a Cisco CallManager SDL trace of a remote server VO30-7835-8 going Out of Service and the devices that were reachable by that server being removed as “available” destinations:

```

RemoteCMOutOfService: Ip address: VO30-7835-8 remoteClusterId
VO30-7835-1-Cluster|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2>
|Delete entries from SsManagerTable, now this table has 75
entries|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>
|Delete entries from FeatActTable, now this table has 70
entries|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>
|Digit analysis: Remove remote pattern /5000020 , PID:
7:34:1|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>
|Digit analysis: Remove remote pattern /5000066 , PID:
7:34:2|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>
.
.
.
|Digit analysis: Remove remote pattern /5001002 , PID:
7:34:106|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>

```

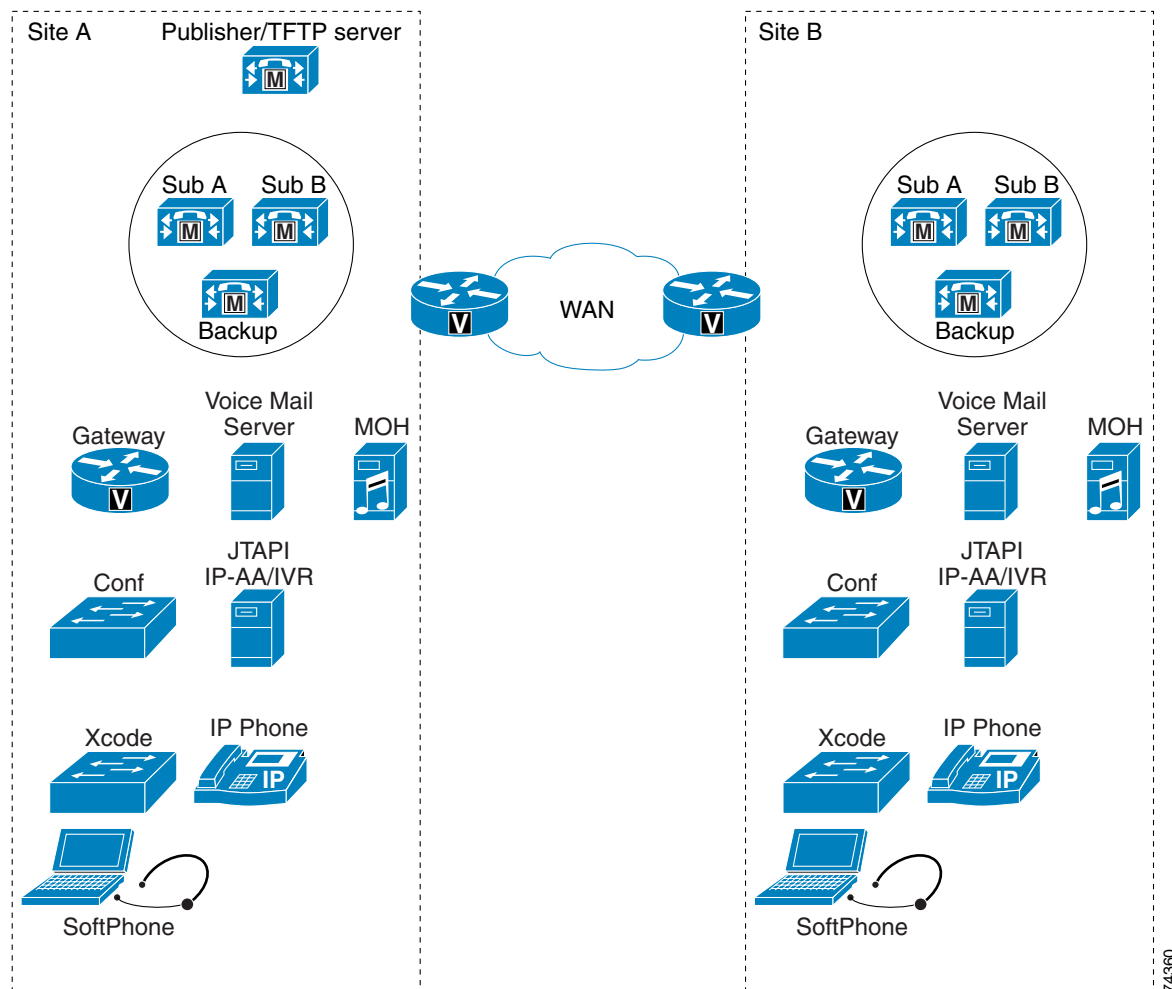
In summary, perform the following tasks to troubleshoot ICCS communication problems:

- Verify the delay between the servers.
- Check all links for errors or dropped packets.
- Verify that QoS is correctly configured.
- Verify that sufficient bandwidth is provisioned for the queues and across the WAN to support all the traffic.

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 can support up to four sites. The maximum number of phones and other devices will be dependant on the quantity and type of servers deployed. The maximum total number of IP phones for all sites is 30,000. (See [Figure 2-5](#).)

Figure 2-5 Example of Local Failover Model



Observe the following guidelines when implementing the local failover model:

- Configure each site to contain at least one primary Cisco CallManager subscriber and one backup subscriber.
- Configure Cisco CallManager *groups* and *device pools* to allow devices within the site to register with only the servers at that site under all conditions.
- Cisco highly recommends that you replicate key services (TFTP, DNS, DHCP, LDAP, and IP Phone Services), all media resources (conference bridges and music on hold), and gateways at each site to provide the highest level of resiliency. You could also extend this practice to include a voicemail system at each site.

- Under a WAN failure condition, sites without access to the publisher database will lose some functionality, as follows:
 - System administration at the local site will not be able to add, modify, or delete any part of the configuration.
 - Extension mobility users will not be able to log in or log out of the IP phones.
 - Changes to Call Forward All will not be allowed.
- Under WAN failure conditions, calls made to phone numbers that are not currently communicating with the subscriber placing the call, will result in either a fast-busy tone or a call forward (possibly to voicemail, dependant on the location of the phone number being forwarded to). During this condition, users should manually dial those numbers via the PSTN.
- Every 10,000 busy hour call attempts (BHCA) between sites that are clustered over the WAN requires 900 kbps of bandwidth for Intra-Cluster Communication Signaling (ICCS). This is a minimum bandwidth requirement, and bandwidth is allocated in multiples of 900 kbps. The ICCS traffic types are classified as either priority or best-effort. Priority ICCS traffic is marked with IP Precedence 3 (DSCP 26 or PHB AF31 in Cisco CallManager releases prior to 4.0 and DSCP 24 or PHB CS3 for release 4.0 and later). Best-effort ICCS traffic is marked with IP Precedence 0 (DSCP 0 or PHB BE).
- The minimum recommended bandwidth between sites that are clustered over the WAN is 1.544 Mbps. This amount allows for the minimum of 900 kbps for ICCS and 644 kbps for SQL, LDAP, and other inter-server traffic.
- 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.
- If remote branches using centralized call processing are connected to the main sites via clustering over the WAN, pay careful attention to the configuration of call admission control to avoid oversubscribing the links used for clustering over the WAN.
 - If the bandwidth is not limited on the links used for clustering over the WAN (that is, if the interfaces to the links are OC-3s or STM-1s and there is no requirement for call admission control), then the remote sites may be connected to any of the main sites because all the main sites should be configured as location <None>. This configuration still maintains hub-and-spoke topology for purposes of call admission control.
 - If you are using the Multiprotocol Label Switching (MPLS) Virtual Private Network (VPN) feature, all sites in Cisco CallManager locations and the remote sites may register with any of the main sites.
 - If bandwidth is limited between the main sites, call admission control must be used between sites, and all remote sites must register with the main site that is configured as location <None>. This main site is considered the hub site, and all other remote sites and clustering-over-the-WAN sites are spokes sites.
- During a software upgrade, all servers in the cluster should be upgraded during the same maintenance period, using the standard upgrade procedures outlined in the software release notes.

Cisco CallManager Provisioning for Local Failover

Provisioning of the Cisco CallManager cluster for the local failover model should follow the design guidelines for capacities outlined in the chapter on [Call Processing, page 8-1](#). If voice or video calls are allowed across the WAN between the sites, then you must configure Cisco CallManager *locations* in

addition to the default location for the other sites, to provide call admission control between the sites. If the bandwidth is over-provisioned for the number of devices, it is still best practice to configure call admission control based on locations. If the locations-based call admission control rejects a call, automatic failover to the PSTN can be provided by the automated alternate routing (AAR) feature.

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

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

IP phones may have shared line appearances between the sites. The ICCS bandwidth provisioned between the sites allows for the additional ICCS traffic that shared line appearances generate. During a WAN outage, call control for each line appearance is segmented, but call control returns to a single Cisco CallManager server once the WAN is restored. During the WAN restoration period, there is additional traffic between the two sites. If this situation occurs during a period of high call volume, the shared lines might not operate as expected during that period. This situation should not last more than a few minutes, but if it is a concern, you can provision additional prioritized bandwidth to minimize the effects.

Gateways for Local Failover

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

You can centralize access to the PSTN gateways if access is not required during a WAN failure and if sufficient additional bandwidth is configured for the number of calls across the WAN. For E911 requirements, additional gateways might be needed at each site.

Voicemail for Local Failover

Cisco Unity or other voicemail systems can be deployed at all sites and integrated into the Cisco CallManager cluster. This configuration provides voicemail access even during a WAN failure and without using the PSTN. Using Voice Mail Profiles, you can allocate the correct voicemail system for the site to the IP phones in the same location. You can configure a maximum of four voicemail systems per cluster that use the SMDI protocol, that are attached directly to the COM port on a subscriber, and that use the Cisco Messaging Interface (CMI).

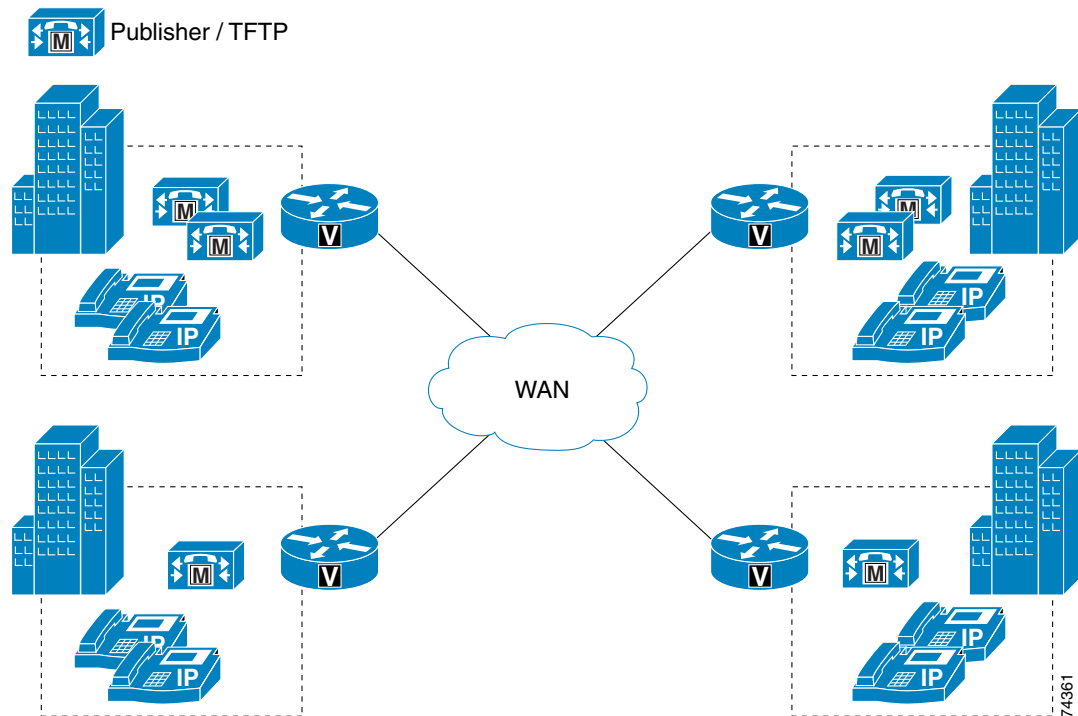
Music on Hold and Media Resources for Local Failover

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

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 up to eight sites, with IP phones and other devices normally registered to a local subscriber when using 1:1 redundancy and the 50/50 load balancing option described in the chapter on [Call Processing, page 8-1](#). Backup subscribers are located across the WAN at one or more of the other sites. (See [Figure 2-6](#).)

Figure 2-6 Remote Failover Model with Four Sites



When implementing the remote failover model, observe all guidelines for the local failover model (see [Local Failover Deployment Model, page 2-21](#)), with the following modifications:

- Configure each site to contain at least one primary Cisco CallManager subscriber and an optional backup subscriber as desired.
- You may configure Cisco CallManager *groups* and *device pools* to allow devices to register with servers over the WAN.
- Signaling or call control traffic requires bandwidth when devices are registered across the WAN with a remote Cisco CallManager server in the same cluster. This bandwidth might be more than the ICCS traffic and should be calculated using the bandwidth provisioning calculations for signalling, as described in [Bandwidth Provisioning, page 3-28](#).

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 VG248, Catalyst 6000, Cisco ATA 188 module, 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 VG248 or ATA 188.



Network Infrastructure

This chapter describes the requirements of the network infrastructure needed to build an IP telephony system in an enterprise environment. [Figure 3-1](#) illustrates the roles of the various devices that form the network infrastructure, and [Table 3-1](#) summarizes the features required to support each of these roles.

IP telephony places strict requirements on IP packet loss, packet delay, and delay variation (or jitter). Therefore, you need to enable most of the Quality of Service (QoS) mechanisms available on Cisco switches and routers throughout the network. For the same reasons, redundant devices and network links that provide quick convergence after network failures or topology changes are also important to ensure a highly available infrastructure

The following sections describe the network infrastructure features as they relate to:

- [LAN Infrastructure, page 3-4](#)
- [WAN Infrastructure, page 3-24](#)
- [Wireless LAN Infrastructure, page 3-40](#)

Figure 3-1 Typical Campus Network Infrastructure

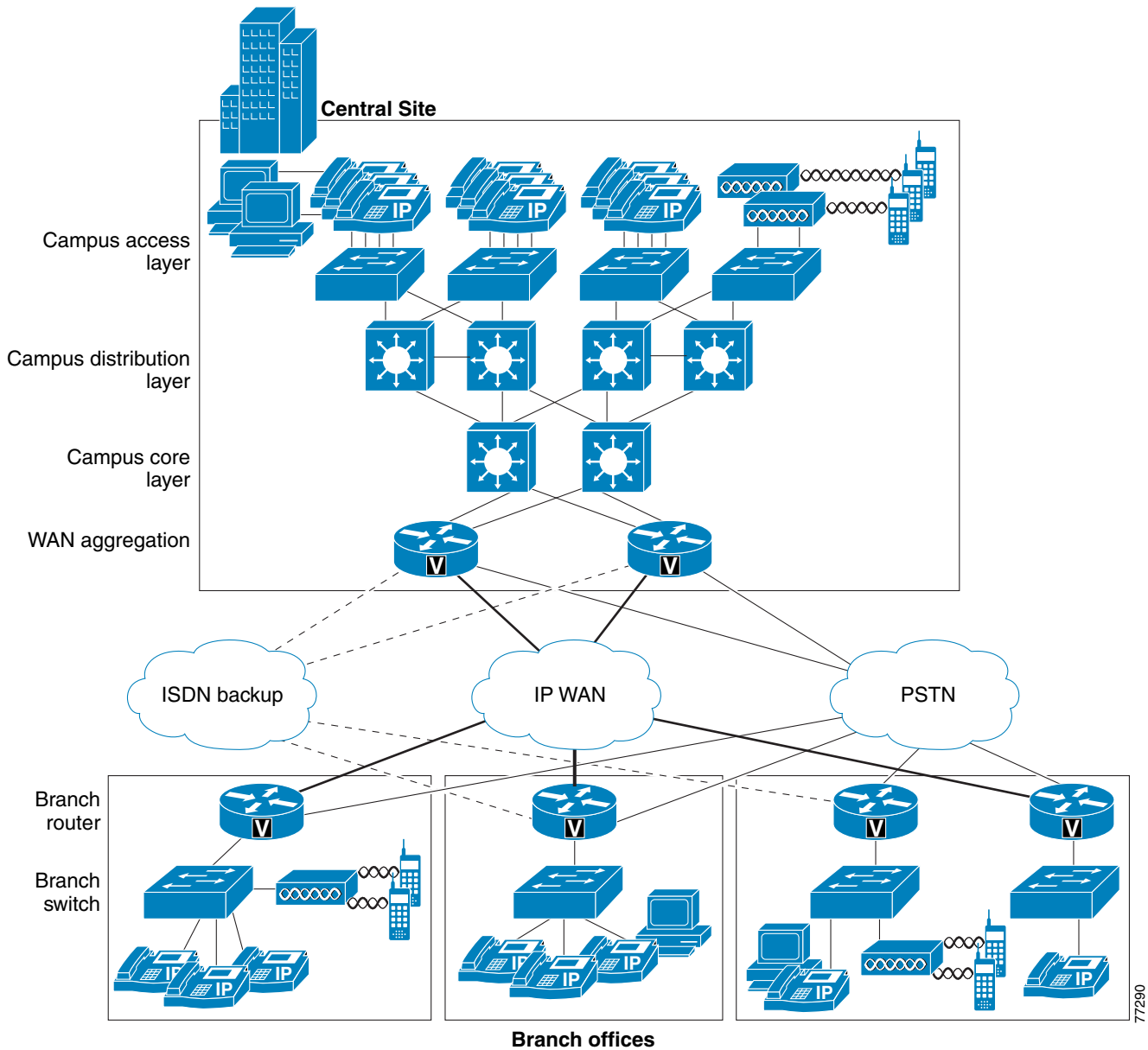


Table 3-1 Required Features for Each Role in the Network Infrastructure

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

Campus LAN infrastructure design is extremely important for proper IP telephony operation on a converged network. Proper LAN infrastructure design requires following basic configuration and design best practices for deploying a highly available network. Further, proper LAN infrastructure design requires deploying end-to-end QoS on the network. The following sections discuss these requirements:

- [LAN Design for High Availability, page 3-4](#)
- [LAN Quality of Service \(QoS\), page 3-21](#)

LAN Design for High Availability

Properly designing a LAN requires building a robust and redundant network from the top down. By structuring the LAN as a layered model (see [Figure 3-1](#)) and developing the LAN infrastructure one step of the model at a time, you can build a highly available, fault tolerant, and redundant network. Once these layers have been designed properly, you can add network services such as DHCP and TFTP to provide additional network functionality. The following sections examine the infrastructure layers and network services:

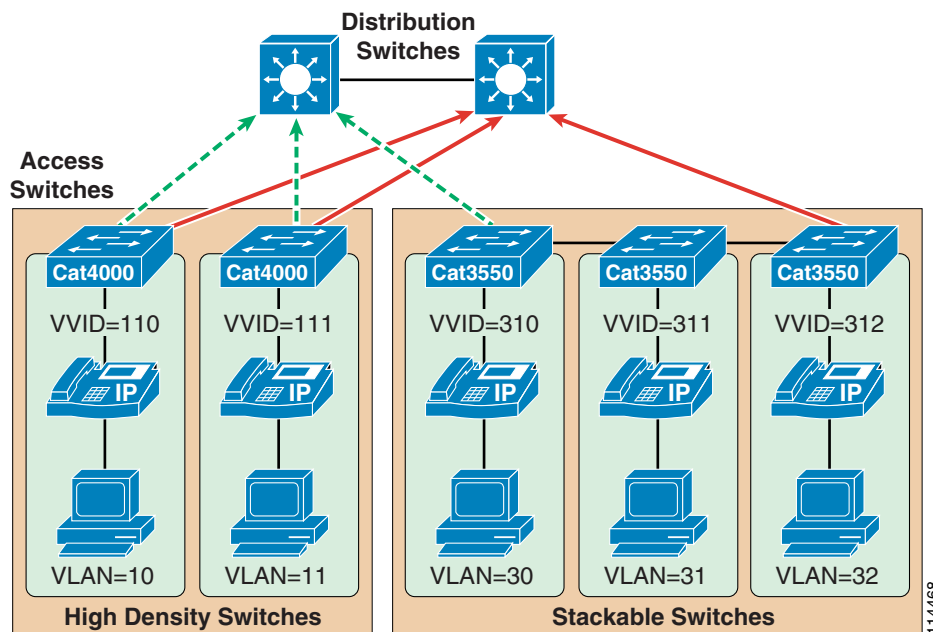
- [Campus Access Layer, page 3-4](#)
- [Campus Distribution Layer, page 3-7](#)
- [Campus Core Layer, page 3-9](#)
- [Network Services, page 3-10](#)

For more information on campus design, refer to the *Gigabit Campus Network Design* white paper at http://www.cisco.com/warp/public/cc/so/neso/lnso/cpsso/gcnd_wp.pdf

Campus Access Layer

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

Proper access layer design starts with assigning a single IP subnet per virtual LAN (VLAN). Typically, a VLAN should not span multiple wiring closet switches; that is, a VLAN should have presence in one and only one access layer switch (see [Figure 3-2](#)). This practice eliminates topological loops at Layer 2, thus avoiding temporary flow interruptions due to Spanning Tree convergence. However, with the introduction of standards-based IEEE 802.1w Rapid Spanning Tree Protocol (RSTP) and 802.1s Multiple Instance Spanning Tree Protocol (MISTP), Spanning Tree can converge at much higher rates. In situations where RSTP and/or MISTP can and have been configured on the access layer switch, there is no need for concern about topological loops. More importantly, confining a VLAN to a single access layer switch also serves to limit the size of the broadcast domain. There is the potential for large numbers of devices within a single VLAN or broadcast domain to generate large amounts of broadcast traffic periodically, which can be problematic. A good rule of thumb is to limit the number of devices per VLAN to about 512, which is equivalent to two Class C subnets (that is, a 23-bit subnet masked Class C address). Typical access layer switches include the stackable Cisco Catalyst 2950, 3500XL, 3550, and 3750, as well as the larger, higher-density Catalyst 4000 and 6000 switches.

Figure 3-2 Access Layer Switches and VLANs for Voice and Data

When you deploy voice, Cisco recommends that you enable two VLANs at the access layer: a native VLAN for data traffic (VLANs 10, 11, 30, 31, and 32 in [Figure 3-2](#)) and a voice VLAN under Cisco IOS or Auxiliary VLAN under CatOS for voice traffic (represented by VVIDs 110, 111, 310, 311, and 312 in [Figure 3-2](#)).

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

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

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

- 802.1Q trunking and 802.1p for proper treatment of Layer 2 CoS packet marking on ports with phones connected
- Multiple egress queues to provide priority queuing of RTP voice packet streams

- The ability to classify or reclassify traffic and establish a network trust boundary
- Inline power capability (Although inline power capability is not mandatory, it is highly recommended for the access layer switches.)
- Layer 3 awareness and the ability to implement QoS access control lists (These features are required if you are using certain IP telephony endpoints, such as a PC running a softphone application, that cannot benefit from an extended trust boundary.)

Spanning Tree Protocol (STP)

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

- PortFast
Enable PortFast on all access ports. The phones, PCs, or servers connected to these ports do not forward bridge protocol data units (BPDUs) that could affect STP operation. PortFast ensures that the phone or PC, when connected to the port, is able to begin receiving and transmitting traffic immediately without having to wait for STP to converge.
- Root guard or BPDU guard
Enable root guard or BPDU guard on all access ports to prevent the introduction of a rogue switch that might attempt to become the Spanning Tree root, thereby causing STP re-convergence events and potentially interrupting network traffic flows. Ports that are set to **errdisable** state by BPDU guard must either be re-enabled manually or the switch must be configured to re-enable ports automatically from the errdisable state after a configured period of time.
- UplinkFast and BackboneFast
Enable these features where appropriate to ensure that, when changes occur on the Layer 2 network, STP converges as rapidly as possible to provide high availability. When using stackable switches such as the Catalyst 2950, 3550, or 3750, enable Cross-Stack UplinkFast (CSUF) to provide fast failover and convergence if a switch in the stack fails.
- UniDirectional Link Detection (UDLD)
Enable this feature to reduce convergence and downtime on the network when link failures or misbehaviors occur, thus ensuring minimal interruption of network service. UDLD detects, and takes out of service, links where traffic is flowing in only one direction. This feature prevents defective links from being mistakenly considered as part of the network topology by the Spanning Tree and routing protocols.



Note

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

Campus Distribution Layer

The distribution layer of the Campus LAN includes the portion of the network from the wiring closet switches to the next-hop switch, and it is the first Layer-2-to-Layer-3 traversal in the LAN. Distribution layer switches typically include Layer 3-enabled Catalyst 4000 and 6000 switches and the Catalyst 3750 for smaller deployments.

At the distribution layer, it is important to provide redundancy to ensure high availability, including redundant links between the distribution layer switches (or routers) and the access layer switches. To avoid creating topological loops at Layer 2, use Layer 3 links for the connections between redundant Distribution switches when possible.

Hot Standby Router Protocol (HSRP)

HSRP should also be enabled at the distribution layer to ensure that all routers are made redundant and that, in the event of a failure, another router can take over. HSRP configuration should incorporate the following:

- Standby track

The **standby track** command indicates that the HSRP should monitor a particular interface(s). If the interface goes down, then the HSRP priority of the box is reduced, typically forcing a failover to another device. This command is used in conjunction with the **standby preempt** command.

- Standby preempt

This command ensures that, when a device's priority becomes higher than all the other HSRP-configured devices in the standby group, that device will take over as the active Layer 3 router for the HSRP standby address.

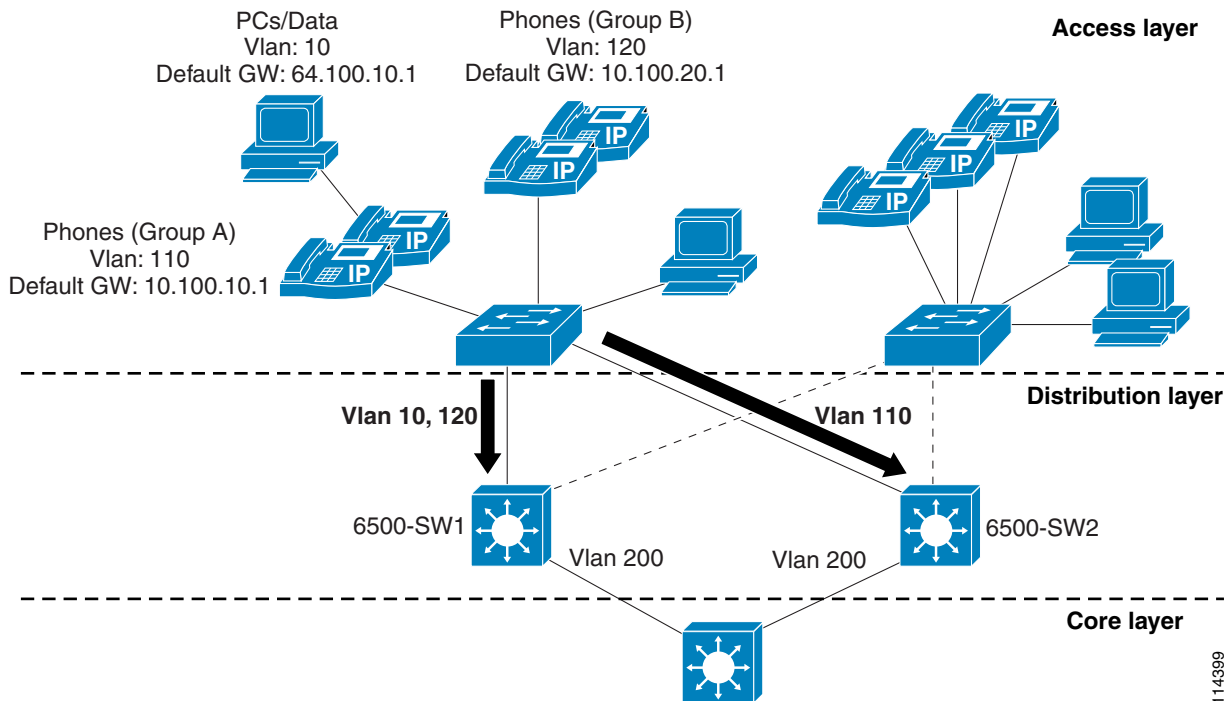
HSRP should also be configured in such a way as to load-balance traffic between both HSRP routers. To provide load balancing, configure each HSRP device as the active HSRP router for one VLAN or interface, and configure the standby router for another VLAN or interface. Evenly distributing active and standby VLANs between both HSRP devices ensures load-balancing. Devices on one VLAN use the active HSRP device as their default gateway, and devices on another VLAN use the same HSRP device as a standby default gateway only if the other HSRP device fails. This type of configuration prevents all network traffic from being sent to a single active router and enables other HSRP devices to help carry the load.

[Figure 3-3](#) shows an example of an HSRP-enabled network. In this figure, the two Catalyst 6500 switches (6500-SW1 and 6500-SW2) have been configured with multiple VLAN interfaces. Assuming that there are no link failures in the network, 6500-SW1 is the standby HSRP router for VLAN 110 (the voice VLAN for Group A phones) and is the active HSRP router for VLAN 10 (the data VLAN) and for VLAN 120 (the voice VLAN for Group B phones). 6500-SW2 is configured in reverse; it is the active HSRP router for VLAN 110 and the standby HSRP router for VLAN 10 and VLAN 120. As configured, both switches are actively in use, and the load can be distributed between the two by evenly distributing all Layer 2 VLANs between them. Each switch is also configured to track its local VLAN 200 interface and, in the event of a VLAN 200 link failure, the other switch will preempt and become the active router for all VLANs. Likewise, if either switch fails, the other switch will handle the traffic for all three VLANs.

The PCs and phones at the access layer in [Figure 3-3](#) have been configured with default gateways that correspond to the HSRP addresses for each of the HSRP groups. Devices in voice VLANs 110 and 120 are pointing to 10.100.10.1 and 10.100.20.1, respectively, as the default gateways, which correspond to the HSRP addresses for the VLAN 110 and 120 interfaces on both switches. Devices in data VLAN 10 are pointing to 64.100.10.1 as the default gateway, which corresponds to the HSRP address of the VLAN 10 interface on both switches. Note that, while traffic flowing from the access layer to the

distribution layer will be distributed between the two switches (as long as there are no failures), no mechanism exists to ensure distribution on the return path. Traffic returning from the core layer and destined for the access layer will follow the shortest and/or least costly routed path.

Figure 3-3 HSRP Network Configuration Example with Standby Preempt and Standby Track



Example 3-1 and Example 3-2 list the configurations for the two Catalyst 6500 switches shown in Figure 3-3.

Example 3-1 Configuration for 6500-SW1

```
interface Vlan 10
  description Data VLAN 10
  ip address 64.100.10.11 255.255.255.0
  standby preempt
  standby ip 64.100.10.1
  standby track Vlan 200

interface Vlan110
  description Voice VLAN 110
  ip address 10.100.10.11 255.255.255.0
  standby preempt
  standby ip 10.100.10.1
  standby track Vlan 200
  standby priority 95

interface Vlan120
  description Voice VLAN 120
  ip address 10.100.20.11 255.255.255.0
  standby preempt
  standby ip 10.100.20.1
  standby track Vlan 200
```

Example 3-2 Configuration for 6500-SW2

```
interface Vlan 10
  description Data VLAN 10
  ip address 64.100.10.12 255.255.255.0
  standby preempt
  standby ip 64.100.10.1
  standby track Vlan 200
  standby priority 95

interface Vlan110
  description Voice VLAN 110
  ip address 10.100.10.12 255.255.255.0
  standby preempt
  standby ip 10.100.10.1
  standby track Vlan 200

interface Vlan120
  description Voice VLAN 120
  ip address 10.100.20.11 255.255.255.0
  standby preempt
  standby ip 10.100.20.1
  standby track Vlan 200
  standby priority 95
```

How quickly HSRP converges when a failure occurs depends on how the HSRP hello and hold timers are configured. By default, these timers are set to 3 and 10 seconds respectively, which means that an hello packet will be sent between the HSRP standby group devices every 3 seconds and that the standby device will become active when an hello packet has not been received for 10 seconds. You can lower these timer settings to speed up the failover or preemption; however, to avoid increased CPU usage and unnecessary standby state flapping, do not set the hello timer below one (1) second or the hold timer below 4 seconds. Note that, if you are using the HSRP tracking mechanism and the tracked link fails, then the failover or preemption occurs immediately regardless of the hello and hold timers.

Routing Protocols

Configure Layer 3 routing protocols, such as Open Shortest Path First (OSPF) and Enhanced Interior Gateway Routing Protocol (EIGRP), at the distribution layer to ensure fast convergence, load balancing, and fault tolerance. Use parameters such as routing protocol timers, path or link costs, and address summaries to optimize and control convergence times as well as to distribute traffic across multiple paths and devices. Cisco also recommends using the **passive-interface** command to prevent routing neighbor adjacencies via the access layer. These adjacencies are typically unnecessary, and they create extra CPU overhead and increased memory utilization because the routing protocol keeps track of them. By using the **passive-interface** command on all interfaces facing the access layer, you prevent routing updates from being sent out on these interfaces and, therefore, neighbor adjacencies are not formed.

Campus Core Layer

The core layer of the Campus LAN includes the portion of the network from the distribution routers or Layer 3 switches to one or more high-end core Layer 3 switches or routers. Layer 3-capable Catalyst 6000 switches are the typical core layer devices, and these core switches can provide connectivity between numerous campus distribution layers.

At the core layer, it is again very important to provide the following types of redundancy to ensure high availability:

- Redundant link or cable paths

Redundancy here ensures that traffic can be rerouted around downed or malfunctioning links.

- Redundant devices

Redundancy here ensures that, in the event of a device failure, another device in the network can continue performing tasks that the failed device was doing.

- Redundant device sub-systems

This type of redundancy ensures that multiple power supplies and Supervisor engines are available within a device so that the device can continue to function in the event that one of these components fails.

Routing protocols at the core layer should again be configured and optimized for path redundancy and fast convergence. There should be no STP in the core because network connectivity should be routed at Layer 3. Finally, each link between the core and distribution devices should belong to its own VLAN or subnet and be configured using a 30-bit subnet mask.

Data Center and Server Farm

Typically, Cisco CallManager cluster servers, including media resource servers, reside in a data center or server farm environment. In addition, centralized gateways and centralized hardware media resources such as conference bridges, DSP or transcoder farms, and media termination points are located in the data center or server farm. Because these servers and resources are critical to voice networks, Cisco recommends distributing all Cisco CallManager cluster servers, centralized voice gateways, and centralized hardware resources between multiple physical switches and, if possible, multiple physical locations within the campus. This distribution of resources ensures that, given a hardware failure (such as a switch or switch line card failure), at least some servers in the cluster will still be available to provide telephony services. In addition, some gateways and hardware resources will still be available to provide access to the PSTN and to provide auxiliary services. Besides being physically distributed, these servers, gateways, and hardware resources should be distributed among separate VLANs or subnets so that, if a broadcast storm or denial of service attack occurs on a particular VLAN, not all voice connectivity and services will be disrupted.

Network Services

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

Domain Name System (DNS)

DNS enables the mapping of host names to IP addresses within a network or networks. DNS server(s) deployed within a network provide a database that maps hostnames to IP addresses. Devices on the network can query the DNS server and receive IP addresses for other devices in the network, thereby facilitating communication between network devices.

Relying on DNS, however, can be problematic. If the DNS server becomes unavailable and a network device is relying on that server to provide a hostname-to-IP-address mapping, communication can and will fail. For this reason, do not rely on DNS for communication between Cisco CallManager and the IP telephony endpoints.

Configure Cisco CallManager(s), gateways, and endpoint devices to use IP addresses rather than hostnames. Cisco does *not* recommend configuration of DNS parameters such as DNS server addresses, hostnames, and domain names. Likewise, configuring HOSTS files is discouraged because the administration of these files in a large IP telephony network with thousands of endpoints and servers can be extremely time-consuming and inefficient. If you eliminate DNS configuration within the IP telephony network, telephony devices and applications do not have to rely on the DNS server.

You must, however, configure the LMHOSTS file. This file provides a mechanism for resolving or mapping server hostnames or NetBios names to IP addresses. The LMHOSTS file must contain a list of server names and corresponding IP address. There should be an entry for every server within the cluster as well as an entry with **127.0.0.1 localhost** (loopback entry). You should copy this file to every server in the cluster. The LMHOSTS file is located in the directory path C:\WINNT\system32\drivers\etc.

[Example 3-3](#) shows a typical LMHOSTS file for a cluster with six servers.

Example 3-3 LMHOSTS File

```
127.0.0.1      localhost      ! The local host entry
64.101.1.7     ccm1-hq-1
64.101.2.7     ccm1-hq-2
64.101.1.8     ccm1-hq-3
64.101.2.8     ccm1-hq-4
64.101.1.21    ccm1-moh-1
64.101.2.21    ccm1-moh-2
```

There are some situations in which configuring and using DNS might be unavoidable. For example, if Network Address Translation (NAT) is required for communication between the IP phones and Cisco CallManager in the telephony network, DNS is required to ensure proper mapping of NAT translated addresses to network host devices. Likewise, some IP telephony disaster recovery network configurations rely on DNS to ensure proper failover of the network during failure scenarios by mapping hostnames to secondary backup site IP addresses.

If either of these two situations exists and DNS must be configured, you must deploy DNS servers in a redundant fashion such that a single DNS server failure will not prevent network communication between IP telephony devices. By providing DNS server redundancy in the event of a single DNS server failure, you ensure that devices relying on DNS to communicate on the network can still receive hostname-to-IP-address mappings from a backup or secondary DNS server.



Note

DNS names within the cluster are resolved only at system initialization (that is, when a server is booted up). As a result, in order for a server within the cluster to resolve a DNS name that has been changed on the DNS server, the Cisco CallManager service must be restarted on all servers within the cluster.

Dynamic Host Configuration Protocol (DHCP)

DHCP is used by hosts on the network to get initial configuration information, including IP address, subnet mask, default gateway, and TFTP server. DHCP eases the administrative burden of manually configuring each host with an IP address and other configuration information. DHCP also provides automatic reconfiguration of network configuration when devices are moved between subnets. The configuration information is provided by a DHCP server located in the network, which responds to DHCP requests from DHCP-capable clients.

You should configure IP telephony endpoints to use DHCP to simplify deployment of these devices. Any RFC 2131 compliant DHCP server can be used to provide configuration information to IP telephony network devices. When deploying IP telephony devices in an existing data-only network, all you have to do is add DHCP voice scopes to an existing DHCP server for these new voice devices. Because IP telephony devices are configured to use and rely on a DHCP server for IP configuration information, you must deploy DHCP servers in a redundant fashion. At least two DHCP servers should be deployed within the telephony network such that, if one of the servers fails, the other can continue to answer DHCP client requests. You should also ensure that DHCP server(s) are configured with enough IP subnet addresses to handle all DHCP-reliant clients within the network.

DHCP Option 150

IP telephony endpoints can be configured to rely on DHCP Option 150 to identify the source of telephony configuration information, available from a server running the Trivial File Transfer Protocol (TFTP).

In the simplest configuration, where a single TFTP server is offering service to all deployed endpoints, Option 150 is delivered as a single IP address pointing to the system's designated TFTP server. The DHCP scope can also deliver two IP addresses under Option 150, for deployments where there are two TFTP servers within the same cluster: the phone would use the second address if it fails to contact the primary TFTP server, thus providing redundancy. To achieve both redundancy and load sharing between the TFTP servers, you can configure Option 150 to provide the two TFTP server addresses in reverse order for half of the DHCP scopes.

**Note**

If the primary TFTP server is available but is not able to grant the requested file to the phone (for example, because the requesting phone is not configured on that cluster), the phone will not attempt to contact the secondary TFTP server.

Cisco highly recommends using a direct IP address (that is, not relying on a DNS service) for Option 150 because doing so eliminates dependencies on DNS service availability during the phone boot-up and registration process.

**Note**

Even though IP phones support a maximum of two TFTP servers under Option 150, you could configure a cluster with more than two TFTP servers. For instance, if a Cisco CallManager system is clustered over a WAN at three separate sites, three TFTP servers could be deployed (one at each site). Phones within each site could then be granted a DHCP scope containing that site's TFTP server within Option 150. This configuration would bring the TFTP service closer to the endpoints, thus reducing latency and ensuring failure isolation between the sites (one site's failure not affecting TFTP service at another site). However, when the configuration file is modified, the publisher must send a new copy of the database to each TFTP server in the cluster. This propagation of the database consumes publisher CPU resources and can slow performance if more than two TFTP servers are deployed in the cluster.

DHCP, Standalone versus Coresident Server

Typically, DHCP should be configured to run on a standalone server because a large number of devices requesting DHCP configuration could cause high CPU and memory usage, which can impact server performance. Therefore, you should not run DHCP on a Cisco CallManager server. In some instances, given a small Cisco CallManager deployment with no more than 1000 devices registering to the cluster, you may run DHCP on a Cisco CallManager server; however, if the server experiences high CPU load, you should move DHCP to a standalone server. If more than 1000 devices are registered to the cluster, DHCP must be run on a standalone server.

DHCP Lease Times

Configure DHCP lease times as appropriate for the network environment. Given a fairly static network in which PCs and telephony devices remain in the same place for long periods of time, Cisco recommends longer DHCP lease times (for example, one week). Shorter lease times require more frequent renewal of the DHCP configuration and increase the amount of DHCP traffic on the network. Conversely, networks that incorporate large numbers of mobile devices, such as laptops and wireless telephony devices, should be configured with shorter DHCP lease times (for example, one day) to prevent depletion of DHCP-managed subnet addresses. Mobile devices typically use IP addresses for

short increments of time and then might not request a DHCP renewal or new address for a long period of time. Longer lease times will tie up these IP addresses and prevent them from being reassigned even when they are no longer being used.

Cisco IP Phones adhere to the conditions of the DHCP lease duration as specified in the DHCP server's scope configuration. Once half the lease time has expired since the last successful DHCP server Acknowledgment, the IP phone will request a lease renewal. This DHCP client Request, once acknowledged by the DHCP server, will allow the IP phone to retain use of the IP scope (that is, the IP address, default gateway, subnet mask, DNS server (optional), and TFTP server (optional)) for another lease period. If the DHCP server becomes unavailable, an IP phone will not be able to renew its DHCP lease, and as soon as the lease expires, it will relinquish its IP configuration and will thus become unregistered from Cisco CallManager until a DHCP server can grant it another valid scope.

In centralized call processing deployments, if a remote site is configured to use a centralized DHCP server (through the use of a DHCP relay agent such as the IP Helper Address in Cisco IOS) and if connectivity to the central site is severed, IP phones within the branch will not be able to renew their DHCP scope leases. In this situation, branch IP phones are at risk of seeing their DHCP lease expire, thus losing the use of their IP address, which would lead to service interruption. Given the fact that phones attempt to renew their leases at half the lease time, DHCP lease expiration can occur as soon as half the lease time since the DHCP server became unreachable. For example, if the lease time of a DHCP scope is set to 4 days and a WAN failure causes the DHCP server to be unavailable to the phones in a branch, those phones will be unable to renew their leases at half the lease time (in this case, 2 days). The IP phones could stop functioning as early as 2 days after the WAN failure, unless the WAN comes back up and the DHCP server is available before that time. If the WAN connectivity failure persists, all phones see their DHCP scope expire after a maximum of 4 days from the WAN failure.

This situation can be mitigated by one of the following methods:

- Set the DHCP scope lease to a long duration (for example, 8 days or more).

This method would give the system administrator a minimum of half the lease time to remedy any DHCP reachability problem. Long lease durations also have the effect of reducing the frequency of network traffic associated with lease renewals.

- Configure co-located DHCP server functionality (for example, run a DHCP server function on the branch's Cisco IOS router).

This approach is immune to WAN connectivity interruption. One effect of such an approach is to decentralize the management of IP addresses, requiring incremental configuration efforts in each branch. (See [DHCP Network Deployments, page 3-13](#), for more information.)

DHCP Network Deployments

There are two options for deploying DHCP functionality within an IP telephony network

- Centralized DHCP Server

Typically, for a single-site campus IP telephony deployment, the DHCP server should be installed at a central location within the campus. As mentioned previously, redundant DHCP servers should be deployed. If the IP telephony deployment also incorporates remote branch telephony sites, as in a centralized multi-site Cisco CallManager deployment, a centralized server can be used to provide DHCP service to devices in the remote sites. This type of deployment requires that you configure the **ip helper-address** on the branch router interface. Keep in mind that, if redundant DHCP servers are deployed at the central site, both servers' IP addresses must be configured as **ip helper-address**. Also note that, if branch-side telephony devices rely on a centralized DHCP server and the WAN link between the two sites fails, devices at the branch site will be unable to send DHCP requests or receive DHCP responses.

**Note**

By default, **service dhcp** is enabled on the Cisco IOS device and does not appear in the configuration. Do not disable this service on the branch router because doing so will disable the DHCP relay agent on the device, and the **ip helper-address** configuration command will not work.

- Centralized DHCP Server and Remote Site Cisco IOS DHCP Server

When configuring DHCP for use in a centralized multi-site Cisco CallManager deployment, you can use a centralized DHCP server to provide DHCP service to centrally located devices. Remote devices could receive DHCP service from a locally installed server or from the Cisco IOS router at the remote site. This type of deployment ensures that DHCP services are available to remote telephony devices even during WAN failures. [Example 3-4](#) lists the basic Cisco IOS DHCP server configuration commands.

Example 3-4 Cisco IOS DHCP Server Configuration Commands

```

service dhcp                                ! Activate DHCP Service on the IOS Device

ip dhcp excluded-address <ip-address>|<ip-address-low> <ip-address-high>
                                           ! Specify an IP Address or IP Address Range to
                                           ! be excluded from the DHCP pool

ip dhcp pool <dhcp-pool name>              ! Specify DHCP pool name
  network <ip-subnet> <mask>                ! Specify DHCP pool IP subnet and mask
  default-router <default-gateway-ip>      ! Specify the Default-gateway
  option 150 ip <tftp-server-ip-1> ...     ! Specify TFTP servers (up to four) -
                                           ! IP phones use only the first two addresses in
                                           ! the array.

```

Trivial File Transfer Protocol (TFTP)

Within a Cisco CallManager system, endpoints (such as IP phones running the SCCP protocol) rely on a TFTP-based process to acquire configuration information. The endpoints request a configuration file whose name is based on the requester's MAC address (for example, for an IP phone with MAC address ABCDEF123456, the file name would be SEPABCDEF123456.cnf.xml). The configuration file includes the version of software that the phone must run and a list of Cisco CallManager servers that the phone should register with.

If the configuration file instructs the phone to run a software file other than the one it currently uses, the phone will request the new version of software from the TFTP server. The phone goes through this process once per software upgrade.

Centralized call processing deployments require remote phones to download configuration files and phone software through the branch's WAN link. When scheduled maintenance involves the downloading of new software, download times are a function of the number of phones requiring upgrades, the file size, and the WAN link's bandwidth and traffic utilization.

For example, in Cisco CallManager Release 3.3(3)SR1, the size of a phone configuration file is approximately 2700 bytes, and the default software load (P00305000101.bin) for a Cisco IP Phone 7960 is 396,804 bytes. If a branch has a WAN bandwidth of 256 Kbps available to download the software, a single phone would require about 12 seconds to download new software during an upgrade. If that same branch has 10 phones requiring the new software, the download process would take about 2 minutes.

Multi-Cluster Campus TFTP Services

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 answer the file transfer requests made by each phone, regardless of which cluster contains the phone. Therefore, this centralized TFTP server must have access to files created and managed by other clusters.

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

With Cisco CallManager Release 3.2 and later, Cisco TFTP servers cache the IP phone configuration files in RAM by default. For those files to be written to a centralized TFTP server, you must disable (turn off) file caching by setting the following service parameters as indicated 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 of a different cluster), it will search for that file in a list of alternate file locations. The centralized TFTP server must be configured to search through the subdirectories associated with the other clusters.

Example 3-5 Alternate TFTP File Locations

A large campus system is deployed using three clusters, and each cluster contains a TFTP server. TFTP1, the TFTP server for Cluster1, is configured as the centralized TFTP server for the campus. The other clusters and TFTP servers are named in sequence as TFTP2 for Cluster2 and TFTP3 for Cluster3. In all subnets, the DHCP scope provides TFTP1's IP address as Option 150.

First, TFTP2 and TFTP3 are configured to write their configuration files to TFTP1's drive, each in a different subdirectory, as follows:

- TFTP2's alternate file location is set to: \\TFTP1_IP\Program Files\Cisco\TFTPpath\TFTP2
- TFTP3's alternate file location is set to: \\TFTP1_IP\Program Files\Cisco\TFTPpath\TFTP3

Second, TFTP1 is configured to search in the alternate file locations as follows:

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



Note In this example, TFTP1_IP represents the IP address of TFTP1. Also, TFTP1 requires that Windows NT subdirectories be created manually for TFTP2 and TFTP3.

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.

TFTP Server Redundancy

Option 150 allows up to two IP addresses to be returned to phones as part of the DHCP scope. The phone tries the first address in the list, and it tries the 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.

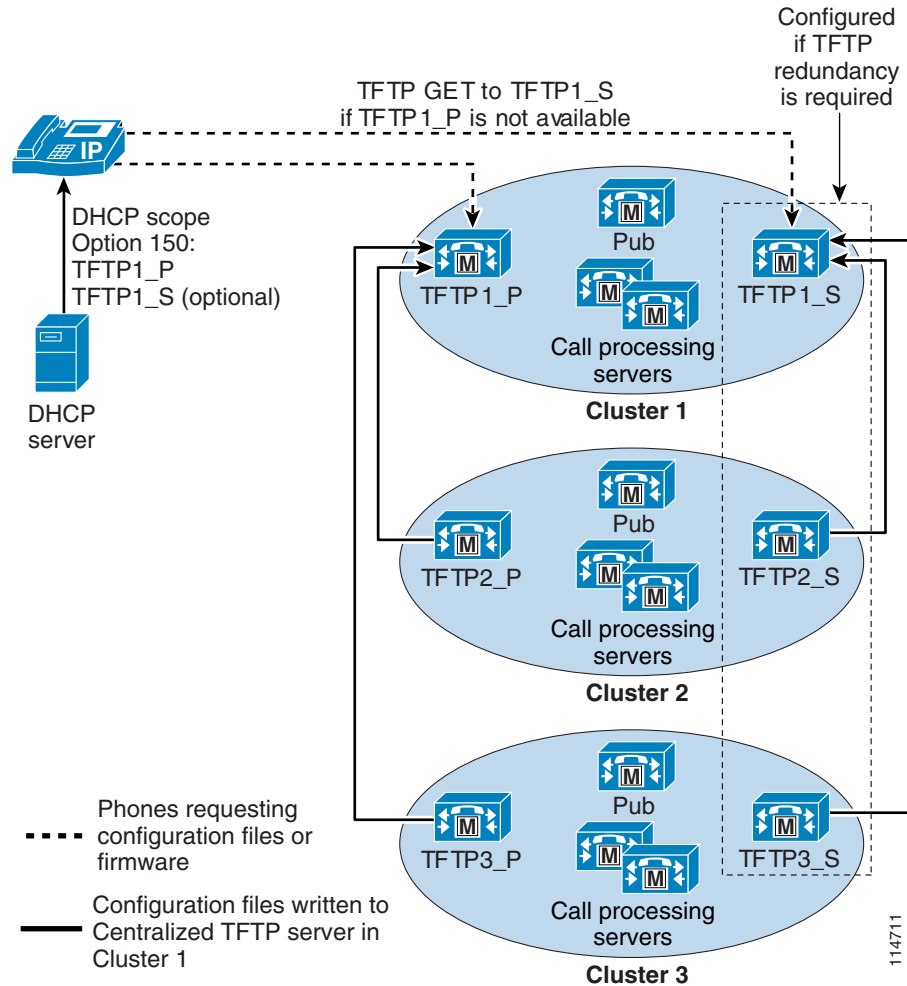
As illustrated in [Figure 3-4](#), two TFTP servers can be configured in a cluster, and each can create and manage separate lists of the same configuration files. In a multi-cluster deployment, each cluster can be configured with two TFTP servers, primary and secondary. The primary TFTP servers can be configured to write files to a centralized primary TFTP server; likewise, the secondary TFTP servers can be configured to write files to a centralized secondary TFTP server. This creates two separate groups (primary and secondary) of TFTP servers configured to ensure redundancy, each with a member serving as the centralized server.

Example 3-6 TFTP Server Redundancy

If we wanted to provide TFTP redundancy for the case described in [Example 3-5](#), 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_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.

Both TFTP1_P and TFTP1_S must be configured as in [Example 3-5](#) to search through the list of alternate file locations.

Figure 3-4 TFTP Server Redundancy with Centralized TFTP Servers for All Clusters**TFTP Load Sharing**

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: TFTP1_P, TFTP1_S
- In subnet 10.1.2.0/24: Option 150: TFTP1_S, TFTP1_P

Under normal operations, a phone in subnet 10.1.1.0/24 will request TFTP services from TFTP1_P, while a phone in subnet 10.1.2.0/24 will request TFTP services from TFTP1_S. If TFTP1_P fails, then phones from both subnets will request TFTP services from TFTP1_S.

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.

Network Time Protocol (NTP)

NTP allows network devices to synchronize their clocks to a network time server or network-capable clock. NTP is critical for ensuring that all devices in a network have the same time. When troubleshooting or managing a telephony network, it is crucial to synchronize the time stamps within all error and security logs, traces, and system reports on devices throughout the network. This synchronization enables administrators to recreate network activities and behaviors based on a common timeline. Billing records and call detail records (CDRs) also require accurate synchronized time.

Cisco CallManager NTP Time Synchronization

Time synchronization is especially critical on Cisco CallManager servers. You should configure automatic NTP time synchronization on all Cisco CallManager servers within the network by performing the following steps:

-
- Step 1** Configure the NTP.conf file.
- The NTP.conf file is located in the C:\WINNT\ directory. The file must be configured with a list of one or more NTP Time servers that can be queried for the time. The file can also be configured to receive NTP Time server updates via NTP broadcasts on the local LAN segment. There must be a broadcast-capable NTP Time server available for Cisco CallManager to receive the time via broadcast messages. [Example 3-7](#) illustrates both methods of configuring the NTP.conf file.
- Step 2** Configure the NTP Service to start automatically at boot-up.
- Under the Services application in Microsoft Windows on each Cisco CallManager server, configure the NTP Service to start automatically at system boot-up.
- Step 3** Use the Services application to start or restart the NTP Service on each Cisco CallManager server.
-

**Note**

If the NTP Service does not appear in the Microsoft Services control panel application on the Cisco CallManager server, install it manually by running C:\Program Files\Cisco\Xntp>**install.bat**.

Example 3-7 NTP.conf Configuration Files

```
server 64.100.21.254
server 64.200.40.10
driftfile %windir%\ntp.drift
```

Or

```
broadcastclient
driftfile %windir%\ntp.drift
```

The driftfile referenced in [Example 3-7](#) is automatically updated via the NTP Service based on information in the NTP messages received from the NTP Time server.

**Note**

By default, NTP updates will not occur if the clock offset or the adjustment between the NTP server and the NTP client is larger than 1000 seconds. To adjust this default behavior, you can add the **tinker panic** *<number of seconds>* command to the NTP.conf file, where the number of seconds determines the amount of slip time that can occur. Setting this value to 0 disables the panic feature, and a clock offset of any value will be accepted.

Cisco IOS and CatOS NTP Time Synchronization

Time synchronization is also important for other devices within the network. Cisco IOS routers and Catalyst switches should be configured to synchronize their time with the rest of the network devices via NTP. This is critical for ensuring that debug, syslog, and console log messages are time-stamped appropriately. Troubleshooting telephony network issues is simplified when a clear timeline can be drawn for events that occur on devices throughout the network.

[Example 3-8](#) illustrates the configuration of NTP time synchronization on Cisco IOS and CatOS devices.

Example 3-8 Cisco IOS and CatOS NTP Configuration

Cisco IOS configuration:

```
ntp server 64.100.21.254
```

CatOS configuration:

```
set ntp server 64.100.21.254
set ntp client enable
```

To ensure proper NTP time synchronization on routers and switches, it may be necessary to configure time zones using the **clock timezone** command (in Cisco IOS) and/or **set timezone** command (in CatOS).

For more information on how to configure NTP for Cisco CallManager, refer to the document *How To Configure Time Synchronization for Cisco CallManager*, available at

<http://www.cisco.com>

Power over Ethernet (PoE)

PoE (or inline power) is 48 Volt DC power provided over standard Ethernet unshielded twisted-pair (UTP) cable. Instead of using wall power, IP phones and other inline powered devices (PDs) such as the Aironet Wireless Access Points can receive power provided by inline power-capable Catalyst Ethernet switches or other inline power source equipment (PSE). Inline power is enabled by default on all inline power-capable Catalyst switches.

Deploying inline power-capable switches with uninterrupted power supplies (UPS) ensures that IP phones continue to receive power during power failure situations. Provided the rest of the telephony network is available during these periods of power failure, then IP phones should be able to continue making and receiving calls. You should deploy inline power-capable switches at the campus access layer within wiring closets to provide inline-powered Ethernet ports for IP phones, thus eliminating the need for wall power.

Cisco PoE is delivered on the same wire pairs used for data connectivity (pins 1, 2, 3, and 6). If existing access switch ports are not capable of inline power, you can use a power patch panel to inject power into the cabling. (In this case pins 4, 5, 7, and 8 are used.) Additionally, power injectors may be used for specific deployment needs.



Caution

The use of power injectors or power patch panels can damage some devices because power is always applied to the Ethernet pairs. PoE switch ports automatically detect the presence of a device that requires PoE before enabling it on a port-by-port basis.

In addition to Cisco PoE inline power, Cisco now supports the IEEE 802.3af PoE standard. Currently, only some access switches and phones comply with 802.3af. Over time, all phones and switches will support 802.3af PoE. The Catalyst 6500, 4500, and 3750 are currently capable of supporting 802.3af. For information about which Cisco IP Phones support the 802.3af PoE standard, see the [Endpoint Features Summary, page 16-30](#).

Category 3 Cabling

The use of Category 3 cabling is supported for IP Communications under the following conditions:

- Phones with a PC port and a PC attached to it (Cisco 7970, 7960, 7940, and 7910+SW IP Phones) should be set to 10 Mb, full-duplex.

This setting requires hard-coding the upstream switch port, the phone switch and PC ports, and the PC NIC port to 10 Mb, full-duplex. No ports should be set to AUTO negotiate. If desired, you can hard-code the phone's PC port to 10 Mb half-duplex, thereby forcing the PC's NIC to negotiate to 10 Mb half-duplex (assuming the PC's NIC is configured to AUTO negotiate). This configuration is acceptable as long as the uplink between the phone and the upstream switch port is set to 10 Mb full-duplex.

- Phones with no PC ports and with 10 Mb switch ports (Cisco 7902, 7905, and 7910 IP Phones) should be allowed to auto-negotiate to 10 Mb, half-duplex.

Because these phones support only 10 Mb Ethernet and their ports cannot be manually configured, the upstream switch port should be set to either AUTO negotiate or 10 Mb, half-duplex. In both cases, these phones will negotiate to 10 Mb, half-duplex.

- Phones with a PC port but no PC attached to it (Cisco 7970, 7960, 7940, 7910+SW, and 7912 IP Phones) can be allowed to negotiate to 10 Mb, half-duplex.

If you leave these phones with the default switch port configuration of AUTO negotiate and configure the upstream switch port to 10 Mb, half-duplex, these phones will revert to 10Mb, half-duplex.



Note

The Cisco 7912 IP Phone should not be used with Category 3 cable when a PC is attached because the switch and PC ports on this phone cannot be forced to 10 Mb, full duplex.

IBM Type 1A and 2A Cabling

The use of IBM Cabling System (ICS) or Token Ring shielded twisted-pair type 1A or 2A cabling is supported for IP Communications under the following conditions:

- Cable lengths should be 100 meters or less.
- Adapters without impedance matching should be used for converting from universal data connector (UDC) to RJ-45 Ethernet standard.



Note

There are only two twisted pairs in the Token Ring cables. Therefore, inline power for IP phones can be supported, but mid-span power insertion cannot (with Cisco Inline Power and 802.3af) because it requires more than two pairs.

Running data over the network is not always a sufficient test of the quality of the cable plant because some non-compliance issues might not be apparent. Therefore, customers might want to perform a cable plant survey to verify that their type 1A and 2A cabling installation is compliant with Ethernet standards.

For more information about the use of IBM cabling, refer to the Product Bulletin *Shielded Twisted-Pair Cabling Support for Cisco Fast Ethernet Products*, available at

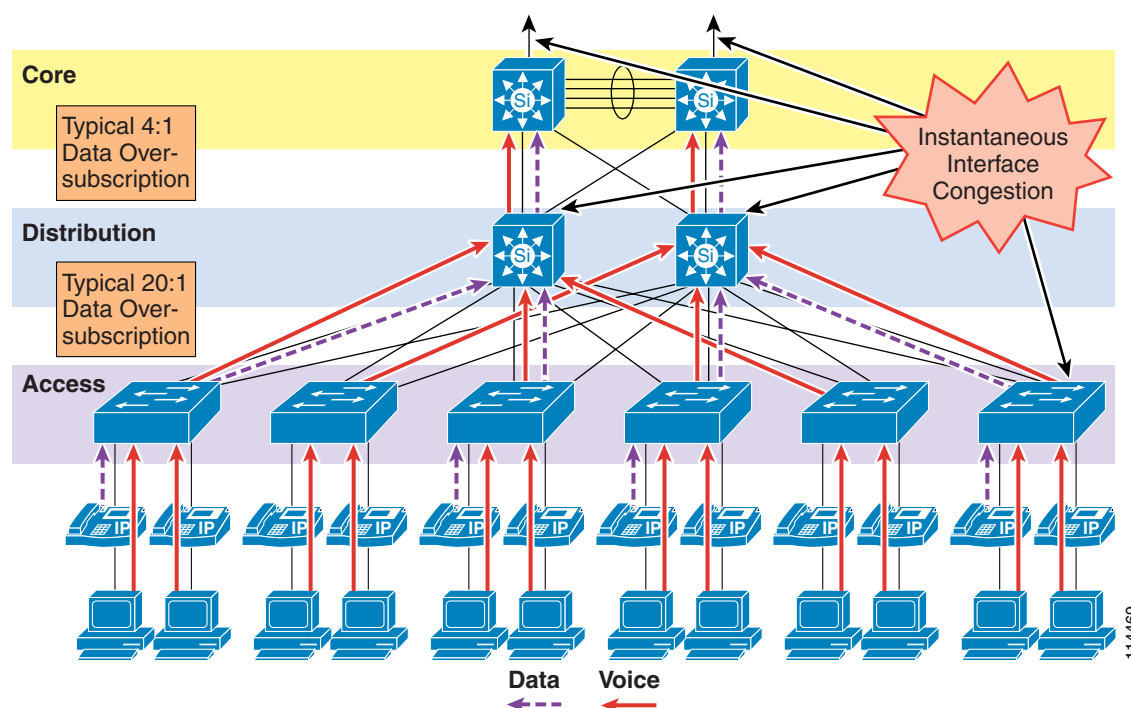
<http://www.cisco.com>

LAN Quality of Service (QoS)

Until recently, quality of service was not an issue in the enterprise campus due to the asynchronous nature of data traffic and the ability of network devices to tolerate buffer overflow and packet loss. However, with new applications such as voice and video, which are sensitive to packet loss and delay, buffers and not bandwidth are the key QoS issue in the enterprise campus.

Figure 3-5 illustrates the typical oversubscription that occurs in LAN infrastructures.

Figure 3-5 Data Traffic Oversubscription in the LAN



This oversubscription, coupled with individual traffic volumes and the cumulative effects of multiple independent traffic sources, can result in the egress interface buffers becoming full instantaneously, thus causing additional packets to drop when they attempt to enter the egress buffer. The fact that campus switches use hardware-based buffers, which compared to the interface speed are much smaller than those found on WAN interfaces in routers, merely increases the potential for even short-lived traffic bursts to cause buffer overflow and dropped packets.

Applications such as file sharing (both peer-to-peer and server-based), remote networked storage, network-based backup software, and emails with large attachments, can create conditions where network congestion occurs more frequently and/or for longer durations. Some of the negative effects of recent worm attacks have been an overwhelming volume of network traffic (both unicast and broadcast-storm based), increasing network congestion. If no buffer management policy is in place, loss, delay, and jitter performance of the LAN may be affected for all traffic.

Another situation to consider is the effect of failures of redundant network elements, which cause topology changes. For example, if a distribution switch fails, all traffic flows will be reestablished through the remaining distribution switch. Prior to the failure, the load balancing design shared the load between two switches, but after the failure all flows are concentrated in a single switch, potentially causing egress buffer conditions that normally would not be present.

For applications such as voice, this packet loss and delay results in severe voice quality degradation. Therefore, QoS tools are required to manage these buffers and to minimize packet loss, delay, and delay variation (jitter).

The following types of QoS tools are needed from end to end on the network to manage traffic and ensure voice quality:

- Traffic classification

Classification involves the marking of packets with a specific priority denoting a requirement for class of service (CoS) from the network. The point at which these packet markings are trusted or not trusted is considered the trust boundary. Trust is typically extended to voice devices (phones) and not to data devices (PCs).

- Queuing or scheduling

Interface queuing or scheduling involves assigning packets to one of several queues based on classification for expedited treatment throughout the network.

- Bandwidth provisioning

Provisioning involves accurately calculating the required bandwidth for all applications plus element overhead.

The following sections discuss the use of these QoS mechanisms in a campus environment:

- [Traffic Classification, page 3-22](#)
- [Interface Queuing, page 3-23](#)
- [Bandwidth Provisioning, page 3-23](#)
- [Impairments to IP Communications if QoS is Not Employed, page 3-23](#)

Traffic Classification

It has always been an integral part of the Cisco network design architecture to classify or mark traffic as close to the edge of the network as possible. Traffic classification is an entrance criterion for access into the various queuing schemes used within the campus switches and WAN interfaces. The IP phone marks its voice control signaling and voice RTP streams at the source, and it adheres to the values presented in [Table 3-2](#). As such, the IP phone can and should classify traffic flows.

[Table 3-2](#) lists the traffic classification requirements for the LAN infrastructure.

Table 3-2 Traffic Classification Guidelines for Various Types of Network Traffic

Traffic Type	Layer 2 Class of Service (CoS)	Layer 3 IP Precedence	Layer 3 Differentiated Services Code Point (DSCP)	Layer 3 Per-Hop Behavior (PHB)
Voice Real-Time Transport Protocol (RTP)	5	5	46	EF
Voice control signaling ¹	3	3	24	CS3
Video conferencing	4	4	34	AF41
Data	0, 1, 2	0, 1, 2	10 to 22	BE to AF23

1. The recommended DSCP/PHB marking for voice control signaling traffic has been changed from 26/AF31 to 24/CS3. A marking migration is planned within Cisco to reflect this change, however many products still mark signaling traffic as 26/AF31. Therefore, in the interim, Cisco recommends that both AF31 and CS3 be reserved for call signaling.

Interface Queuing

After packets have been marked with the appropriate tag at Layer 2 (CoS) and Layer 3 (DSCP or PHB), it is important to configure the network to schedule or queue traffic based on this classification, so as to provide each class of traffic with the service it needs from the network. By enabling QoS on campus switches, you can configure all voice traffic to use separate queues, thus virtually eliminating the possibility of dropped voice packets when an interface buffer fills instantaneously.

Although network management tools may show that the campus network is not congested, QoS tools are still required to guarantee voice quality. Network management tools show only the average congestion over a sample time span. While useful, this average does not show the congestion peaks on a campus interface.

Transmit interface buffers within a campus tend to congest in small, finite intervals as a result of the bursty nature of network traffic. When this congestion occurs, any packets destined for that transmit interface are dropped. The only way to prevent dropped voice traffic is to configure multiple queues on campus switches. For this reason, Cisco recommends always using a switch that has at least two output queues on each port and the ability to send packets to these queues based on QoS Layer 2 and/or Layer 3 classification. Cisco Catalyst 6000, 4000, 3750, 35XX, and 2950 switches all support two or more output queues per port.

Bandwidth Provisioning

In the campus LAN, bandwidth provisioning recommendations can be summarized by the motto, *Over provision and under subscribe*. This motto implies careful planning of the LAN infrastructure so that the available bandwidth is always considerably higher than the load and there is no steady-state congestion over the LAN links.

The addition of voice traffic onto a converged network does not represent a significant increase in overall network traffic load; the bandwidth provisioning is still driven by the demands of the data traffic requirements. The design goal is to avoid extensive data traffic congestion on any link that will be traversed by telephony signaling or media flows. Contrasting the bandwidth requirements of a single G.711 voice call (approximately 86 kbps) to the raw bandwidth of a FastEthernet link (100 Mbps) indicates that voice is not a source of traffic that causes network congestion in the LAN, but rather it is a traffic flow to be protected from LAN network congestion.

Impairments to IP Communications if QoS is Not Employed

If QoS is not deployed, packet drops and excessive delay and jitter can occur, leading to impairments of the telephony services. When media packets are subjected to drops, delay, and jitter, the user-perceivable effects include clicking sound, harsh-sounding voice, extended periods of silence, and echo.

When signaling packets are subjected to the same conditions, user-perceivable impairments include unresponsiveness to user input (such as delay to dial tone), continued ringing upon answer, and double dialing of digits due to the user's belief that the first attempt was not effective (thus requiring hang-up and redial). More extreme cases can include endpoint re-initialization, call termination, and the spurious activation of SRST functionality at branch offices (leading to interruption of gateway calls).

These effects apply to all deployment models. However, single-site (campus) deployments tend to be less likely to experience the conditions caused by sustained link interruptions because the larger quantity of bandwidth typically deployed in LAN environments (minimum links of 100 Mbps) allows for some residual bandwidth to be available for the IP Communications system.

In any WAN-based deployment model, traffic congestion is more likely to produce sustained and/or more frequent link interruptions because the available bandwidth is much less than in a LAN (typically less than 2 Mbps), so the link is more easily saturated. The effects of link interruptions impact the users, whether or not the voice media traverses the packet network.

WAN Infrastructure

Proper WAN infrastructure design is also extremely important for proper IP telephony operation on a converged network. Proper infrastructure design requires following basic configuration and design best practices for deploying a WAN that is as highly available as possible and that provides guaranteed throughput. Furthermore, proper WAN infrastructure design requires deploying end-to-end QoS on all WAN links. The following sections discuss these requirements:

- [WAN Design and Configuration Best Practices, page 3-24](#)
- [WAN Quality of Service \(QoS\), page 3-26](#)

WAN Design and Configuration Best Practices

Properly designing a WAN requires building fault-tolerant network links and planning for the possibility that these links might become unavailable. By carefully choosing WAN topologies, provisioning the required bandwidth, and approaching the WAN infrastructure as another layer in the network topology, you can build a fault-tolerant and redundant network. The following sections examine the required infrastructure layers and network services:

- [Deployment Considerations, page 3-24](#)
- [Guaranteed Bandwidth, page 3-25](#)
- [Best-Effort Bandwidth, page 3-26](#)

Deployment Considerations

WAN deployments for voice networks must follow a hub-and-spoke topology, with a central hub site and multiple remote spoke sites connected into the central hub site. In this scenario, each remote or spoke site is one WAN link hop away from the central or hub site and two WAN link hops away from all other spoke sites. This topology ensures that call admission control, provided by Cisco CallManager or a gatekeeper at the central site, works properly in keeping track of the bandwidth available between any two sites in the WAN. In addition, multiple hub-and-spoke deployments can be interconnected via WAN links.

See the chapter on [IP Telephony Deployment Models, page 2-1](#), for more information about centralized and distributed multi-site deployment models as well as Multiprotocol Label Switching (MPLS) implications for these deployment models.

WAN links should, when possible, be made redundant to provide higher levels of fault-tolerance. Redundant WAN links provided by different service providers or located in different physical ingress/egress points within the network can ensure backup bandwidth and connectivity in the event that

a single link fails. In non-failure scenarios, these redundant links may be used to provide additional bandwidth and offer load balancing of traffic on a per-flow basis over multiple paths and equipment within the WAN.

Voice and data should remain converged at the WAN, just as they are converged at the LAN. QoS provisioning and queuing mechanisms are typically available in a WAN environment to ensure that voice and data can interoperate on the same WAN links. Attempts to separate and forward voice and data over different links can be problematic in many instances because the failure of one link typically forces all traffic over a single link, thus diminishing throughput for each type of traffic and in most cases reducing the quality of voice. Furthermore, maintaining separate network links or devices makes troubleshooting and management difficult at best.

Because of the potential for WAN links to fail or to become oversubscribed, Cisco recommends deploying non-centralized resources as appropriate at sites on the other side of the WAN. Specifically, media resources, DHCP servers, voice gateways, and call processing applications such as Survivable Remote Site Telephony (SRST) and Cisco CallManager Express (CME) should be deployed at non-central sites when and if appropriate, depending on the site size and how critical these functions are to that site. Keep in mind that de-centralizing voice applications and devices can increase the complexity of network deployments, the complexity of managing these resources throughout the enterprise, and the overall cost of a the network solution; however, these factors can be mitigated by the fact that the resources will be available during a WAN link failure.

When deploying voice in a WAN environment, Cisco recommends that you use the lower-bandwidth G.729 codec for any voice calls that will traverse WAN links because this practice will provide bandwidth savings on these lower-speed links. Furthermore, media resources such as MoH should be configured to use multicast transport mechanism when possible because this practice will provide additional bandwidth savings.

Finally, recommendation G.114 of the International Telecommunication Union (ITU) states that the one-way delay in a voice network should be less than or equal to 150 milliseconds. It is important to keep this in mind when implementing low-speed WAN links within a network. Topologies, technologies, and physical distance should be considered for WAN links so that one-way delay is kept at or below this 150-millisecond recommendation.

Guaranteed Bandwidth

Because voice is typically deemed a critical network application, it is imperative that bearer and signaling voice traffic always reaches its destination. For this reason, it is important to choose a WAN topology and link type that can provide guaranteed dedicated bandwidth. The following WAN link technologies can provide guaranteed dedicated bandwidth:

- Leased Lines
- Frame Relay
- Asynchronous Transfer Mode (ATM)
- ATM/Frame-Relay Service Interworking
- Multiprotocol Label Switching (MPLS)
- Cisco Voice and Video Enabled IP Security VPN (IPSec V3PN)

These link technologies, when deployed in a dedicated fashion or when deployed in a private network, can provide guaranteed traffic throughput. All of these WAN link technologies can be provisioned at specific speeds or bandwidth sizes. In addition, these link technologies have built-in mechanisms that help guarantee throughput of network traffic even at low link speeds. Features such as traffic shaping,

fragmentation and packet interleaving, and committed information rates (CIR) can help ensure that packets are not dropped in the WAN, that all packets are given access at regular intervals to the WAN link, and that enough bandwidth is available for all network traffic attempting to traverse these links.

Best-Effort Bandwidth

There are some WAN topologies that are unable to provide guaranteed dedicated bandwidth to ensure that network traffic will reach its destination, even when that traffic is critical. These topologies are extremely problematic for voice traffic, not only because they provide no mechanisms to provision guaranteed network throughput, but also because they provide no traffic shaping, packet fragmentation and interleaving, queuing mechanisms, or end-to-end QoS to ensure that critical traffic such as voice will be given preferential treatment.

The following WAN network topologies and link types are examples of this kind of best-effort bandwidth technology:

- The Internet
- DSL
- Cable
- Satellite
- Wireless

In most cases, none of these link types can provide the guaranteed network connectivity and bandwidth required for critical voice and voice applications. However, these technologies might be suitable for personal or telecommuter-type network deployments. At times, these topologies can provide highly available network connectivity and adequate network throughput; but at other times, these topologies can become unavailable for extended periods of time, can be throttled to speeds that render network throughput unacceptable for real-time applications such as voice, or can cause extensive packet losses and require repeated retransmissions. In other words, these links and topologies are unable to provide guaranteed bandwidth, and when traffic is sent on these links, it is sent best-effort with no guarantee that it will reach its destination. For this reason, Cisco recommends that you do *not* use best-effort WAN topologies for voice-enabled networks that require enterprise-class voice services and quality.



Note

There are some new QoS mechanisms for DSL and cable technologies that can provide guaranteed bandwidth; however, these mechanisms are not typically deployed by service providers, and these services are still significantly oversubscribed.

WAN Quality of Service (QoS)

Before placing voice and video traffic on a network, it is important to ensure that there is adequate bandwidth for all required applications. Once this bandwidth has been provisioned, voice priority queuing must be performed on all interfaces. This queuing is required to reduce jitter and possible packet loss if a burst of traffic oversubscribes a buffer. This queuing requirement is similar to the one for the LAN infrastructure.

Next, the WAN typically requires additional mechanisms such as traffic shaping to ensure that WAN links are not sent more traffic than they can handle, which could cause dropped packets.

Finally, link efficiency techniques can be applied to WAN paths. For example, link fragmentation and interleaving (LFI) can be used to prevent small voice packets from being queued behind large data packets, which could lead to unacceptable delays on low-speed links.

The goal of these QoS mechanisms is to ensure reliable, high-quality voice by reducing delay, packet loss, and jitter for the voice traffic. [Table 3-3](#) lists the QoS features and tools required for the WAN infrastructure to achieve this goal.

Table 3-3 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 • Optional: Voice-Adaptive Traffic Shaping (VATS) • Optional: Voice-Adaptive Fragmentation (VAF)	• Traffic Shaping • LLQ • Optional: VATS
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
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

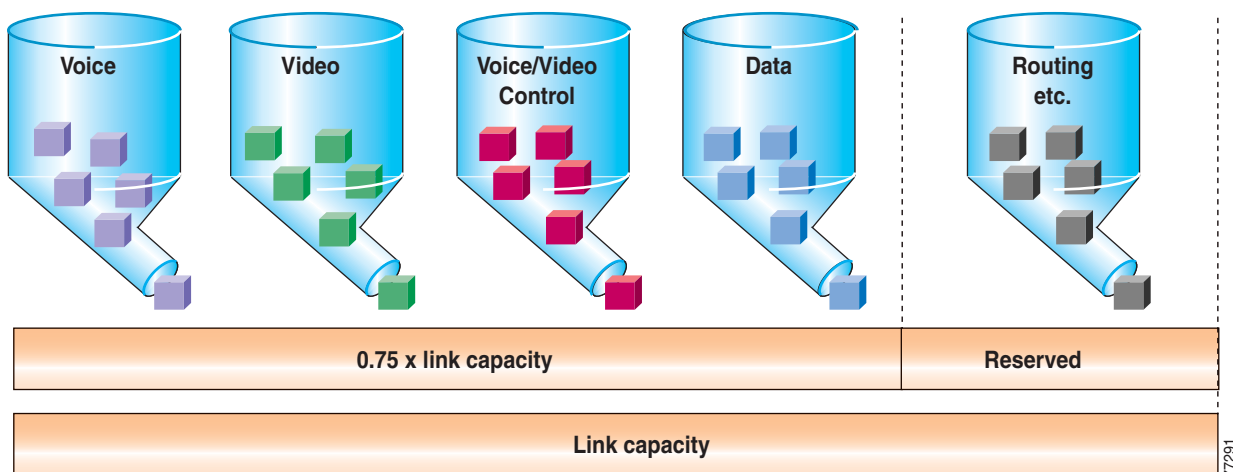
The following sections highlight some of the most important features and techniques to consider when designing a WAN to support both voice and data traffic:

- [Bandwidth Provisioning, page 3-28](#)
- [Traffic Prioritization, page 3-34](#)
- [Link Efficiency Techniques, page 3-36](#)
- [Traffic Shaping, page 3-38](#)

Bandwidth Provisioning

Properly provisioning the network bandwidth is a major component of designing a successful IP network. You can calculate the required bandwidth by adding the bandwidth requirements for each major application (for example, voice, video, and data). This sum then represents the minimum bandwidth requirement for any given link, and it should not exceed approximately 75% of the total available bandwidth for the link. This 75% rule assumes that some bandwidth is required for overhead traffic, such as routing and Layer 2 keepalives. Figure 3-6 illustrates this bandwidth provisioning process.

Figure 3-6 Link Bandwidth Provisioning



In addition to using no more than 75% of the total available bandwidth for data, voice, and video, the total bandwidth configured for all LLQ priority queues should typically not exceed 33% of the total link bandwidth. Provisioning more than 33% of the available bandwidth for the priority queue can be problematic for a number of reasons. First, provisioning more than 33% of the bandwidth for voice can result in increased CPU usage. Because each voice call will send 50 packets per second (with 20 ms samples), provisioning for large numbers of calls in the priority queue can lead to high CPU levels due to high packet rates. In addition, if more than one type of traffic is provisioned in the priority queue (for example, voice and video), this configuration defeats the purpose of enabling QoS because the priority queue essentially becomes a first-in, first-out (FIFO) queue. A larger percentage of reserved priority bandwidth effectively dampens the QoS effects by making more of the link bandwidth FIFO. Finally, allocating more than 33% of the available bandwidth can effectively starve any data queues that are provisioned. Obviously, for very slow links (less than 192 kbps), the recommendation to provision no more than 33% of the link bandwidth for the priority queue(s) might be unrealistic because a single call could require more than 33% of the link bandwidth. In these situations, and in situations where specific business needs cannot be met while holding to this recommendation, it may be necessary to exceed the 33% rule.

From a traffic standpoint, an IP telephony call consists of two parts:

- The voice carrier stream, which consists of Real-Time Transport Protocol (RTP) packets that contain the actual voice samples.
- The call control signaling, which consists of packets belonging to one of several protocols, according to the endpoints involved in the call (for example, H.323, MGCP, SCCP, or (J)TAPI). Call control functions are, for instance, those used to set up, maintain, tear down, or redirect a call.

Bandwidth provisioning should include not only the voice stream traffic but also the call control traffic. In fact, in multi-site WAN deployments, the call control traffic (as well as the voice stream) must traverse the WAN, and failure to allocate sufficient bandwidth for it can adversely affect the user experience.

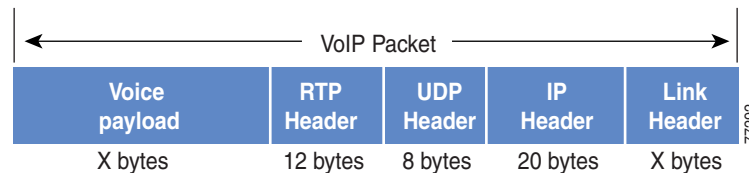
The next three sub-sections describe the bandwidth provisioning recommendations for the following types of traffic:

- Voice bearer traffic in all multi-site WAN deployments (see [Provisioning for Voice Bearer Traffic, page 3-29](#))
- Call control traffic in multi-site WAN deployments with centralized call processing (see [Provisioning for Call Control Traffic with Centralized Call Processing, page 3-31](#))
- Call control traffic in multi-site WAN deployments with distributed call processing (see [Provisioning for Call Control Traffic with Distributed Call Processing, page 3-33](#))

Provisioning for Voice Bearer Traffic

As illustrated in [Figure 3-7](#), a voice-over-IP (VoIP) packet consists of the payload, IP header, User Datagram Protocol (UDP) header, Real-Time Transport Protocol (RTP) header, and Layer 2 Link header. At the default packetization rate of 20 ms, VoIP packets have a 160-byte payload for G.711 or a 20-byte payload for G.729. When Secure Real-Time Transport Protocol (SRTP) encryption is used, the payload for each packet is increased by 4 bytes. At the default packetization rate of 20 ms, SRTP VoIP packets have a 164-byte payload for G.711 or a 24-byte payload for G.729. The IP header is 20 bytes, the UDP header is 8 bytes, and the RTP header is 12 bytes. The link header varies in size according to the Layer 2 media used.

Figure 3-7 Typical VoIP Packet



The bandwidth consumed by VoIP streams is calculated by adding the packet payload and all headers (in bits), then multiplying by the packet rate per second (default of 50 packets per second). [Table 3-4](#) details the bandwidth per VoIP flow at a default packet rate of 50 packets per second (pps). [Table 3-4](#) does not include Layer 2 header overhead and does not take into account any possible compression schemes, such as compressed Real-Time Transport Protocol (cRTP). You can use the Service Parameters menu in Cisco CallManager Administration to adjust the packet rate.

[Table 3-4](#) lists the bandwidth consumed by the voice payload and IP header only, at a default packet rate of 50 packets per second (pps) and at a rate of 33.3 pps for both non-encrypted and encrypted payloads.

Table 3-4 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 (SRTP)	20 ms	164	50.0	81.6 kbps
G.711	30 ms	240	33.3	74.7 kbps

Table 3-4 Bandwidth Consumption for Voice Payload and IP Header Only (continued)

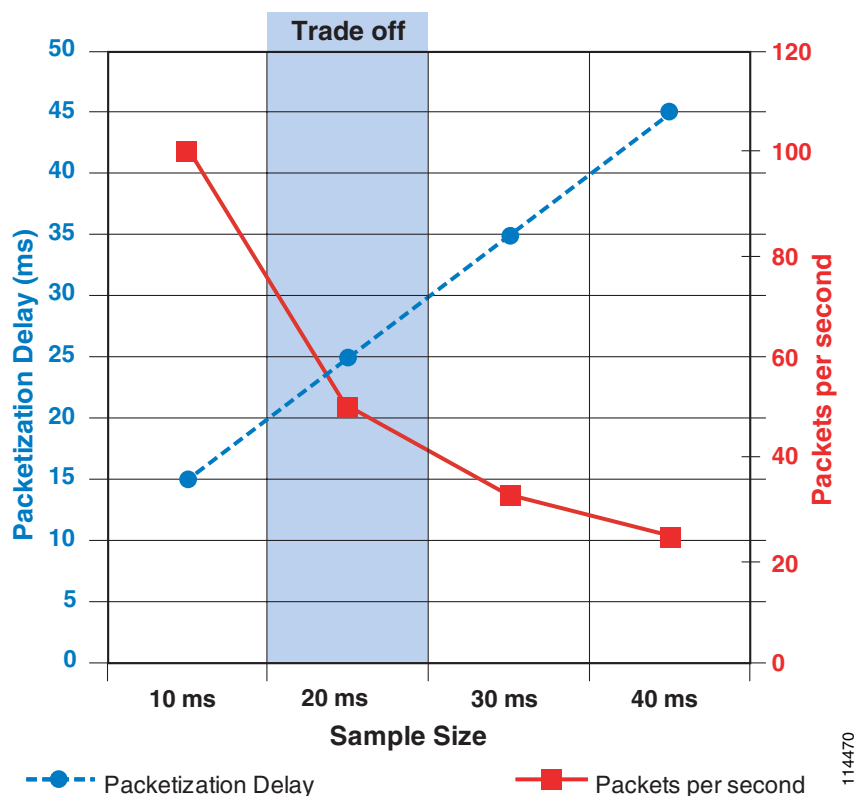
CODEC	Sampling Rate	Voice Payload in Bytes	Packets per Second	Bandwidth per Conversation
G.711 (SRTP)	30 ms	244	33.3	75.8 kbps
G.729A	20 ms	20	50.0	24.0 kbps
G.729A (SRTP)	20 ms	24	50.0	25.6 kbps
G.729A	30 ms	30	33.3	18.7 kbps
G.729A (SRTP)	30 ms	34	33.3	19.8 kbps

A more accurate method for provisioning is to include the Layer 2 headers in the bandwidth calculations. [Table 3-5](#) lists the amount of bandwidth consumed by voice traffic when the Layer 2 headers are included in the calculations.

Table 3-5 Bandwidth Consumption with Layer 2 Headers Included

CODEC	Header Type and Size						
	Ethernet 14 Bytes	PPP 6 Bytes	ATM 53-Byte Cells with a 48-Byte Payload	Frame Relay 4 Bytes	MLPPP 10 Bytes	MPLS 4 Bytes	WLAN 24 Bytes
G.711 at 50.0 pps	85.6 kbps	82.4 kbps	106.0 kbps	81.6 kbps	84.0 kbps	81.6 kbps	89.6 kbps
G.711 (SRTP) at 50.0 pps	87.2 kbps	84.0 kbps	106.0 kbps	83.2 kbps	85.6 kbps	83.2 kbps	N/A
G.711 at 33.3 pps	78.4 kbps	76.3 kbps	84.8 kbps	75.7 kbps	77.3 kbps	75.7 kbps	81.1 kbps
G.711 (SRTP) at 33.3 pps	79.5 kbps	77.4 kbps	84.8 kbps	76.8 kbps	78.4 kbps	76.8 kbps	N/A
G.729A at 50.0 pps	29.6 kbps	26.4 kbps	42.4 kbps	25.6 kbps	28.0 kbps	25.6 kbps	33.6 kbps
G.729A (SRTP) at 50.0 pps	31.2 kbps	28.0 kbps	42.4 kbps	27.2 kbps	29.6 kbps	27.2 kbps	N/A
G.729A at 33.3 pps	22.4 kbps	20.3 kbps	28.3 kbps	19.7 kbps	21.3 kbps	19.7 kbps	25.1 kbps
G.729A (SRTP) at 33.3 pps	23.5 kbps	21.4 kbps	28.3 kbps	20.8 kbps	22.4 kbps	20.8 kbps	N/A

While it is possible to configure the sampling rate above 30 ms, doing so usually results in very poor voice quality. As illustrated in [Figure 3-8](#), as sampling size increases, the number of packets per second decreases, resulting in a smaller impact to the CPU of the device. Likewise, as the sample size increases, IP header overhead is lower because the payload per packet is larger. However, as sample size increases, so does packetization delay, resulting in higher end-to-end delay for voice traffic. The trade-off between packetization delay and packets per second must be considered when configuring sample size. While this trade-off is optimized at 20 ms, 30 ms sample sizes still provide a reasonable ratio of delay to packets per second; however, with 40 ms sample sizes, the packetization delay becomes too high.

Figure 3-8 Voice Sample Size: Packets per Second vs. Packetization Delay

Provisioning for Call Control Traffic with Centralized Call Processing

In a centralized call processing deployment, the Cisco CallManager cluster and the applications (such as voice mail) are located at the central site, while several remote sites are connected through an IP WAN. The remote sites rely on the centralized Cisco CallManagers to handle their call processing.

The following considerations apply to this deployment model:

- Each time a remote branch phone places a call, the control traffic traverses the IP WAN to reach the Cisco CallManager at the central site, even if the call is local to the branch.
- The signaling protocols that may traverse the IP WAN in this deployment model are SCCP (encrypted and non-encrypted), H.323, MGCP, and TAPI. All the control traffic is exchanged between a Cisco CallManager at the central site and endpoints or gateways at the remote branches.

As a consequence, the area where you must provision bandwidth for control traffic lies between the branch routers and the WAN aggregation router at the central site.

The control traffic that traverses the WAN in this scenario can be split into two categories:

- Quiescent traffic, which consists of keep-alive messages periodically exchanged between the branch IP phones and Cisco CallManager, regardless of phone activity.
- Call-related traffic, which consists of signaling messages exchanged between the branch IP phones and/or gateways and the Cisco CallManager at the central site when a call needs to be set up, torn down, forwarded, and so on.

To obtain an estimate of the generated call control traffic, it is therefore necessary to make some assumptions regarding the average number of calls per hour made by each branch IP phone. In the interest of simplicity, the calculations in this section assume an average of 10 calls per hour per phone.

**Note**

If this average number does not satisfy the needs of your specific deployment, you can calculate the recommended bandwidth by using the advanced formulas provided in [Advanced Formulas, page 3-33](#).

Given the assumptions made, and initially considering the case of a remote branch with no signaling encryption configured, the recommended bandwidth needed for call control traffic can be obtained from the following formula:

Equation 1: Recommended Bandwidth Needed for Control Traffic without Signaling Encryption.

$$\text{Bandwidth (bps)} = 265 * (\text{Number of IP phones and gateways in the branch})$$

Equation 1 and all other formulas within this section include a 25% over-provisioning factor. Control traffic has a bursty nature, with peaks of high activity followed by periods of low activity. For this reason, assigning just the minimum bandwidth required to a control traffic queue can result in undesired effects such as buffering delays and, potentially, packet drops during periods of high activity. The default queue depth for a Class-Based Weighted Fair Queuing (CBWFQ) queue in Cisco IOS equals 64 packets. The bandwidth assigned to this queue determines its servicing rate. Assuming that the bandwidth configured is the average bandwidth consumed by this type of traffic, it is clear that, during the periods of high activity, the servicing rate will not be sufficient to "drain" all the incoming packets out of the queue, thus causing them to be buffered. Note that, if the 64-packet limit is reached, any subsequent packets are either assigned to the best-effort queue or are dropped. It is therefore advisable to introduce an over-provisioning factor to absorb and smooth the variations in the traffic pattern and to minimize the risk of a temporary buffer overrun. This is equivalent to increasing the servicing rate of the queue.

If encryption is configured, the recommended bandwidth is affected because encryption increases the size of signaling packets exchanged between Cisco CallManager and the endpoints. The following formula takes into account the impact of signaling encryption:

Equation 2: Recommended Bandwidth Needed for Control Traffic with Signaling Encryption.

$$\text{Bandwidth with signaling encryption (bps)} = 415 * (\text{Number of IP phones and gateways in the branch})$$

If we now take into account the fact that the smallest bandwidth that can be assigned to a queue on a Cisco IOS router is 8 kbps, we can summarize the values of minimum and recommended bandwidth for various branch office sizes, as shown in [Table 3-6](#).

Table 3-6 Recommended Bandwidth for Call Control Traffic With and Without Signaling Encryption

Branch Office Size (Number of IP Phones and Gateways)	Recommended Bandwidth for Control Traffic (no encryption)	Recommended Bandwidth for Control Traffic (with encryption)
1 to 10	8 kbps	8 kbps
20	8 kbps	9 kbps
30	8 kbps	13 kbps
40	11 kbps	17 kbps
50	14 kbps	21 kbps
60	16 kbps	25 kbps
70	19 kbps	29 kbps

Table 3-6 Recommended Bandwidth for Call Control Traffic With and Without Signaling Encryption

Branch Office Size (Number of IP Phones and Gateways)	Recommended Bandwidth for Control Traffic (no encryption)	Recommended Bandwidth for Control Traffic (with encryption)
80	21 kbps	33 kbps
90	24 kbps	38 kbps
100	27 kbps	42 kbps
110	29 kbps	46 kbps
120	32 kbps	50 kbps
130	35 kbps	54 kbps
140	37 kbps	58 kbps
150	40 kbps	62 kbps

**Note**

The values in [Table 3-6](#) 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 Signaling Encryption.

$$\text{Bandwidth (bps)} = (53 + 21 * \text{CH}) * (\text{Number of IP phones and gateways in the branch})$$

Equation 4: Recommended Bandwidth for a Remote Branch with Signaling Encryption.

$$\text{Bandwidth with signaling encryption (bps)} = (73.5 + 33.9 * \text{CH}) * (\text{Number of IP phones and gateways in the branch})$$

Provisioning for Call Control Traffic with Distributed Call Processing

In distributed call processing deployments, several sites are connected through an IP WAN. Each site contains a Cisco CallManager cluster and can follow either the single-site model or the centralized call processing model. A gatekeeper may be used for call admission control between sites.

The following considerations apply to this deployment model:

- The signaling protocol used to place a call across the WAN is H.323 or SIP.
- Control traffic is exchanged between the Cisco IOS gatekeeper and the Cisco CallManager clusters at each site, as well as between the Cisco CallManager clusters themselves.

Therefore, bandwidth for control traffic must be provisioned on the WAN links between Cisco CallManagers as well as between each Cisco CallManager and the gatekeeper. Because the topology is limited to hub-and-spoke, with the gatekeeper typically located at the hub, the WAN link that connects each site to the other sites usually coincides with the link that connects the site to the gatekeeper.

The control traffic that traverses the WAN belongs to one of the following categories:

- Quiescent traffic, which consists of registration messages periodically exchanged between each Cisco CallManager and the gatekeeper
- Call-related traffic, which in turn consists of two types of traffic:
 - Call admission control traffic, exchanged between the Cisco CallManagers and the gatekeeper before a call can be set up and after it has been torn down
 - H.225 or H.245 signaling traffic, exchanged between two Cisco CallManagers when a call needs to be set up, torn down, forwarded, and so on

Because the total amount of control traffic depends on the number of calls that are set up and torn down at any given time, it is necessary to make some assumptions about the call patterns and the link utilization. The WAN links that connect each of the spoke sites to the hub site are normally provisioned to accommodate different types of traffic (for example, data, voice, and video). Using a traditional telephony analogy, we can view the portion of the WAN link that has been provisioned for voice as a number of *virtual tie lines*.

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.

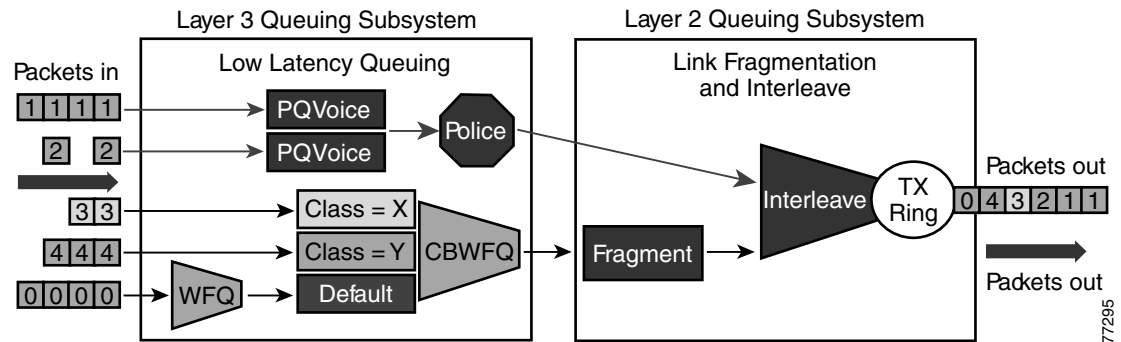
$$\text{Recommended Bandwidth (bps)} = 116 * (\text{Number of virtual tie lines})$$

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

In choosing from among the many available prioritization schemes, the major factors to consider include the type of traffic involved and the type of media on the WAN. For multi-service traffic over an IP WAN, Cisco recommends low-latency queuing (LLQ) for all links. This method supports up to 64 traffic classes, with the ability to specify, for example, priority queuing behavior for voice and interactive video, minimum bandwidth class-based weighted fair queuing for voice control traffic, additional minimum bandwidth weighted fair queues for mission critical data, and a default best-effort queue for all other traffic types.

Figure 3-9 shows an example prioritization scheme.

Figure 3-9 Optimized Queuing for VoIP over the WAN

Cisco recommends the following prioritization criteria for LLQ:

- The criterion for *voice* to be placed into a priority queue is the differentiated services code point (DSCP) value of 46, or a per-hop behavior (PHB) value of EF.
- The criterion for *video conferencing* traffic to be placed into a priority queue is a DSCP value of 34, or a PHB value of AF41. However, due to the larger packet sizes of video traffic, these packets should be placed in the priority queue only on WAN links that are faster than 768 Kbps. Link speeds below this value require packet fragmentation, but packets placed in the priority queue are not fragmented, thus smaller voice packets could be queued behind larger video packets. For links speeds of 768 Kbps or lower, video conferencing traffic should be placed in a separate class-based weighted fair queue (CBWFQ).

**Note**

One-way video traffic, such as the traffic generated by streaming video applications for services such as video-on-demand or live video feeds, should always use a CBWFQ scheme because that type of traffic has a much higher delay tolerance than two-way video conferencing traffic

- As the WAN links become congested, it is possible to starve the *voice control* signaling protocols, thereby eliminating the ability of the IP phones to complete calls across the IP WAN. Therefore, voice control protocols, such as H.323, MGCP, 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 24 or a PHB value of CS3.

**Note**

Cisco has begun to change the marking of voice control protocols from DSCP 26 (PHB AF31) to DSCP 24 (PHB CS3). However many products still mark signaling traffic as DSCP 26 (PHB AF31); therefore, in the interim, Cisco recommends that you reserve both AF31 and CS3 for call signaling.

- In some cases, certain data traffic might require better than best-effort treatment. This traffic is referred to as *mission-critical data*, and it is placed into one or more queues that have the required amount of bandwidth. The queuing scheme within this class is first-in-first-out (FIFO) with a minimum allocated bandwidth. Traffic in this class that exceeds the configured bandwidth limit is placed in the default queue. The entrance criterion for this queue could be a Transmission Control Protocol (TCP) port number, a Layer 3 address, or a DSCP/PHB value.
- All remaining traffic can be placed in a default queue for best-effort treatment. If you specify the keyword **fair**, the queuing algorithm will be weighted fair queuing (WFQ).

Link Efficiency Techniques

The following link efficiency techniques improve the quality and efficiency of low-speed WAN links.

Compressed Real-Time Transport Protocol (cRTP)

You can increase link efficiency 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. cRTP operates on a per-hop basis. Use cRTP on a particular link only if that link meets *all* of the following conditions:

- Voice traffic represents more than 33% 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.

If the link fails to meet any one of the preceding conditions, then cRTP is not effective and you should not use it on that link. Another important parameter to consider before using cRTP is router CPU utilization, which is adversely affected by compression and decompression operations.

cRTP on ATM and Frame Relay Service Inter-Working (SIW) links requires the use of Multilink Point-to-Point Protocol (MLP).

Note that cRTP compression occurs as the final step before a packet leaves the egress interface; that is, after LLQ class-based queueing has occurred. Beginning in Cisco IOS Release 12.(2)2T and later, cRTP provides a feedback mechanism to the LLQ class-based queueing mechanism that allows the bandwidth in the *voice* class to be configured based on the compressed packet value. With Cisco IOS releases prior to 12.(2)2T, this mechanism is not in place, so the LLQ is unaware of the compressed bandwidth and, therefore, the *voice* class bandwidth has to be provisioned as if no compression is taking place. [Table 3-7](#) shows an example of the difference in *voice* class bandwidth configuration given a 512-kbps link with G.729 codec and a requirement for 10 calls.

Note that [Table 3-7](#) assumes 24 kbps for non-cRTP G.729 calls and 10 kbps for cRTP G.729 calls. These bandwidth numbers are based on voice payload and IP/UDP/RTP headers only. They do not take into consideration Layer 2 header bandwidth. However, actual bandwidth provisioning should also include Layer 2 header bandwidth based on the type WAN link used.

Table 3-7 *LLQ Voice Class Bandwidth Requirements for 10 Calls with 512 kbps Link Bandwidth and G.729 Codec*

Cisco IOS Release	With cRTP Not Configured	With cRTP Configured
Prior to 12.2(2)T	240 kbps	240 kbps ¹
12.2(2)T or later	240 kbps	100 kbps

1. 140 kbps of unnecessary bandwidth must be configured in the LLQ *voice* class.

It should also be noted that, beginning in Cisco IOS Release 12.2(13)T, cRTP can be configured as part of the voice class with the Class-Based cRTP feature. This option allows cRTP to be specified within a class, attached to an interface via a service policy. This new feature provides compression statistics and bandwidth status via the **show policy interface** command, which can be very helpful in determining the offered rate on an interface service policy class given the fact that cRTP is compressing the IP/RTP headers.

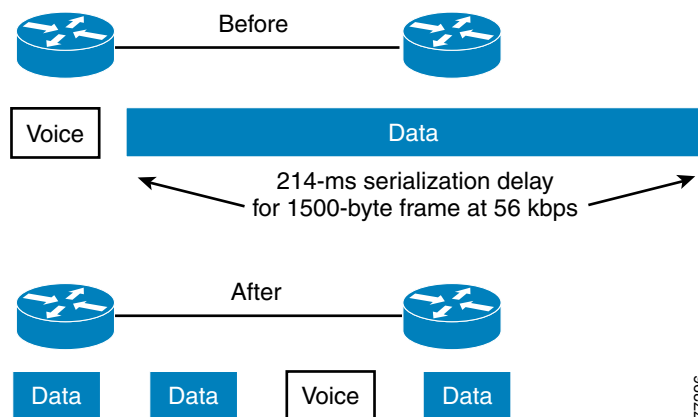
For additional recommendations about using cRTP with a Voice and Video Enabled IPsec VPN (V3PN), refer to the V3PN documentation available at

<http://cisco.com/go/srnd>

Link Fragmentation and Interleaving (LFI)

For low-speed links (less than 768 kbps), use of link fragmentation and interleaving (LFI) mechanisms is required for acceptable voice quality. This technique limits jitter by preventing voice traffic from being delayed behind large data frames, as illustrated in [Figure 3-10](#). The two techniques that exist for this purpose are Multilink Point-to-Point Protocol (MLP) LFI (for Leased Lines, ATM, and SIW) and FRF.12 for Frame Relay.

Figure 3-10 Link Fragmentation and Interleaving (LFI)



Voice-Adaptive Fragmentation (VAF)

In addition to the LFI mechanisms mentioned above, voice-adaptive fragmentation (VAF) is another LFI mechanism for Frame Relay links. VAF uses FRF.12 Frame Relay LFI; however, once configured, fragmentation occurs only when traffic is present in the LLQ priority queue or when H.323 signaling packets are detected on the interface. This method ensures that, when voice traffic is being sent on the WAN interface, large packets are fragmented and interleaved. However, when voice traffic is not present on the WAN link, traffic is forwarded across the link unfragmented, thus reducing the overhead required for fragmentation.

VAF is typically used in combination with voice-adaptive traffic shaping (see [Voice-Adaptive Traffic Shaping \(VATS\)](#), [page 3-39](#)). VAF is an optional LFI tool, and you should exercise care when enabling it because there is a slight delay between the time when voice activity is detected and the time when the LFI mechanism engages. In addition, a configurable deactivation timer (default of 30 seconds) must expire after the last voice packet is detected and before VAF is deactivated, so during that time LFI will occur unnecessarily. VAF is available in Cisco IOS Release 12.2(15)T and later.

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.

Figure 3-11 illustrates the main reasons why traffic shaping is needed when transporting voice and data on the same IP WAN.

Figure 3-11 Traffic Shaping with Frame Relay and ATM

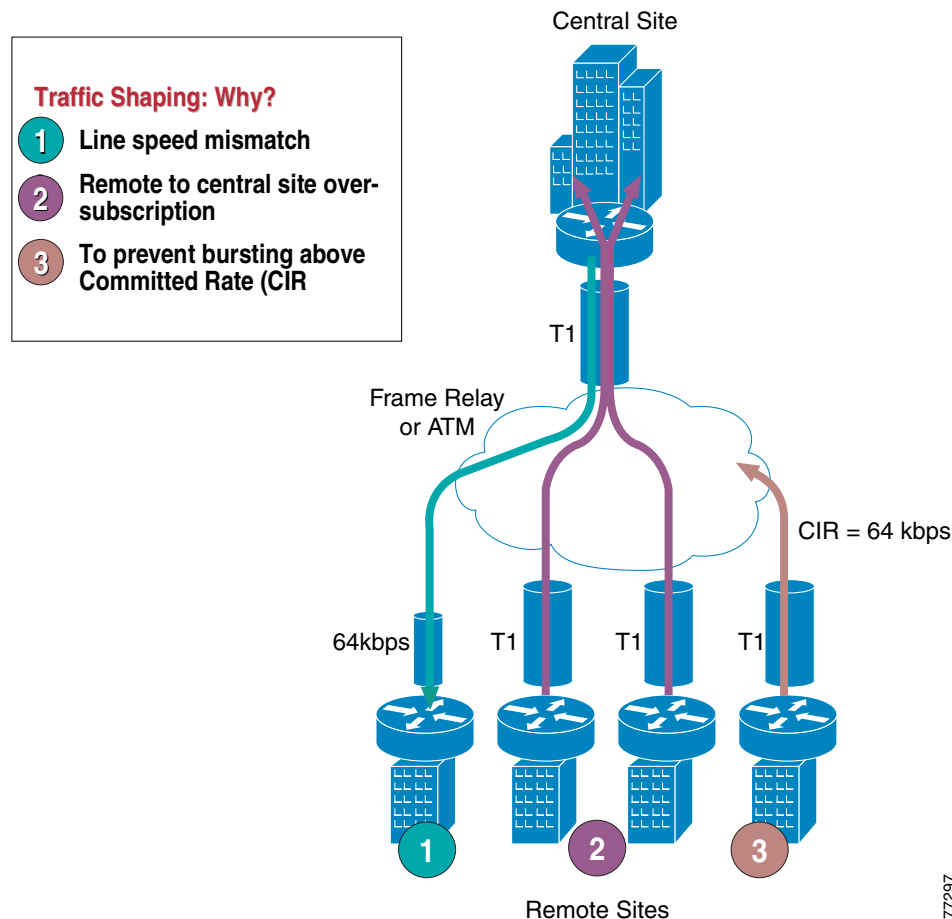


Figure 3-11 shows three different scenarios:

1. 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 might become congested and degrade voice performance.

2. Oversubscription of the link between the central site and the remote sites

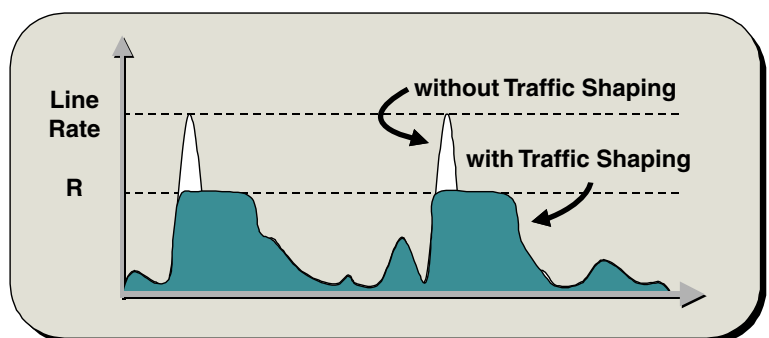
It is common practice in Frame Relay or ATM networks to oversubscribe bandwidth when aggregating many remote sites to a single central site. For example, there may be multiple remote sites that connect to the WAN with a T1 interface, yet the central site has only a single T1 interface. While this configuration 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.

3. 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 might 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 with no regard to traffic classification, possibly having a negative affect 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. [Figure 3-12](#) illustrates this mechanism with a generic example, where R is the rate with traffic shaping applied.

Figure 3-12 Traffic Shaping Mechanism



Voice-Adaptive Traffic Shaping (VATS)

VATS is an optional dynamic mechanism that shapes traffic on Frame Relay permanent virtual circuits (PVCs) at different rates based on whether voice is being sent across the WAN. The presence of traffic in the LLQ voice priority queue or the detection of H.323 signaling on the link causes VATS to engage. Typically, Frame Relay shapes traffic to the guaranteed bandwidth or CIR of the PVC at all times. However, because these PVCs are typically allowed to burst above the CIR (up to line speed), traffic shaping keeps traffic from using the additional bandwidth that might be present in the WAN. With VATS enabled on Frame Relay PVCs, WAN interfaces are able to send at CIR when voice traffic is present on the link. However, when voice is not present, non-voice traffic is able to burst up to line speed and take advantage of the additional bandwidth that might be present in the WAN.

When VATS is used in combination with voice-adaptive fragmentation (VAF) (see [Link Fragmentation and Interleaving \(LFI\)](#), [page 3-37](#)), all non-voice traffic is fragmented and all traffic is shaped to the CIR of the WAN link when voice activity is detected on the interface.

As with VAF, exercise care when enabling VATS because activation can have an adverse effect on non-voice traffic. When voice is present on the link, data applications will experience decreased throughput because they are throttled back to well below CIR. This behavior will likely result in packet drops and delays for non-voice traffic. Furthermore, after voice traffic is no longer detected, the deactivation timer (default of 30 seconds) must expire before traffic can burst back to line speed. It is

important, when using VATS, to set end-user expectations and make them aware that data applications will experience slowdowns on a regular basis due to the presence of voice calls across the WAN. VATS is available in Cisco IOS Release 12.2(15)T and later.

For more information on the Voice-Adaptive Traffic Shaping and Fragmentation features and how to configure them, refer to the documentation at

http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122newft/122t/122t15/ft_vats.htm

Wireless LAN Infrastructure

Wireless LAN infrastructure design becomes important when IP telephony is added to the wireless LAN (WLAN) portions of a converged network. With the addition of wireless IP telephony endpoints such as the Cisco Wireless IP Phone 7920, voice traffic has moved onto the WLAN and is now converged with the existing data traffic there. Just as with wired LAN and wired WAN infrastructure, the addition of voice in the WLAN requires following basic configuration and design best-practices for deploying a highly available network. In addition, proper WLAN infrastructure design requires understanding and deploying QoS on the wireless network to ensure end-to-end voice quality on the entire network. The following sections discuss these requirements:

- [WLAN Design and Configuration, page 3-40](#)
- [WLAN Quality of Service \(QoS\), page 3-47](#)

For more information about WLAN design, refer to the *Cisco Wireless LAN SRND* guide, available at

<http://www.cisco.com/go/srnd>

For more information about the Cisco 7920 Wireless IP Phone, refer to the *Cisco Wireless IP Phone 7920 Design and Deployment Guide*, available at

<http://www.cisco.com/go/srnd>

WLAN Design and Configuration

Properly designing a WLAN requires, first and foremost, ensuring that the existing wired network is deployed in a highly available, fault-tolerant and redundant manner. Next, an understanding of wireless technology is required. Finally, by configuring and deploying wireless access points (APs) and wireless telephony endpoints in an effective way, you can build a flexible, secure, redundant, and highly scalable network.

The following sections examine the WLAN infrastructure layers and network services:

- [Wireless Infrastructure Considerations, page 3-41](#)
- [Wireless AP Configuration and Design, page 3-44](#)
- [Wireless Security, page 3-45](#)

Wireless Infrastructure Considerations

The following sections provide guidelines and best practices for designing the WLAN infrastructure:

- [VLANs, page 3-41](#)
- [Roaming, page 3-41](#)
- [Wireless Channels, page 3-41](#)
- [Wireless Interference, page 3-43](#)
- [Multicast on the WLAN, page 3-44](#)

VLANs

Just as with a wired LAN infrastructure, when deploying voice in a wireless LAN, you should enable at least two virtual LANs (VLANs) at the Access Layer. The Access Layer in a wireless LAN environment includes the access point (AP) and the first-hop access switch. On the AP and access switch, you should configure both a native VLAN for data traffic and a voice VLAN (under Cisco IOS) or Auxiliary VLAN (under CatOS) for voice traffic. This voice/auxiliary VLAN must be separate from all the other wired voice VLANs in the network. In addition, as with voice endpoints on wired LANs, wireless voice endpoints should be addressed using RFC 1918 private subnet addresses. When deploying a wireless infrastructure, Cisco also recommends configuring a separate management VLAN for the management of WLAN APs. This management VLAN should not have a WLAN appearance; that is, it should not have an associated service set identifier (SSID) and it should not be directly accessible from the WLAN.

Roaming

Another very important consideration for wireless infrastructure is wireless endpoint roaming. When wireless devices roam at Layer 2, they keep their IP address and network configuration. For this reason, roaming can occur fairly quickly (in 100 to 400 ms). All that is required is re-authentication, if Cisco LEAP or Extensible Authentication Protocol (EAP) is used, and the passing of Inter-Access Point Protocol (IAPP) messages between the last AP and the new AP to indicate that the endpoint has roamed. Layer 2 roaming is typically unnoticeable to the end user.

When devices roam at Layer 3, they move from one AP to another AP and cross a subnet boundary. With the release of the new Cisco Catalyst 6500 Series Wireless LAN Services Module (WLSM), the Cisco 7920 Wireless IP Phone now supports call-survivable Layer 3 roaming while using Static WEP. Cisco Centralized Key Management (Cisco CKM) enables the Cisco 7920 phone to achieve full Layer 3 mobility while using LEAP. For details about the Cisco WLSM, refer to the product documentation available at <http://www.cisco.com>.



Note

If Cisco Catalyst 4000 Series switches are used as Layer 3 devices at the distribution layer, a minimum of a Supervisor Engine 2+ (SUP2+) or Supervisor Engine 3 (SUP3) module is required. The Supervisor Engine 1 or 2 (SUP1 or SUP2) modules can cause roaming delays. The Cisco Catalyst 2948G, 2948G-GE-TX, 2980G, 2980G-A, and 4912 switches are also known to introduce roaming delays. Cisco does not recommend using these switches in a wireless voice network.

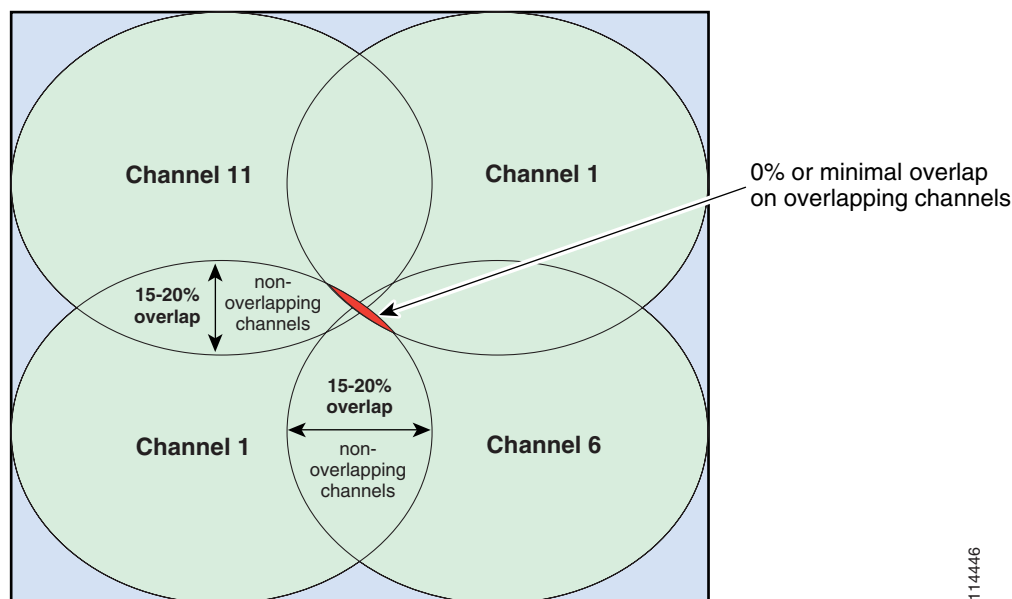
Wireless Channels

Wireless endpoints and APs communicate via radios on particular channels. When communicating on one channel, wireless endpoints typically are unaware of traffic and communication occurring on other non-overlapping channels. Optimal channel configuration for 2.4 GHz 802.11b requires a five-channel spread to prevent interference or overlap between channels. In North America, channels 1, 6, and 11 are

the three usable non-overlapping channels for APs and wireless endpoint devices. In Europe the non-overlapping usable channels for 802.11b are 1, 6, and 11 or 12 or 13. In Japan these channels are 1, 6, and 11 or 12, 13, or 14.

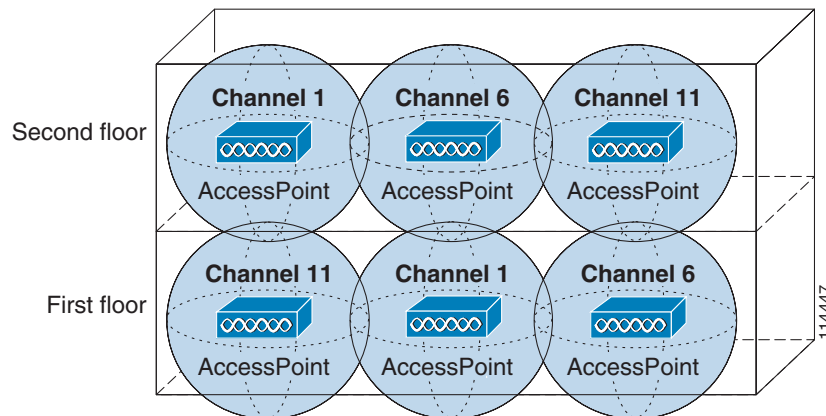
AP coverage should be deployed so that minimal or no overlap occurs between APs configured with the same channel (for example, see Channel 1 in [Figure 3-13](#)). However, proper AP deployment and coverage on *non-overlapping* channels (1, 6, and 11 in North America) require an overlap of 15 to 20%. This amount of overlap ensures smooth roaming for wireless endpoints as they move between AP coverage cells. Overlap of less than 15 to 20% can result in slower roaming times and poor voice quality; while overlap of more than 15 to 20% can result in too frequent or constant roaming. [Figure 3-13](#) illustrates appropriate AP overlap for both overlapping and non-overlapping channels.

Figure 3-13 Wireless 802.11b Channel Overlap



Deploying wireless devices in a multi-story building such as an office high-rise or hospital introduces a third dimension to wireless AP and channel coverage planning. The 2.4 GHz wave form of 802.11b can pass through floors and ceilings as well as walls. For this reason, not only is it important to consider overlapping cells or channels on the same floor, but it is also necessary to consider channel overlap between adjacent floors. With only three channels, proper overlap can be achieved only through careful three-dimensional planning. [Figure 3-14](#) shows the potential for channel overlap when considering the three-dimensional aspects of 802.11b wireless coverage.

114446

Figure 3-14 Wireless 802.11b Channel Overlap Considerations (in Three Dimensions)**Note**

Careful deployment of APs and channel configuration within the wireless infrastructure are imperative for proper wireless network operation. For this reason, Cisco requires that a complete and thorough site survey be conducted before deploying wireless networks in a production environment. The survey should include verifying non-overlapping channel configurations, AP coverage, and required data and traffic rates; eliminating rogue APs; and identifying and mitigating the impact of potential interference sources.

Wireless Interference

Interference sources within a wireless environment can severely limit endpoint connectivity and channel coverage. In addition, objects and obstructions can cause signal reflection and multipath distortion. Multipath distortion occurs when traffic or signaling travels in more than one direction from the source to the destination. Typically, some of the traffic arrives at the destination before the rest of the traffic, which can result in delay and bit errors in some cases. You can reduce the affects of multipath distortion by eliminating or reducing interference sources and obstructions, and by using diversity antennas so that only a single antenna is receiving traffic at any one time. Interference sources should be identified during the site survey and, if possible, eliminated. At the very least, interference impact should be alleviated by proper AP placement and the use of location-appropriate directional or omni-directional diversity radio antennas.

Possible interference sources include:

- Other APs on overlapping channels
- Other 2.4 GHz appliances, such as 2.4 GHz cordless phones, personal wireless network devices, sulphur plasma lighting systems, microwave ovens, rogue APs, and other WLAN equipment that takes advantage of the license-free operation of the 2.4 GHz band
- Metal equipment, structures, and other metal or reflective surfaces such as metal I-beams, filing cabinets, equipment racks, wire mesh or metallic walls, fire doors and fire walls, concrete, and heating and air conditioning ducts
- High-power electrical devices such as transformers, heavy-duty electric motors, refrigerators, elevators, and elevator equipment

Multicast on the WLAN

Cisco does not recommend forwarding multicast traffic on a WLAN with voice devices because of the following reasons:

- Multicast packets are buffered on the AP if devices associated with the AP are in power-save mode. When devices such as the Cisco Wireless IP Phone 7920 go into power-save mode, an AP buffers all multicast packets until the next active interval for these devices. This buffering causes packet delay and affects all devices associated with the AP even if they are not in power-save mode. This condition can be extremely problematic for real-time multicast applications such as music on hold and streaming video.
- Multicast packets on the WLAN are unacknowledged and are not retransmitted if lost or corrupted. AP and wireless endpoint devices use acknowledgements on the link layer to ensure reliable delivery. When packets are not received or acknowledged, they are retransmitted. This retransmission does not occur for multicast traffic on the WLAN. Given that wireless networks have a higher instance of bit error than wired networks, this lack of retransmission results in a larger number of lost packets when compared to a wired LAN.

Before enabling multicast applications on the wireless network, Cisco recommends testing these applications to ensure that performance and behavior are acceptable.

Wireless AP Configuration and Design

Proper AP selection, deployment, and configuration are essential to ensure that the wireless network handles voice traffic in a way that provides high-quality voice to the end users.

AP Selection

Cisco recommends the following APs for deploying wireless voice:

- Aironet 350 Series AP
- Aironet 1100 Series AP
- Aironet 1200 Series AP

For these APs, Cisco IOS Release 12.2(13)JA3 or later should be used. Cisco does *not* recommend the VxWorks operating system for APs when deploying wireless voice because new features are not being added to VxWorks, but some of those new features are required for voice deployments.

AP Deployment

When deploying Cisco access points (APs), do not associate more than 15 to 25 devices to a single AP at any given time. This number will vary depending on usage profiles. The number of devices on an AP affects the amount of time each device has access to the medium. As the number of devices increases, the traffic contention increases. Associating more than 15 to 25 devices to an AP can result in poor AP performance and slower response times for associated devices.

AP Configuration

When deploying wireless voice, observe the following specific AP configuration requirements:

- Enable Address Resolution Protocol (ARP) caching. ARP caching is required on the AP because it enables the AP to answer ARP requests for the wireless endpoint devices without requiring the endpoint to leave power-save or idle mode. This feature results in extended battery life for the wireless endpoint devices.

- Match the transmit power on the AP to that on the wireless voice endpoints.

When possible, the transmit power on the AP and the voice endpoints should match. Matching transmit power on the AP and voice endpoints helps eliminate the possibility of one-way audio traffic. If the transmit power configuration on the APs must vary, the transmit power on all voice endpoints should be configured to match the AP with the highest transmit power.



Note Beginning with version 1.0(8) of the Cisco 7920 Wireless IP Phone firmware, the phone will take advantage of the Dynamic Transmit Power Control (DTPC) feature by automatically adjusting its transmit power based on the Limit Client Power (mW) setting of the current AP.

- Set the data rate to 11 Mbps.

Configuring the maximum 11 Mbps data rate ensures the best level of throughput for voice devices and the largest number of active calls per AP.

- Manually configure the RF channel selection. (Do *not* use the option **Search for Least Congested Channel**).

To control wireless network channels and eliminate channel overlap, it is important to configure a channel number manually on each AP based on its location.

- Assign a Service Set Identifier (SSID) to each VLAN configured on the AP.

SSIDs enable endpoints to select the wireless VLAN they will use for sending and receiving traffic. These wireless VLANs and SSIDs map to wired VLANs. For voice endpoints, this mapping ensures priority queuing treatment and access to the voice VLAN on the wired network.

- Enable **QoS Element for Wireless Phones** on the AP.

This feature ensures that the AP will provide QoS Basic Service Set (QBSS) information elements in beacons. The QBSS element provides an estimate of the channel utilization on the AP, and Cisco wireless voice devices use it to help make roaming decisions and to reject call attempts when loads are too high.

- Configure two QoS policies on the AP, and apply them to the VLANs and interfaces.

Configure a voice policy and a data policy with default classifications for the respective VLANs to ensure that voice traffic is given priority queuing treatment. (See [Interface Queuing, page 3-47](#), for more information).

Wireless Security

Another important consideration for a wireless infrastructure is security. Wireless endpoints, including wireless phones, can connect to the wireless network using one of the following security mechanisms:

- Cisco LEAP

Cisco LEAP requires the wireless endpoint to provide a user name and password to authenticate with the network. Once this authentication occurs, a dynamic key is generated, and traffic to and from the wireless device is encrypted. This method requires an EAP-compliant Remote Authentication Dial-In User Service (RADIUS) authentication server such as the Cisco Secure Access Control Server (ACS), which provides access to a user database for authenticating the wireless devices. Cisco LEAP is the recommended security mechanism for use with wireless voice because it requires the highest level of security for access to voice VLANs.

- Static Wire Equivalent Privacy (WEP)

Static WEP requires the exchange of a statically configured 40-bit or 128-bit character key between the wireless endpoint and the AP. If the keys match, the wireless device is given access to the network. Be aware that there are known weaknesses in the WEP encryption algorithm. These weaknesses, coupled with the complexity of configuring and maintaining static keys, can make this security mechanism undesirable in many cases.

- Open authentication

This method requires no authentication and provides no security for traffic traveling between the wireless endpoint device and the wireless network. This method requires only that the wireless device be configured with the proper SSID for the wireless VLAN to which the device wishes to connect. Cisco does not typically recommend this method for wireless voice because it requires no authentication for access to voice VLANs and provides no encryption for voice traffic.

Cisco LEAP Authentication and ACS Deployment Models

As indicated previously, Cisco LEAP is the preferred method of wireless device authentication (especially voice devices) because it provides the most secure and robust mechanism for access to the network and voice VLAN(s). Because an EAP-compliant RADIUS server is required, Cisco recommends the use of Cisco Secure ACS for Windows Server Version 3.1 or later.

When deploying Cisco LEAP for wireless authentication and encryption, carefully consider the placement of the ACS within the network, and select one of the following ACS deployment models:

- Centralized ACS

ACS server or servers are located in a centralized place within the network and are used to authenticate all wireless devices and users within the network.

- Remote ACS

In networks where remote locations are separated from the central site by low-speed or congested WAN links, an ACS server can be located at the remote site and remote wireless devices or users can be authenticated by this server locally, thus eliminating the potential for delayed authentication via a centralized ACS across the WAN link.

- Local and fallback RADIUS server on the Cisco AP

In networks where remote locations are separated from a central site by low-speed WAN links, local wireless devices can authenticate against local Cisco IOS APs. APs running Cisco IOS Release 12.2(11)JA or later can authenticate Cisco LEAP users and devices locally without relying on an external ACS. A single AP can support up to 50 users with this functionality. This feature can be used in lieu of a centralized or local ACS, or in the case of a WAN or ACS failure in which the remote site users are unable to contact a local ACS or the central site ACS.

When choosing a deployment model for the ACS, it is imperative to make authentication services redundant so that the ACS does not become a single point of failure when wireless devices attempt to access the network. For this reason, each ACS server should replicate its database to a secondary server. Furthermore, it is always a good idea to provide a local ACS or an on-AP RADIUS server at remote sites to ensure that remote wireless devices can still authenticate in the event of a WAN failure.

In addition to ACS server placement, it is also important to consider the implications of user database location in relation to the ACS server. Because the ACS server must access the user database to authenticate wireless devices, the location of the user database affects the amount of time the authentication will take. If the user database is a Microsoft Active Directory (AD) server located on the network, the ACS must send an authentication request to the AD server and wait for a response. To ensure the fastest response times for wireless voice endpoints attempting to authenticate to the network, Cisco recommends defining users locally on the ACS server. Remote databases have unknown response times and can adversely affect authentication times.

WLAN Quality of Service (QoS)

Just as QoS is necessary for LAN and WAN wired network infrastructure in order to ensure high voice quality, QoS is also required for wireless LAN infrastructure. Because of the bursty nature of data traffic and the fact that real-time traffic such as voice is sensitive to packet loss and delay, QoS tools are required to manage wireless LAN buffers, limit radio contention, and minimize packet loss, delay, and delay variation.

However, unlike most wired networks, wireless networks are a shared medium, and wireless endpoints do not have dedicated bandwidth for sending and receiving traffic. While wireless endpoints can mark traffic with 802.1p CoS, DSCP, and PHB, the shared nature of the wireless network means limited admission control and access to the network for these endpoints.

Wireless QoS involves the following main areas of configuration:

- [Traffic Classification, page 3-47](#)
- [Interface Queuing, page 3-47](#)
- [Bandwidth Provisioning, page 3-48](#)

Traffic Classification

As with wired network infrastructure, it is important to classify or mark pertinent wireless traffic as close to the edge of the network as possible. Because traffic marking is an entrance criterion for queuing schemes throughout the wired and wireless network, marking should be done at the wireless endpoint device whenever possible. Marking or classification by wireless network devices should be identical to that for wired network devices, as indicated in [Table 3-2](#).

In accordance with traffic classification guidelines for wired networks, the Cisco Wireless IP Phone 7920 marks voice media traffic or RTP traffic with DSCP 46 (or PHB EF) and voice signaling traffic (SCCP) with DSCP 26 (or PHB AF31). Once this traffic is marked, it can be given priority or better than best-effort treatment and queuing throughout the network. All wireless voice devices should be capable of marking traffic in this manner. All other traffic on the wireless network should be marked as best-effort or with some intermediary classification as outlined in wired network marking guidelines.

Interface Queuing

Once marking has occurred, it is necessary to enable the wired network APs and devices to provide QoS queuing so that voice traffic types are given separate queues to reduce the chances of this traffic being dropped or delayed as it traversed the wireless LAN. Queuing on the wireless network occurs in two directions, upstream and downstream. Upstream queuing concerns traffic traveling from the wireless endpoint up to the AP and from the AP up to the wired network. Downstream queuing concerns traffic traveling from the wired network to the AP and down to the wireless endpoint.

Unfortunately, there is little upstream queuing available in a wireless network. While wireless devices such as the Cisco Wireless IP Phone 7920 can provide queuing upstream as the packets leave the device, there is no mechanism in place to provide queuing among all clients on the wireless LAN because wireless networks are a shared medium. Therefore, although voice media packets might receive priority treatment leaving the wireless endpoint, these packets must contend with all the other packets that other wireless devices may be attempting to send. For this reason, it is extremely important to follow the guideline of no more than 15 to 25 wireless clients per AP. Going beyond the upper limit of this guideline can result in additional voice packet delay and jitter.

As for downstream QoS, Cisco APs currently provide up to eight queues for downstream traffic being sent to wireless clients. The entrance criterion for these queues can be based on a number of factors including DSCP, access control lists (ACLs), and VLAN. Although eight queues are available, Cisco recommends using only two queues when deploying wireless voice. All voice media and signaling traffic should be placed in the highest-priority queue, and all other traffic should be placed in the best-effort queue. This ensures the best possible queuing treatment for voice traffic.

To set up this two-queue configuration, create two QoS policies on the AP. Name one policy **voice** and configure it with the class of service **Voice <10 ms Latency (6)** as the **Default Classification for all packets on the Vlan**. Name the other policy **data** and configure it with the class of service **Best Effort (0)** as the **Default Classification for all packets on the Vlan**. Then assign the **data** policy to the incoming and outgoing radio interface for the data VLAN(s), and assign the **voice** policy to the incoming and outgoing radio interfaces for the voice VLAN(s). With the QoS policies applied at the VLAN level, the AP is not forced to examine every packet coming in or going out to determine the type of queuing it should receive. This configuration ensures that all voice media and signaling are given priority queuing treatment in a downstream direction.

Bandwidth Provisioning

Another QoS requirement for wireless networking is the appropriate provisioning of bandwidth. Bandwidth provisioning involves the bandwidth between the wired and wireless networks as well as the number of simultaneous voice calls that an AP can handle. Wireless APs typically connect to the wired network via a 100 Mbps link to an Access Layer switch port. While the ingress Ethernet port on the AP can receive traffic at 100 Mbps, the maximum throughput on an 802.11b wireless network is 11 Mbps. After taking into account the half-duplex nature of the wireless medium and the overhead of wireless headers, the practical throughput on the 802.11b wireless network is about 7 Mbps. This mismatch in throughput between the wired and wireless network can result in packet drops when traffic bursts occur in the network.

Rather than allowing traffic bursts to send excessive traffic toward the AP only to have it dropped by the AP, it is a good idea to rate-limit or police this traffic to a rate that the wireless network can handle. Forcing the AP to drop excessive traffic causes increased CPU utilization and congestion at the AP. Instead, limiting the traffic rate to 7 Mbps on the link between the wired Access Layer switch and the wireless AP ensures that traffic is dropped at the Access Layer switch, thus removing the burden from the AP. For more information on rate-limiting traffic sent to the AP, see the QoS recommendations in the section on the [Cisco Wireless IP Phone 7920G, page 16-27](#). Note that, depending on the wireless network deployment, the practical throughput might be less than 7 Mbps, especially if more than the recommended number of devices are associated to a single AP.

Based on wireless voice network testing, Cisco has determined that a single wireless AP can support up to seven (7) G.711 voice calls or eight (8) G.729 voice calls. If these limits are exceeded, voice quality will suffer and voice calls might be dropped. While there is no true call admission control mechanism or method for provisioning wireless bandwidth for voice traffic, Cisco 7920 Wireless IP Phones can provide a simplified version of call admission control or bandwidth provisioning based on channel utilization information received from APs on the network. This information can be sent by the AP to the phone via a beacon that includes the QoS Basic Service Set (QBSS). The QBSS provides an estimation of the percentage of time the RF channel is in use by that AP. The higher the QBSS element value, the higher the channel utilization and the less likely the channel and AP can provide sufficient bandwidth for additional wireless voice devices. If the QBSS element value is 45 or higher, then any calls attempted by the wireless IP phone will be rejected with a "Network Busy" message and/or fast busy tone. In addition, the wireless IP phone considers the QBSS element in its roaming algorithm and will not roam to an AP that is sending beacons with a QBSS element of 45 or higher.

**Note**

The QBSS value is simply an estimation of the channel utilization for a particular AP. The real channel utilization can be much higher than indicated. For this reason, the wireless voice device might still be able to place a voice call on an AP that has already reached the limit of 7 or 8 calls, thus still resulting in dropped calls or poor voice quality.

The QBSS information element is sent by the AP only if **QoS Element for Wireless Phones** has been enable on the AP. (Refer to [Wireless AP Configuration and Design](#), page 3-44.)



Voice Gateways

Gateways provide a number of methods for connecting an IP telephony network to the Public Switched Telephone Network (PSTN), a legacy PBX, or key systems. Gateways range from specialized, entry-level and stand-alone voice gateways to high-end, feature-rich integrated router and Cisco Catalyst 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:

- [Understanding Cisco Gateways, page 4-1](#)
- [Gateway Selection, page 4-2](#)
- [QSIG Support, page 4-17](#)
- [Fax and Modem Support, page 4-18](#)

Understanding Cisco Gateways

Cisco access gateways allow Cisco CallManager to communicate with non-IP telecommunications devices. There are two types of Cisco access gateways, analog and digital.

Cisco Access Analog Gateways

There are two categories of Cisco access analog gateways, trunk gateways and station gateways.

- Access analog station gateways

Analog station gateways connect Cisco CallManager to Plain Old Telephone Service (POTS) analog telephones, interactive voice response (IVR) systems, fax machines, and voice mail systems. Station gateways provide Foreign Exchange Station (FXS) ports.

- Access analog trunk gateways

Analog trunk gateways connect Cisco CallManager to PSTN central office (CO) or PBX trunks. Trunk gateways provide Foreign Exchange Office (FXO) ports for access to the PSTN, PBXs, or key systems, and E&M (recEive and transMit, or ear and mouth) ports for analog trunk connection to a legacy PBX. Whenever possible, use digital gateways to minimize any answer and disconnect supervision issues. Analog Direct Inward Dialing (DID) and Centralized Automatic Message Accounting (CAMA) are also available for PSTN connectivity.

Cisco Access Digital Trunk Gateways

A Cisco access digital trunk gateway connects Cisco CallManager to the PSTN or to a PBX via digital trunks such as Primary Rate Interface (PRI), Basic Rate Interface (BRI), or T1 Channel Associated Signaling (CAS). Digital T1 PRI trunks may also be used to connect to certain legacy voice mail systems.

Gateway Selection

When selecting an IP telephony gateway, consider the following factors:

- [Core Feature Requirements, page 4-2](#)
- [Gateway Protocols, page 4-3](#)
- [Gateway Protocols and Core Feature Requirements, page 4-5](#)
- [Site-Specific Gateway Requirements, page 4-10](#)

Core Feature Requirements

Gateways used in IP telephony applications must meet the following core feature 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
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. For more information, see [Fax and Modem Support, page 4-18](#)
- Cisco CallManager redundancy support
Cisco 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.

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 4-10](#)).

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.

Protocol selection depends on site-specific requirements and the installed base of equipment. For example, most remote branch locations have Cisco 2600XM or 3700 series routers installed. These routers support H.323 and MGCP 0.1 with Cisco IOS Release 12.2.11(T) and Cisco CallManager Release 3.1 or later. For gateway configuration, MGCP might be preferred to H.323 due to simpler configuration or support for call survivability during a Cisco CallManager failover from a primary to a secondary Cisco CallManager. On the other hand, H.323 might be preferred over MGCP because of the robustness of the interfaces supported.

Simplified Message Desk Interface (SMDI) is a standard for integrating voice mail systems to PBXs or Centrex systems. Connecting to a voice mail system via SMDI and using either analog FXS or digital T1 PRI would require either SCCP or MGCP protocol because H.323 devices do not identify the specific line being used from a group of ports. Use of H.323 gateways for this purpose means the Cisco Message Interface cannot correctly correlate the SMDI information with the actual port or channel being used for an incoming call.

In addition, the Cisco CallManager deployment model being used can influence gateway protocol selection. (Refer to the chapter on [IP Telephony Deployment Models](#), page 2-1.)

[Table 4-1](#) shows which gateways support a given protocol. Each of these protocols follows a slightly different methodology to provide support for the core gateway requirements. [Gateway Protocols and Core Feature Requirements](#), page 4-5, describes how each protocol provides these feature requirements.

Table 4-1 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)
VG224	Yes, FXS only. Also supports conferencing and transcoding for VG224 beginning with Cisco IOS Release 12.3(T).	Yes, FXS only	Not supported
VG248	No	No	Yes ²
DE-30+, DT-24+ ³	Yes	No	No
Cisco 827-V4	No	Yes, supported for FXS	No
Cisco ATA 188	Yes, FXS only	Yes, FXS only	Yes, FXS only

Table 4-1 Supported Gateway Protocols and Cisco IP Telephony Gateways (continued)

Cisco Gateway	MGCP 0.1	H.323	SCCP
Cisco 1751 and 1760	Yes	Yes	Yes, conferencing and transcoding
Cisco 2600 and 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 2800	Yes, beginning with Cisco IOS Release 12.3.8T4	Yes, beginning with Cisco IOS Release 12.3.8T4	Yes, beginning with Cisco IOS Release 12.3.8T4
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
Cisco 3700	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 3800	Yes, beginning with Cisco IOS Release 12.3.11T	Yes, beginning with Cisco IOS Release 12.3.11T	Yes, beginning with Cisco IOS Release 12.3.11T
Cisco 5300	No	Yes	No
Cisco AS5350	No	Yes	No
Cisco AS5400			
Cisco AS5850	No	Yes	No
Cisco 7200 and 7301	No	Yes	No
Catalyst 4000 WS-X4604-GWY Gateway Module	Yes	Yes	No

Table 4-1 Supported Gateway Protocols and Cisco IP Telephony Gateways (continued)

Cisco Gateway	MGCP 0.1	H.323	SCCP
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

1. The VG200 is no longer available for purchase and has been replaced by the Cisco 2610XM 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

**Note**

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

Gateway Protocols and Core Feature Requirements

This section describes how each protocol (SCCP, H.323, and MGCP) supports the following gateway feature requirements:

- [DTMF Relay, page 4-5](#)
- [Supplementary Services, page 4-6](#)
- [Cisco CallManager Redundancy, page 4-8](#)

DTMF Relay

Dual-Tone Multifrequency (DTMF) is a signaling method that uses specific pairs of frequencies within the voice band for signals. A 64 kbps pulse code modulation (PCM) voice channel can carry these signals without difficulty. However, when using a low bite-rate codec for voice compression, the potential exists for DTMF signal loss or distortion. An out-of-band signaling method for carrying DTMF tones across a Voice over IP (VoIP) infrastructure provides an elegant solution for these codec-induced symptoms.

SCCP Gateways

The SCCP gateways, such as the Cisco VG248, carry DTMF signals out-of-band using Transmission Control Protocol (TCP) port 2002. Out-of-band DTMF is the default gateway configuration mode for the VG248.

H.323 Gateways

The H.323 gateways, such as the Cisco 3700 series products, can communicate with Cisco CallManager using the enhanced H.245 capability for exchanging DTMF signals out-of-band. The following is an example out-of-band DTMF configuration on a Cisco IOS gateway:

```
dial-peer voice 100 voip
destination-pattern 555...
session target ipv4:10.1.1.1
CODEC g729ar8
dtmf-relay h245-alphanumeric
preference 0
```

MGCP Gateway

The Cisco IOS-based VG224, 2600XM, 2800, 3700, and 3800 platforms use MGCP for Cisco CallManager communication. Within the MGCP protocol is the concept of *packages*. The MGCP gateway loads the DTMF package upon start-up. The MGCP gateway sends *symbols* over the control channel to represent any DTMF tones it receives. Cisco CallManager then interprets these signals and passes on the DTMF signals, out-of-band, to the signaling endpoint. The global configuration command for DTMF relay is:

```
mgcp dtmf-relay CODEC all mode out-of-band
```

You must enter additional configuration parameters in the Cisco CallManager MGCP gateway configuration interface.

The Catalyst 6000, DE-30+, and DT-24+ all support MGCP with Cisco CallManager Release 3.1 and later. DTMF relay is enabled by default and does not need additional configuration.

Supplementary Services

Supplementary services provide user functions such as hold, transfer, and conferencing. These are considered fundamental requirements of any voice installation. Each gateway evaluated for use in an IP telephony network should provide support for supplementary services natively, without the use of a software media termination point (MTP).

SCCP Gateways

The Cisco VG248 and ATA 188 gateways provide full supplementary service support. The SCCP gateways use the Gateway-to-Cisco CallManager signaling channel and SCCP to exchange call control parameters.

H.323 Gateways

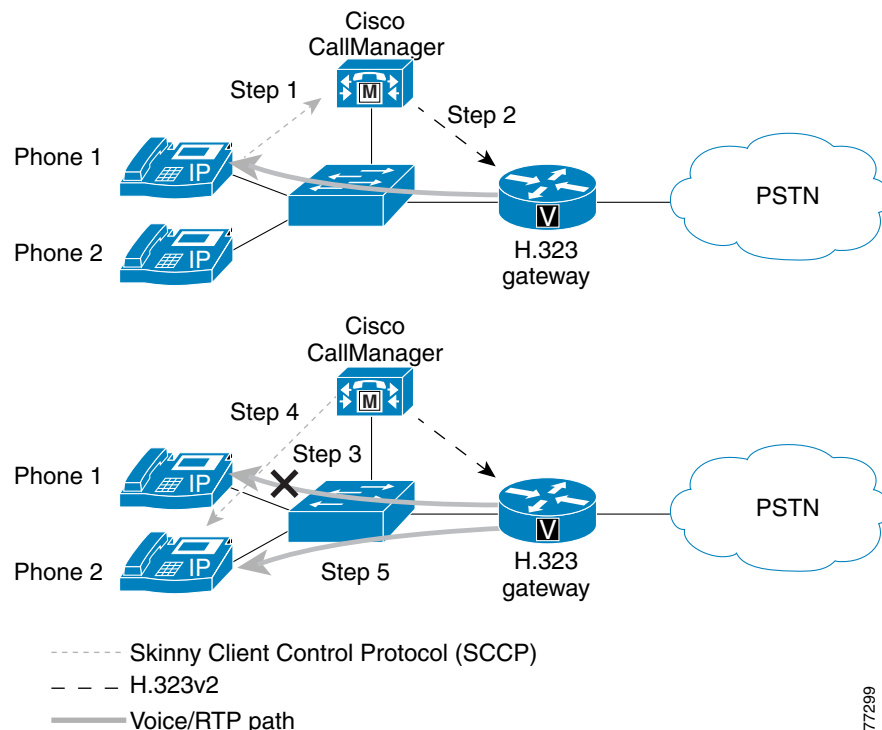
H.323v2 implements Open/Close LogicalChannel and the emptyCapabilitySet features. The use of H.323v2 by H.323 gateways, beginning in Cisco IOS Release 12.0(7)T and Cisco CallManager Release 3.0 and later, eliminates the requirement for an MTP to provide supplementary services. With Cisco CallManager Release 3.1 and later, a transcoder is allocated dynamically only if required during a call to provide access to G.711-only devices while still maintaining a G.729 stream across the WAN. Full support for H.323v2 is available in Cisco IOS Release 12.1.1T.

Once an H.323v2 call is set up between a Cisco IOS gateway and an IP phone, using the Cisco CallManager as an H.323 proxy, the IP phone can request to modify the bearer connection. Because the Real-Time Transport Protocol (RTP) stream is directly connected to the IP phone from the Cisco IOS gateway, a supported voice codec can be negotiated.

Figure 4-1 and the following steps illustrate a call transfer between two IP phones:

1. If IP Phone 1 wishes to transfer the call from the Cisco IOS gateway to Phone 2, it issues a transfer request to Cisco CallManager using SCCP.
2. Cisco CallManager translates this request into an H.323v2 CloseLogicalChannel request to the Cisco IOS gateway for the appropriate SessionID.
3. The Cisco IOS gateway closes the RTP channel to Phone 1.
4. Cisco CallManager issues a request to Phone 2, using SCCP, to set up an RTP connection to the Cisco IOS gateway. At the same time, Cisco CallManager issues an OpenLogicalChannel request to the Cisco IOS gateway with the new destination parameters, but using the same SessionID.
5. After the Cisco IOS gateway acknowledges the request, an RTP voice bearer channel is established between Phone 2 and the Cisco IOS gateway.

Figure 4-1 H.323 Gateway Supplementary Service Support

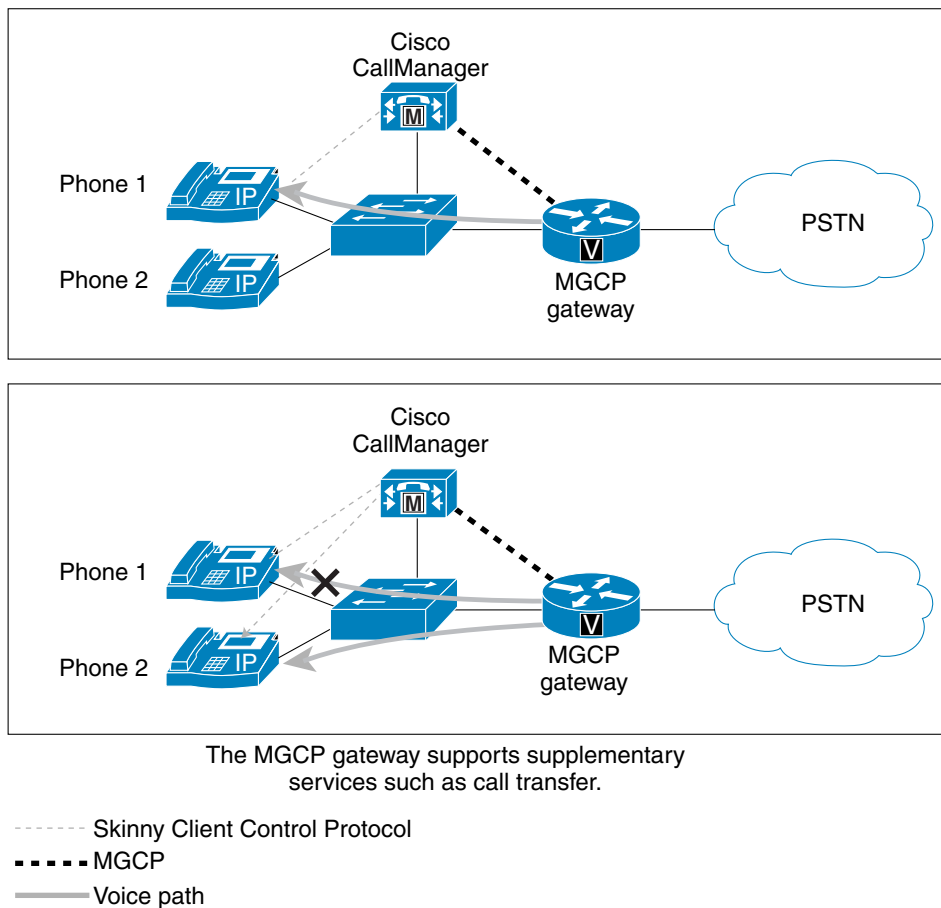


MGCP Gateway

The MGCP gateways provide full support for the hold, transfer, and conference features through the MGCP protocol. Because MGCP is a master/slave protocol with Cisco CallManager controlling all session intelligence, Cisco CallManager can easily manipulate MGCP gateway voice connections. If an IP telephony endpoint (for example, an IP phone) needs to modify the session (for example, transfer the call to another endpoint), the endpoint would notify Cisco CallManager using SCCP. Cisco CallManager then informs the MGCP gateway, using the MGCP User Datagram Protocol (UDP) control connection, to terminate the current RTP stream associated with the Session ID and to start a new media session with the new endpoint information. Figure 4-2 illustrates the protocols exchanged between the MGCP gateway, endpoints, and Cisco CallManager.

Figure 4-2 MGCP Gateway Supplementary Service Support

Direct call from MGCP gateway to IP phone.
MTP is not required.



77300

Cisco CallManager Redundancy

An integral piece of the IP telephony architecture is the provisioning of low-cost, distributed PC-based systems to replace expensive and proprietary legacy PBX systems. This distributed design lends itself to the robust fault tolerant architecture of clustered Cisco CallManagers. Even in its most simplistic form (a two-system cluster), a secondary Cisco CallManager should be able to pick up control of all gateways initially managed by the primary Cisco CallManager.

SCCP Gateways

Upon boot-up, the Cisco VG248 and ATA 188 gateways are provisioned with Cisco CallManager server information. When these gateways initialize, a list of Cisco CallManagers is downloaded to the gateways. This list is prioritized into a primary Cisco CallManager and secondary Cisco CallManager. In the event that the primary Cisco CallManager becomes unreachable, the gateway registers with the secondary Cisco CallManager.

H.323 Gateways

Using several enhancements to the **dial-peer** and **voice class** command sets in Cisco IOS Release 12.1(2)T, Cisco H.323 gateways support redundant Cisco CallManagers. A new command, **H.225 tcp timeout <seconds>**, has been added. This command tracks the time it takes for the H.323 gateway to establish an H.225 control connection for H.323 call setup. If the H.323 gateway cannot establish an H.225 connection to the primary Cisco CallManager, it tries a second Cisco CallManager defined in another **dial-peer** statement. The H.323 gateway shifts to the **dial-peer** statement with next highest **preference** setting. The following commands allow you to configure Cisco CallManager redundancy for a H.323 gateway:

```
dial-peer voice 101 voip
  destination-pattern 1111
  session target ipv4:10.1.1.101
  preference 0
  voice class h323 1
dial-peer voice 102 voip
  destination-pattern 1111
  session target ipv4:10.1.1.102
  preference 1
  voice class h323 1
voice class h323 1
  h225 tcp timeout <1-30 sec>
```



Note

Cisco CallManager redundancy does not imply call survivability. If the primary Cisco CallManager fails, any call between an IP phone and a phone connected via an H.323 gateway is terminated.

MGCP Gateway

MGCP gateways also have the ability to fail over to a secondary Cisco CallManager in the event of communication loss with the primary Cisco CallManager. When the failover occurs, active calls are preserved.

Within the MGCP gateway configuration file, the primary Cisco CallManager is identified using the **call-agent <hostname>** command, and a list of secondary Cisco CallManager is added using the **ccm-manager redundant-host** command. Keepalives with the primary Cisco CallManager are through the MGCP application-level keepalive mechanism, whereby the MGCP gateway sends an empty MGCP notify (NTFY) message to Cisco CallManager and waits for an acknowledgement. Keepalive with the backup Cisco CallManagers is through the TCP keepalive mechanism.

If the primary Cisco CallManager becomes available at a later time, the MGCP gateway can “re-home,” or switch back to the original Cisco CallManager. This re-homing can occur either immediately, after a configurable amount of time, or only when all connected sessions have been released. This is enabled through the following global configuration commands:

```
ccm-manager redundant-host <hostname1 | ipaddress1> <hostname2 | ipaddress2>
[no] call-manager redundancy switchback [immediate|graceful|delay <delay_time>]
```

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 site-specific and core gateway requirements are a good start to help narrow the possible choices. Once you have defined the required features, you can make a gateway selection for each of the pertinent configurations, whether they are single-site enterprise deployments of various sizes and complexities or multi-site enterprise deployments.

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, refer to the documentation at

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/ccmcomp.htm

Cisco Analog Gateways

Table 4-2 lists supported interface types for Cisco analog gateways using H.323 or Session Initiation Protocol (SIP), and Table 4-3 lists supported interface types for Cisco analog gateways using Media Gateway Control Protocol (MGCP).

Table 4-2 Supported Analog H.323 and SIP 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	Yes	No	No	No	No	No
1751 and 1760	Yes	Yes	Yes	Yes	Yes	Future
VG200	Yes	Yes	Yes	No	Yes	No
VG248	No	No	No	No	No	No
VG224	Yes	N/A	N/A	N/A	N/A	N/A
2600 Series	Yes	Yes	Yes	Yes	Yes	12.2.11T
2800 Series	12.3.8T4	Yes	Yes	Yes	Yes	Yes
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
3800 Series	12.3.11T	Yes	Yes	Yes	Yes	Yes
ICS 7750	Yes	Yes	Yes	Yes	Yes	No
Catalyst 4000 Access Gateway Module (AGM)	Yes	Yes	No	No	No	No
6608 and 6624	N/A	N/A	N/A	N/A	N/A	N/A
Communication Media Module (CMM) 24FXS	12.2.13ZP	N/A	N/A	N/A	N/A	N/A

Table 4-2 Supported Analog H.323 and SIP Features (continued)

Cisco Gateway	Interface Type					
	FXS	FXO	E&M	FXO, Battery Reversal	Analog DID	CAMA 911
CMM-6T1/E1	N/A	N/A	N/A	N/A	N/A	N/A
7x00 family	N/A	N/A	N/A	N/A	N/A	N/A

Table 4-3 Supported Analog MGCP Features

Cisco Gateway	Interface Type					
	FXS	FXO	E&M	FXO, Battery Reversal	Analog DID	CAMA 911
Analog Telephone Adapter (ATA)	Yes	N/A	N/A	N/A	N/A	N/A
827-4V	No	No	N/A	N/A	N/A	N/A
1751 and 1760	12.2.11YU	12.2.11YU	No	Yes	No	No
VG200	Yes	Yes	No	Yes	No	No
VG248	No	No	No	No	No	No
VG224	12.3.7T	No	No	No	No	No
2600 Series	12.2.4T	12.2.4T	No	Yes	No	No
2800 Series	12.3.8T4	12.3.8T4	No	Yes	No	No
3600 Series	12.2.4T	12.2.4T	No	Yes	No	No
3700 Series	12.2.8T	12.2.8T	No	Yes	No	No
3800 Series	12.3.11T	12.3.11T	No	Yes	No	No
ICS 7750	Yes	Yes	No	No	No	No
Catalyst 4000 Access Gateway Module (AGM)	12.2.13T	12.2.13T	No	No	No	No
6608 and 6624	Yes	No	No	No	No	No
Communication Media Module (CMM) 24FXS	12.2.13ZP	N/A	N/A	N/A	N/A	N/A
CMM-6T1/E1	N/A	N/A	N/A	N/A	N/A	N/A
7x00 family	N/A	N/A	N/A	N/A	N/A	N/A

Cisco Digital Gateways

Table 4-4 through Table 4-7 list supported interface types for Cisco digital gateways using H.323 or Session Initiation Protocol (SIP). Table 4-8 lists supported interface types for Cisco digital gateways using Media Gateway Control Protocol (MGCP).

Table 4-4 Supported Digital H.323 and SIP Features for BRI, T1 CAS, T1 FGB, 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 FGB	T1 FGD	T1 QSIG
Analog Telephone Adapter (ATA)	No	No	No	No	No	No	No	No
827-4V	No	No	No	No	No	No	No	No
1751 and 1760	No	Yes	12.2.8YH	No	12.2.4YB	No	No	12.2.4YB
VG200	Yes	Yes	No	No	Yes	No	Yes	No
VG248	No	No	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2600 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
2800 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
3600 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
3700 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
3800 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
ICS 7750	Yes	Yes	No	No	Yes	No	12.2.15ZL	No
Catalyst 4000 Access Gateway Module (AGM)	Yes	No	Yes	No	Yes	No	12.2.13T	Yes
6608 and 6624	N/A	N/A	N/A	N/A	No	No	No	No
Communication Media Module (CMM) 24FXS	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CMM-6T1/E1	N/A	N/A	N/A	N/A	Yes	No	No	Yes
7x00 family	N/A	N/A	N/A	N/A	Yes	No	Yes	Yes

Table 4-5 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

Table 4-5 Supported Digital H.323 and SIP Features for T1 PRI DMS-100, 4ESS, and 5ESS (continued)

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)
VG200	Yes	No	Yes	No	Yes	No
VG248	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A
2600 Series	Yes	Future	Yes	Future	Yes	Future
2800 Series	Yes	Future	Yes	Future	Yes	Future
3600 Series	Yes	Future	Yes	Future	Yes	Future
3700 Series	Yes	Future	Yes	Future	Yes	Future
3800 Series	Yes	Future	Yes	Future	Yes	Future
ICS 7750	Yes	No	Yes	No	Yes	No
Catalyst 4000 Access Gateway Module (AGM)	Yes	Future	Yes	Future	Yes	Future
6608 and 6624	No	No	No	No	No	No
Communication Media Module (CMM) 24FXS	N/A	N/A	N/A	N/A	N/A	N/A
CMM-6T1/E1	Yes	Yes	Yes	Yes	Yes	Yes
7x00 family	Yes	Future	Yes	Future	Yes	Future

Table 4-6 Supported Digital H.323 and SIP Features for T1 PRI NI2, NFAS, and Network Specific Facilities (NSF) Service

Cisco Gateway	Interface Type					
	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
VG200	Yes	Yes	No	No	No	No
VG248	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A
2600 Series	Yes	Yes	Yes	Yes	Yes	Yes
2800 Series	Yes	Yes	Yes	Yes	Yes	Yes
3600 Series	Yes	Yes	Yes	Yes	Yes	Yes
3700 Series	Yes	Yes	Yes	Yes	Yes	Yes
3800 Series	Yes	Yes	Yes	Yes	Yes	Yes

Table 4-6 Supported Digital H.323 and SIP Features for T1 PRI NI2, NFAS, and Network Specific Facilities (NSF) Service

Cisco Gateway	Interface Type					
	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)
ICS 7750	Yes	Yes	Yes	Yes	Yes	No
Catalyst 4000 Access Gateway Module (AGM)	Yes	Yes	Future	Future	Future	Future
6608 and 6624	No	No	No	No	No	No
Communication Media Module (CMM) 24FXS	N/A	N/A	N/A	N/A	N/A	N/A
CMM-6T1/E1	Yes	Yes	Future	Future	Future	No
7x00 family	Yes	Yes	No	No	No	No

Table 4-7 Supported Digital H.323 and SIP Features for E1 and J1

Cisco Gateway	Interface Type						
	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
VG200	No	Yes	Yes	Yes	Yes	No	12.2.13T
VG248	No	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2600 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2800 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
3800 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ICS 7750	No	No	12.2.15ZL	Yes	Yes	No	No
Catalyst 4000 Access Gateway Module (AGM)	Future	Future	Yes	Yes	Yes	Yes	No
6608 and 6624	No	No	No	No	No	No	No
Communication Media Module (CMM) 24FXS	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CMM-6T1/E1	No	No	Yes	Yes	Yes	Yes	N/A
7x00 family	Yes	No	Yes	Yes	Yes	Yes	No

Table 4-8 Supported Digital MGCP Features

Cisco Gateway	Interface Type					
	BRI	T1 CAS (E&M)	T1 PRI	T1 QSIG	E1 PRI	E1 QSIG
Analog Telephone Adapter (ATA)	N/A	N/A	N/A	N/A	N/A	N/A
827-4V	N/A	N/A	N/A	N/A	N/A	N/A
1751 and 1760	No	12.2.11YU	12.2.11YU	Future	12.2.11YU	Future
VG200	No	Yes	Yes	Yes	Yes	Yes
VG248	N/A	N/A	N/A	N/A	N/A	N/A
VG224	N/A	N/A	N/A	N/A	N/A	N/A
2600 Series	12.3.4T	Yes ¹	Yes ¹	Yes ¹	Yes ¹	Yes ¹
2800 Series	12.3.8T4	Yes ¹	Yes ¹	Yes ¹	Yes ¹	Yes ¹
3600 Series	12.3.4T	Yes ¹	Yes ¹	Yes ¹	Yes ¹	Yes ¹
3700 Series	12.3.4T	Yes ¹	Yes ¹	Yes ¹	Yes ¹	Yes ¹
3800 Series	12.3.11T	Yes ¹	Yes ¹	Yes ¹	Yes ¹	Yes ¹
ICS 7750	12.3.2XA	Yes	Yes	Yes	Yes	Yes
Catalyst 4000 Access Gateway Module (AGM)	No	Yes	Yes	Yes	Yes	Yes
6608	N/A	Yes	Yes	Yes	Yes	Yes
6624	N/A	N/A	N/A	N/A	N/A	N/A
Communication Media Module (CMM) 24FXS	N/A	N/A	N/A	N/A	N/A	N/A
CMM-6T1/E1	N/A	Yes	Yes	Yes	Yes	Yes
7x00 family	N/A	No	No	No	No	No

1. AIM-VOICE-30 modules require Cisco IOS Release 12.2.13T to support MGCP.

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.

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 4-9](#).

Table 4-9 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 2800 Series	BRI and T1/E1 QSIG	12.3.8T4
Cisco 3700	T1/E1 QSIG	12.2.8T
Cisco 3800	BRI and T1/E1 QSIG	12.3.11T
Cisco 7200 and 7301	T1/E1 QSIG	12.1.2T
Cisco 5300	T1/E1	12.0.7T
Cisco AS5350	T1/E1	12.2.2T
Cisco AS5400		

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, 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 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 2600XM, 3700, VG224, 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 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 and 2691 (with FXS)
- 2800 (with FXS)
- 3725 and 3745 (with FXS)
- 3800 (with FXS)
- VG200 (with FXS)
- VG224
- 1751 and 1760

- Communication Media Module (CMM) FXS card

Non-IOS Gateways:

- VG248
- ATA 188
- 6624

Digital Gateways

Cisco IOS Gateways:

- 2600XM and 2691
- 2800
- 3725 and 3745
- 3800
- VG200
- VG224
- 1751 and 1760
- 7200 and 7500
- AS5300, 5350, 5400, and 5850
- Communication Media Module (CMM)

Non-IOS Gateways:

- 6608



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 3700 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 VG224, VG248, 2600XM, 2800, 3700, 3800, 5300, and Catalyst 6000.

Table 4-10 lists the protocol combinations that currently support fax and modem interoperability.

Table 4-10 Fax and Modem Features Supported with Various Combinations of Call Control Protocols

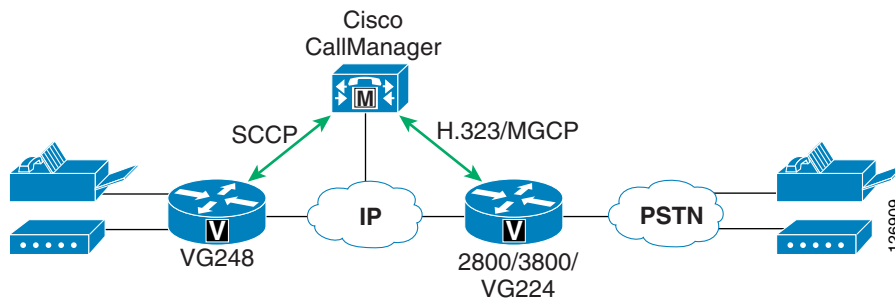
Protocol Combinations	Modem Relay	Modem Pass-Through	T.38 Fax Relay	Cisco Fax Relay	Fax Pass-Through
Cisco CallManager using MGCP combined with Cisco CallManager using H.323	Yes	Yes	Yes	Yes	Yes
Cisco CallManager using MGCP combined with Cisco CallManager using MGCP	Yes	Yes	Yes	Yes	Yes
SCCP combined with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

**Note**

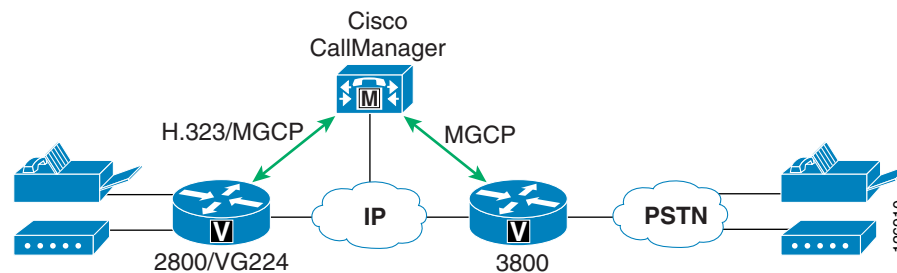
Cisco ATA 188, VG248 and Catalyst 6000 platforms currently do not support T.38 fax relay. When these platforms connect to Cisco AS5350 or AS5400 gateways, only fax pass-through is supported for fax applications.

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 2800 or 3800) with a non-IOS gateway (such as a Cisco VG248), as illustrated in [Figure 4-3](#).

Figure 4-3 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 4-4](#).

Figure 4-4 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 4-11 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 3700, or between 2600XM and AS5300. In these cases, as long as both platforms support a given feature, those platforms will interoperate.

Table 4-11 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 ATA 188)	N/A	Supported (except on ATA 188)	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 VG248 (VGC10d8002407)                               |
|-----|
| Advanced settings |
|-----|
| Allow last good configuration (enabled) |
| SRST policy (disabled) |
| SRST provider () |
| Call preservation (enabled: no timeout) |
| Media receive timeout (disabled) |
| Busy out off hook ports (disabled) |
| DTMF tone dur ----- 100ms) |
| Echo cancelli| Passthrough signalling |e: use DSP) |
| Passthrough s|-----|) |
| Hook flash ti| legacy | default>) |
| Hook flash re| IOS mode | |
| Fax relay max ----- 14400 bps) |
| Fax relay playout delay (default: 300) |
|-----|
|-----|

```

```

-----
|                               Cisco VG248 (VGC10d8002407)                               |
|-----|
| Advanced settings |
|-----|
| Allow last good configuration (enabled) |
| SRST policy (disabled) |
| SRST provider () |
| Call preservation (enabled: no timeout) |
| Media receive timeout (disabled) |
| Busy out off hook ports (disabled) |
| DTMF tone duration (default: 100ms) |
| Echo cancelling policy (alternate: use DSP) |
| Passthrough signalling (IOS mode) |
| Hook flash timer (<country default>) |
| Hook flash reject period (none) |
| Fax relay maximum speed (default: 14400 bps) |
| Fax relay playout delay (default: 300) |
|-----|
|-----|

```


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 4-5](#) and [Figure 4-6](#) 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 4-5 Gateway Configuration in Cisco CallManager

Cisco CallManager Administration
For Cisco IP Telephony Solutions

Gateway Configuration [Back to Find/List Gateways](#)

Product : Cisco Catalyst 6000 T1 VoIP Gateway
 Gateway : 60/DS1-0060A00016473E7B4
 Device Protocol: Digital Access PRI
 Registration: Unknown
 IP Address: Not Found

Status: Ready

MAC Address*	00016473E7B4
Description	Port1 -Lennon C&K-01
Device Pool*	Default
Media Resource Group List	< None >
Network Hold Audio Source	< None >
User Hold Audio Source	< None >
Calling Search Space	< None >
Location	< None >
Load Information	D0040301U004
Channel Selection Order*	Top Down
PCM Type*	u-law
Protocol Side*	User
Caller ID DN	
Calling Party Selection*	Originator
Channel IE Type*	Use Number when IB
MCOM Channel Number Extension Bit Set to Zero**	☐
Interface Identifier Present**	☐

9776 1

Figure 4-6 Gateway Configuration in Cisco CallManager (continued)

Adaptive Call Control Enable ☐

SNMP Community String

Debug Port Enable* ☒

Hold Tone Silence Duration*

Port Used for Voice Calls* ☒

Port Used for Modem Calls* ☒

Port Used for Fax Calls* ☒

Fax and Modem Parameters

Fax Relay Enable* ☒

Fax Error Correction Mode Override* ☒

Maximum Fax Rate*

Fax Payload Size*

Non Standard Facilities Country Code*

Non Standard Facilities Vendor Code*

Fax/Modem Packet Redundancy* ☐

V.21 Flag Sequence Detection Count*

NSE Type*

Playout Delay Parameters

Initial Playout Delay*

Minimum Playout Delay*

Maximum Playout Delay*

* indicates required item
 ** applicable to DMS-100 protocol only
 *** applicable to DMS-100 protocol and DMS-250 protocol only
 **** may be required to force ringback from some PBXs

[Back to Find/List Gateways](#)

97762

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 ATA 188, VG248, 6608, and 6624 gateways, but it is supported on most of the high-performance Cisco IOS voice platforms such as the Cisco 2800 and 3800.

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

- [Loose Gateway Controlled with Network Services Engine \(NSE\), page 4-27](#)
- [Gateway Controlled with Capability Exchange Through H.245 or Session Definition Protocol \(SDP\), page 4-28](#)
- [Call-Agent-Controlled T.38 with H.323 Annex D and MGCP, page 4-29](#)

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  
!
```



Cisco CallManager Trunks

Cisco CallManager Release 4.0 introduced support for Session Initiation Protocol (SIP) trunks; prior to Release 4.0, Cisco CallManager supported only H.323 trunks. This chapter discusses design considerations for Cisco CallManager Release 4.1, but much of the discussion also applies to releases 4.0 and 3.3.

The H.323 trunk currently provides connectivity to other Cisco CallManager clusters and other H.323 devices such as gateways. H.323 trunks support most of the audio and video codecs that Cisco CallManager supports for intracluster communications, with the exception of wideband audio, wideband video, and H.264 video.

H.323 trunks use the Empty Capabilities Set (ECS) to provide supplementary call services such as hold/resume and transfer. This method is a standard H.245 mechanism to stop or close a media stream (or channel) and start or open it to the same or a different endpoint address. This method allows Cisco CallManager to keep a call active while still being able to control the source and destination of the media streams on the fly.

For example, consider a call between two clusters (A and B) using the H.323 trunk. When a user in cluster A places a user in cluster B on hold, the media streams between the two users are closed and the user in cluster B is connected to a music on hold (MoH) server in cluster A. The MoH server is instructed to send media (the music file) to the user. When the user in cluster A resumes the call, the MoH stream is closed and the two-way media streams are reopened between the two users. (Cisco CallManager does not support H.450 for supplementary call services.) In this case, MoH is an example of an ECS operation. Because the H.323 trunk does not support multicast MoH, the media resource group list (MRGL) for the H.323 trunks must contain only unicast MoH sources. (For details, see [Music on Hold, page 7-1.](#))

The bandwidth used for calls on H.323 trunks can be controlled by the use of *regions* configured in Cisco CallManager and assigned to each trunk. A region limits the amount of bandwidth allocated for calls by specifying the audio codec type and video bandwidth for that region. Calls between that region and another region must be within the specified bandwidth limit. If the device making the call over the H.323 trunk is in a more restrictive region or does not support a particular codec such as video, then it is a subset of codecs that are allowed for that call. All dual tone multifrequency (DTMF) signaling on an H.323 trunk is provided out-of-band using H.245.

SIP trunks provide connectivity to other SIP devices such as gateways, proxies, voicemail systems, and other Cisco CallManager clusters. SIP trunks have the following characteristics that differ from H.323 trunks:

- G.711 A-law and mu-law codec support
- No video support
- Requires a media termination point (MTP) per call, which also provides the RFC 2833 DTMF support

H.323 Trunks

The following major types of H.323 trunks can be configured in a Cisco CallManager:

- [Intercluster Trunk \(Non-Gatekeeper Controlled\)](#), page 5-2
- [Intercluster Trunk \(Gatekeeper Controlled\)](#), page 5-2
- [H.225 Trunk \(Gatekeeper Controlled\)](#), page 5-3

Intercluster Trunk (Non-Gatekeeper Controlled)

This trunk is the simplest and is used for connecting to other Cisco CallManager clusters in either a multi-cluster single campus or a distributed call processing deployment. This trunk does not use a gatekeeper for call admission control, although it may use locations configured in Cisco CallManager if bandwidth control is required.

When defining this type of trunk, you may define up to three remote Cisco CallManager servers in the same destination cluster. The trunk will automatically load-balance across all defined servers. In the remote cluster, it is important to configure a corresponding intercluster trunk (non-gatekeeper controlled) that has a Cisco CallManager redundancy group containing the same servers that were defined as remote Cisco CallManager servers in the first cluster. A similar configuration is required in each Cisco CallManager cluster connected by the intercluster trunks.

For example, if Cluster 1 has a trunk to Cluster 2 and Cluster 2 has a trunk to Cluster 1, the following configurations would be needed:

- Cluster 1
 - Servers B, C, and D are configured as members of the Cisco CallManager redundancy group defined in the device pool associated with the trunk to Cluster 2.
 - The non-gatekeeper controlled trunk has Cluster 2's remote servers D, E, and F configured.
- Cluster 2
 - Servers D, E, and F are configured as members of the Cisco CallManager redundancy group defined in the device pool associated with the trunk to Cluster 1.
 - The non-gatekeeper controlled trunk has Cluster 1's remote servers B, C, and D configured.

Intercluster Trunk (Gatekeeper Controlled)

The intercluster gatekeeper controlled trunk should be used instead of the non-gatekeeper controlled trunk for a larger number of clusters. The advantages of using the gatekeeper controlled trunk are mainly the overall administration of the cluster and failover times. Non-gatekeeper controlled trunks generally require that a full mesh of trunks be configured, which can become an administrative burden as the number of clusters increases. In addition, if a subscriber server in a cluster becomes unreachable, there will be a 5-second (default) timeout while the call is attempted. If an entire cluster is unreachable, the number of attempts before either call failure or rerouting over the PSTN will depend on the number of remote servers defined for the trunk and on the number of trunks in the route list or route group. If there are many remote servers and many non-gatekeeper controlled trunks, the call delay can become excessive.

With a gatekeeper controlled trunk, you configure only one trunk that can then communicate via the gatekeeper with all other clusters registered to the gatekeeper. If a cluster or subscriber becomes unreachable, the gatekeeper automatically directs the call to another subscriber in the cluster or rejects

the call if no other possibilities exist, thus allowing the call to be rerouted over the PSTN (if required) with little incurred delay. With a single Cisco gatekeeper, it is possible to have 100 clusters all registering a single trunk each, with all clusters being able to call each other. With non-gatekeeper controlled trunks, this same topology would require 99 trunks configured in each cluster. The gatekeeper controlled intercluster trunk should be used for communicating only with other Cisco CallManagers because the use of this trunk with other H.323 devices might cause problems with supplementary services. In addition, a gatekeeper controlled intercluster trunk must be used for backward compatibility with Cisco CallManager prior to Release 3.2.

H.225 Trunk (Gatekeeper Controlled)

The H.225 gatekeeper controlled trunk is essentially the same as the intercluster gatekeeper controlled trunk except that it has the capability of working with Cisco CallManager clusters Release 3.2 and later) as well as other H.323 devices such as gateways, conferencing systems, and clients. This capability is achieved through a discovery mechanism on a call-by-call basis. (See [H.323 Operation in Cisco CallManager, page 5-8](#), for details of this discovery process.) This type of trunk is the recommended H.323 trunk if all Cisco CallManager clusters are at least Release 3.2.

Gatekeeper Trunk Redundancy, Resilience, and Load Balancing

Redundancy can be achieved in several ways, depending on the requirements of the design. The simplest method is to configure a gatekeeper controlled trunk and assign up to three subscribers in the Cisco CallManager redundancy group associated with the device pool assigned to that trunk. This configuration will cause all servers to register with the same gatekeeper in the same zone with the same technology prefix. However, the H.323 trunk name that is used for the h323_id will have a suffix of "_n" where *n* is the node number in the cluster. This ID is automatically generated and cannot be changed. You configure a single trunk, but the gatekeeper registers multiple trunks, one from each subscriber in the Cisco CallManager redundancy group.

If you have additional redundancy requirements, it is possible to configure another gatekeeper controlled trunk with a different name and different subscribers in the Cisco CallManager redundancy group, but with all the other parameters identical to the first trunk. This second trunk will cause additional subscribers to register with the gatekeeper.

Cisco recommends assigning device pools that contain a Cisco CallManager redundancy group consisting of the two servers that make up the standard subscriber pair. (See [Call Processing Subscriber, page 8-5](#), for more information on subscriber redundancy.) For complete redundancy in a full cluster, four trunks would be needed using four different device pools, resulting in eight subscribers registering with the gatekeeper. (The same result could be achieved with three trunks and larger redundancy groups.)

During registration, several parameters are passed between Cisco CallManager and the gatekeeper. Cisco CallManager uses an ephemeral User Datagram Protocol (UDP) port for gatekeeper Registration Admission Status (RAS) messages on a per-trunk basis. This port would normally be UDP 1719. However, Cisco CallManager must be able to determine which trunk is the destination for a particular RAS message, therefore it uses a range of UDP ports and assigns them dynamically.

During the registration process, a trunk registers the following information for the other subscribers in its Cisco CallManager redundancy group:

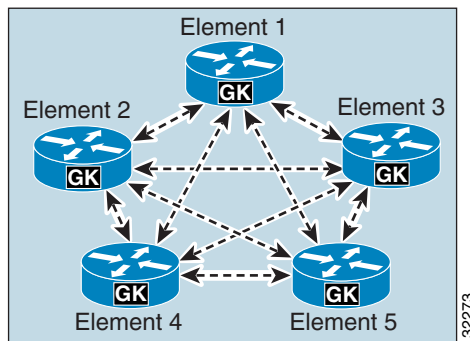
- H.225 call signaling port
- h323_id
- CanMapAlias support

- Technology prefix
- H.225 call signaling address

If the recommended clustered gatekeepers are used, the gatekeeper will return a list of alternate gatekeeper addresses that may be used if the primary gatekeeper fails or does not have sufficient available resources.

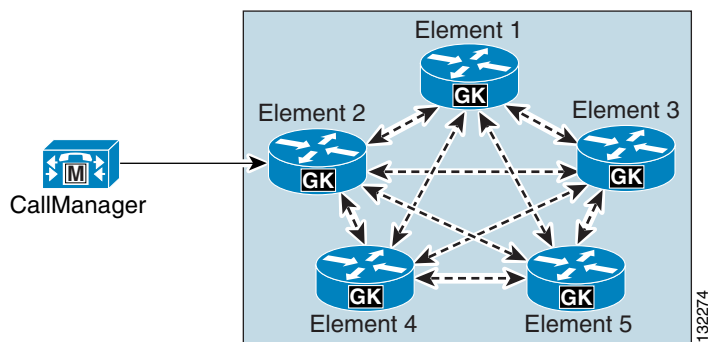
Figure 5-1 shows a cluster of gatekeepers that use Gatekeeper Update Protocol (GUP) to communicate. (See the chapter on [Call Processing](#), page 8-1, for more information on gatekeepers.)

Figure 5-1 Gatekeeper Cluster

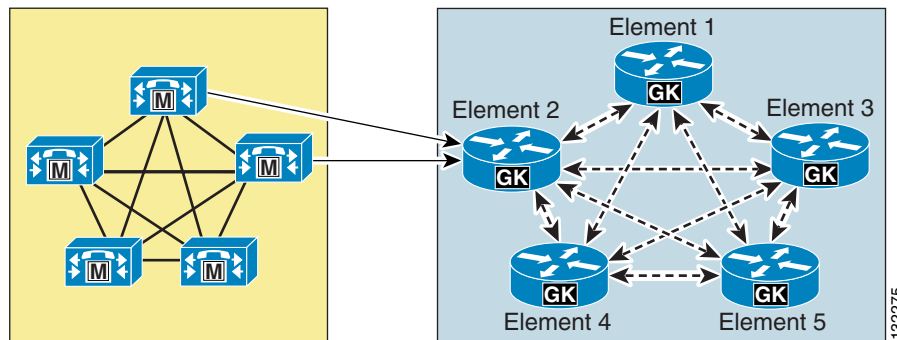


If an H.323 trunk has only a single subscriber in its Cisco CallManager redundancy group, there will be only one connection between the configured gatekeeper in Cisco CallManager and the gatekeeper cluster, as illustrated in Figure 5-2.

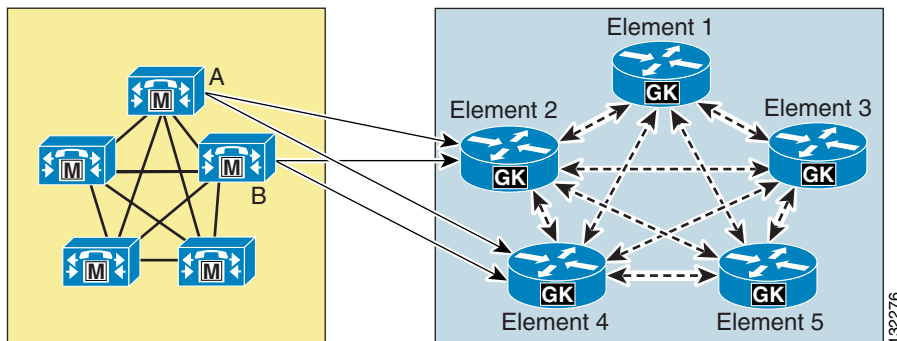
Figure 5-2 H.323 Trunk with a Single Cisco CallManager Subscriber



If there are multiple subscribers in the Cisco CallManager redundancy group associated with the trunk, additional connections will be established between the Cisco CallManager cluster and the gatekeeper cluster, as illustrated in Figure 5-3.

Figure 5-3 *H.323 Trunk with Multiple Cisco CallManager Subscribers*

This approach provides redundancy for subscriber failures as well as gatekeeper failures after registration because the alternate gatekeeper is communicated when the trunk registers. This approach does not, however, provide redundancy if the configured gatekeeper is unavailable at initial registration or following a reset because the list of alternate gatekeepers is dynamic and not stored in the database. To provide an additional level of redundancy as well as load balancing, an additional gatekeeper from the gatekeeper cluster is configured in Cisco CallManager. For example, if the original trunk is registered with Element 2, the additional gatekeeper could be configured as Element 4, as illustrated in Figure 5-4.

Figure 5-4 *Additional Gatekeeper Configured for Load Balancing and Additional Redundancy*

The Cisco CallManager configuration for the example in Figure 5-4 would contain the following components:

- Two gatekeepers for Element 2 and Element 4
- Two H.323 trunks defined with a Cisco CallManager redundancy group containing subscriber servers A and B

Using this approach, the Cisco CallManager cluster will still be able to register when either Element 2 or Element 4 is not reachable during initial registration (that is, during power-up or trunk reset).

Load balancing of calls inbound to the Cisco CallManager cluster is done automatic by default because the gatekeeper randomly selects one of the subscribers registered within the zone. If this is not the desired behavior, you can use the **gw-priority** configuration command in the gatekeeper to modify this default behavior, as illustrated in Example 5-1.

Example 5-1 Using the gw-priority Command to Direct Calls to a Particular Trunk

```
gatekeeper
zone local SJC cisco.com 10.0.1.10
zone prefix SJC 1408..... gw-priority 10 sjc-trunk_2
zone prefix SJC 1408..... gw-priority 9 sjc-trunk_3
zone prefix SJC 1408..... gw-default-priority 0
gw-type-prefix 1#* default-technology
arq reject-unknown-prefix
no shutdown
endpoint ttl 60
```

In [Example 5-1](#), the H.323 trunk was configured as sjc-trunk in Cisco CallManager, and the “_2” and “_3” suffixes are appended automatically by the Cisco CallManager subscribers to indicate which node number they are in the cluster. Therefore, this example uses node 2 as the first choice, which should be the highest-priority Cisco CallManager in the CallManager redundancy group for this trunk. Node 3 is the second choice in this case.

The use of **gw-default-priority 0** is optional. It was used in this example to disable the use of any other trunk that might accidentally be configured to register in this zone.

Outbound calls from a Cisco CallManager cluster can be load-balanced in any of the following ways:

- A single H.323 trunk in a route group will always use the highest-priority subscriber available in the Cisco CallManager redundancy group. Lower-priority subscribers are used only when higher-priority subscribers are unavailable.
- Multiple H.323 trunks in a circular route group provide equal call loading across all H.323 trunks in the group.

The following examples list how to configure load balancing in various scenarios.

All calls originate from a single subscriber in a cluster:

- Single H.323 trunk in a route group

Calls spread across four primary subscribers in a cluster:

- Four H.323 trunks with four Cisco CallManager redundancy groups, all contained within a circular route group
- Cisco CallManager redundancy groups defined as follows:
 - Subscriber A, Subscriber B
 - Subscriber C, Subscriber D
 - Subscriber E, Subscriber F
 - Subscriber G, Subscriber H

Where subscribers A, C, E, and G are all primaries and subscribers B, D, F, and H are the backups.

Calls spread across eight subscribers in a cluster:

- Eight H.323 trunks with eight different Cisco CallManager redundancy groups, each containing only one subscriber, and all contained within a circular route group
- Cisco CallManager redundancy groups defined as follows:
 - Subscriber A
 - Subscriber B
 - Subscriber C
 - Subscriber D

- Subscriber E
- Subscriber F
- Subscriber G
- Subscriber H

H.323 Trunks with Media Termination Points

Media termination points (MTPs) are generally not required for normal operation of the H.323 trunk. They are, however, required for communication with devices that are H.323 Version 1 or that do not support the Empty Capabilities Set (ECS) for supplementary services.

To test whether or not an MTP is required, use the following simple procedure:

1. Place a call from a phone via the H.323 trunk to the other device. This call should work normally.
2. Place the call on hold, then resume it. If the call drops, then it is highly likely that an MTP is required to ensure interoperability between Cisco CallManager and the other device.

MTPs are very useful for terminating media streams from other devices that make calls over the H.323 trunk and for re-originating the media streams with the same voice payload; however, in such cases the IP address is changed to that of the MTP. With this fact in mind, you can utilize MTPs in the following scenarios:

- If the phones, gateways, and other devices within your enterprise all use RFC 1918 private addresses, you might still want to connect to other systems on a public network without using Network Address Translation (NAT) for all your voice and video devices. If the Cisco CallManager subscriber that communicates to the public network is using a public IP address, the signaling will be routed. If all MTPs are also using public addresses, the media from the devices with RFC 1918 addresses will be terminated on the MTP and then originated again, but this time with a public address that is routable on the public network. This approach allows tens of thousands of devices with RFC 1918 addresses to communicate with the public network. This same method can be used to conceal the real IP addresses of devices in an enterprise network when communicating with other enterprises or service providers.
- Trust boundaries can be established to traverse firewalls or to allow access through an access control list (ACL). Normally, for media to traverse a firewall, you could either use an Application Layer Gateway (ALG) or fixup to provide access dynamically for the media streams or you could allocate a wide range of addresses and ports for use by all voice devices that need to communicate across the firewall. All calls that use the H.323 trunk and traverse a firewall or ACL will have media that is sourced from the MTP(s), which may use either a single IP address or a small range of IP addresses.

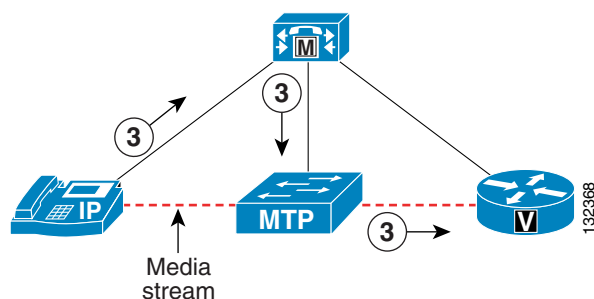
With both of these methods, if the **MTP Required** box is checked, the default behavior is to allow calls on the H.323 trunks even if MTP resources are unavailable or exhausted. This default behavior might result in no voice path for the call, but the behavior can be changed by setting the Cisco CallManager service parameter **Fail Call if MTP allocation fails** under the H.323 section to **True**.

SIP Trunks with Media Termination Points

Cisco CallManager sends and receives dual tone multifrequency (DTMF) signals using out-of-band signaling mechanisms provided by the protocol being used. Because the signaling is always between Cisco CallManager and the endpoints while media flows directly between the endpoints, this behavior is slightly different in Session Initiation Protocol (SIP), where the common method for sending DTMF is by using RFC 2833. RFC 2833 sends the DTMF digits as a specific type of RTP payload in the media stream; the DTMF digits are not sent as tones encoded by the codec. In order for the Cisco CallManager to detect and send DTMF via the SIP trunk, the MRGL associated with the SIP trunk must contain an MTP that supports RFC 2833. (See [Media Termination Point \(MTP\)](#), page 6-9 for more information.)

Figure 5-5 shows a DTMF signal for the digit 3 being sent out-of-band via Skinny Client Control Protocol (SCCP) to Cisco CallManager. It is then sent via SCCP out-of-band to the MTP, which then inserts an RFC 2833 RTP packet into the media stream to indicate a DTMF digit of 3. The same events would happen in reverse if an RFC 2833 RTP packet were received by the MTP. The packet would be extracted from the media stream and sent out-of-band to Cisco CallManager.

Figure 5-5 DTMF Signaling



H.323 Operation in Cisco CallManager

This section provides information on how the H.323 protocol is used and implemented in Cisco CallManager, and it explains how and why certain features work the way they do.

The most important point to understand is which subscribers run the call signaling daemons. These daemons are pieces of code that make and receive H.323 calls. They are usually referred to as H.225 daemons, or H.225Ds. H.225 is part of the H.323 protocol and is mainly responsible for call control. H.245 is the other major component of H.323 that is responsible for the media control of a call.

The subscribers listed in the Cisco CallManager redundancy group for a particular H.323 device determine which subscribers run the daemons and when. This point is a very important because calls sent to an incorrect subscriber might be rejected or handled by a different H.225D. For example, this situation would occur if a Cisco IOS H.323 gateway is configured with dial peers that send calls to subscriber C in a Cisco CallManager cluster but the Cisco CallManager redundancy group for that gateway has only subscribers A and B in its list. In such a case, the call will fail or be handled by an H.323 trunk daemon if one happens to be configured on the subscriber.

The following scenarios describe where and when H.225Ds are created on subscribers:

- H.323 client

The H.225D is active on only the highest-priority subscriber available in the Cisco CallManager redundancy group associated with the H.323 client.

If the H.323 client is gatekeeper controlled, the RasAggregator device registers from only the highest-priority subscriber available in the Cisco CallManager redundancy group associated with the gatekeeper controlled H.323 client.

The RasAggregator is a special device that registers in gatekeeper zones for the purpose of providing two specific features:

- If H.323 clients are using DHCP, you cannot use them with Cisco CallManager if Cisco CallManager is using DNS because they require H.323 clients to support dynamic DNS. With the RasAggregator, Cisco CallManager can obtain the IP address of a specific H.323 client that is registered with the gatekeeper whenever a call is placed. The gatekeeper registration is done using standard RAS ARQ messages that contain the E.164 address of the H.323 client. The gatekeeper resolves the E.164 address and provides the IP address back to Cisco CallManager in an ACF message.
- The RasAggregator also ensures that all calls by the H.323 clients are made through Cisco CallManager and not directly between the clients themselves, thus ensuring that dialing rules and codec restrictions are enforced.

- H.323 gateway

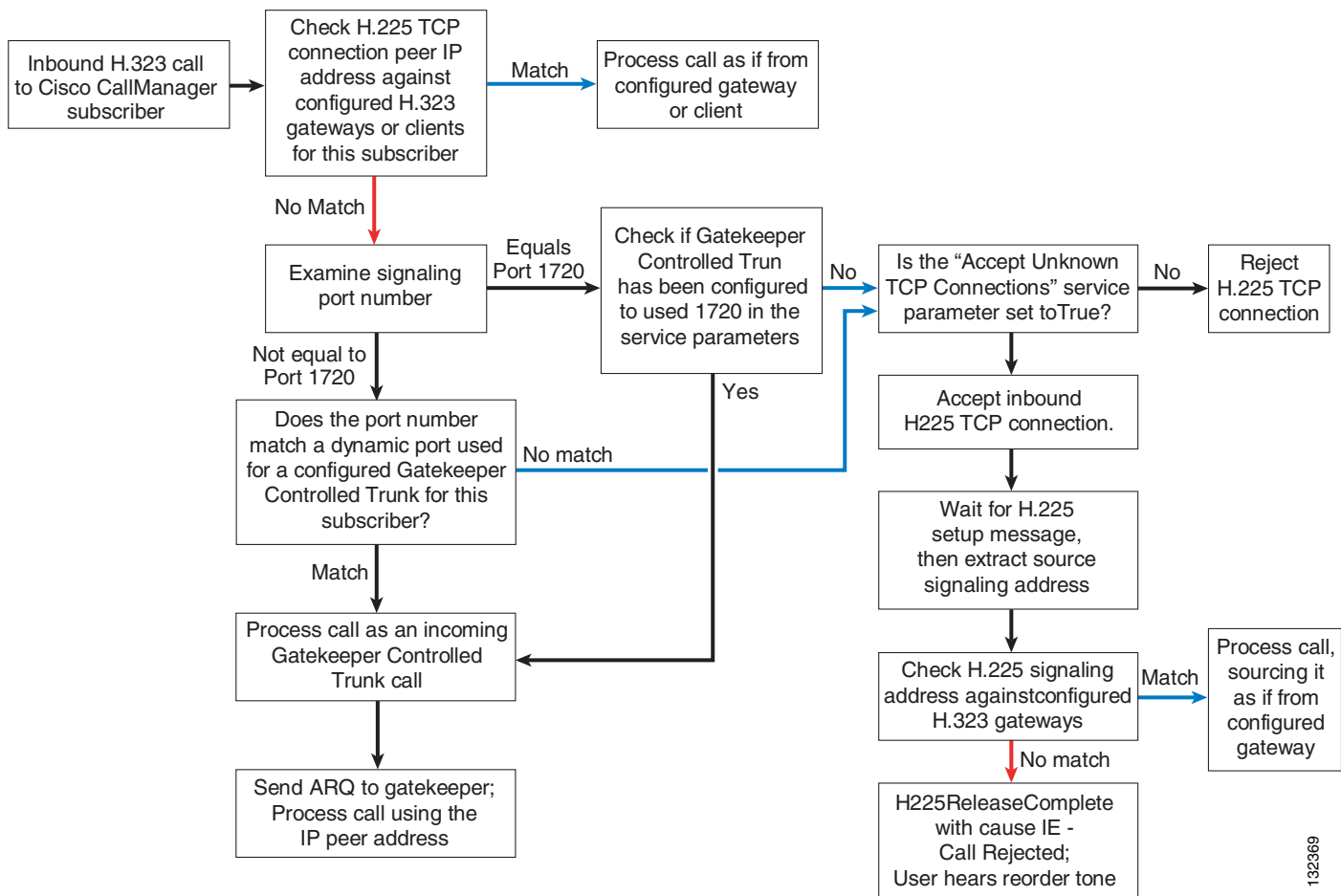
The H.225D is active on all subscribers in the Cisco CallManager redundancy group associated with the H.323 gateway.

- H.323 trunks

The H.225D is active on all subscribers in the Cisco CallManager redundancy group associated with the H.323 trunk.

A RAS daemon registers the trunk with the gatekeeper from all subscribers in the associated Cisco CallManager redundancy group.

When an incoming H.323 call is made to a subscriber in a Cisco CallManager cluster, various decisions are made to determine if the call is accepted or rejected and which H.225D will receive the call if it is accepted. [Figure 5-6](#) shows how this process works.

Figure 5-6 Process for Determining if an H.323 Call is Accepted or Rejected

Cisco CallManager H.323 protocol includes the following additional features:

- Protocol Auto Detect

This feature provides the ability to determine, on a call-by-call basis, if the calling device is from Cisco CallManager Release 3.2 or later. Whenever a call is received, Cisco CallManager looks for an H.225 User-to-User Information Element (UUIE) that indicates if the other end is another Cisco CallManager. If it is, it will always use the Intercluster Trunk Protocol. If no UUIE is found, it will use the configured protocol for that device. This feature enables an H.225 gatekeeper controlled trunk to switch between Intercluster Trunk Protocol and H.225 on a call-by-call basis, allowing a mixture of Cisco CallManager clusters and other H.323 devices to use the gatekeeper. Intercluster Trunk Protocol is the same as H.225 except for several differences that enable specific features to work correctly between Cisco CallManager clusters.

- Tunneled Q.SIG or H.323 Annex M1

With the release of Cisco CallManager 4.1(3), this feature can be enabled on all H.323 trunks. It allows specific H.323 Annex M1 features to be implemented between Cisco CallManager clusters and other verified systems that also support H.323 Annex M1. These features include:

- Path replacement
- Message waiting indication (MWI)

132369

- Callback

- Alternate Endpoints

When registering with a gatekeeper that supports this feature, such as a Cisco MCM Gatekeeper, Cisco CallManager can inform the gatekeeper of alternate destinations for calls to the H.323 trunk. These alternate endpoints or destinations are sent to the calling device by the gatekeeper when this H.323 trunk is called. They are the other subscribers listed in the Cisco CallManager redundancy group associated with the H.323 trunk that registers with the gatekeeper.

- Alternate Gatekeeper

When an H.323 trunk registers with a gatekeeper that supports this feature (for example, a Cisco gatekeeper cluster), Cisco CallManager is dynamically informed about other gatekeepers that can process registrations, call admission requests, and other RAS functions in the event that this gatekeeper fails or exhausts its own resources.

- CanMapAlias

When an H.323 trunk sends an admission request (ARQ) to the gatekeeper, it might receive a different E.164 number in the admission confirmation message (ACF), indicating that the original called number should be replaced with this new one. This feature requires a route server using Gatekeeper Transaction Message Protocol (GKTMP) to communicate with Cisco gatekeepers.



Note CanMapAlias is supported for the called number only.

- Bandwidth Requests

H.323 trunks can update the gatekeeper with bandwidth information to indicate a change in the requested bandwidth allocated to a specific call. This feature is disabled by default and is controlled by setting the Cisco CallManager service parameter **BRQ Enabled** to **True**, under the H.323 section. This feature is especially important when video is used on an H.323 trunk because the original bandwidth request is for the maximum amount allowed. Enabling this feature ensures that call admission control uses the actual bandwidth negotiated during call setup.



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 TDM circuit (coding/decoding), packetization of a stream, streaming audio (annunciation), and so forth.

This chapter focuses on the following topics pertaining to media resources:

- [Voice Termination Resources, page 6-1](#)
- [Conferencing, Transcoding, and MTP Resources, page 6-7](#)
- [Conferencing Guidelines and Application Scenarios, page 6-17](#)
- [Software MTP Resources, page 6-23](#)
- [Transcoding Guidelines and Application Scenarios, page 6-23](#)

For details about Music on Hold (MoH) media resources, refer to the chapter on [Music on Hold, page 7-1](#).

Voice Termination Resources

Voice termination applies to a call that has two call legs, one leg on a time-division multiplexing (TDM) interface and the second leg on a Voice over IP (VoIP) connection. The TDM leg must be terminated by hardware that performs coding/decoding and packetization of the stream. This termination function is performed by digital signal processor (DSP) resources residing in the same hardware module, blade, or platform. All DSP hardware on Cisco TDM gateways is capable of terminating voice streams, and certain hardware is also capable of performing other media resource functions such as conferencing or transcoding (see [Conferencing, page 6-8](#) and [Transcoding, page 6-8](#)).

[Table 6-2](#) through [Table 6-6](#) show the number of calls that each hardware platform can support, which is determined by the type of DSP chipset and the number of DSPs on the hardware. The hardware has either fixed DSP resources that cannot be upgraded or changed, or it has modular DSP resources that can be upgraded. For modular (upgradeable) hardware, [Table 6-2](#) through [Table 6-6](#) also list the maximum number of DSPs per hardware module.

The number of supported calls depends on the computational complexity of the codec used for a call and also on the complexity mode configured on the DSP. Cisco IOS enables you to configure a complexity mode on the hardware module. Some hardware platforms have two complexity modes, medium complexity and high complexity, while other hardware platforms have medium and high complexity as well as flex mode.

Medium and High Complexity Mode

To determine how many calls a module can support, locate the module in [Table 6-2](#) through [Table 6-6](#), and determine the number of DSPs it can provide as well as the desired codec type. For example, an NM-HD-2VE module with three C2510 DSPs configured in flex mode can support eight G.729 calls on each DSP, for a total of 24 calls using flex mode and G.729 codecs. Using G.711 codecs in flex mode, the same hardware could support 48 calls.

As indicate in [Table 6-1](#), if a codec is supported in medium complexity mode, it is also supported in high complexity mode, albeit with fewer supported calls.

You can configure each DSP separately as either medium complexity, high complexity, or flex mode (C5510 only). The DSP treats all calls according to its configured complexity, regardless of the actual complexity of the codec of the call. A resource with configured complexity equal or higher than the actual complexity of the incoming call must be available, or the call will fail. For example, if a call requires a high-complexity codec but the DSP resource is configured for medium complexity mode, the call will fail. However, if a medium-complexity call is attempted on a DSP configured for high complexity mode, then the call will succeed and Cisco IOS will allocate a high-complexity mode resource.

To determine the maximum number of calls supported, find the appropriate row in [Table 6-2](#) through [Table 6-6](#) that contains the desired hardware. Check the medium and high complexity columns in [Table 6-1](#) to verify which complexity modes are capable of handling the desired codecs. Then read the maximum number of supported calls for each DSP at the desired complexity mode.

Flex Mode

Flex mode, available only on hardware platforms that use the C5510 chipset, eliminates the requirement to specify the codec complexity at configuration time. A DSP in flex mode accepts a call of any supported codec type, as long as it has available processing power. The overhead of each call is tracked dynamically via a calculation of processing power in millions of instructions per second (MIPS). Cisco IOS performs a MIPS calculation for each call received and subtracts MIPS credits from its budget whenever a new call is initiated. The number of MIPS consumed by a call depends on the codec of the call, as shown in the Flex Mode column of [Table 6-1](#). The DSP will allow a new call as long as it has remaining MIPS credits greater than or equal to the MIPS required for the incoming call. The Flex Mode column of [Table 6-1](#) groups the supported codecs by the number of MIPS for a single call (either 15, 30, or 40 MIPS per call) and shows the MIPS budget available for various hardware.

Flex mode has an advantage when calls of multiple codecs must be supported on the same hardware because flex mode can support more calls than when the DSPs are configured as medium or high complexity. However, flex mode does allow oversubscription of the resources, which introduces the risk of call failure if all resources are used. With flex mode it is possible to have fewer DSP resources than with physical TDM interfaces.

For example, each DSP has a budget of 240 MIPS, for a total budget of 720 MIPS per NM-HD-2VE module. For an NM-HDV2 module, the budget per DSP is also 240 MIPS, but refer to [Table 6-2](#) to determine the total number of MIPS available based on your choice and the number of PVDMs.

Compared to medium or high complexity mode, flex mode has the advantage of supporting the most G.711 calls per DSP. In medium complexity mode a DSP can support 8 G.711 calls, while flex mode supports 16 G.711 calls.

**Note**

The product documentation uses various nomenclatures for the DSP designations. For example, the C5510 is also referred to as the C2510. In addition, the C prefix is sometimes replaced by TI, as in TI5510 or TI549.

DSP Resources for Voice Termination

Table 6-2 through Table 6-6, organized by DSP chipset, provide information on DSP support by platform, DSP density, and the number of voice terminations supported per DSP. Table 6-1 lists the codecs supported by the hardware modules in each complexity mode.

Table 6-1 **Codecs Supported by Each Complexity Mode**

Medium Complexity	High Complexity	Flex Mode
G.711 (A-law, mu-law)	G.711 (A-law, mu-law)	<i>At 15 MIPS per call:</i>
Fax/modem pass-through	Fax/modem pass-through	• G.711 (A-law, mu-law)
Clear channel	Clear channel	• Fax/modem pass-through
G.726 (32K, 24K, 16K)	G.726 (32K, 24K, 16K)	• Clear channel
GSM-FR	GSM-FR	<i>At 30 MIPS per call:</i>
Fax relay	Fax relay	• G.726 (32K, 24K, 16K)
G.729 (a, ab)	G.729	• GSM-FR
	G.729 (a, b, ab)	• Fax relay
	G.728	• G.729
	G.723.1 (32K, 24K, 16K)	• G.729 (a, b, ab)
	G.723.1a (5.3K, 6.3K)	<i>At 40 MIPS per call:</i>
	GSM-EFR	• G.728
	Modem relay	• G.723.1 (32K, 24K, 16K)
		• G.723.1a (5.3K, 6.3K)
		• GSM-EFR
		• Modem relay

Hardware based on the C5510 chipset supports medium and high complexity modes as well as flex mode, as indicated in Table 6-2.

Table 6-2 DSP Resources on Cisco IOS Hardware Platforms with C5510 Chipset

Hardware Module or Chassis	DSP Configuration	Number of Calls per DSP, and Maximum Number of Calls per Module		
		Medium Complexity 8 calls per DSP	High Complexity 6 calls per DSP	Flex Mode
VG-224	Fixed 4 DSPs	N/A	24 per platform Supported codecs: - G.711 (A-law, mu-law) - G.729a	N/A
NM-HD-1V	Fixed 1 DSP	4 calls per NM	4 calls per NM	240 MIPS per NM
NM-HD-2V	Fixed 1 DSP	8 calls per NM	6 calls per NM	240 MIPS per NM
NM-HD-2VE	Fixed 3 DSPs	24 per NM	18 per NM	720 MIPS per NM
NM-HDV2	1 to 4 of:			
NM-HDV2-2T1/ E1	PVDM2-8 ¹	4 per PVDM	3 per PVDM	120 MIPS per PVDM
	PVDM2-16	8 per PVDM	6 per PVDM	240 MIPS per PVDM
	PVDM2-32	16 per PVDM	12 per PVDM	480 MIPS per PVDM
NM-HDV2-1T1/ E1	PVDM2-48	24 per PVDM	18 per PVDM	720 MIPS per PVDM
	PVDM2-64	32 per PVDM	24 per PVDM	960 MIPS per PVDM
	½, 1, 2, 4, 8 DSPs			
2801	1 to 2 of:			
2811	PVDM2-8 ¹	4 per PVDM	3 per PVDM	120 MIPS per PVDM
	PVDM2-16	8 per PVDM	6 per PVDM	240 MIPS per PVDM
	PVDM2-32	16 per PVDM	12 per PVDM	480 MIPS per PVDM
	PVDM2-48	24 per PVDM	18 per PVDM	720 MIPS per PVDM
	PVDM2-64	32 per PVDM	24 per PVDM	960 MIPS per PVDM
2821	1 to 3 of:			
2851	PVDM2-8 ¹	4 per PVDM	3 per PVDM	120 MIPS per PVDM
	PVDM2-16	8 per PVDM	6 per PVDM	240 MIPS per PVDM
	PVDM2-32	16 per PVDM	12 per PVDM	480 MIPS per PVDM
	PVDM2-48	24 per PVDM	18 per PVDM	720 MIPS per PVDM
	PVDM2-64	32 per PVDM	24 per PVDM	960 MIPS per PVDM
3825	1 to 4 of:			
3845	PVDM2-8 ¹	4 per PVDM	3 per PVDM	120 MIPS per PVDM
	PVDM2-16	8 per PVDM	6 per PVDM	240 MIPS per PVDM
	PVDM2-32	16 per PVDM	12 per PVDM	480 MIPS per PVDM
	PVDM2-48	24 per PVDM	18 per PVDM	720 MIPS per PVDM
	PVDM2-64	32 per PVDM	24 per PVDM	960 MIPS per PVDM

1. The PVDM2-8 has a single half-capacity C2510.

Hardware that is based on the C5421 chipset may have DSPs configured as either medium complexity or high complexity. [Table 6-3](#) lists the call density per DSP, and [Table 6-1](#) lists the codecs supported in each complexity mode.

Table 6-3 DSP Resources on Cisco IOS Hardware Platforms with C5421 Chipset

Hardware Module	DSP Configuration	Number of Calls per DSP, and Maximum Number of Calls per Module	
		Medium Complexity 8 calls per DSP	High Complexity 8 calls per DSP
NM-HDA-4FXS	Fixed 2 DSPs	16 per NM	8 per NM
	<i>or</i> Fixed and 1 of DSP-HDA-16 4 DSPs	16 per NM	16 per NM
AIM-VOICE-30	Fixed	30 or 60 per AIM	16 or 30 per AIM
AIM-ATM-VOICE-30	4 DSPs		

Hardware that is based on the C549 chipset may have DSPs configured as either medium complexity or high complexity. [Table 6-4](#) lists the call density per DSP, and [Table 6-1](#) lists the codecs supported in each complexity mode.

Table 6-4 DSP Resources on Cisco IOS Hardware Platforms with C549 Chipset

Hardware Module	DSP Configuration	Number of Calls per DSP, and Maximum Number of Calls per Module	
		Medium Complexity 4 calls per DSP	High Complexity 2 calls per DSP
NM-HDV	1 to 5 of PVDM-12	12, 24, 36, or 60 per NM	6, 12, 18, or 30 per NM
NM-HDV-FARM	3 DSPs		
1751 ¹	1 to 2 of:		
1761	PVDM-256K-4	4 or 8 per NM	2 or 4 per NM
	PVDM-256K-8	8 or 16 per NM ²	4 or 8 per NM
	PVDM-256K-12	12 or 24 per NM	6 or 12 per NM
	PVDM-256K-16HD	16 or 32 per NM	8 or 16 per NM
	PVDM-256K-20HD	20 per NM	10 per NM
	1,2,3,4, or 5 DSPs		

Table 6-4 DSP Resources on Cisco IOS Hardware Platforms with C549 Chipset (continued)

Hardware Module	DSP Configuration	Number of Calls per DSP, and Maximum Number of Calls per Module	
		Medium Complexity 4 calls per DSP	High Complexity 2 calls per DSP
PA-VXA-1TE1-24+	Fixed		
PA-VXA-1TE1-30+	7 DSPs	28 per PA	14 per PA
PA-VXB-2TE1+	8 DSPs	32 per PA	16 per PA
PA-VXC-2TE1+	12 DSPs	48 per PA	24 per PA
PA-MCX-2TE1	30 DSPs	120 per PA	60 per PA
PA-MCX-4TE1	Fixed	Dependant on PA-VX(x)	Dependant on PA-VX(x)
PA-MCX-8TE1	No on-board DSPs		

1. The 1751 supports a maximum of 8 DSPs (32 channels), and you can order these modules in multiples of 2 PVDMs as long as the total is less than 32 channels. The part number indicates the number of channels.
2. The multichannel port adaptors utilize unused DSPs from PA-VXAs, PA-VXBs, or PA-VXCs across the mix backplane.

Hardware that is based on the C542 chipset supports the following codecs:

- G.711 (A-law, mu-law)
- Fax/modem pass-through
- Clear channel
- G.726 (32K, 24K, 16K)
- GSM-FR
- Fax relay
- G.729
- G.729 (a, b, ab)
- G.728
- G.723.1 (32K, 24K, 16K)
- G.723.1a (5.3K, 6.3K)
- GSM-EFR
- Modem relay

Table 6-5 lists the call density per DSP.

Table 6-5 DSP Resources on Cisco IOS Hardware Platforms with C542 Chipset

Hardware Module ¹	DSP Configuration	Codec for Calls:
NM-1V	Fixed	1 per DSP
	2 DSPs	2 per NM
NM-2V	Fixed	1 per DSP
	4 DSPs	4 per NM

1. These modules do not have any complexity mode but support all codecs equally.

Table 6-6 lists non-IOS hardware for DSP resources. All non-IOS hardware platforms have a fixed configuration of DSPs, as indicated in Table 6-6.

Table 6-6 DSP Resources on Non-IOS Hardware Platforms

Hardware Module or Platform	DSP Configuration	Number of Calls per DSP, and Maximum Number of Calls per Module	Codecs Supported
WS-6608-T1 WS-6608-E1	Fixed 64 of C549	2 per DSP 256 per module ¹	G.711 A-law, mu-law G.729a
WS-6624-FXS	Fixed 12 of C549	2 per DSP 24 per module	G.711 A-law, mu-law G.729a
VG-248	Fixed 24 of C549	2 per DSP 48 per platform	
WS-SVC-CMM-ACT	Fixed 4 of Broadcom 1500	32 per DSP 128 per module	G.711 (10-30 ms) G.729 (10-60 ms) G.723 (30-60 ms)
WS-SVC-CMM-6T1	Fixed 12 of C5441	15 per DSP 144 per module	G.711 (10, 20, 30 ms) G.729 (10, 20, 30, 40, 50, 60 ms)
WS-SVC-CMM-6E1	Fixed 12 of C5441	15 per DSP 180 per module	G.711 (10, 20, 30 ms) G.729 (10, 20, 30, 40, 50, 60 ms)
WS-SVC-CMM-24FXS	Fixed 3 of C5441	15 per DSP 24 per module	G.711 A-law, mu-law G.729 G.729a
ATA-188 ²	Fixed Komodo 3880	2 per platform	G.711 A-law, mu-law G.729

1. A maximum of 192 calls are possible with T1 and 240 calls with E1, based on the number of physical ports.

2. The ATA module does not define complexity. It supports G.711, G.729, and G.723 only.

Conferencing, Transcoding, and MTP Resources

This section describes the following types of media resources:

- [Conferencing, page 6-8](#)
- [Transcoding, page 6-8](#)
- [Media Termination Point \(MTP\), page 6-9](#)
- [Hardware Resources for MTP, Conferencing, and Transcoding, page 6-10](#)
- [Software Conferencing, page 6-14](#)
- [Annunciator, page 6-15](#)
- [Cisco IP Voice Media Streaming Application, page 6-16](#)

Conferencing

A conference bridge is a resource that joins multiple participants into a single call. It can accept any number of connections for a given conference, up to the maximum number of streams allowed for a single conference on that device. There is a one-to-one correspondence between media streams connected to a conference and participants connected to the conference. The conference bridge mixes the streams together and creates a unique output stream for each connected party. The output stream for a given party is the composite of the streams from all connected parties minus their own input stream. Some conference bridges mix only the three loudest talkers on the conference and distribute that composite stream to each participant (minus their own input stream if they are one of the talkers).

There are two main types of conference bridge resources:

- Software conference bridge

A software unicast conference bridge is a standard conference mixer that is capable of mixing G.711 audio streams and Cisco Wideband audio streams. Any combination of Wideband or G.711 A-law and mu-law streams may be connected to the same conference. The number of parties that can be supported on a given conference depends on the server where the conference bridge software is running and on the configuration for that device.

- Hardware conference bridge

A hardware conference bridge has all the capabilities of a software conference bridge. In addition, some hardware conference bridges can support multiple low bit-rate (LBR) stream types such as G.729, GSM, or G.723. This capability allows some hardware conference bridges to handle mixed-mode conferences. In a mixed-mode conference, the hardware conference bridge transcodes G.729, GSM, and G.723 streams into G.711 streams, mixes them, and then encodes the resulting stream into the appropriate stream type for transmission back to the user. Some hardware conference bridges support only G.711 conferences.

All conference bridges that are under the control of Cisco CallManager use Skinny Client Control Protocol (SCCP) to communicate with Cisco CallManager.

Cisco CallManager allocates a conference bridge from a conferencing resource that is registered with the Cisco CallManager cluster. Both hardware and software conferencing resources can register with Cisco CallManager at the same time, and Cisco CallManager can allocate and use conference bridges from either resource. Cisco CallManager does not distinguish between these types of conference bridges when it processes a conference allocation request.

[Table 6-6](#) lists the types of hardware and software conference resources available, as well as the maximum number of conference participants and bridges per resource and codec type.

Transcoding

A transcoder is a device that takes the output stream of one codec and converts it in real time (transcodes it) into an input stream for a different codec type. In other words, a transcoder converts a stream of one compression type into a stream of another compression type.

A transcoder can convert a G.711 voice stream into a low bit-rate (LBR) compressed voice stream, such as G.729a. This conversion is critical for enabling applications such as Cisco IP Interactive Voice Response (IVR), voice messaging, and conference calls over low-speed IP WANs. In addition, a transcoder provides the capabilities of a media termination point (MTP) and can be used to enable supplementary services for H.323 endpoints when required.

[Table 6-8](#) lists the maximum number of MTPs and transcoding sessions for each codec type supported on the indicated platforms.

To scale IP telephony systems in large enterprise environments, hardware conferencing is required. DSP resources in the voice modules for the hardware platforms listed in [Table 6-6](#) provide the hardware conferencing features needed for large systems, and they support the following characteristics:

- Through the use of media resource groups (MRGs) and media resources group lists (MRGLs), Cisco CallManager permits sharing of the hardware conference ports among the Cisco CallManager servers in a cluster.
- Cisco CallManager uses Skinny Client Control Protocol (SCCP) to communicate with the hardware conference bridges, but the gateway services on those platforms may use Media Gateway Control Protocol (MGCP).

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.

Media Termination Point (MTP)

A media termination point (MTP) is an entity that accepts two full-duplex G.711 streams. It bridges the media streams together and allows them to be set up and torn down independently. The streaming data received from the input stream on one connection is passed to the output stream on the other connection, and vice versa. In addition, the MTP can be used to transcode A-law to mu-law and vice versa, or it can be used to bridge two connections that utilize different packetization periods (different packet sizes). MTPs are also used to provide further processing of a call, such as RFC 2833 support.

MTPs can be used to extend supplementary services to H.323 endpoints that do not support the H.323v2 OpenLogicalChannel and CloseLogicalChannel request features with the EmptyCapabilitiesSet feature. When needed, an MTP is allocated and connected into a call on behalf of an H.323 endpoint. Once inserted, the media streams are connected between the MTP and the H.323 device, and these connections are present for the duration of the call. The media streams connected to the other side of the MTP can be connected and disconnected as needed to implement features such as hold, transfer, and so forth.



Note

Implementations prior to Cisco CallManager Release 3.2 required MTPs to provide supplementary services for H.323 endpoints, but Cisco CallManager Release 3.2 and later no longer require MTP resources to provide this functionality.

There are two main types of MTPs:

- Software MTP

A software MTP is a device that is implemented by installing the Cisco IP Voice Media Streaming Application on a server. When the installed application is configured as an MTP application, it registers with a Cisco CallManager node and informs Cisco CallManager of how many MTP resources it supports. A software MTP device supports only G.711 streams.

A Cisco IOS Enhanced software device may be implemented on a Cisco IOS router by configuring a software-only MTP under a DSP farm. This DSP farm may be used only as a pure MTP and does not require any hardware DSPs on the router.

- Hardware MTP

A hardware MTP is a DSP-based resource that resides on hardware external to Cisco CallManager. Transcoders have MTP capabilities, and a hardware MTP is really a transcoder used as an MTP. The hardware device registers with a Cisco CallManager node as a transcoder and informs Cisco CallManager of how many resources it supports. Hardware MTPs that have transcoder functionality can accept streams other than G.711.

MTPs for SIP RFC 2833 Support

SIP endpoints have the ability, as defined by the SIP specifications, to send DTMF digits and call progress tones in-band in the media stream. The DTMF tones are sent in the RTP stream as modified RTP packets. Skinny Client Control Protocol (SCCP) endpoints do not have this capability as of Cisco CallManager Release 4.1. Therefore, to integrate a Cisco CallManager system with a SIP system, you must insert an MTP in the media stream to convert SIP in-band signaling into SCCP out-of-band signaling, and vice versa.

As of Cisco CallManager Release 4.1, both software and hardware are capable of providing RFC 2833 support. The Ad-hoc Conferencing and Transcoding (ACT) Port Adapter, minimum code version 12.3(8)XY2, and the Cisco IP Voice Media Streaming Application both support RFC 2833.

You must provision MTPs for calls that are to traverse a SIP trunk. If two Cisco CallManager clusters are connected via a SIP trunk, then each cluster requires its own MTP resources for this purpose. A software MTP can support 512 streams when implemented on a dual-CPU server dedicated for this purpose.

Hardware Resources for MTP, Conferencing, and Transcoding

The resources for implementing these features may be hardware or software based. Software resources can reside on a Cisco CallManager server or on a Cisco IOS platform, and hardware platforms can be either a Cisco IOS or Cisco Catalyst platform. This section summarizes the codec type and number of streams supported by all the available resources. These resources are configured on the gateway as a DSP farm and used by Cisco CallManager via the SCCP protocol. The gateway is responsible for control and allocation of the resources.

The support of these features depends on the hardware platform, Cisco CallManager version, and Cisco IOS version. [Table 6-7](#) lists the minimum software releases required for the modules and platforms to support MTP, conferencing, and transcoding. Use [Table 6-8](#) to determine the support of media resource modules in various Cisco IOS platforms. The tables list only the modules that support media resources.



Note

The Cisco VG200, 2620, 2621, and 3620 do not support the NM-HDV-FARM, nor do they support MTP, conferencing, or transcoding. The Cisco 2801 has no NM slot. DSP farm services are not supported for Cisco Survivable Remote Site Telephony (SRST) or Cisco CallManager Express

Table 6-7 Minimum Software Releases for Conferencing, Transcoding, and MTP

Cisco Platform	Module	MTP	Conferencing and Transcoding
2691 and 2600XM 2811, 2821, and 2851 3640, 3640A, and 3660 3725 and 3745 3825 and 3845	NM-HD-1V NM-HD-2V NM-HD-2VE	Cisco CallManager 4.0(1) Cisco IOS: 12.3(8)T4 12.3(11)T for 3800 Series	Cisco CallManager 3.3(4) or 4.0(1) Cisco IOS: 12.3(11)T for 3800 Series 12.3(8)T4 for all others listed
2691 and 2600XM 2811, 2821, and 2851 3725 and 3745 3825 and 3845	NM-HDV2 NM-HDV2-1T1/E1 NM-HDV2-2T1/E1	Cisco CallManager 4.0(1) Cisco IOS: 12.3(8)T4 12.3(11)T for 3800 Series	Cisco CallManager 3.3(4), 4.0(1) Cisco IOS: 12.3(11)T for 3800 Series 12.3(8)T4 for all others listed
VG200 2691 and 2600XM 2650 and 2651 2811, 2821, and 2851 3640, 3640A, and 3660 3725 and 3745 3825 and 3845	NM-HDV NM-HDV-FARM	N/A	Cisco CallManager 3.2(2c) Cisco IOS: 12.1(5)YH for VG200 with HDV-FARM 12.2(13)T for listed 2600 Series, 3600 Series, and VG200 with HDV.
1751, 1751V, and 1761	PVDM-256K	N/A	Cisco CallManager 3.3(4) Cisco IOS 12.3(2)XE
2801	PVDM2-8 PVDM2-16 PVDM2-32 PVDM2-48 PVDM2-64	Cisco CallManager 4.1 Cisco IOS 12.3(11)T	Cisco CallManager 4.1 Cisco IOS 12.3(11)T No conferencing on PVDM2-8.
2811, 2821, and 2851	PVDM2-8 PVDM2-16 PVDM2-32 PVDM2-48 PVDM2-64	Cisco CallManager 3.3(5), 4.1 Cisco IOS 12.3(8)T4	Cisco CallManager 3.3(5), 4.1 Cisco IOS 12.3(8)T4 No conferencing on PVDM2-8.

As described in the following sections, the allocation of resources is handled differently on C549 platforms than it is on C5510 platforms.

Allocating Resources to Voice Termination or a DSP Farm

You can configure the DSP resources on a module as either voice termination of a voice trunk group or as a DSP farm. Within a DSP farm, you can allocate the DSPs to either conferencing or transcoding/MTP. [Table 6-8](#) shows the number of participants that each conference bridge can support and the number of sessions supported per DSP for conferencing, transcoding, or MTP. Use [Table 6-2](#) through [Table 6-6](#) to determine the number of DSPs available on a module, and use [Table 6-9](#) to determine the maximum number of NM modules that a chassis can support. From these tables, you can determine the necessary hardware configuration.

Table 6-8 *Number of Supported Conferencing, Transcoding, and MTP Sessions per DSP*

Module	Conference		Transcode G.711 to or from:			MTP
	G.711 (A-law, mu-law)	G.729 G.729a,b,ab	G.711 (A-law, mu-law)	G.729a G.729ab GSM-FR	G.729 G.729b GSM-EFR	G.711 (A-law, mu-law)
NM-HDV2 ¹ NM-HD-1V, -2V, -2VE 8 participants per conference	8	2	8	8	6	16
NM-HDV NM-HDV-FARM 6 participants per conference	1	1	4	4	4	N/A
PVDM-256K 1751 6 participants per conference	1 Maximum of 5 per chassis	1	2 Maximum of 16 per chassis	2 Maximum of 16 per chassis	2	N/A
PVDM-256K 1760 6 participants per conference	1 Maximum of 30 per chassis	1	2 Maximum of 20 per chassis	2 Maximum of 20 per chassis	2	N/A
PVDM2-8 PVDM2-16 PVDM2-32 PVDM2-48 PVDM2-64	8	2	8	8	6	16
WS-6608-T1 ¹ WS-6608-E1 3 to 32 participants per conference	32 per port	32 per port G.729a only	24 per port	24 per port	24 per port	24 per port
WS-SVC-CMM-ACT ² 8 participants per conference	128	128	128	128	128	256
Cisco IP Voice Media Streaming Application ³ 3 to 64 participants per conference	48 per coresident server	N/A	N/A	N/A	N/A	N/A

1. Does not support RFC 2833 and may not be used for RFC 2833 integration.
2. Supports RFC 2833 and may be used for SIP integration.
3. Supports RFC 2833 and may be used for SIP integration. The IP Voice Media Streaming Application does not use DSPs, but it is included in this table for comparison purposes.

Resource Allocation on the NM-HDV (C549-Based Hardware)

You configure each DSP individually, and each DSP functions independently of the others. The conferencing, transcoding, and MTP resources must be allocated to different DSPs, and a single DSP can support only one of these functions at a time. The configuration specifies which function each DSP will perform.

An NM-HDV module may be associated with only a single Cisco CallManager.

Resource Allocation on the NM-HDV2, NM-HD-xx, and PVDM2 (C5510-Based Hardware)

Hardware resources based on the C5510 chipset are allocated using DSP profiles that define the resource type within the profile. The profile is associated with a DSP farm so that DSPs are not defined as specific resource types.

DSPs from a DSP farm may be used by multiple Cisco CallManagers.

Calculating DSP Requirements for NM-HDV

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 sampling rate of 10 ms 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.

$$100 \text{ pps} * (\text{number of voice terminations}) + 600 \text{ pps} * (\text{number of conferences}) + 200 \text{ pps} * (\text{number of transcoding sessions}) < 6600 \text{ pps}$$

Platform Support of DSP Resources

Cisco 2800 and 3800 Series

Although the Cisco 2800 and 3800 Series routers all have two AIM slots, they do not support the AIM-VOICE-30 or AIM-ATM-VOICE-30 cards because the functionality of these cards is provided instead by PVDM2s that are installed on the motherboard.

Network Modules

You can install the NM-HDV2, NM-HD-xx, and NM-HDV modules in the Cisco IOS platforms listed in [Table 6-9](#), up to the maximum numbers indicated.

All three families of modules in [Table 6-9](#) may be installed in a single chassis. However, the conferencing and transcoding features cannot be used simultaneously on both the NM-HDV family and either of the other two families (NM-HD-xx or NM-HDV2).

The NM-HDV (TI-549), NM-HD-xx, and NM-HDV2 (TI-5510) cannot be used simultaneously for conferencing and transcoding within a single chassis.

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 6-9](#) shows the maximum number of modules that each type of hardware platform can support.

Table 6-9 Maximum Number of Modules per Cisco IOS Platform

Cisco IOS Platform	Number of Slots	Modules Supported by Platform		
		NM-HDV2	NM-HD-1V NM-HD-2V NM-HD-2VE	NM-HDV NM-HDV-FARM
VG200 ¹	1	N	N	Y
2600 ²	1			
3620 ³	2			
2600XM, 2691	1	Y	Y	Y
3640	3	N	N	Y
3660	6	N	Y	Y
3725	2	Y	Y	Y
3745	4	Y	Y	Y
2811	1	Y	Y	Y
2821	1	Y	Y	Y
2851	1	Y	Y	Y
3825	2	Y	Y	Y
3845	4	Y	Y	Y

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. Although it has two NM slots, the Cisco 3620 Router supports only a single NM-HDV module.

Software Conferencing

The software conference service is the Cisco IP Voice Media Streaming Application, and it runs on any approved platform for Cisco CallManager. (See the chapter on [Call Processing](#), page 8-1, for a list of approved platforms.) This application can be configured to operate and register with Cisco CallManager as any of the following device types: software conference bridge, media termination point (MTP), annunciator, or music on hold (MoH) server.

The Cisco IP Voice Media Streaming Application uses the IP Voice Media Streaming Driver to do real-time voice media streaming. This application also obtains the configuration information both for itself and for the IP Voice Media Streaming Driver. Because the software conference bridge supports only G.711 codecs, it is best suited for single-site deployments, where transcoding is not needed for conference calls. It is a relatively scalable solution for a software conferencing service that does not require scheduling features.

Software conference bridges have the following characteristics:

- Support for G.711 (A-law and mu-law) and Cisco Wideband
- Maximum of 64 participants in Ad Hoc conferences, or 128 in a Meet Me conference
- All low bit-rate (LBR) calls must be transcoded prior to joining the conference call

**Note**

If calls using codecs other than G.711 want to join a software conference bridge, a separate transcoder resource must be available.

Annunciator

An annunciator is a software function of the Cisco IP Voice Media Streaming Application that provides the ability to stream spoken messages or various call progress tones from the system to a user. It is capable of sending multiple one-way RTP streams to devices such as Cisco IP phones or gateways, and it uses SCCP messages to establish the RTP stream. The device must be capable of SCCP to utilize this feature. Tones and announcements are predefined by the system. The announcements support localization and also may be customized by replacing the appropriate .wav file. The annunciator is capable of supporting G.711 A-law and mu-law, G.729, and Wideband codecs without any transcoding resources.

The following features require an annunciator resource:

- Cisco Multilevel Precedence Preemption (MLPP)

This feature has streaming messages that it plays in response to the following call failure conditions.

 - Unable to preempt due to an existing higher-precedence call.
 - A precedence access limitation was reached.
 - The attempted precedence level was unauthorized.
 - The called number is not equipped for preemption or call waiting.
- Integration via SIP trunk

SIP endpoints have the ability to generate and send tones in-band in the RTP stream. Because SCCP devices do not have this ability, an annunciator is used in conjunction with an MTP to generate or accept DTMF tones when integrating with a SIP endpoint. The following types of tones are supported:

 - Call progress tones (busy, alerting, and ringback)
 - DTMF tones
- Cisco IOS gateways and intercluster trunks

These devices require support for call progress tone (ringback tone).

- System messages

During the following call failure conditions, the system plays a streaming message to the end user:

- A dialed number that the system cannot recognize
- A call that is not routed due to a service disruption
- A number that is busy and not configured for preemption or call waiting

- Conferencing

During a conference call, the system plays a barge-in tone to announce that a participant has joined or left the bridge.

An annunciator is automatically created in the system when the Cisco IP Voice Media Streaming Application is activated on a server. If the Media Streaming Application is deactivated, then the annunciator is also deleted. A single annunciator instance can service the entire Cisco CallManager cluster if it meets the performance requirements (see [Annunciator Performance, page 6-16](#)); otherwise, you must configure additional annunciators for the cluster. Additional annunciators can be added by activating the Cisco IP Voice Media Streaming Application on other servers within the cluster.

The annunciator registers with a single Cisco CallManager at a time, as defined by its device pool. It will automatically fail over to a secondary Cisco CallManager if a secondary is configured for the device pool. Any announcement that is playing at the time of an outage will not be maintained.

An annunciator is considered a media device, and it can be included in media resource groups (MRGs) to control which annunciator is selected for use by phones and gateways.

Annunciator Performance

By default, the annunciator is configured to support 48 simultaneous streams, which is the maximum recommended for an annunciator running on the same server (coresident) with the Cisco CallManager service. If the server has only 10 Mbps connectivity, lower the setting to 24 simultaneous streams.

A standalone server without the Cisco CallManager service can support up to 255 simultaneous announcement streams, and a high-performance server with dual CPUs and a high-performance disk system can support up to 400 streams. You can add multiple standalone servers to support the required number of streams.

Cisco IP Voice Media Streaming Application

The Cisco IP Voice Media Streaming Application provides the following resources in software:

- Music on Hold (MoH)
- Annunciator
- Software conference bridge
- Media termination point (MTP)

When the Media Streaming Application is activated, one of each of the above resources is automatically configured. If the annunciator, software conference bridge, or MTP are not needed, Cisco recommends that you disable them by disabling the Run Flag service parameter for the Cisco IP Voice Media Streaming Application.

Give careful consideration to situations that require multiple resources and to the load they place on the Media Streaming Application. Each resource has a service parameter that controls the maximum number of connections it can handle, with an associated default setting. You may run all four resources on the

same server if you have not changed the default settings. However, if your deployment requires more than the default number of any one resource, configure that resource to run on its own dedicated server (without any other resources or the Cisco CallManager service running on that server).

Conferencing Guidelines and Application Scenarios

This section provides conferencing resource guidelines for the following Cisco IP Telephony deployment models:

- [Conferencing Guidelines for All Deployment Models, page 6-17](#)
- [Conferencing Guidelines for Single-Site Deployments, page 6-17](#)
- [Conferencing Guidelines for Multi-Site WAN Deployments with Centralized Call Processing, page 6-18](#)
- [Conferencing Guidelines for Multi-Site WAN Deployments with Distributed Call Processing, page 6-21](#)

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.
- Each type of conferencing resource has different limits on the maximum number of participants in a single bridge (see [Table 6-8](#)). If multiple resources are defined in an MRGL, the first available resource is used. Two Cisco Callmanager service parameters control the maximum number of participants: Maximum Ad Hoc Conferences, which can be 3 to 64, and Maximum Meet Me Conference Unicast, which can be 1 to 128. The default setting for both is 4. The maximum number of participants for the conference resource overrides the service parameter settings.

To maximize the conference bridge size, set the service parameters to match the smallest bridge size of the resources in the MRGL. Otherwise, maximize bridge size by using only resources of homogenous characteristics, and set the service parameters to match.
- CTI applications and the Drop Any Party feature do not support more than 16 participants. Although some Ad Hoc conference resources can support more than 16 participants, these applications display only the 16 most recent participants.

Conferencing Guidelines for Single-Site Deployments

In a single-site deployment, no voice traffic travels over the IP WAN, and only one type of codec (usually G.711) is used. Therefore, this type of deployment does not require transcoding resources for conference calls, so you can use software conferencing. If more conferencing capacity is needed, you can add hardware conferencing resources.

The following guidelines apply to 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.

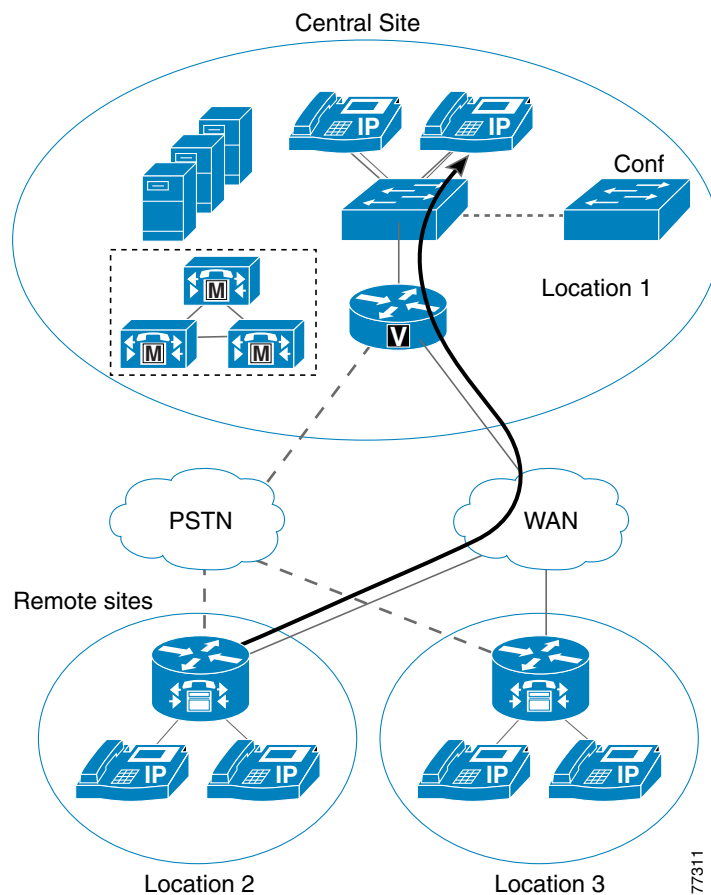
Conferencing Guidelines for Multi-Site WAN Deployments with Centralized Call Processing

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.
 - Centrally located resources can cause local calls to traverse the WAN, so you must consider the effects on bandwidth consumption. (See the chapter on [Call Admission Control](#), page 9-1.)
- 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 and any devices that do not support G.729. All Cisco devices are currently capable of supporting both codec types. Cisco Customer Response Solution (CRS), Release 3.0 and later, is capable of supporting G.729 or G.711 but not both on the same installation.

All centralized call processing deployments use the locations mechanism in Cisco CallManager to provide call admission control. Cisco CallManager uses locations in conjunction with regions to define the characteristics of a network link. Regions define the type of compression used on the link, and locations define the amount of available bandwidth for the link. [Figure 6-1](#) depicts a typical centralized call processing model with locations that are constrained by the amount of bandwidth allocated for voice calls. Because of the bandwidth limitations, the regions in this configuration use a low-bit-rate (LBR) codec such as G.729. Each LBR call uses approximately 24 kbps instead of the 80 kbps required for a G.711 call.

Figure 6-1 Locations-Based Call Admission Control for Centralized Call Processing Deployments

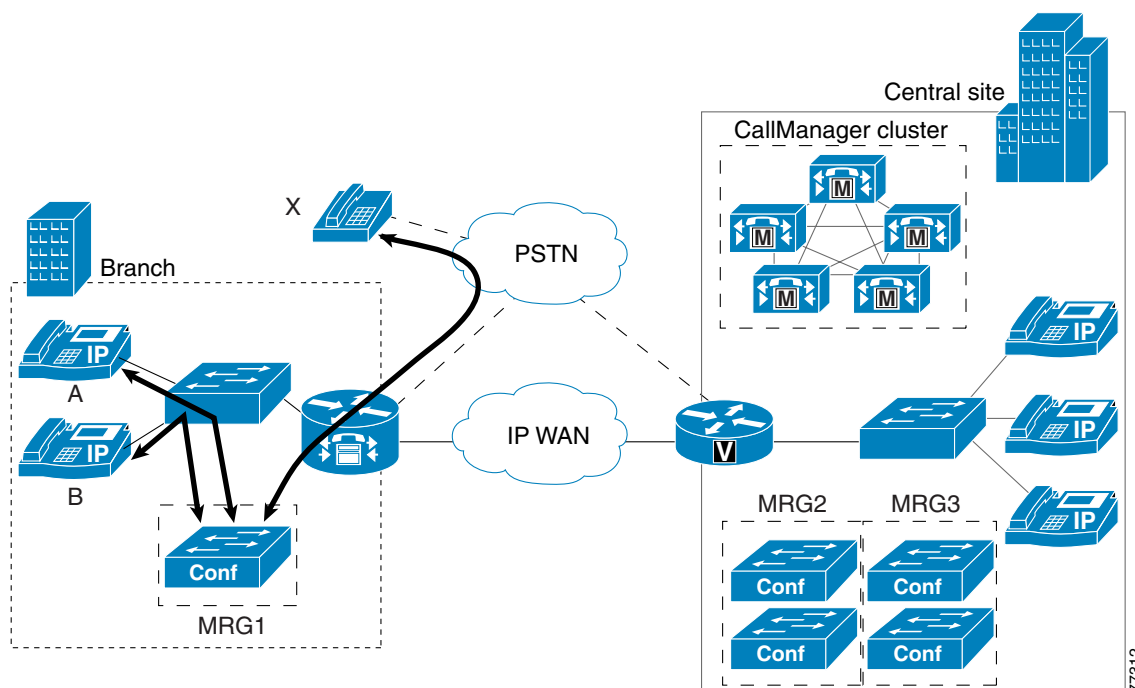


77311

Media Resource Groups and Lists

Through the use of media resource groups and lists in Cisco CallManager, you can provide conferencing resources at the remote sites, thus minimizing bandwidth usage on the WAN. (See [Figure 6-2](#).) Call admission control based on locations is still required to ensure that WAN bandwidth is not oversubscribed.

Figure 6-2 Localized Conferencing Resources in a Centralized Call Processing Deployment



In [Figure 6-2](#), the phones at the remote branch site are assigned a media resource group list (MRGL) that contains the media resource group MRG1, which in turn contains the conference bridge resource at that site. This configuration allows for conferencing within that site without using any WAN bandwidth. For example, assume Phone X calls Phone A, and Phone A conferences Phone B. At this point, conferencing resources are requested by Cisco CallManager to host the conference call. Because of the MRG and MRGL configuration, Cisco CallManager selects the conference bridge at the branch site for this conference call.

Media Resource Allocation in Cisco CallManager

It is imperative to understand how Cisco CallManager allocates resources in a media resource group (MRG). As noted previously, an MRG is part of a media resource group list (MRGL), and the MRGL is associated with a device pool. When allocating resources, Cisco CallManager first selects the appropriate MRG according to the order specified in the MRGL. Within an MRG, resources are listed alphabetically by name regardless of the order in which you added the resources to the MRG, and Cisco CallManager allocates resources from the MRG in the order listed (that is, alphabetically). Therefore, if you require prioritized allocation of a resource (such as hardware conference bridges before software conference bridges), you have to name those resources properly in the MRGs and/or configure the appropriate MRGLs to define the desired order.

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

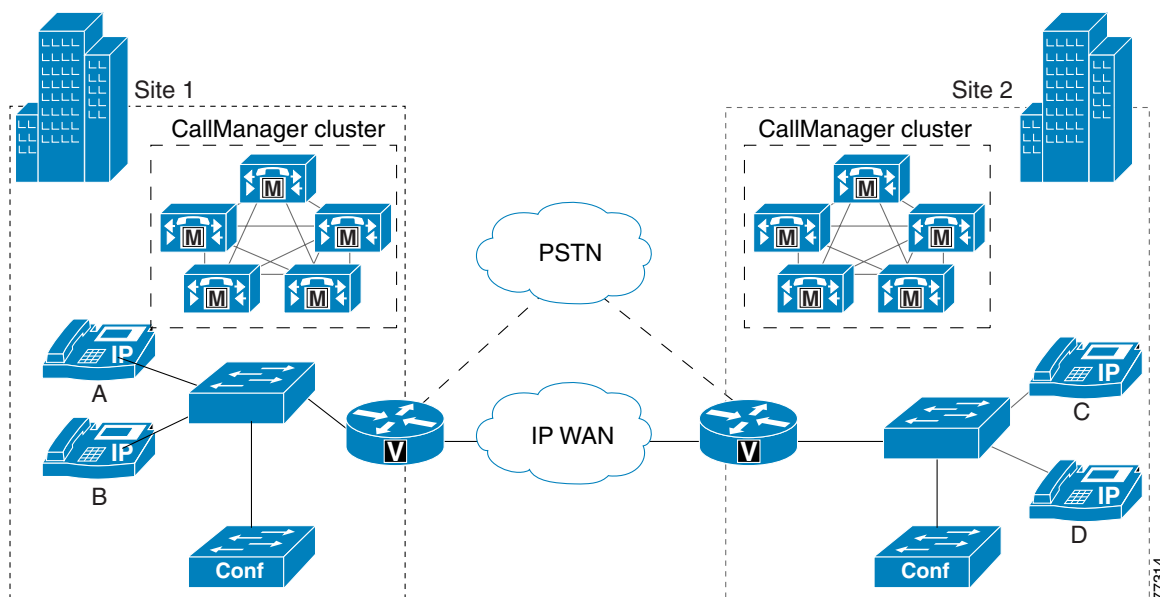
In distributed call processing deployments, several sites are connected through an IP WAN. Each site contains its own call processing entity, such as a Cisco CallManager cluster, and can in turn follow the single-site model or the centralized call processing model. A gatekeeper may be used for call admission control between sites. Also in this case, WAN bandwidth is typically limited, so inter-site calls are usually configured to use a low bit-rate codec (such as G.729) when traversing the WAN.

The following guidelines apply to distributed call processing deployments:

- 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. (See [Figure 6-3](#).)
- 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 the chapter on [Call Admission Control](#), page 9-1.)
- Within a single cluster, the maximum number of conference participants is determined by the type of resource and the service parameter settings. However, with participants from multiple clusters, that maximum number can be exceeded in 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 its maximum number of participants.

[Figure 6-3](#) provides an example of how conferencing resource allocation works across multiple Cisco CallManager clusters.

Figure 6-3 Localized Resources for a Distributed Call Processing Deployment



In [Figure 6-3](#), two Cisco CallManager clusters at Site 1 and Site 2 are connected through an IP WAN and intercluster trunks. Each site has its own conferencing resources that are assigned to individual IP phones through MRGs and MRGLs.

Assume that phone A at Site 1 calls phone C at Site 2 and then conferences phone D at Site 2. Because the conference initiator is controlled by the Cisco CallManager cluster at Site 1, the conference bridge allocated to this call is one located at Site 1. This means that there are now two streams traversing the WAN, one between phone A and phone C and another between phone A and phone D. If, on the other hand, the conference had been initiated by phone C, which is controlled by the Cisco CallManager cluster at Site 2, the call would have been assigned the conference bridge located at Site 2, and only one stream (between phone C and phone A) would have traversed the WAN.

Note that, when a conference call crosses multiple Cisco CallManager clusters, it is possible to exceed the maximum number of participants defined in each cluster. Consider the following example, based on [Figure 6-3](#):

Phone A at Site 1 uses a conference bridge located at Site 1 to set up a six-party conference call that includes phone C at Site 2. The service parameters for both clusters are set so that conference sizes are limited to six participants. Phone C can now set up another conference call for another five parties and “join” the two conferences together using a conference bridge located at Site 2 (assuming that there is sufficient bandwidth to accommodate all voice streams).



Note

If all resources in the MRGL are exhausted, Cisco CallManager looks for media resources in the default, or <None>, MRG.

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.

Transcoding Guidelines and Application Scenarios

This section examines where and when the MTP and transcoding resources are used within the following three enterprise IP Telephony deployment models plus a fourth application scenario:

- [Single-Site Deployments, page 6-23](#), consist of one or more call processing agents within a single site, with no voice traffic carried over the IP WAN.
- [Multi-Site WAN Deployments with Centralized Call Processing, page 6-23](#), consist of a single call processing agent that provides service for many sites connected through an IP WAN.
- [Multi-Site WAN Deployments with Distributed Call Processing, page 6-25](#), consist of call processing agents residing at each of several remote sites connected through an IP WAN.
- [IP PSTN Access, page 6-26](#), is another scenario that requires MTP resources, and it applies to all of the preceding deployment models.

Single-Site Deployments

In a single-site deployment, there is no need for transcoding because there are no low-speed links to justify the use of a low bit-rate (LBR) codec. Some MTP resources might be required in the presence of a significant number of devices that are not compliant with H.323v2, such as older versions of Microsoft NetMeeting or certain video devices. Another case in which MTP resources might be needed is that of a single site accessing the PSTN via an IP PSTN provider (see [IP PSTN Access, page 6-26](#)).

Multi-Site WAN Deployments with Centralized Call Processing

In a centralized call processing deployment, the Cisco CallManager cluster and the applications (such as voice mail and IVR) are located at the central site, while several remote sites are connected through an IP WAN. The remote sites rely on the centralized Cisco CallManagers to handle their call processing.

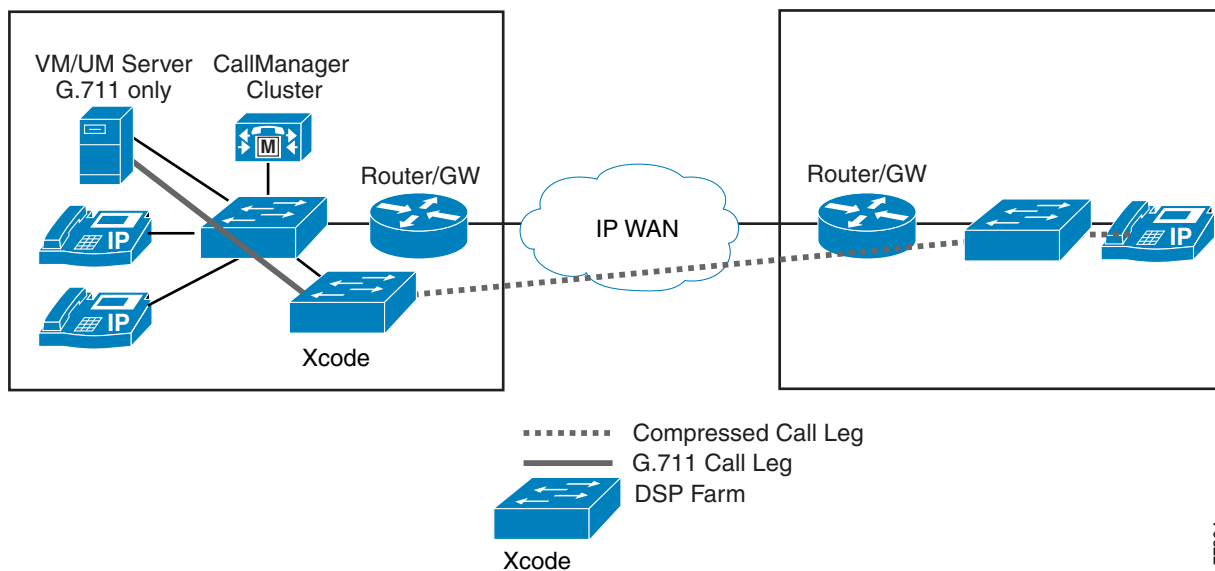
Because WAN bandwidth is typically limited, calls are configured to use a low bit-rate codec, such as G.729, when traversing the WAN. See [Figure 6-4](#).

Voice compression between IP phones is easily configured through the use of *regions* and *locations* in Cisco CallManager. A region defines the type of compression (for example, G.711 or G.729) used by the devices in that region, and a location specifies the total amount of bandwidth available for calls to and from devices at that location.

With current versions of Cisco applications, it is possible and recommended to avoid use of transcoding resources. The Cisco CRS (Release 3.0 or later) applications can support either G.711 or G.729, but not both simultaneously. If both codecs are required, then you can use two separate servers. If you are using a single server with G.711, then transcoding resources become necessary.

Prior to Cisco CRS Release 3.0, the applications supported only G.711, which requires use of transcoding when two codec types are used in a centralized call processing deployment.

Figure 6-4 Transcoding for the WAN with Centralized Call Processing



Cisco CallManager uses media resource groups (MRGs) to enable sharing of MTP and transcoding resources among the Cisco CallManager servers within a cluster. In addition, when using an LBR codec (for example, G.729) for calls that traverse different regions, the transcoding resources are used only if one (or both) of the endpoints is unable to use the LBR codec. This means that, even with G.711-only applications at the central site and H.323 gateways at the remote sites, the gateways are not always required to go through a transcoder.



Note

For Cisco CallManager releases prior to Release 3.3(2), the MTP and transcoding resources must be located at the central site with the Cisco CallManager cluster because they cannot be configured in a *location* for call admission control.

In summary, Cisco CallManager provides the following features:

- Optimum utilization of the transcoding resources through resource sharing across the Cisco CallManager cluster
- Efficient utilization of the resources through dynamic allocation of transcoding resources only when required by one or both of the endpoints

Multi-Site WAN Deployments with Distributed Call Processing

In distributed call processing deployments, several sites are connected through an IP WAN. Each site contains a Cisco CallManager cluster that can, in turn, follow the single-site model or the centralized call processing model. A gatekeeper may be used for call admission control between sites.

Because WAN bandwidth is typically limited, calls between sites are configured to use an LBR codec (such as G.729) when traversing the WAN. H.323v2 intercluster trunks are used to connect Cisco CallManager clusters. Cisco CallManager also supports compressed voice call connections through the MTP service if a hardware MTP is used. (See [Figure 6-5](#).)

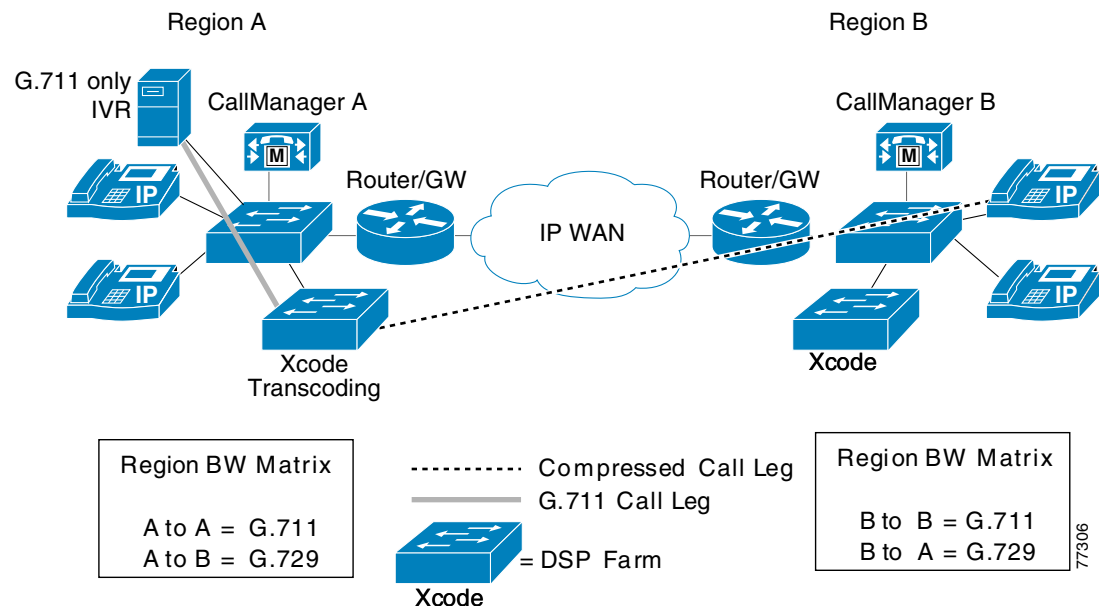
A distributed call processing deployment might need transcoding and MTP services in the following situations:

- With current versions of Cisco applications, it is possible and recommended to avoid use of transcoding resources. The Cisco CRS (Release 3.0 or later) applications can support either G.711 or G.729, but not both simultaneously. If both codecs are required, then you can use two separate servers. If you are using a single server with G.711, then transcoding resources become necessary.

Prior to Cisco CRS Release 3.0, the applications supported only G.711, which requires use of transcoding when two codec types are used in a distributed call processing deployment.

- Some endpoints (for example, video endpoints) do not support the H.323v2 features.

Figure 6-5 Intercluster Call Flow with Transcoding



Cisco CallManager uses media resource groups (MRGs) to enable sharing of MTP and transcoding resources among the Cisco CallManager servers within a cluster. In addition, for calls across intercluster trunks, MTP and transcoding resources are used only when needed, thus eliminating the need to configure the MTP service for applications that do not support LBR codecs.

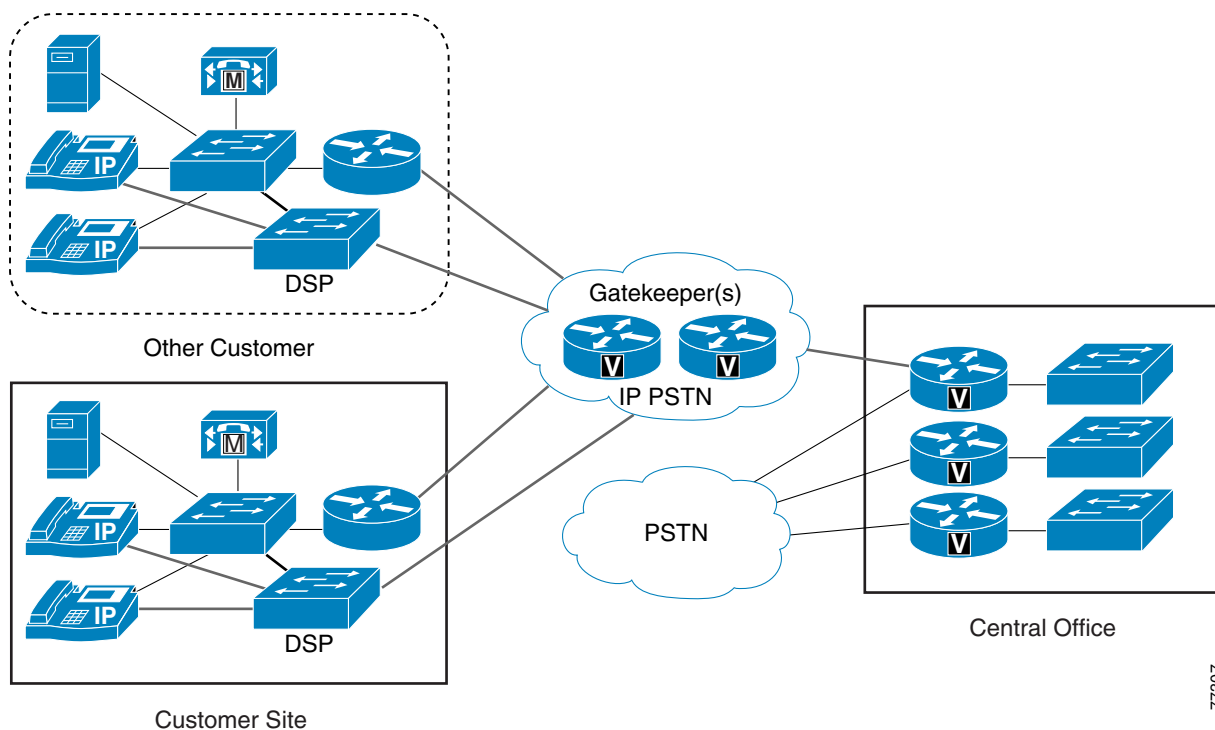
The following characteristics apply to distributed call processing deployments:

- Only the intercluster calls that require transcoding will use the MTP service. For example, if both endpoints of a call are capable of using a G.729 codec, no transcoding resources will be used.
- Sharing MTP resources among servers within a cluster provides more efficient resource utilization.

IP PSTN Access

The fourth application scenario for MTP and transcoding resources involves a service provider that provides its customers with access to an IP PSTN instead of the traditional PSTN. In such a scenario, the gatekeepers reside in the service provider network. In order to simplify the dial plan, each customer is required to use an MTP to anchor its calls, so that the individual IP addresses assigned to the endpoints can be hidden. The service provider's central office can then relay through the traditional PSTN and/or provide IP connectivity to other customers. [Figure 6-6](#) illustrates this deployment model.

Figure 6-6 IP PSTN Access Example



Note that the customer site in [Figure 6-6](#) can use any of the previous three deployment models: single site, multi-site WAN with centralized call processing, or multi-site WAN with distributed call processing.

The H.323 trunk from the customer site to the IP PSTN must be configured with MTP so that the endpoint IP addresses remain masked. Thus, all external calls use the MTP resources. However, MTP resources can be shared within the Cisco CallManager cluster to achieve more efficient use of the resources.

77307



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
- Annunciator resources
- Music on hold resources

For more information about media resources in general, see the chapter on [Media Resources, page 6-1](#).

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

- [Deployment Basics of MoH, page 7-2](#)
- [Basic MoH and MoH Call Flows, page 7-5](#)
- [MoH Configuration Considerations and Best Practices, page 7-8](#)
- [Hardware and Capacity Planning for MoH Resources, page 7-12](#)
- [Implications for MoH With Regard to IP Telephony Deployment Models, page 7-14](#)
- [Detailed Unicast and Multicast MoH Call Flows, page 7-19](#)

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 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 MoH uses a separate source stream for each user or connection. As more endpoint devices go on hold via a user or network event, the number of MoH streams increases. 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>

Recommended Unicast/Multicast Gateways

The following recommended gateways support both unicast and multicast MoH:

- Cisco 6624 and 6608 gateway modules with MGCP and Cisco CallManager Release 3.3(3) or later
- Cisco Communication Media Module (CMM) with MGCP or H.323 and Cisco CallManager Release 4.0, Cisco IOS Release 12.2(13)ZP3 or later, and Catalyst OS Release 8.1(1) or later
- Cisco 2600, 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 a fixed or live audio source:

- Telex P-800 USB

A USB sound device supported in Cisco CallManager Release 3.3(3) with Windows 2000 (OS 2000 version 2.5). This device is supported on all Cisco MCS-78xxH or MCS-78xxI servers with 2.2 GHz or greater processor.

- Open AW-840 (PCI)

A Peripheral Component Interconnect (PCI) sound device compatible with all Cisco MCS-78xx servers up to and including the MCS-78xx-1266.

- Soundblaster PCI 16

A Peripheral Component Interconnect (PCI) sound device compatible with all Cisco MCS-7835s prior to MCS-7835-1266.

- Cisco MCS-7815 Sound Card

The Cisco MCS-7815 comes with a built-in sound card that is supported as an MoH server fixed audio source device.

**Note**

Prior to using a fixed audio source to transmit music on hold, you should consider the legalities and the ramifications of re-broadcasting copyrighted 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.

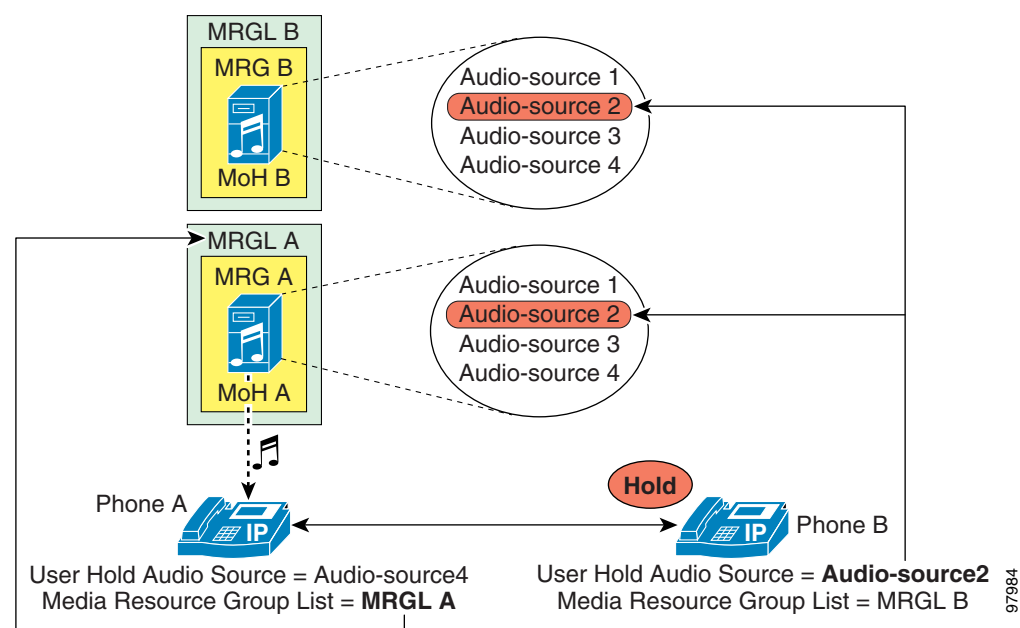
Basic MoH

The basic operation of MoH in a Cisco IP Telephony environment consists of a holder and a 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 7-1](#), 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 7-1 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 (or Network) Audio Source configured at the directory or line level. If this level is not defined, Cisco CallManager looks at the User (or Network) Audio Source configured on the holder device. If this level is not defined, Cisco CallManager looks at the User (or Network) 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 for both User and Network Hold Audio Sources, 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

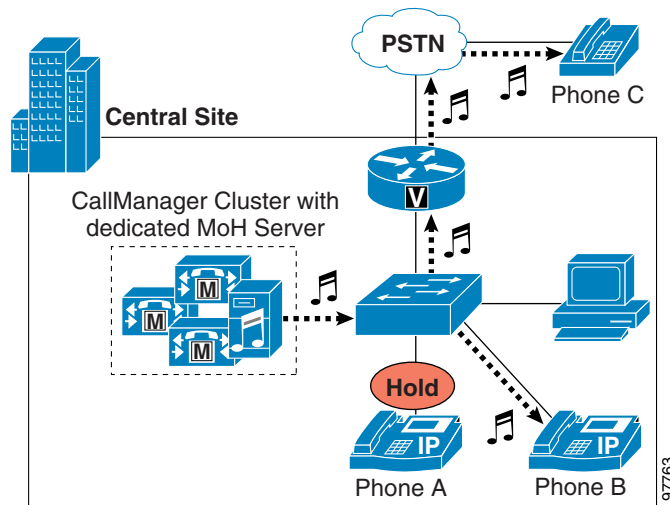
When attempting to determine the MRGL for a particular holdee, Cisco CallManager looks at the MRGL configured at the device level. If this level is not defined, Cisco CallManager looks at the MRGL configured for the device pool of the holdee device. If this level is not defined, then Cisco CallManager 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

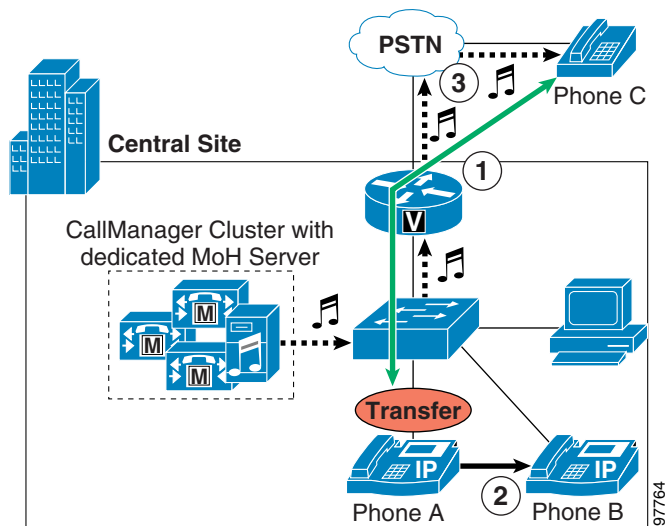
Figure 7-2 shows these two types of call flows. If phone A is in a call with phone B and phone A (holder) pushes the Hold softkey, then a music stream is sent from the MoH server to phone B (holdee). The music stream can be sent to holdees within the IP network or holdees on the PSTN, as is the case 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 7-2 Basic User Hold Example

Network hold includes the following types:

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

Figure 7-3 shows the call transfer call flow. When phone A receives a call from 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 7-3 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 7-19](#).

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 via a sound card). For exceptions to this limit, refer to the sections on [Using Multiple Fixed or Live Audio Sources, page 7-9](#) and [Multicast MoH from Branch Router Flash, page 7-16](#).

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 prevent the packets from streaming further than the local subnet.

[Figure 7-4](#) 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 beyond 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 time to live (TTL) greater than one to ensure that it is able to reach phone B when phone A presses the Hold softkey.

Diagram illustrating the effect of Max Hop (TTL) on MCast traffic. The diagram shows a network topology where MCast traffic from a Radio Station to a CallManager Cluster is blocked by a router. The router is labeled "Hold" and has a red "X" over it. A red arrow points to the router with the text "Max Hop (TTL)=1". The network is divided into a "Central Site" and a "Radio Station". The Central Site contains a "CallManager Cluster with dedicated MoH Server" and a "Media Server". The Radio Station contains a "Radio Station" icon. The MCast Address is 239.1.1.1 and the RTP Port is 16384. The diagram shows that the traffic is blocked at the router, preventing it from reaching the CallManager Cluster.



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

Unicast and Multicast in the Same Cisco CallManager Cluster

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

- 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 deploying a single MoH server for both unicast and multicast MoH, 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**

When configuring the unicast MRG, do not be confused by the fact that 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.

When choosing a method for deploying both multicast and unicast MoH in the same cluster, an important factor to consider is the number of servers required. When using a single MoH server for both unicast and multicast, fewer MoH servers are required throughout the cluster. Deploying separate multicast and unicast MoH servers will obviously require more servers within the cluster.

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 7-1](#) 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 7-1 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: 40 MoH sessions Standalone MoH server: 200 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: 100 MoH sessions Standalone MoH server: 250 MoH sessions ¹

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

The Maximum Half Duplex Streams and the Maximum Multicast Connections fields on the MoH Server configuration page must be configured to match the capacities indicated in [Table 7-1](#). These fields are set by default to 250 and 30, respectively, but should be configured according to the type of server platform and type of server deployment (coresident or standalone) as indicated in the table. Failure to match the recommended capacity numbers could lead to under-utilization of the server resources or failure of the server to handle the network load.

**Note**

The maximum session limits listed in [Table 7-1](#) 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 is required for each configured codec type, no matter how many devices of that type 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 one coresident MoH server (MCS 7835 or 7845) would be required to handle this load, as shown by the following calculation:

$$[(250 \text{ sessions per MCS 7835 or 7845 standalone server}) * (2 \text{ standalone servers})] + [(100 \text{ sessions per MCS 7835 or 7845 coresident server}) * (1 \text{ coresident server})] = 600 \text{ sessions}$$

By comparison, a multicast-only MoH environment with 36 unique MoH audio streams, for example, would require only one coresident MoH server (MCS 7815, 782x, or 7830), as shown by the following calculation:

$$(40 \text{ sessions per MCS 7815, 782x, or 7830 coresident server}) * (1 \text{ coresident server}) > 36 \text{ sessions}$$

The 36 unique multicast streams can be provisioned in any one of the following ways:

- 36 unique audio sources streamed using a single codec
- 18 unique audio sources streamed using only 2 codecs
- 12 unique audio sources streamed using only 3 codecs
- 9 unique audio source streamed using all 4 codecs

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.

**Note**

Because you can configure only 51 unique audio sources per Cisco CallManager cluster and because there are only four possible codecs for MoH streams, the maximum number of multicast streams per MoH server is 204.

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 2-1.

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 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.

If the sound quality of an MoH stream becomes an issue when using the G.729 codec across the WAN, you can use the G.711 codec for MoH audio streams across the WAN while still using G.729 for voice calls. In order to send MoH streams across the WAN with the G.711 codec but voice calls across the WAN with the G.729 codec, place all MoH servers in a Cisco CallManager region by themselves, and configure that region to use G.711 between itself and all other regions. Thus, when a call is placed between two phones on either side of a WAN, the G.729 codec is used between their respective regions.

However, when the call is placed on hold by either party, the MoH audio stream is encoded using G.711 because G.711 is the configured codec to use between the MoH server's region and the region of the phone placed on hold.

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 downstream WAN interfaces by configuring the low-latency queuing (LLQ) voice priority queue with additional bandwidth. Because MoH streams are uni-directional, only the voice priority queues of the downstream interfaces (from the central site to remote sites) must be over-provisioned. 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 ($4 * 24 \text{ kbps per G.729 audio stream} = 96 \text{ kbps}$).

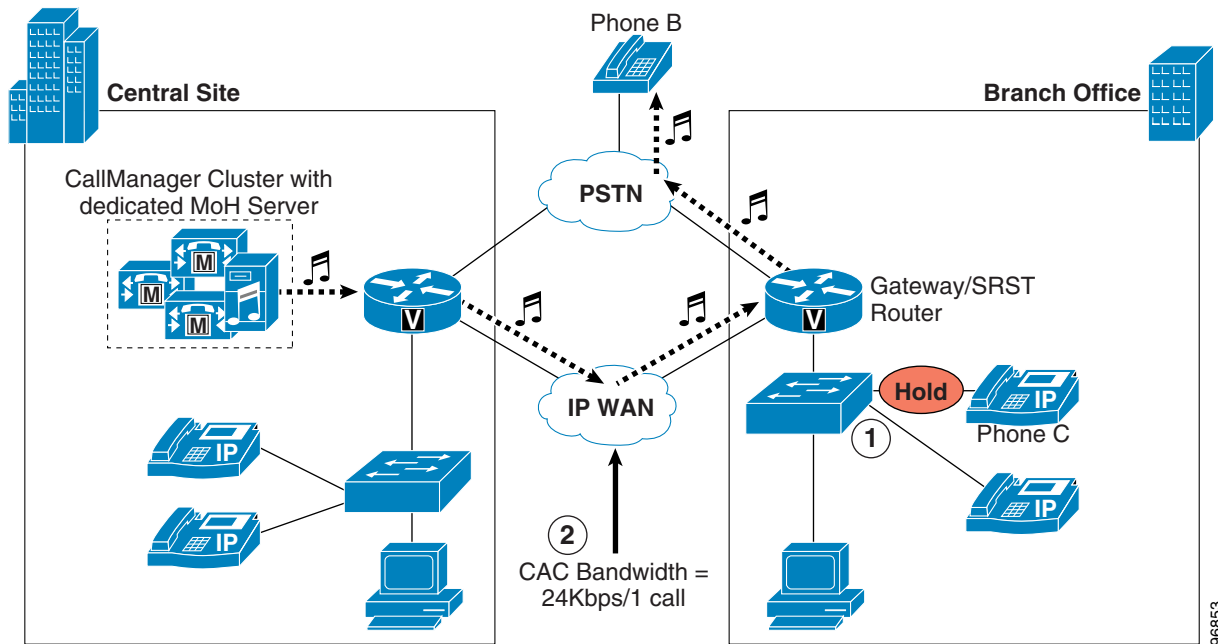
Figure 7-5 shows an example of call admission control and MoH in a centralized multi-site deployment. For this example, assume that IP phone C is in a call with a PSTN phone (phone B). At this point, no bandwidth has been consumed on the WAN. When phone C pushes the Hold softkey (step 1), phone B receives 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 downstream 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. However, 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 7-5 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 7-1 illustrates the commands to use in the Cisco IOS router configuration (under the SRST section) to enable multicast MoH from the router flash.

Example 7-1 Enabling Multicast MoH from Branch Router Flash

```
SRST-router(config)#call-manager-fallback
SRST-router(config-cm-fallback)#ip source-address 10.1.1.1
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 7-1**, 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 as specified by the **ip source-address** command under the SRST configuration. 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.

**Note**

An SRST license is required regardless of whether the SRST functionality will actually be used. The license is required because the configuration for streaming MoH from branch router flash is done under the SRST configuration mode and, even if SRST functionality will not be used, at least one **max-ephones** and one **max-dn** must be configured. These configuration commands are required in addition to the commands listed in [Example 7-1](#).

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 7-1](#).) 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 one of the following methods:

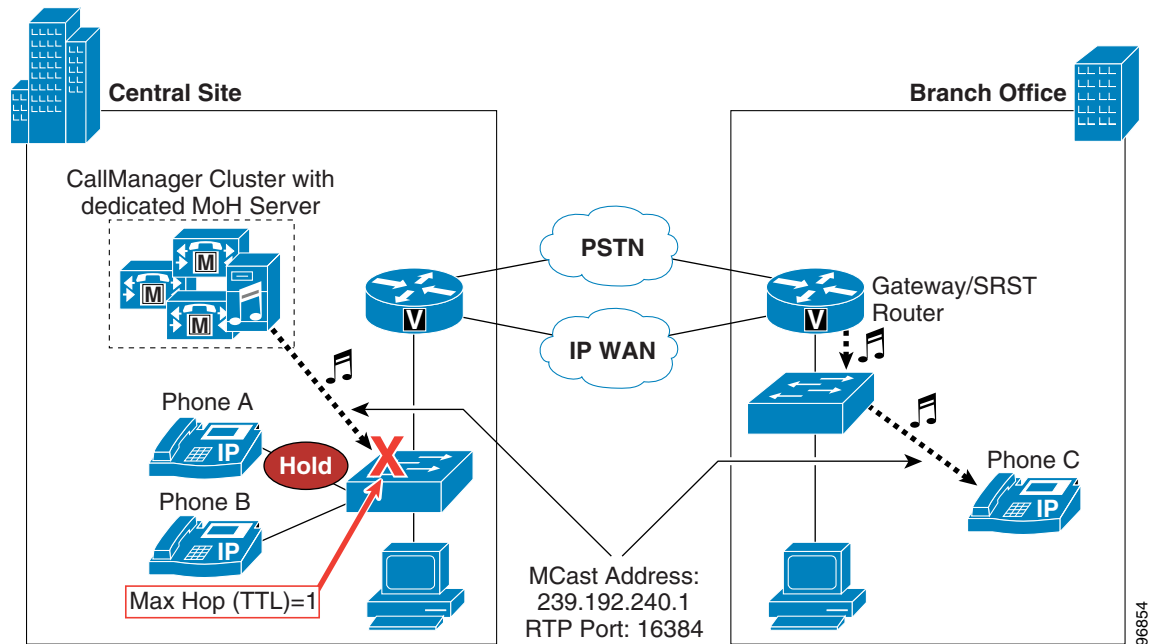
- Configure a maximum hop count

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.
- Configure an 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.
- Disable multicast routing on the WAN interface

Do not configure multicast routing on the WAN interface, thus ensuring that multicast streams are not forwarded into the WAN.

[Figure 7-6](#) 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 7-6 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 7-9), 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 use of the 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.

Unlike with centralized multi-site deployments, in situations where G.711 might be required for MoH audio streams traveling across a WAN, MoH audio streams cannot be forced to G.711 in a distributed multi-site deployment. Even when MoH servers are placed in a separate Cisco CallManager region and the G.711 codec is configured between this region and the intercluster trunk's region, the codec of the original voice call is maintained when a call between the two clusters is placed on hold by either phone. Because these intercluster calls are typically encoded using G.729 for bandwidth savings, a MoH stream from either cluster will also be encoded using G.729.

Furthermore, 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.

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 7-7](#) 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 7-7](#) and [Figure 7-8](#) 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 7-7 Detailed Multicast MoH Call Flow

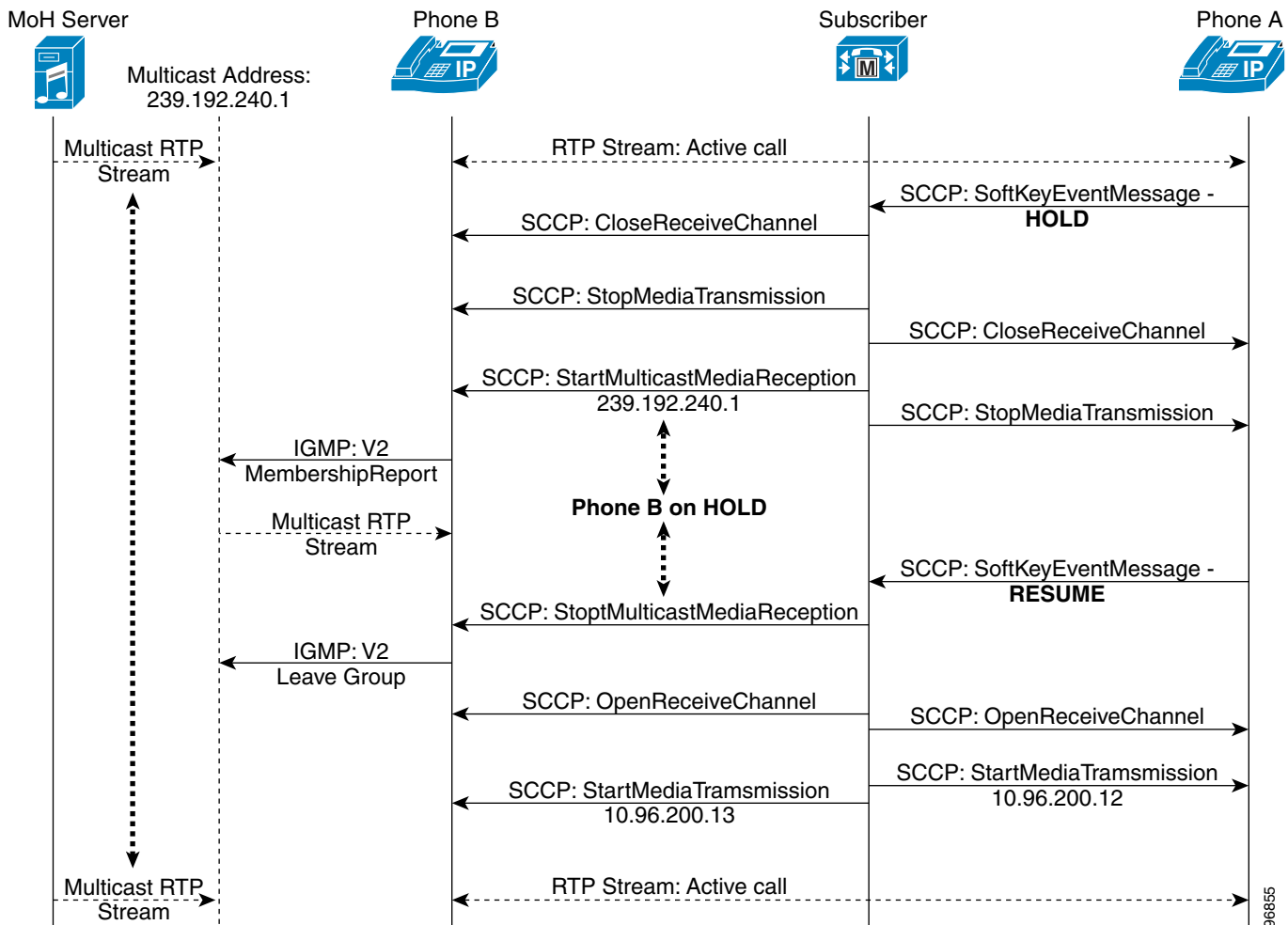
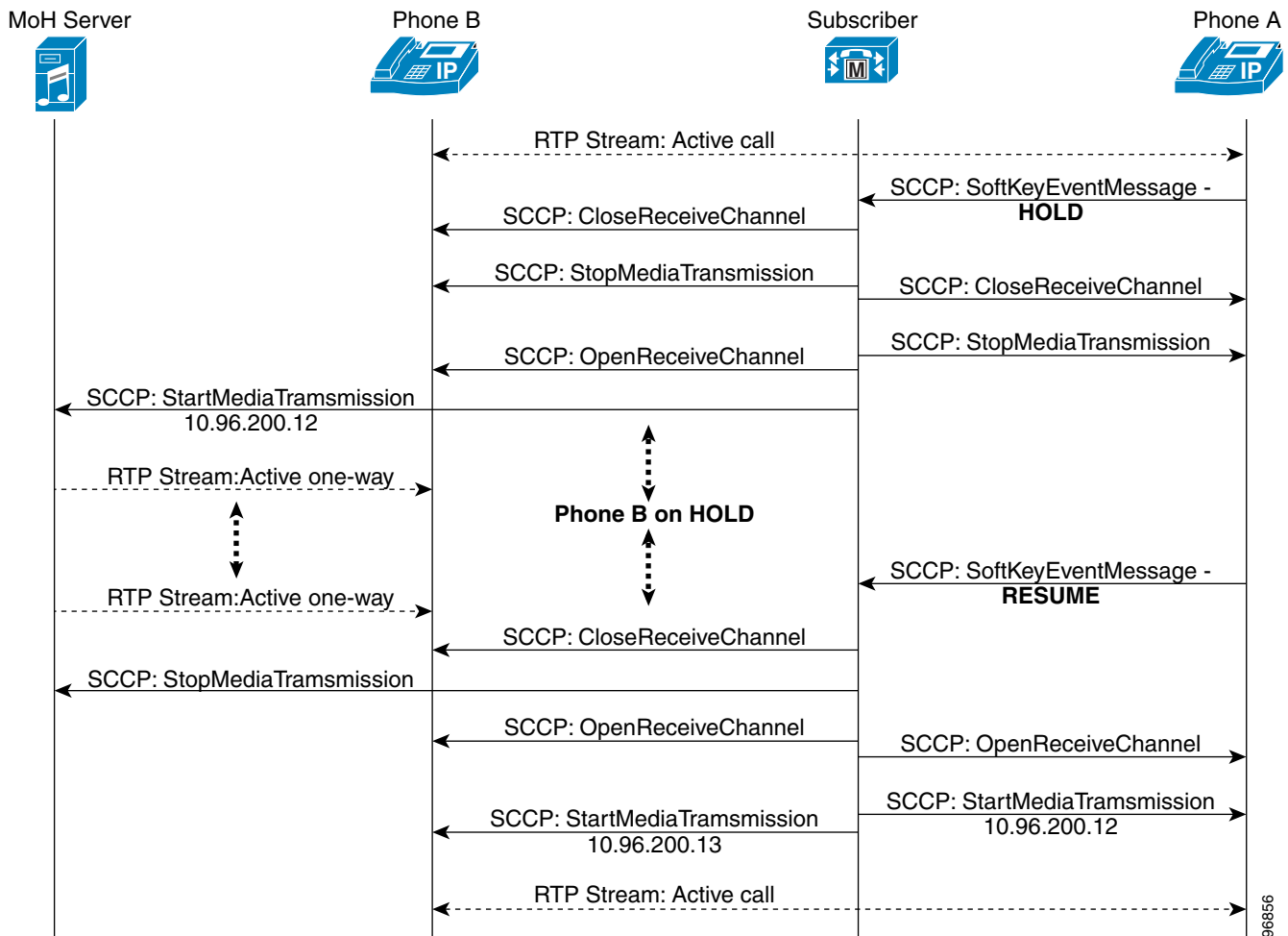


Figure 7-8 depicts a unicast MoH call flow. In this call flow diagram, when the Hold softkey is pressed at phone A, Cisco CallManager instructs both phone A and phone B to Close Receive Channel and Stop Media Transmission. This action effectively stops the RTP two-way audio stream. Up to this point, 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 7-8 Detailed Unicast MoH Call Flow



96856



Call Processing

This chapter provides design guidance for scalable and resilient call processing systems, and it focuses on Cisco CallManager and H.323 gatekeeper design.

The Cisco CallManager architecture enables a group of physical servers to work together as a single IP PBX system. This grouping of servers is known as a *cluster*. A cluster of Cisco CallManager servers may be distributed across an IP network, within design limitations, allowing for spatial redundancy and, hence, resilience to be designed into the IP Communications system.

Cisco Gatekeeper may also be paired with other standby gatekeepers or may also be clustered for higher performance and resilience.

The chapter also discusses how to choose the appropriate hardware and deployment scenario based on specific requirements that include the following:

- Scale — The number of users, gateways, applications, and so forth
- Performance — The call rate
- Resilience — The amount of redundancy

Cisco CallManager Cluster Guidelines

This section describes the various functions performed by the servers that form a Cisco CallManager cluster, and it provides guidelines for deploying the servers in ways that achieve the desired scale, performance, and resilience.

Hardware Platforms

Cisco CallManager clusters utilize various types of servers, depending on the scale, performance, and redundancy required. They range from non-redundant, single-processor servers to highly redundant, multi-processor units.

Cisco CallManager is supported only on specific hardware platforms. For a list of the currently supported hardware configurations, refer to the documentation available at

<http://www.cisco.com/go/swonly>

The server platforms have particular memory requirements, which are documented in Product Bulletin 2864, *Physical Memory Recommendations For Cisco CallManager Version 4.0 and Later*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_bulletin0900aecd80284099.html

Table 8-1 lists the general types of servers you can use in a cluster, along with their main characteristics.

Table 8-1 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

Servers should be deployed in an environment that provides high availability, not only for the IP network but also for power and cooling. Servers should be powered from an uninterruptible power supply (UPS) if building power does not have the required availability. Servers with dual power supplies could also be plugged into two different power sources to avoid the failure of one power circuit causing the server to fail.

Connectivity to the IP network should also ensure maximum performance and availability. The Cisco CallManager servers should be connected to the Ethernet at 100 Mbps full-duplex. If 100 Mbps is not available on smaller deployments, then use 10 Mbps full-duplex. On many of the newer servers, 1000 Mbps full-duplex is now also an option. You can achieve this configuration by forcing the network interface card (NIC) to the required speed and duplex setting and by manually configuring the Ethernet switch port as well.



Note

A mismatch will occur if either the server port or the Ethernet switch port is left in AUTO mode and the other port is configured manually. The best practice is to configure both the server port and the Ethernet switch port manually.

Server platforms with dual Ethernet NICs can support NIC Teaming. This feature will depend on the manufacturer, and it allows a server to be connected to the Ethernet via two NICs and, hence, two cables. In some circumstances, this feature can be useful to provide additional levels of redundancy; however, careful design of the IP infrastructure is required prior to successful deployment. For additional information regarding the installation of teaming drivers on Cisco and Hewlett-Packard (HP) server platforms, refer to the documentation at

http://www.cisco.com/univercd/cc/td/doc/product/voice/iptel_os/driver/hp_team7.htm

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. Cisco does not recommend placing all members of a cluster on the same VLAN or switch.

- For redundancy, the members of the cluster should be deployed in the following manner to minimize the impact of any failures in the infrastructure or building:
 - Different access switches connected to the same distribution or core switch
 - Different access switches attached to different distribution or core switches
 - Different buildings within the same LAN or MAN
- If the cluster spans an IP WAN, follow the guidelines for clustering over an IP WAN as specified in the section on [Clustering Over the IP WAN](#), page 2-16.

Cisco CallManager Cluster Services

Within a Cisco CallManager cluster, there are servers that provide unique services. Each of these services can coexist with others on the same physical server. For example, in a small system it is possible to have a single server be a publisher, backup subscriber, music on hold (MoH) server, TFTP server, CTI Manager, and conference bridge. As the scale and performance requirements of the cluster increase, many of these services should be moved to a single, dedicated physical server.

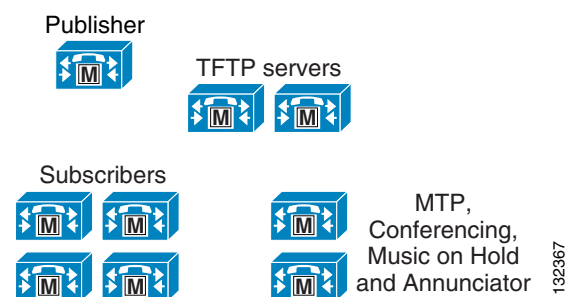
With the release of Cisco CallManager 3.3 and 4.x, a cluster may contain as many as 20 servers, of which a maximum of eight may run the Cisco CallManager service that provides call processing. The other servers may be configured as a dedicated database publisher, dedicated Trivial File Transfer Protocol (TFTP) server, or music on hold (MoH) servers. Media streaming applications (conference bridge or media termination point) may also be installed on a separate server that registers with the cluster.

When deploying a cluster with Cisco MCS 7815 or equivalent servers, there is a limit of two servers in a cluster – one as the publisher, TFTP server, and backup call processing server, and the other as the primary call processing server. The maximum number of users supported in this configuration is 300. When deploying a two-server cluster with higher-capacity servers, Cisco recommends that you do not exceed 1250 users in the cluster. Above 1250 users, a dedicated publisher is recommended.

It is also possible to deploy a single-server cluster. With an MCS 7825 or equivalent server, the limit is 500 users; with a higher-availability server, the cluster should not exceed 1000 users. In a single-server configuration, there is no redundancy unless Survivable Remote Site Telephony (SRST) is also deployed to provide service during periods when the Cisco CallManager is not available. Cisco does not recommend a single-server deployment for production environments. The load balancing option is not available when the publisher is a backup call processing subscriber.

[Figure 8-1](#) illustrates a typical Cisco CallManager cluster.

Figure 8-1 Typical Cisco CallManager Cluster



Intracuster Communications

There are two primary kinds of communication within a Cisco CallManager cluster (intracuster communications). (See [Figure 8-2](#).) The first is a mechanism for distributing the database that contains all the device configuration information. The configuration database is stored on a publisher server, and a read-only copy is replicated to the subscriber members of the cluster. Changes made on the publisher are communicated to the subscriber databases, ensuring that the configuration is consistent across the members of the cluster, as well as facilitating spatial redundancy of the database.

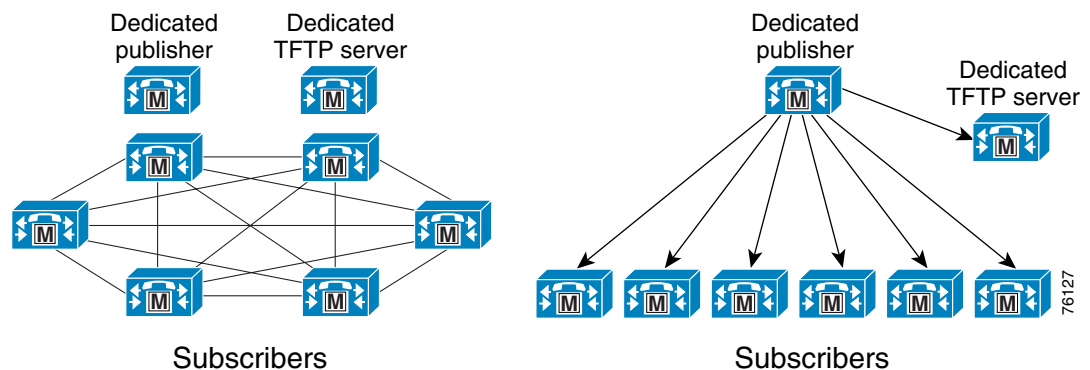
The second type of intracuster communication is the propagation and replication of run-time data such as registration of devices, locations bandwidth, and shared media resources. This information is shared across all members of a cluster running the Cisco CallManager Service, and it assures the optimum routing of calls between members of the cluster and associated gateways.

Lightweight Directory Access Protocol (LDAP) directory information is also replicated between all servers in a cluster. The LDAP directory is replicated by the publisher to all other servers. Cisco CallManager may be integrated into a corporate LDAP directory, such as Microsoft's Active Directory or Netscape Directory. The replication is then dependant on the integration method deployed and is outside the scope of this document. For more information on directory integration, refer to the chapter on [Directory Access and Integration](#), page 14-1.

Figure 8-2 Intracuster Communications

Intra-Cluster Communication Signaling (ICCS)

SQL and directory replication



Publisher

The *publisher* is required in all clusters, and there can currently be only one per cluster. This server is the first to be installed and provides the database and directory services to all other members in the cluster. The publisher server is the only server that has read and write access to the configuration database. When configuration changes are made, other members of the cluster have a read-only copy of the database. On larger systems with more than 1000 users, Cisco recommends that no call processing take place on the publisher server to prevent administrative operations from affecting the telephony users.

Servers in the cluster will attempt to use the publisher's database when initializing. If the publisher is not available, they will use the local read-only copy from their hard drives.

If the system is running but the publisher is not available, the following operations will not be available:

- Configuration changes
- Call forwarding changes
- Extension Mobility login or logout operations

Extension Mobility will not function without the publisher because it requires access to the read/write database. Therefore, Cisco recommends running this service on the publisher only.

The choice of hardware platform for the publisher is based on the scale and performance of the cluster. Cisco recommends that the publisher be of the same performance capability as the call processing subscribers. Ideally the publisher should also be a high-availability server to minimize the impact of a hardware failure.

Call Processing Subscriber

When installing the Cisco CallManager software, you can define two types of servers – publisher and subscriber. These terms are used to define the database relationship during installation. Once the software is installed, the only services enabled are the database services. All subscribers will subscribe to the publisher to obtain a read-only copy of the database and directory information.

A call processing subscriber is a server that has the Cisco CallManager Service enabled. Once this service is enabled, the server is able to perform call processing functions. Devices such as phones, gateways, and media resources can register and make calls only to servers with this service enabled. Prior to the Cisco CallManager Release 3.3, only six servers in a cluster were allowed with this service enabled. All subsequent releases support up to eight servers in a cluster with the Cisco CallManager Service enabled.

Dependant on the redundancy scheme chosen (see [Call Processing Redundancy, page 8-5](#)), the call processing subscriber will be either a primary (or active) subscriber or a backup (or standby) subscriber. In the load balancing option, the subscriber can be both a primary and backup subscriber. When planning the design of a cluster, you should generally dedicate the call processing subscribers to this function. In larger scale or higher performance clusters, the publisher and TFTP functions should not be enabled on the call processing servers. Call processing subscribers normally operate in either dedicated pairs or shared pairs, depending on the redundancy scheme adopted. One-to-one redundancy uses dedicated pairs, while two-to-one redundancy uses two pairs of servers that share one server from each pair (the backup server).

The choice of hardware platform depends on the scale, performance, redundancy, and cost of the servers. Scale and performance are covered in the section on [Cisco CallManager Platform Capacity Planning, page 8-11](#), and redundancy is covered in the section on [Call Processing Redundancy, page 8-5](#).

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.

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 one at a time.
 - Step 3** Upgrade the backup subscribers one at a time. 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 one at a time, 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.

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.

Call Processing Subscriber Redundancy

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

Figure 8-3 Basic Redundancy Schemes

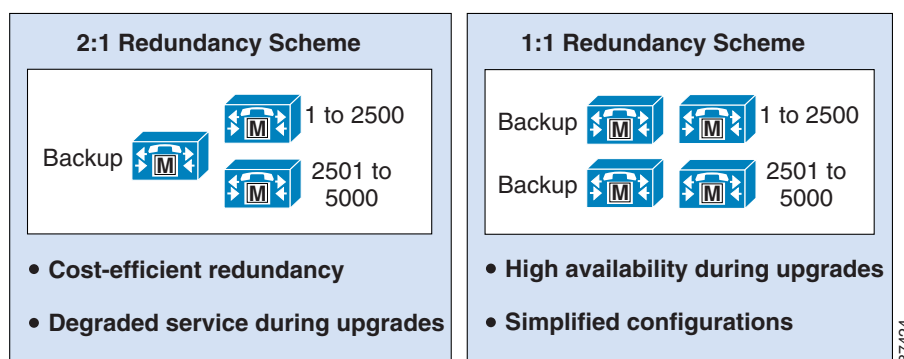
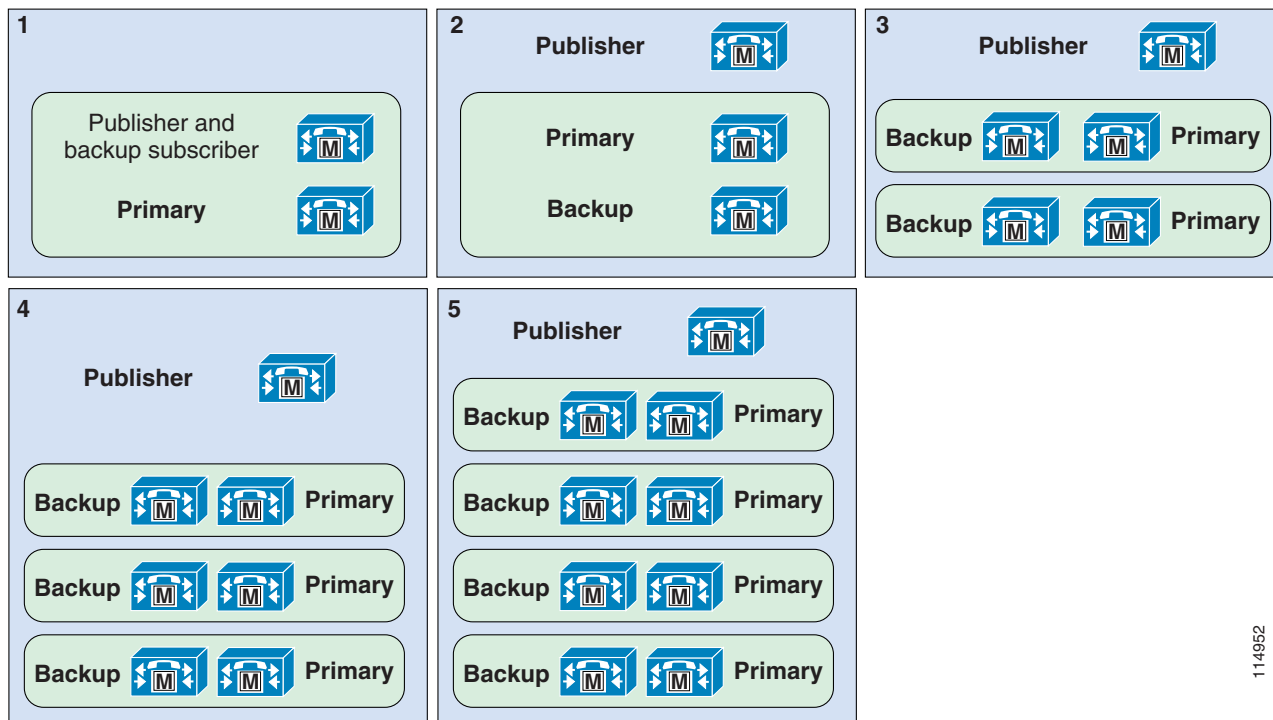


Figure 8-3 illustrates the two basic redundancy schemes available. In each case the backup server must be capable of handling the capacity of at least a single primary call processing server failure. In the 2:1 redundancy scheme, the backup might have to be capable of handling the failure of a single call processing server or potentially both primary call processing servers, depending on the requirements of a particular deployment. Sizing the capacity of the servers and choice of hardware platforms is covered in the section on [Cisco CallManager Platform Capacity Planning, page 8-11](#).

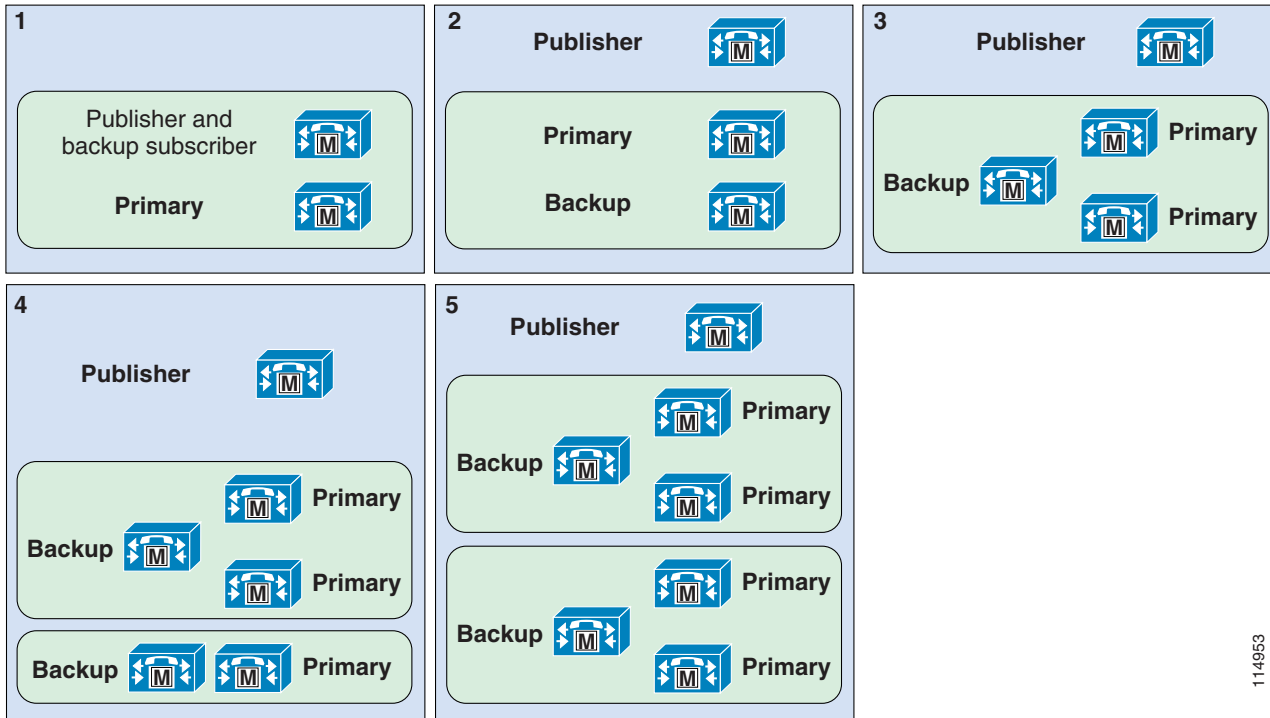
87424

Figure 8-4 1:1 Redundancy Configuration Options

114952

In [Figure 8-4](#) the five options shown all indicate 1:1 redundancy. Option 1 is used for clusters supporting less than 1250 users. Options 2 through 5 illustrate increasingly scalable clusters. The exact scale will depend on the hardware platforms chosen or required.

Note that the illustrations show only publisher and call processing subscribers.

Figure 8-5 2:1 Redundancy Configuration Options

114953

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, calculate the capacity of a cluster without load balancing and then distribute the load across the primary and backup subscribers based on devices and call volume. To allow for failure of the primary or the backup, the total load on the primary and secondary subscribers should not exceed that of a single subscriber.

Cisco recommends the equal distribution of all devices and call volumes as much as possible across all active subscribers. For instance, distributing the gateways, trunks, voicemail ports, and users between all subscribers equally, minimizes the impact of any outage.

TFTP Server

The TFTP server performs two main functions:

- The serving of files for services such as MoH, configuration files for devices such as phones and gateways, binary files for the upgrade of phones as well as some gateways, and various security files.
- Generation of configuration and security files. Most files generated by the Cisco TFTP service are signed and in some cases encrypted before being available for download.

The TFTP service can be enabled on any server in the cluster. However, in a cluster with more than 1250 users, other services might be impacted by configuration changes that can cause the TFTP service to regenerate configuration files. Therefore, Cisco recommends that you dedicate a specific server to the TFTP service in a cluster with more than 1250 users, with Extension Mobility, or with other features that cause configuration changes.

The TFTP server is used by phones and MGCP gateways to obtain configuration information. As the scale of the cluster increases and/or additional redundancy is required, Cisco recommends that you deploy two TFTP servers to provide for load balancing and redundancy. When configuring the TFTP options in DHCP or statically, you can normally define a IP address array, or more than one IP address of a TFTP server. Therefore, you can assign half of the devices to use TFTP server A as the primary and TFTP server B as the backup, and the other half to use TFTP server B as the primary and TFTP server A as the backup. To improve performance on dedicated TFTP servers, you can set service parameters to increase the number of simultaneous TFTP sessions allowed on the server.

When upgrading a Cisco CallManager cluster, Cisco recommends that you upgrade the TFTP servers after the publisher but before any other server. This upgrade order ensures that any new binaries and configuration changes are available before the upgrade of other services in the cluster. If you are adding a specific binary or firmware load for a phone or gateway manually, be sure to copy the file to each TFTP server in the cluster.

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.

Cisco recommends that you use the same hardware platform for the TFTP servers as used for the call processing subscribers.

CTI Manager

CTI Manager is required on a cluster for applications that use TAPI or JTAPI Computer Telephony Integration (CTI). The CTI Manager acts as a broker between the CTI application and the Cisco CallManager service. It provides authentication of the application and enables control or monitoring of authorized devices. The CTI application communicates with a primary CTI Manager and, in the event of a failure, will switch to a backup CTI Manager. The CTI Manager should be enabled only on call processing subscribers, thus allowing for a maximum of eight CTI Managers in a cluster (six prior to Cisco CallManager Release 3.3). To provide maximum resilience, performance, and redundancy, Cisco recommends that you load-balance CTI applications across the various CTI Managers in the cluster.

Generally, it is good practice to associate devices that will be controlled or monitored by an application with the same server pair used for the CTI Manager. For example, an interactive voice response (IVR) application requires four CTI ports. They would be provisioned as follows, assuming the use of 1:1 redundancy and 50/50 load balancing:

- Two CTI Ports would have a Cisco CallManager redundancy group of server A as the primary and server B as the backup (or secondary). The other two ports would have a Cisco CallManager redundancy group of server B as the primary and server A as the backup.
- The IVR application would be configured to use the CTI Manager on server A as the primary and server B as the backup.

The above example allows for redundancy in case of failure of the CTI Manager on server A and also allows for the IVR call load to be spread across two servers. This approach also minimizes the impact of a Cisco CallManager server failure.

Media Resources

Media resources such as conferencing and music on hold may be provided by software services running on the same physical server as the Cisco CallManager service.

Media resources include:

- Music on Hold (MoH) — Provides multicast or unicast music to devices that are placed on hold or temporary hold, transferred, or added to a conference. (See [Music on Hold, page 7-1.](#))
- Annunciator service — Provides announcements in place of tones to indicate incorrectly dialed numbers or call routing unavailability. (See [Annunciator, page 6-15.](#))
- Conference bridges — Provide software-based conferencing for ad-hoc and meet-me conferences. (See [Conferencing, page 6-8.](#))
- Media termination point (MTP) services — Provide features for H.323 clients, H.323 trunks, and Session Initiation Protocol (SIP) trunks. (See [Media Termination Point \(MTP\), page 6-9.](#))

Because of the additional processing and network requirements for media, it is essential to follow all guidelines for running media resources within a cluster. Generally, Cisco recommends non-dedicated servers for multicast MoH and the annunciator, but we recommend dedicated servers for large-scale software-based conferencing and MTPs unless those services are within the design guideline detailed in the chapters on [Media Resources, page 6-1](#), and [Music on Hold, page 7-1](#).

Voice Activity Detection

Cisco also 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

Extension Mobility

When Extension Mobility is used in a cluster, the expected capacity and performance can impact the choice of the publisher hardware platform. As a user logs in or logs out from a phone, the configuration must be updated in the configuration database, the TFTP service has to regenerate the configuration file, and then the device is reset to make the change. Most of this takes place on the publisher.

Currently supported limits for Extension Mobility are as follows:

- Cisco CallManager Release 3.3
 - 2000 sequential logins/logouts per hour
- Cisco CallManager Release 4.0
 - 2000 sequential logins/logouts per hour
- Cisco CallManager Release 4.1
 - 2000 sequential logins/logouts per hour
 - An MCS 7845 publisher will support 4500 sequential logins/logouts per hour.

Cisco CallManager Platform Capacity Planning

Many types of devices can register with Cisco CallManager; for example, IP phones, voicemail ports, CTI (TAPI or JTAPI) devices, gateways, and DSP resources such as transcoding and conferencing. Each of these devices requires resources from the server platform with which it is registered. The required resources can include memory, processor usage, and disk I/O. Each device then consumes additional server resources during transactions, which are normally in the form of calls. For example, a device that makes only 6 calls per hour consumes fewer resources than a device making 12 calls per hour.

The recommendations provided in this section are based on calculations made using the Cisco CallManager Capacity Calculator with default trace levels and CDRs enabled. Higher levels of performance can be achieved by disabling, reducing, or reconfiguring other functions that are not directly related to processing calls. Increasing some of these functions can also have an impact on the call processing capabilities of the system. These functions include tracing, call detail recording, highly complex dial plans, and other services that are coresident on the server. Highly complex dial plans can include multiple line appearances as well as large numbers of partitions, calling search spaces, route patterns, translations, route groups, hunt groups, pickup groups, route lists, extensive use of call forwarding, coresident services, and other coresident applications. All of these functions can consume additional resources within the Cisco CallManager server.

To improve performance, the following techniques can provide useful options:

- Install additional certified memory in the server, up to the maximum supported for the particular platform. Doubling the RAM in MCS 7825 and MCS 7835 or equivalent servers is recommended in large configurations for that server class. Verification using a performance monitor will indicate if this memory upgrade is required. As the server approaches maximum utilization of physical memory, the operating system will start to swap to disk. This swapping is a sign that additional physical memory should be installed. For specific server memory recommendations, refer to Product Bulletin 2864, *Physical Memory Recommendations For Cisco CallManager Version 4.0 and Later*, available at http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_bulletin0900aecd80284099.html
- On supported hardware, make sure the battery-backed write cache (BBWC) is enabled with 50% write capability. This feature might require the addition of a battery to the Small Computer System Interface (SCSI) controller card.
- On MCS 7845 or equivalent servers, set the trace files location to the F: drive. This setting is a Cisco CallManager service parameter.
- Reformat the trace file partition to use smaller block sizes. (Refer to the release notes for the version of the operating system you have installed.) A utility is provided for this purpose.

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 increase to allow additional time for the configuration to initialize. For details on the service parameters, refer to the online help for *Service Parameters* in Cisco CallManager Administration.

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.
- 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.
- Each cluster can support a maximum of 600 H.323 (gateway and client) and digital MGCP devices.

Call Processing with Cisco CallManager Release 3.3 and Later

The following guidelines apply to Cisco CallManager Release 3.3 and later:

- 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 high-performance server, or a maximum 10,000 per cluster if they are equally balanced among the servers.
- Each cluster can support up to 30,000 IP phones.
- Each cluster can support a maximum of 600 H.323 (gateway and client) and digital MGCP devices.

Capacity Calculations

In prior software releases, Cisco has utilized various schemes to allow the capacity of a system to be calculated using device weights, BHCA multipliers, and dial plan weights. With Cisco CallManager Release 4.x, this simple scheme has been replaced by a capacity tool to allow for more accurate planning of the system. (See [Cisco CallManager Capacity Tool, page 8-13](#).) The capacity planning tool is currently available only to Cisco employees and Cisco partners. If your system does not meet the following guidelines, or if you consider the system to be more complex and would like to verify the capacity, please contact your Cisco Systems Engineer or Cisco Technical Assistance Center (TAC).

If your system meets the following requirements, you should not need to verify your configuration with the Cisco CallManager Capacity Tool:

- System contains less than 25% of the maximum number of users for the server platforms
- Average number of lines per phone does not exceed 1.1
- Average busy hour call attempts (BHCA) per user is below 6 calls per hour
- No more than 20% trunking (5 users per trunk) either via gateways or trunks
- No more than 5% voicemail ports (20 users per voicemail port)
- No more than 20 MoH streams per server

- No CTI, JTAPI, or TAPI devices
- No more than 5% conference bridges (20 users per port on a bridge)
- No transcoders
- Only MTPs are used to support the maximum calls required for trunking
- Less than 20 locations
- System contains nothing else that is not defined above and is not an IP Phone, IP Communicator, gateway, media resource, voicemail port, or trunk

The maximum number of users Cisco CallManager can support depends on the server platform, as indicated in [Table 8-2](#).

Table 8-2 Maximum Number of Devices per Server Platform

Server Platform Characteristics	Maximum Users per Server ¹	High-Availability Server ²	High-Performance Server
Cisco MCS-7845 (All models)	7500	Yes	Yes
Cisco MCS-7835 (All models)	2500	Yes	No
Cisco MCS-7825 (All models)	1000	No	No
Cisco MCS-7815 (All models) ³	300	No	No

1. A platform that is not a high-availability server can support a maximum of 500 users in a non-redundant installation.
2. A high-availability server supports redundancy for both the power supplies and the hard disks.
3. MCS-7815 servers support only N+1 redundancy and may not be a member of a cluster.

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

<http://www.cisco.com/go/swonly>



Note

The maximum supported non-redundant installation on platforms that are not highly available is 500 IP phones.

Cisco CallManager Capacity Tool

The Cisco CallManager Capacity Tool requires various pieces of information to provide a calculation of the minimum size and type of servers required for a system. The information includes the type and quantity of devices, such as IP phones, gateways, and media resources. For each device type, the capacity tool also requires the average BHCA and the average utilization time. For example, if all IP phones make an average of 5 calls per hour and the average call lasts 3 minutes, then the BHCA is 5 and the utilization is 0.25. (Five calls of 3 minutes each equals 15 minutes per hour on the phone, which is 0.25 hour.)

In addition to the device information, the capacity tool also requires information regarding the dial plan, such as route patterns and translation patterns. When all the details have been entered, the capacity tool will calculate how many primary servers of the desired server type are required, as well as the number of clusters if the required capacity exceeds a single cluster.

The capacity tool replaces the previous mechanisms known as device weights, BHCA multipliers, call type multipliers, and dial plan weights.

Use the guidelines in the following sections to provide information for the capacity calculator.

Phone Calculations

Quantity

This value is the total number of Skinny Client Control Protocol (SCCP) phones that will be configured in the cluster. The quantity includes all Cisco 7900 Series IP Phones, VG248 ports, IP Communicator, and other third-party SCCP endpoint devices. The quantity must include all configured phones even if they are not active.

BHCA

This value is the average BHCA of all phones. If you have shared lines across multiple phones, the BHCA should include one call for each and every phone that shares that line. (That is, one call to a shared line will be calculated as multiple calls.) If you have different groups of phones that generate different BHCAs, use the following method to provide a BHCA value for use in the Cisco CallManager Capacity Tool.

For example, assume there are two classes of user with:

- 100 phones at 20 BHCA = 2,000 BHCA total
- 5000 phones at 4 BHCA = 20,000 BHCA total

The total BHCA for all phone devices in this case is 22,000.

Then divide the total BHCA by the total number of phone devices, to yield:

$$\text{Average BHCA per phone device} = 22,000 / 5,100 = 4.31 \text{ BHCA}$$

Utilization

This value is the average call utilization per phone, and it includes all the time that a call is on the phone. The actual utilization of a phone can exceed 100% if that phone allows multiple calls per line or has multiple line appearances that also are frequently utilized. Utilization is measured as a percentage of an hour. For example, a phone that is on a call for 3 minutes in the busiest hour is considered 5% utilized. The same method for calculating BHCA can also be used to calculate the average utilization for various groups of phone. If the phone has shared lines, then only the expected utilization for that phone should be calculated and not the actual utilization for the line that is shared. The Cisco CallManager Capacity Tool currently accepts an average utilization for all phones.

Lines Appearances

This value is the average number of lines for all the phones. If a phone has multiple appearances of the same DN in different partitions, it will be counted as multiple line appearances. Shared lines should be counted as one line, but ensure that the correct BHCA and utilization are calculated. Multiple calls per line do not increase the line appearance count, but they do affect the BHCA and utilization calculations.

Gateways

Quantity of Gateways

The quantity of gateways is split into several entries because it depends on gateway type, as follows:

- MGCP T1/E1 gateways

This value is the total number of gateways that need to be configured in the Cisco CallManager database. For example, a Cisco IOS MGCP gateway can have multiple T1s or E1s, but they would be added as a single gateway. A Cisco WS-6608 module would be added as a gateway per port that is configured as a T1 or E1 gateway, up to the maximum of eight per module.

- MGCP analogue gateways

This value is the total number of analogue gateways to be added in the Cisco CallManager database. Generally, an entire module (WS-6624 or Cisco CallManager) or hardware platform (Cisco IOS Router platform) would be added as a single gateway.

- H.323 gateways

An entire module or hardware platform would be added as a single gateway. This quantity does not include H.323 gateways that are not defined in Cisco CallManager but are used via H.323 trunks.

Number of DS0s

This value is the total number of DS0s or analogue ports supported by each type of gateway. The number of DS0s is split between the different types of gateways as follows:

- T1 CAS,
 $24 * (\text{Number of T1 CAS spans})$
- T1 (E1) PRI
 $(23 \text{ or } 30) * (\text{Total number of PRIs})$
- H.323 gateways
 $(\text{Total number of DS0s}) / (\text{Number of calls supported on all digital, analogue, or IP interfaces})$

BHCA

The BHCA is the average for all DS0s or analogue ports on a gateway during the busiest hour. The method for calculating this average is the same as the method used for the phone BHCA.

Utilization

This value is the average utilization of all DS0s or analogue ports during the busiest hour.

EM Profiles

Extension Mobility (EM) profiles contain line appearances that should also be included in the phone calculations. They will not affect the quantity of devices, but they will increase the average number of line appearances per phone. The BHCA and utilization of an EM user should already have been calculated for the phone they will log into.

H.323 and SIP Trunks

Quantity of Trunks

This value is the total number of trunks that will be configured in the Cisco CallManager database. A gatekeeper controlled trunk is not affected by the possible number of destinations and, hence, will be counted once for each configured gatekeeper controlled trunk. This situation is also the case for Session Initiation Protocol (SIP) trunks using a SIP proxy.

Quantity of Calls

This value is the total number of concurrent calls expected to be allowed on all trunks. The number of allowed calls will generally be controlled by locations or gatekeeper call admission control. Keep in mind that regions and codecs can also affect the total number of calls allowed.

Utilization

This value is the average utilization of all calls across all trunks. It is a percentage of the busiest hour, where 75% utilization means that calls will be active for 45 minutes per hour.

MTP Requirements

MTP requirements must also be considered separately in the calculator. If you have MTP required on an H.323 trunk (it is not optional on a SIP trunk), then an MTP resource is required for every concurrent call on that trunk.

Gatekeeper Considerations

A single Cisco IOS gatekeeper can provide call routing and call admission control for up to 100 Cisco CallManager clusters in a distributed call processing environment. Multiple gatekeepers can be configured to support thousands of Cisco CallManager clusters. 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, 2800, 3600, 3700, 3800, and 7200 Series routers all support the gatekeeper feature. You can configure Cisco IOS gatekeepers in a number of different ways for redundancy, load balancing, and hierarchical call routing. This section considers the design requirements for building a gatekeeper network, but it does not deal with the call admission control or dial plan resolution aspects, which are covered in the chapters on [Call Admission Control, page 9-1](#), and [Dial Plan, page 10-1](#), respectively.

For additional information regarding gatekeepers, refer to the *Cisco IOS H.323 Configuration Guide*, available at

http://www.cisco.com/en/US/products/sw/iosswrel/ps5207/products_configuration_guide_book09186a00801fcee1.html

Hardware Platform Selection

The choice of gatekeeper platform is based on the number of calls per second and the number of concurrent calls. A higher number of calls per second requires a more powerful CPU, such as a Cisco 3700, 3800, or 7200 Series Router. A higher number of concurrent calls requires more memory. For the latest information on platform selection, contact your Cisco partner or Cisco Systems Engineer (SE).

Gatekeeper Redundancy

With gatekeepers providing all call routing and admission control for intercluster communications, 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.

Hot Standby Router Protocol (HSRP)

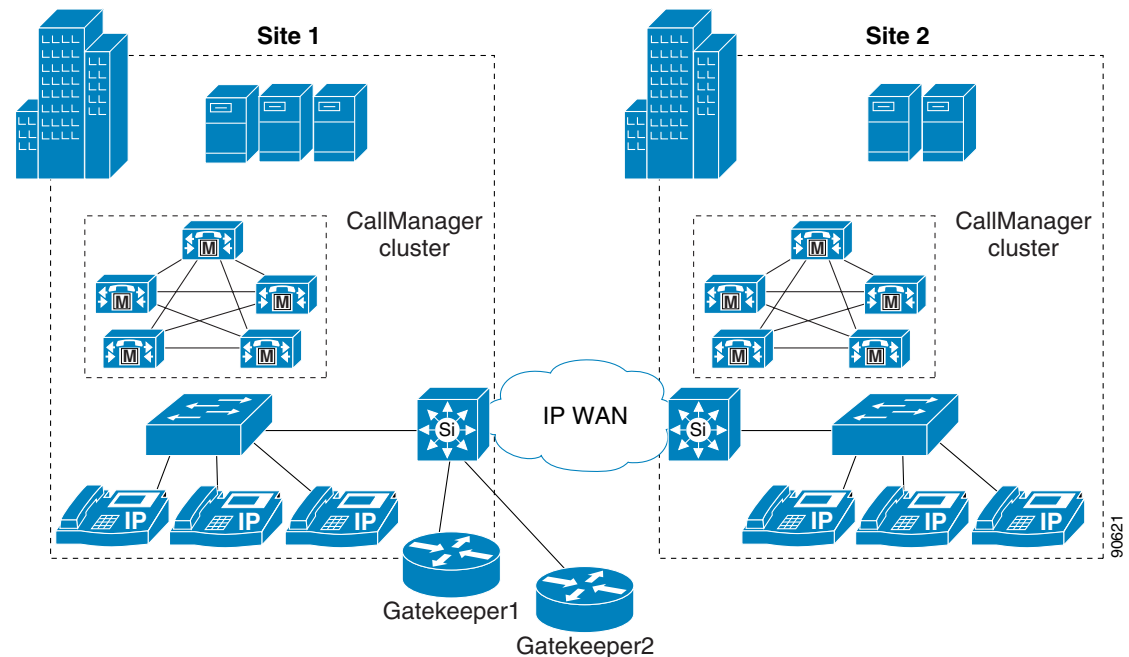
Hot Standby Router Protocol (HSRP) is the only option for gatekeeper redundancy with Cisco CallManager prior to Release 3.3. HSRP does not provide the features required to build a redundant and scalable gatekeeper network, therefore it should be used only in Cisco CallManager environments prior to Release 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 previous state information is available after 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 can be substantial because the endpoints have to re-register with the HSRP standby gatekeeper before calls can be placed. The failover time depends on the settings of the registration timers.

Figure 8-6 shows a network configuration using HSRP for gatekeeper redundancy.

Figure 8-6 Gatekeeper Redundancy Using HSRP



Example 8-1 shows the configuration for Gatekeeper 1, and Example 8-2 shows the configuration for Gatekeeper 2 in Figure 8-6. Both configurations are identical except for the HSRP configuration on the Ethernet interface.

Example 8-1 Configuration for Gatekeeper 1

```

interface Ethernet0/0
 ip address 10.1.10.2 255.255.255.0
 standby ip 10.1.10.1
 standby priority 110

gatekeeper
 zone local GK-Site1 customer.com 10.1.10.1
 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 8-2 Configuration for Gatekeeper 2

```

interface Ethernet0/0
 ip address 10.1.10.3 255.255.255.0
 standby ip 10.1.10.1

gatekeeper
 zone local GK-Site1 customer.com 10.1.10.1
 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 8-1](#) and [Example 8-2](#):

- Each router has **standby** commands configured for HSRP and to identify the virtual IP address shared by each. Gatekeeper 1 is configured as the primary with the command **standby priority 110**.
- Each Cisco CallManager cluster has a local zone configured on each router to support Cisco CallManager trunk registrations. Note that the IP address defined on the first zone should match the virtual IP address used by HSRP.
- 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 because the **bandwidth total** command does not work 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.

For additional and advanced HSRP information, refer to the online documentation at the following locations:

- <http://www.cisco.com/univercd/cc/td/doc/cisintwk/ics/cs009.htm>
- <http://www.cisco.com/warp/public/619/3.html>
- <http://www.cisco.com/warp/public/473/62.shtml>

Gatekeeper Clustering (Alternate Gatekeeper)

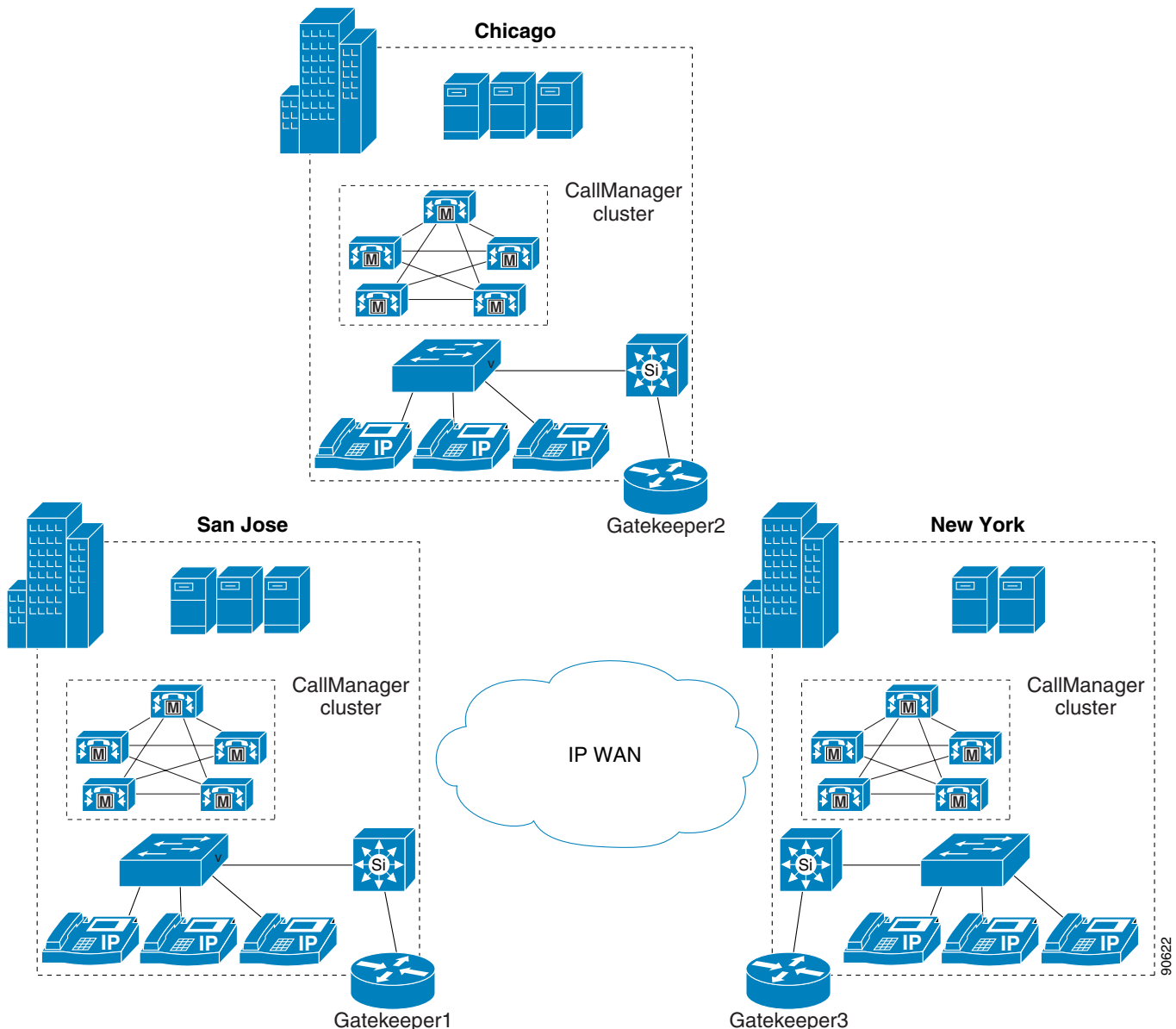
Gatekeeper clustering (alternate gatekeeper) enables the configuration of a "local" gatekeeper cluster, with each gatekeeper acting as primary for some Cisco CallManager trunks and an alternate for others. Gatekeeper Update Protocol (GUP) is used to exchange state information between gatekeepers in a 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 or later, stateful redundancy and load balancing is available. Gatekeeper clustering provides the following features:

- 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

Figure 8-7 shows three sites with Cisco CallManager distributed call processing and three distributed gatekeepers configured in a local cluster.

Figure 8-7 Gatekeeper Clustering

In [Figure 8-7](#), Gatekeeper 2 is the backup for Gatekeeper 1, Gatekeeper 3 is the backup for Gatekeeper 2, and Gatekeeper 1 is the backup for Gatekeeper 3.

[Example 8-3](#) shows the configuration for Gatekeeper 1 (SJC), and [Example 8-4](#) 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 8-3 Gatekeeper Clustering Configuration for Gatekeeper 1

```
gatekeeper
zone local SJC cisco.com 10.1.1.1
zone local CHC_GK1 cisco.com
zone local NYC_GK1 cisco.com
!
```

```

zone cluster local SJC_Cluster SJC
  element SJC_GK2 10.1.2.1 1719
  element SJC_GK3 10.1.3.1 1719
!
zone cluster local CHC_Cluster CHC_GK1
  element CHC 10.1.2.1 1719
  element CHC_GK3 10.1.3.1 1719
!
zone cluster local NYC_Cluster NYC_GK1
  element NYC 10.1.3.1 1719
  element NYC_GK2 10.1.2.1 1719
!
zone prefix SJC 40852.....
zone prefix NYC_GK1 21251.....
zone prefix CHC_GK1 72067.....
gw-type-prefix 1#* default-technology
load-balance cpu 80 memory 80
bandwidth interzone SJC 192
bandwidth interzone NYC_GK1 160
bandwidth interzone CHC_GK1 160
arq reject-unknown-prefix
no shutdown

```

Example 8-4 Gatekeeper Clustering Configuration for Gatekeeper 2

```

gatekeeper
zone local CHC cisco.com 10.1.2.1
zone local SJC_GK2 cisco.com
zone local NYC_GK2 cisco.com
!
zone cluster local CHC_Cluster CHC
  element CHC_GK3 10.1.3.1 1719
  element CHC_GK1 10.1.1.1 1719
!
zone cluster local SJC_Cluster SJC_GK2
  element SJC 10.1.1.1 1719
  element SJC_GK3 10.1.3.1 1719
!
zone cluster local NYC_Cluster NYC_GK2
  element NYC_GK1 10.1.1.1 1719
  element NYC 10.1.3.1 1719
!
zone prefix SJC_GK2 40852.....
zone prefix NYC_GK2 21251.....
zone prefix CHC 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 8-3](#) and [Example 8-4](#):

- 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 because the **bandwidth total** command does not work 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 8-5](#) shows the output from the **show gatekeeper endpoints** command on Gatekeeper 1 (SJC).

Example 8-5 Output for Gatekeeper Endpoints

```

                                GATEKEEPER ENDPOINT REGISTRATION
                                =====
CallSignalAddr  Port  RASignalAddr  Port  Zone Name      Type      Flags
-----
10.1.1.12      1307  10.1.1.12     1254  SJC            VOIP-GW
H323-ID: SJC-to-GK-trunk_1
10.1.1.12      4422  10.1.1.12     4330  SJC            VOIP-GW
H323-ID: SJC-to-GK-trunk_2
10.1.2.12      4587  10.1.2.12     4330  CHC_GK1        VOIP-GW    A
H323-ID: CHC-to-GK-trunk_1
10.1.3.21      2249  10.1.3.21     1245  NYC_GK1        VOIP-GW    A
H323-ID: NYC-to-GK-trunk_1
Total number of active registrations = 4

```

Directory Gatekeeper Redundancy

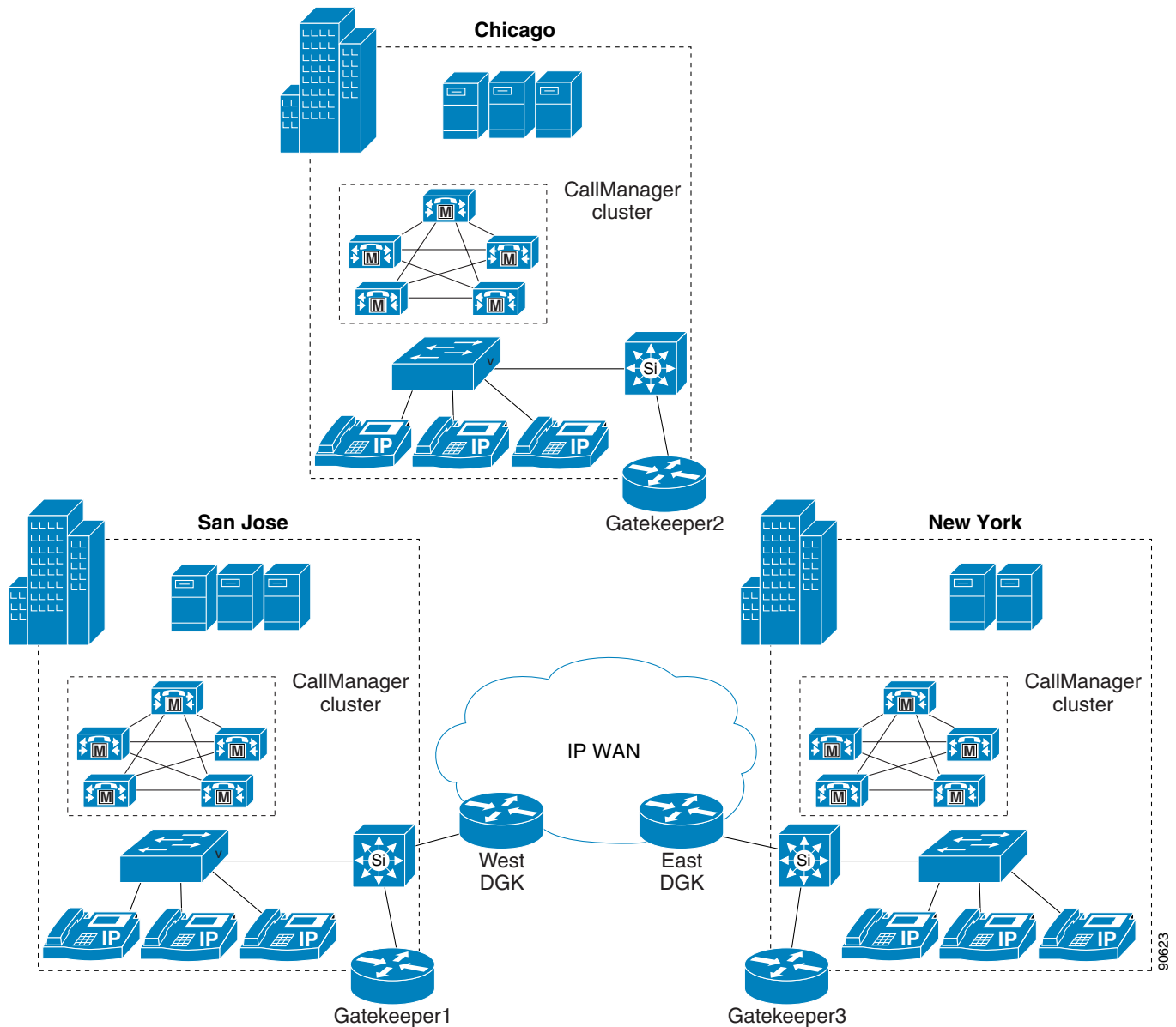
You can implement directory gatekeeper redundancy by using HSRP or by configuring multiple identical directory gatekeepers. When a gatekeeper is configured with multiple remote zones using the same zone prefix, the gatekeeper can use either of the following methods:

- 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 multiple 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 8-8 illustrates a Cisco CallManager distributed call processing environment with two active directory gatekeepers.

Figure 8-8 Redundant Directory Gatekeepers



Example 8-6 and Example 8-7 show the configurations for the two directory gatekeepers in Figure 8-8.

Example 8-6 Configuration for West Directory Gatekeeper

```
gatekeeper
zone local DGKW customer.com 10.1.10.1
zone remote SJC customer.com 10.1.1.1
zone remote CHC customer.com 10.1.2.1
zone remote NYC customer.com 10.1.3.1
zone prefix SJC 408.....
zone prefix CHC 720.....
```

```
zone prefix NYC 212.....
lrq forward-queries
no shutdown
```

Example 8-7 Configuration for East Directory Gatekeeper

```
gatekeeper
zone local DGKE customer.com 10.1.12.1
zone remote SJC customer.com 10.1.1.1
zone remote CHC customer.com 10.1.2.1
zone remote NYC customer.com 10.1.3.1
zone prefix SJC 408.....
zone prefix CHC 720.....
zone prefix NYC 212.....
lrq forward-queries
no shutdown
```

The following notes also apply to [Example 8-6](#) and [Example 8-7](#):

- 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 (*) could be used in the zone prefix to simplify configuration, but the use of dots (.) is more specific. 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 8-8](#), [Example 8-9](#), and [Example 8-10](#) show the configurations for Gatekeepers 1 to 3 in [Figure 8-8](#).

Example 8-8 Configuration for Gatekeeper 1 (SJC)

```
zone local SJC customer.com 10.1.1.1
zone remote DGKW customer.com 10.1.10.1
zone remote DGKE customer.com 10.1.12.1
zone prefix SJC 408.....
zone prefix DGKW .....
zone prefix DGKE .....
bandwidth remote 192
gw-type-prefix 1# default-technology
arq reject-unknown-prefix
no shutdown
```

Example 8-9 Configuration for Gatekeeper 2 (CHC)

```
gatekeeper
zone local GK-CHC customer.com 10.1.2.1
zone remote DGKE customer.com 10.1.12.1
zone remote DGKW customer.com 10.1.10.1
zone prefix CHC 720.....
```

```

zone prefix DGKE .....
zone prefix DGKW .....
bandwidth remote 160
gw-type-prefix 1# default-technology
arq reject-unknown-prefix
no shutdown

```

Example 8-10 Configuration for Gatekeeper 3 (NYC)

```

gatekeeper
zone local NYC customer.com 10.1.3.1
zone remote DGKE customer.com 10.1.12.1
zone remote DGKW customer.com 10.1.10.1
zone prefix 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 8-8](#), [Example 8-9](#), and [Example 8-10](#):

- 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.

Interoperability of Cisco CallManager and CallManager Express

This section explains the requirements for interoperability and internetworking of Cisco CallManager with Cisco CallManager Express (CME, formerly called Cisco IOS Telephony Services, or ITS) using H.323 protocol in a multi-site IP telephony deployment. This section highlights the supplementary services such as call transfers and forwards between phones controlled by Cisco CallManager and phones controlled by CME.

Interoperability between Cisco CallManager and CME requires the following minimum software releases:

- Cisco CallManager Express Release 3.1 or later, with Cisco IOS Release 12.3(7)T or later and the IP VOICE feature set
- Cisco CallManager Release 3.3(3) or later

Cisco CME 3.1 adds two new features:

- Cisco CallManager auto-detection

This feature automatically detects remote Cisco CallManager endpoints and invokes an H.323-to-H.323 hairpin on a call transferred or forwarded to/from a Cisco CallManager endpoint. In addition, the auto-detection feature handles signaling decisions such as ringback generation between IP phones controlled by Cisco CallManager and CME.

- H.323-to-H.323 call hairpinning support

H.323-to-H.323 hairpinning is required because Cisco CallManager does not support H.450 specifications.



Note

Cisco CME 3.0 and earlier versions of ITS do not support interoperability with Cisco CallManager using H.323 call hairpinning. These earlier versions used loopback-dn ports to achieve interoperability with Cisco CallManager.

The following sections provide guidelines for achieving interoperability between Cisco CallManager and CME:

- [Multi-Site IP Telephony Deployments Using Cisco CallManager and CME, page 8-26](#)
- [Multi-Site IP Telephony Deployments Using Cisco CallManager, CME, and an H.450 Tandem Gateway, page 8-28](#)

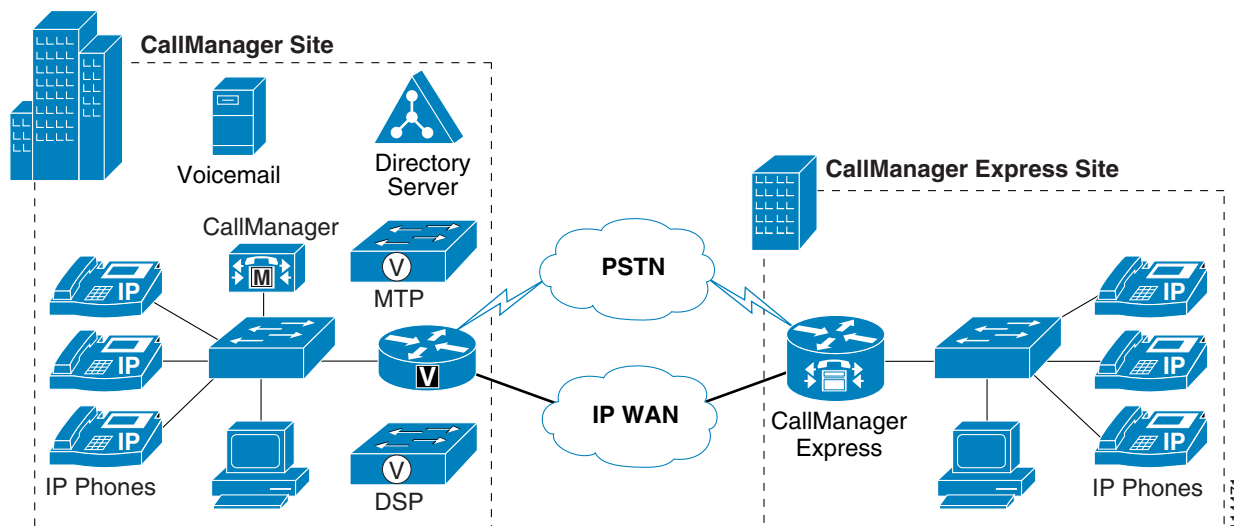
For more information about CME, refer to the product documentation for Cisco CallManager Express, available online at

<http://www.cisco.com>

Multi-Site IP Telephony Deployments Using Cisco CallManager and CME

Cisco CallManager can communicate directly with CME using an H.323 interface. [Figure 8-9](#) shows an IP Telephony deployment with Cisco CallManager networked directly with Cisco CME.

Figure 8-9 Multi-Site IP Telephony Deployment with Cisco CallManager and CME



Best Practices

Follow these guidelines and best practices when using the deployment model illustrated in [Figure 8-9](#):

- Configure a gatekeeper-controlled intercluster trunk (ICT) on Cisco CallManager. Cisco CallManager registers with the gatekeeper through the ICT, and CME registers with the gatekeeper as an H.323 gateway.
- Enable the use of media termination points (MTPs) when using an intercluster trunk. Cisco CME does not initiate any Terminal Capabilities Set (TCS) signaling, and use of an MTP will prevent any TCS exchange between Cisco CallManager and CME.
- In Cisco CallManager, set the service parameter **Send H225 user info message to H225 info for Ringback**.
- Use the Cisco CallManager dial plan configuration (route patterns, route lists, and route groups) to send calls to the intercluster trunk that connects to CME.
- Use Cisco CallManager device pools and regions to configure the G.711 codec within a site and the G.729 codec for remote CME sites.
- Configure the **allow-connection h323 to h323** command on CME to allow H.323-to-H.323 call hairpin connections. The Cisco CallManager auto-detection feature in CME 3.1 disables all H.450-based communications on the interface connected to Cisco CallManager. To complete the call transfer or forward between Cisco CallManager and CME phones, an H.323-to-H.323 connection is required.
- Define VoIP dial peers to forward the calls, destined for Cisco CallManager endpoints, to the gatekeeper.
- Configure the G.729 codec for the VoIP dial peers routing calls to remote sites.

[Example 8-11](#) lists a sample configuration for the CME in this deployment model.

Example 8-11 Configuration for Cisco CME 3.1

```
voice service voip
  allow-connections h323 to h323
  h323
dial-peer voice 1 voip          /* To Cisco CallManager endpoints */
  destination-pattern xxxx
  session target ras           /* "ras" if gatekeeper is used, otherwise "ipv4:y.y.y.y" */
  dtmf-relay h245-alphanumeric
  codec g729r8                 /* Voice class can also be used */
  no vad
```

Using the guidelines presented here ensures that basic calls can be completed between Cisco CallManager and CME phones. However, all calls transferred or forwarded to/from a Cisco CallManager phone in this deployment model are hairpinned. Hairpinning requires end-to-end signaling between all devices involved in the call flow. The call is, therefore, more vulnerable to jitter, delay, and network outages. Also, if the call traverses the WAN, it utilizes unnecessary bandwidth. To avoid these potential problems, Cisco recommends deploying an H.450 tandem gateway at the front end of the Cisco CallManager cluster, as described in the next section.

Multi-Site IP Telephony Deployments Using Cisco CallManager, CME, and an H.450 Tandem Gateway

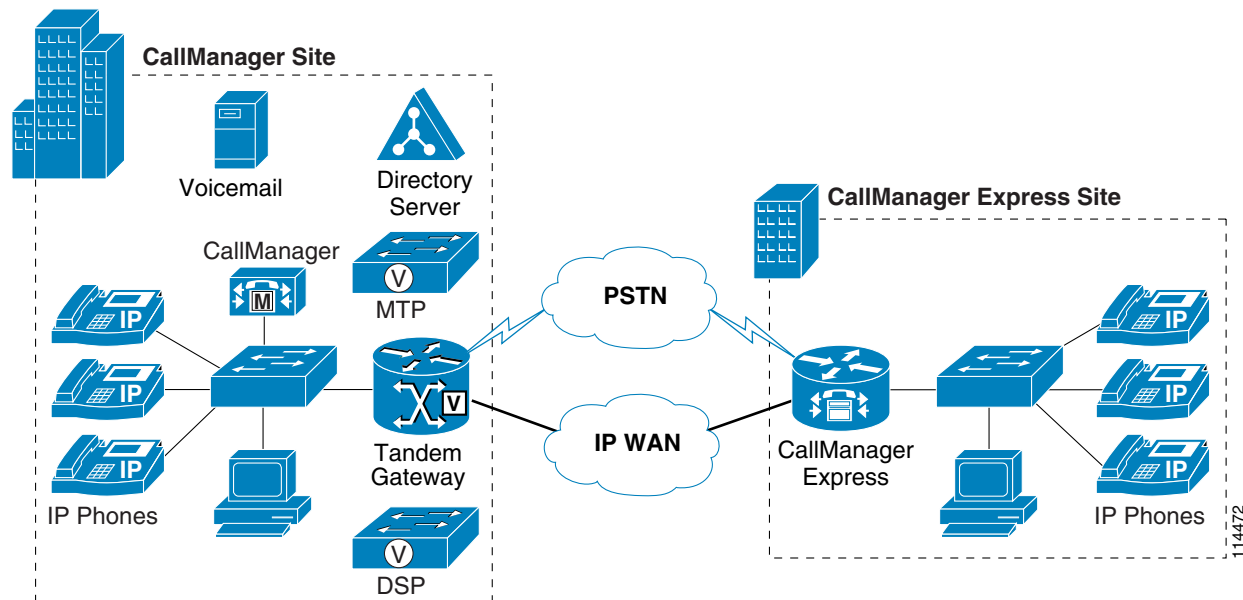
A tandem gateway is a separate router that provides a proxy (or front end) for a system that does not support H.450, such as Cisco CallManager. The tandem gateway can be placed between Cisco CallManager and a CME router, and it provides H.323-to-H.323 call connections to terminate and re-originate calls for transfers and forwards to endpoints that do not support H.450.

The H.450 tandem gateway can also act as the PSTN gateway for the remote Cisco CME systems as well as for Cisco CallManager. A separate PSTN gateway is not required in such cases.

The H.450 tandem gateway must run a Cisco IOS release that is compliant with Cisco CME 3.1 and that supports H.450, such as Cisco IOS Release 12.3(7)T or later with the IP VOICE feature set.

Figure 8-10 shows an IP Telephony deployment with Cisco CallManager connected to Cisco CME through an H.450 tandem gateway.

Figure 8-10 Multi-Site IP Telephony Deployment with Cisco CallManager, CME, and an H.450 Tandem Gateway



Best Practices

Follow these guidelines and best practices when using the deployment model illustrated in Figure 8-10:

- Configure a non-gatekeeper controlled intercluster trunk (ICT) between Cisco CallManager and the H.450 tandem gateway.
- Enable the use of media termination points (MTPs) when using an intercluster trunk. Cisco CME does not initiate any Terminal Capabilities Set (TCS) signaling, and use of an MTP will prevent any TCS exchange between Cisco CallManager and CME.
- In Cisco CallManager, set the service parameter **Send H225 user info message** to **H225 info for Ringback**.
- Use the Cisco CallManager dial plan configuration (route patterns, route lists, and route groups) to send calls to the intercluster trunk that connects to the H.450 tandem gateway.

- Register Cisco CME and the H.450 tandem gateway as H.323 gateways on the gatekeeper.
- Configure the **allow-connection h323 to h323** command on the H.450 tandem gateway to allow H.323-to-H.323 call hairpin connections. Remote CME routers need not have this command enabled.
- Define VoIP dial peers on the H.450 tandem gateway to route calls to the Cisco CallManager and CME endpoints.
- Define VoIP dial peers on CME to forward calls, destined for Cisco CallManager endpoints, to the H.450 tandem gateway.
- Use the same voice codecs for all dial peers on the H.450 tandem gateway that route calls to/from Cisco CallManager and CME. Codec re-negotiation is not supported. Therefore, if G.729 is used over the WAN, configure G.729 codecs for all VoIP dial peers that route calls to remote sites. Also, use Cisco CallManager device pools and regions to configure G.729 codecs between the Cisco CallManager IP phones and the H.450 tandem gateway.
- You can use the H.450 tandem gateway as an MGCP gateway for PSTN connections. If an H.323 gateway is required for PSTN connections, then you must use a separate gateway.

[Example 8-12](#) lists a sample configuration for the H.450 tandem gateway.

Example 8-12 Configuration for H.450 Tandem Gateway

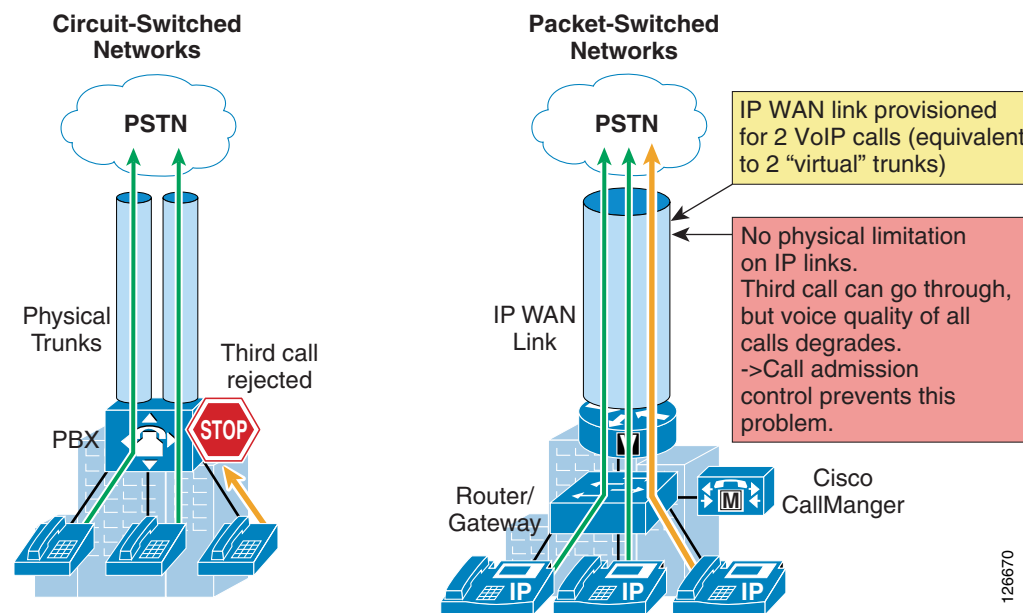
```
voice service voip
  allow-connections h323 to h323
  supplementary-service h450.12
  h323
dial-peer voice 1 voip          /* To Cisco CallManager endpoints */
  destination-pattern xxxx
  session target ipv4:y.y.y.y
  dtmf-relay h245-alphanumeric
  codec g729r8
  no vad
dial-peer voice 1 voip          /* To Cisco CallManager endpoints */
  destination-pattern zzzz
  session target ras           /* "ras" if gatekeeper is used, otherwise "ipv4:a.b.c.d" */
  dtmf-relay h245-alphanumeric
  codec g729r8
  no vad
```




Call Admission Control

The call admission control function is an essential component of any Cisco IP Communications system that involves multiple sites connected through an IP WAN. In order to better understand what call admission control does and why it is needed, consider the example in [Figure 9-1](#).

Figure 9-1 Why Call Admission Control is Needed



As shown on the left side of [Figure 9-1](#), traditional TDM-based PBXs operate within circuit-switched networks, where a circuit is established each time a call is set up. As a consequence, when a legacy PBX is connected to the PSTN or to another PBX, a certain number of physical trunks must be provisioned. When calls have to be set up to the PSTN or to another PBX, the PBX selects a trunk from those that are available. If no trunks are available, the call is rejected by the PBX and the caller hears a network-busy signal.

Now consider the Cisco IP Communications system shown on the right side of [Figure 9-1](#). Because it is based on a packet-switched network (the IP network), no circuits are established to set up an IP Communications call. Instead, the IP packets containing the voice samples are simply routed across the IP network together with other types of data packets. Quality of Service (QoS) is used to differentiate the voice packets from the data packets, but bandwidth resources, especially on IP WAN links, are not infinite. Therefore, network administrators dedicate a certain amount of "priority" bandwidth to voice

traffic on each IP WAN link. However, once the provisioned bandwidth has been fully utilized, the Cisco IP Communications system must reject subsequent calls to avoid oversubscription of the priority queue on the IP WAN link, which would cause quality degradation for all voice calls. This function is known as call admission control, and is essential to guarantee good voice quality in a multi-site deployment.

This chapter examines the following main aspect of call admission control:

- [Call Admission Control Elements, page 9-2](#)

This section describes the different call admission control mechanisms available across the various components of a Cisco IP Communications system, such as Cisco CallManager locations, Cisco IOS gatekeeper, RSVP, and the IP-to-IP gateway.

- [Call Admission Control Design, page 9-20](#)

This section shows how to apply and combine the mechanisms described in the previous section, based on the IP WAN topology (simple hub-and-spoke, two-tier hub-and-spoke, MPLS, or complex topologies) and also based on the Cisco CallManager deployment model adopted (centralized, distributed, or hybrid centralized/distributed).

Call Admission Control Elements

This section provides design and configuration guidelines for the following call admission control elements within a Cisco IP Communications system:

- [Cisco CallManager Locations, page 9-2](#)
- [Gatekeeper, page 9-6](#)
- [RSVP, page 9-7](#)
- [IP-to-IP Gateway, page 9-14](#)

Cisco CallManager Locations

Cisco CallManager provides a simple mechanism known as *locations* for implementing call admission control in centralized call processing deployments. When configuring a device in Cisco CallManager, you can assign the device to a location and allocate a certain amount of bandwidth for calls to or from each location. (See [Figure 9-2](#).) The locations configured in Cisco CallManager are virtual locations and not real, physical locations. Cisco CallManager has no knowledge of the physical locations of devices. Therefore, if a device is moved from one physical location to another, the system administrator must perform a manual update of its location information so that Cisco CallManager can correctly calculate bandwidth allocation for that device. Each device is in location <None> by default. Location <None> is a special unnamed location that has unlimited audio and video bandwidth. If the devices at a branch site are configured in the <None> location, all the phone calls to and from that branch device will be admitted by the call admission control mechanism.

Cisco CallManager uses the location's available audio and video bandwidths to control the number of audio calls and video calls to and from that location. If the location's audio and video bandwidths are set to **Unlimited**, there will be unlimited bandwidth available for that location, and every audio or video call to and from that location will be permitted by Cisco CallManager. On the other hand, if the bandwidth values are set to a certain number of kilobits per second (kbps), Cisco CallManager will allow calls in and out of that location as long as the aggregate bandwidth used by all active calls is less than or equal to the configured values. The video bandwidth for the location can also be set to **None**, which denies every video call to and from this location but does not affect the video calls within the same location.

For video calls, the video location bandwidth takes into account both the video and the audio portions of the call. Hence, for a video call, no bandwidth is deducted from the audio bandwidth pool.

The locations-based call admission control mechanism also takes into account mid-call changes of call type. For example, if an inter-site video call is established, Cisco CallManager will subtract the appropriate amount of video bandwidth from the respective locations. If this video call changes to an audio call via a transfer to a device that is not video-capable, Cisco CallManager will return the allocated bandwidth to the video pool and allocate the appropriate amount of bandwidth from the audio pool. Calls that change from audio to video will cause the opposite change of bandwidth allocation.



Note

Prior to Cisco CallManager Release 3.1, there was only one primary (active) Cisco CallManager server in the cluster when locations-based call admission control is used. With Cisco CallManager Release 3.1 and later, the locations bandwidth is shared among all Cisco CallManager subscribers in the cluster, thus enabling the locations mechanism to work with any size cluster.

Figure 9-2 Defining a Location in Cisco CallManager

The screenshot displays the Cisco CallManager Administration web interface. The top navigation bar includes links for System, Route Plan, Service, Feature, Device, User, Application, and Help. The main header area features the Cisco CallManager Administration logo and the Cisco Systems logo. The page title is "Location Configuration". On the right side, there are three links: "Add a New Location", "Back to Find/List Locations", and "Dependency Records". The main content area shows the configuration for a location named "Branch_1". The status is "Ready". Below the status, there are four buttons: "Copy", "Update", "Delete", and "Resync Bandwidth". The configuration is divided into three sections: "Location Information", "Audio Calls Information", and "Video Calls Information". In the "Location Information" section, the "Location Name*" is set to "Branch_1". In the "Audio Calls Information" section, the "Audio Bandwidth*" is set to "256" kbps, with radio buttons for "Unlimited" and "256". A note below states: "If the audio quality is poor or choppy, lower the bandwidth setting. For ISDN use multiples of 56 kbps or 64 kbps." In the "Video Calls Information" section, the "Video Bandwidth*" is set to "384" kbps, with radio buttons for "None", "Unlimited", and "384". A note at the bottom states: "* indicates required item".

Figure 9-2 shows the configuration for the location Branch_1, with 256 kbps of available audio bandwidth and 384 kbps of available video bandwidth. This location can support up to three G.711 audio calls (at 80 kbps per call) or ten G.729 audio calls (at 24 kbps per call), or any combination of both that does not exceed 256 kbps. The location can also support various numbers of video calls depending on the video and audio codecs being used. For example, one video call with G.711 audio codec and H.261 or H.263 video codec at a requested rate of 320 kbps consumes 384 kbps of bandwidth. So do three video calls with G.729 audio codec and H.261 or H.263 video codec at a requested rate of 120 kbps.

126907

**Note**

With Cisco CallManager Release 4.0 and later, Cisco IP phones can make encrypted phone calls. The payload of an encrypted call is slightly larger than that of a non-encrypted call, so if you are planning to send encrypted calls across your network, you should adapt the bandwidth provisioning in the IP WAN routers to accommodate this additional bandwidth. The other factor to consider when provisioning queues in IP WAN routers is the overhead introduced by Layer 2 headers. For more details about bandwidth calculations for encrypted calls and the various Layer 2 technologies, refer to [Bandwidth Provisioning, page 3-23](#).

The devices that can be placed into locations include:

- IP phones
- CTI ports
- H.323 clients
- CTI route points (only if they terminate media)
- Conference bridges
- Music on hold (MoH) servers
- Gateways
- Trunks

The location configured on a CTI route point is taken into account by Cisco CallManager only if an application has registered with that route point to handle media. If the CTI route point is used exclusively to route calls and does not terminate media, the configured location is ignored by Cisco CallManager.

**Note**

Transcoders and MTPs cannot be assigned to a location, so they should normally be placed at the central site. However, with Cisco CallManager Releases 3.2(3), 3.3(3), 4.0(1), and later, it is also possible to place a transcoder at a remote site as long as it is co-located with the G.711-only device. Essentially, Cisco CallManager assumes that the call involving a transcoder is always a G.729 call, and it subtracts 24 kbps of bandwidth from the relevant locations.

Call admission control does not apply to calls between devices within the same location because Cisco CallManager assumes that those devices reside on the same LAN, which has unlimited available bandwidth.

When a call is placed from one location to another, Cisco CallManager deducts the appropriate amount of bandwidth at both locations. For example, if a device in Branch_1 places a G.711 call to a device in Branch_2, Cisco CallManager deducts 80 kbps from the available bandwidth at both locations. When the call has completed, Cisco CallManager returns the bandwidth to the affected locations. If there isn't enough bandwidth at either Branch_1 or Branch_2 location, the call will be denied by Cisco CallManager and the caller will hear the network busy tone. If the calling device is an IP phone with a display, that device also displays the message "Not Enough Bandwidth."

Starting with Cisco CallManager Release 3.3, when an intra-cluster call is denied due to lack of bandwidth in either the calling or called party's location, Cisco CallManager can automatically reroute the call to the destination via the PSTN by means of the Automated Alternate Routing (AAR) feature. Cisco CallManager uses the called number and the called party's External Phone Number Mask to derive the full E.164 destination number. The AAR group configured on the calling party's line determines the access code to be prefixed to the E.164 number, and the calling device's AAR calling search space determines how to choose the local egress PSTN gateway.

**Note**

AAR is invoked only when the locations-based call admission control denies the call due to lack of bandwidth. AAR is not invoked when the IP WAN is unavailable or other connectivity issues cause the called device to become unregistered with Cisco CallManager. In such cases, the calls will be redirected to the target specified in the Call Forward Busy field of the called device. For more information about AAR, refer to [IP Telephony Deployment Models, page 2-1](#), and [Automated Alternate Routing, page 10-19](#).

When an external call is placed from or to one location within the Cisco CallManager cluster, Cisco CallManager will also perform the locations-based call admission control calculation. If there is sufficient bandwidth to support the call, Cisco CallManager will admit the call and allocate the appropriate amount of bandwidth at the location. If there isn't sufficient bandwidth in the location, the call will be denied by Cisco CallManager.

[Figure 9-3](#) shows the configuration for an IP phone assigned to location Branch_1. Cisco CallManager admits each call placed to or from this IP phone if there is sufficient bandwidth available at Branch_1 to support the call. If Branch_1 does not have sufficient bandwidth when the IP phone makes an outbound call, Cisco CallManager will reroute the call via AAR and send the call out the Branch_1 PSTN gateway.

Figure 9-3 *Assigning a Device to a Location*

The screenshot displays the Cisco CallManager Administration web interface. The top navigation bar includes links for System, Route Plan, Service, Feature, Device, User, Application, and Help. The main header reads "Cisco CallManager Administration For Cisco IP Telephony Solutions" with the Cisco Systems logo on the right. The page title is "Phone Configuration". On the right side, there are links: "Add a new phone", "Add/Update Speed Dials", "Subscribe/Unsubscribe Services", "Dependency Records", and "Back to Find/List Phones".

On the left, under "Directory Numbers", the "Base Phone" section shows two lines: "Line 1 - 3921001 in Branch_1_Partition" and "Line 2 - Add new DN". The main content area shows the configuration for "Phone: SEP003094C3063C (SEP003094C3063C)". The registration status is "Unknown" and the IP address is "Not Found". The status is "Ready". There are buttons for "Copy", "Update", "Delete", and "Reset Phone".

Below this is the "Phone Configuration (Model = Cisco 7970)" section, which includes "Device Information". The fields are as follows:

MAC Address*	003094C3063C
Description	SEP003094C3063C
Owner User ID	(Select User ID)
Device Pool*	Default (View details)
Calling Search Space	Branch_1_CSS
AAR Calling Search Space	Branch_1_AAR_CSS
Media Resource Group List	< None >
User Hold Audio Source	< None >
Network Hold Audio Source	< None >
Location	Branch_1

126908

Table 9-1 lists the amount of bandwidth requested by the Cisco CallManager locations algorithm for various call speeds.

Table 9-1 Amount of Bandwidth Requested by the Locations Algorithm

Call Speed	Location Bandwidth Value
G.711 audio call (64 kbps)	80 kbps
G.729 audio call (8 kbps)	24 kbps
128-kbps video call	128 kbps
384-kbps video call	384 kbps
512-kbps video call	512 kbps
768-kbps video call	768 kbps

Gatekeeper

A Cisco IOS gatekeeper can provide call routing and call admission control between devices such as Cisco CallManager, Cisco CallManager Express, or H.323 gateways connected to legacy PBXs. It uses the H.323 Registration Admission Status (RAS) protocol to communicate with these devices and route calls across the network.

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 simple hub-and-spoke topologies. In some cases, when multiple gatekeepers are deployed, it is possible to support two-tier hub-and-spoke topologies. Refer to the section on [Call Admission Control Design, page 9-20](#), for detailed topology examples.

The Cisco 2600, 3600, 3700, 2800, 3800, and 7200 Series routers all support the gatekeeper feature. You can configure Cisco IOS gatekeepers in a number of different ways for redundancy, load balancing, and hierarchical call routing. This section focuses on the call admission control aspect of the gatekeeper feature. For redundancy and scalability considerations, refer to the Gatekeeper section in the Call Processing chapter. For call routing considerations, refer to [Call Routing in Cisco IOS with a Gatekeeper, page 10-30](#).

In order to understand the call admission control capabilities of a Cisco IOS gatekeeper, it is helpful to recall the concept of gatekeeper *zones*. A zone is a collection of H.323 devices, such as endpoints, gateways, or Multipoint Control Units (MCUs), that register with a gatekeeper. There can be only one active gatekeeper per zone, and you can define up to 100 local zones on a single gatekeeper.

A local zone is a zone that is actively handled by that gatekeeper – that is, all H.323 devices assigned to that zone register with that gatekeeper.

When multiple gatekeepers are deployed in the same network, a zone is configured as a local zone on only one gatekeeper. On the other gatekeepers, that zone is configured as a remote zone. This configuration instructs the gatekeeper to forward calls destined for that zone to the gatekeeper that "owns it" (that is, the gatekeeper on which that zone is configured as a local zone).

Use the **bandwidth** command to manage the number of calls that the gatekeeper will allow, thus providing call admission control functionality. This command has several options, but the most relevant are the following:

- The **interzone** option controls the amount of bandwidth for all calls in or out of a given local zone.
- The **session** option controls the amount of bandwidth per call for a given local zone.
- The **remote** option controls the total amount of bandwidth to or from all remote zones.

For example, suppose your network has two Cisco CallManager clusters connected to the headquarter side via DS-3 WAN links. Also assume that you want to allow twenty G.711 audio calls and five 384-kbps video calls on each link and that you want all video calls to be restricted to 384 kbps. In this case, you can use the **bandwidth interzone** command to restrict the bandwidth for each zone to the total of twenty G.711 calls plus five 384-kbps video calls, and you can use the **bandwidth session** command to restrict each call to a maximum of 384 kbps, as shown in the following example:

```
gatekeeper
zone local ccm-1 customer.com 10.10.10.10
zone local ccm-2
bandwidth interzone ccm-1 6400
bandwidth session ccm-1 768
bandwidth interzone ccm-2 6400
bandwidth session ccm-2 768
```

The bandwidth value used by gatekeeper is double the bit-rate of the call, excluding Layer 2, IP, and RTP overhead. For example, a G.711 audio call that uses 64 kbps would be denoted as 128 kbps in the gatekeeper, and a 384-kbps video call would be denoted as 768 kbps. Table 9-2 shows the bandwidth values used by the gatekeeper feature for some of the most popular call speeds.

Table 9-2 Gatekeeper Bandwidth Settings for Various Call Speeds

Call Speed	Gatekeeper Bandwidth Value
G.711 audio call (64 kbps)	128 kbps
G.729 audio call (8 kbps)	16 kbps
128-kbps video call	256 kbps
384-kbps video call	768 kbps
512-kbps video call	1024 kbps
768-kbps video call	1536 kbps

RSVP

The Resource Reservation Protocol (RSVP) is the first significant industry-standard protocol for dynamically setting up end-to-end QoS across a heterogeneous network. RSVP, which runs over IP, enables an application to reserve network bandwidth dynamically. Using RSVP, applications can request a certain level of QoS for a data flow across a network. Because of its distributed and dynamic nature, RSVP is capable of reserving bandwidth across any network topology, thus providing a real topology-aware call admission control mechanism.

Most of the Cisco IOS router platforms support RSVP, and it can be introduced in a Cisco IP Communications network by using the Cisco Multiservice IP-to-IP Gateway feature described in the section on [IP-to-IP Gateway](#), page 9-14.

An in-depth analysis of the RSVP protocol is outside the scope of this document. For more detailed information, refer to the relevant Request For Comments (RFC) documents published by the Internet Engineering Task Force (IETF) and available at

<http://www.ietf.org>

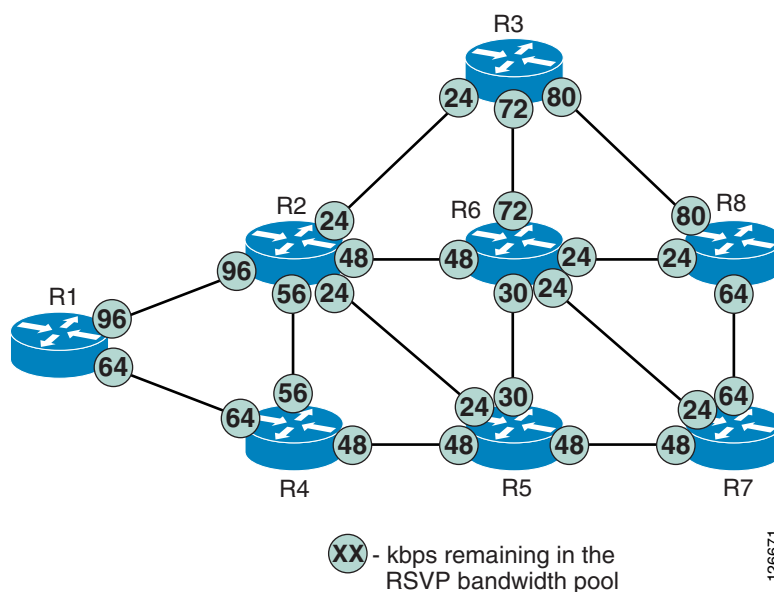
This section focuses on describing the overall principles of the RSVP behavior, the different RSVP operation models in Cisco IOS, and some design best-practices for using RSVP for voice and video applications in an enterprise network.

RSVP Principles

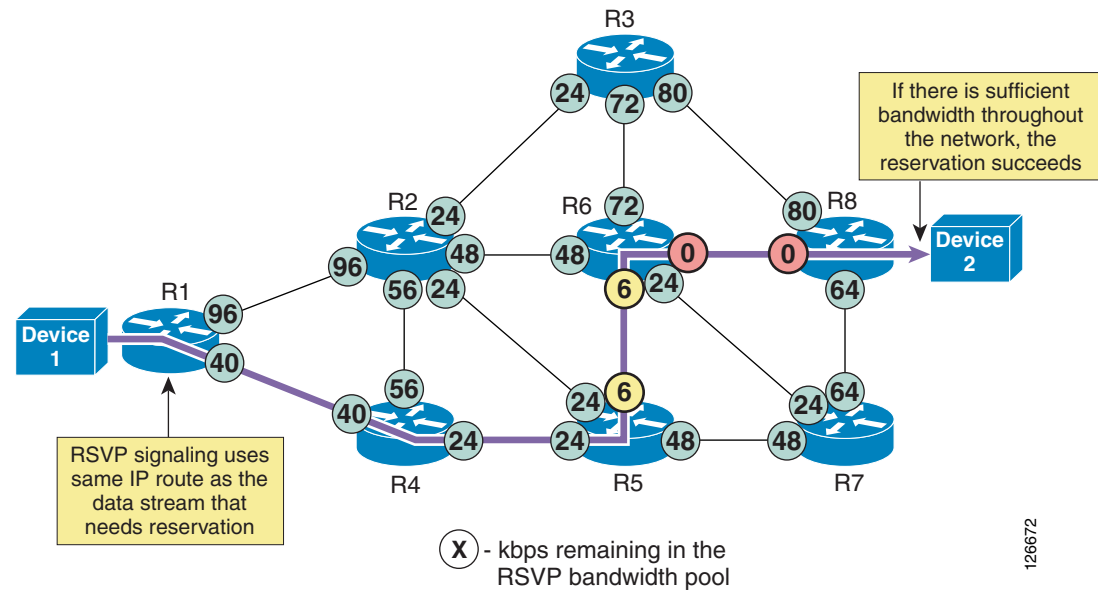
To better understand the basic principles of how RSVP performs call admission control and bandwidth reservation in a network, consider the simple example depicted in [Figure 9-4](#). This example does not analyze the exact message exchanges and protocol behaviors, but rather focus on the end results from a functionality perspective.

Assume that RSVP is enabled on each router interface in the network shown in [Figure 9-4](#) and that the numbers shown in the circles represent the amount of available RSVP bandwidth remaining on each interface.

Figure 9-4 Sample Network to Show RSVP Principles



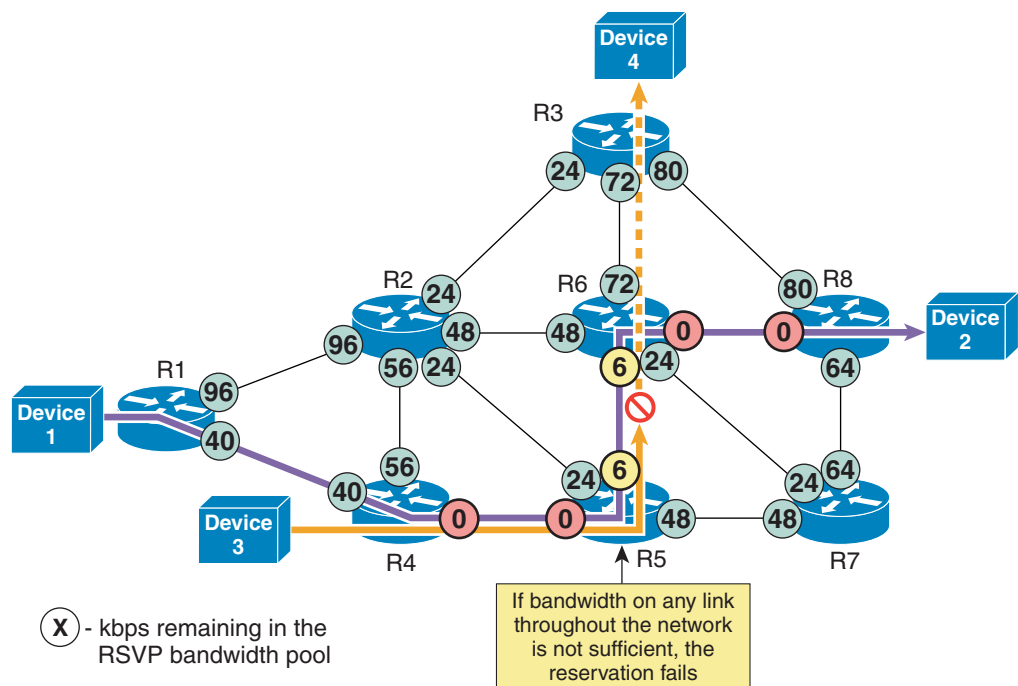
Now consider an RSVP-enabled application that wants to reserve a certain amount of bandwidth for a data stream between two devices. This scenario is depicted in [Figure 9-5](#), which shows a particular data stream that requires 24 kbps of bandwidth from Device 1 to Device 2.

Figure 9-5 *RSVP Signaling for a Successful Reservation*

The following considerations apply to [Figure 9-5](#):

- RSVP does not perform its own routing; instead it uses underlying routing protocols to determine where it should carry reservation requests. As routing changes paths to adapt to topology changes, RSVP adapts its reservation to the new paths wherever reservations are in place.
- The RSVP protocol attempts to establish an end-to-end reservation by checking for available bandwidth resources on all RSVP-enabled routers along the path from Device 1 to Device 2. As the RSVP messages progress through the network, the available RSVP bandwidth gets decremented by 24 kbps on the involved router interfaces, as shown in [Figure 9-5](#).
- The available bandwidth on all interfaces is sufficient to accept the new data stream, so the reservation succeeds and the application is notified.
- RSVP reservations are unidirectional (in this case, the reservation is established from Device 1 to Device 2, and not vice versa). In the presence of bidirectional applications such as voice and videoconferencing, two reservations must be established, one in each direction.
- RSVP provides transparent operation through router nodes that do not support RSVP. If there are any routers along the path that are not RSVP-enabled, they simply ignore the RSVP messages, and a reservation can still be established. However, in order to have an end-to-end QoS guarantee, you have to ensure that there is no possibility of bandwidth congestion on the links controlled by the non-RSVP routers.

After a reservation has been successfully established between Device 1 and Device 2, now assume that another application requests 24 kbps between Device 3 and Device 4, as depicted in [Figure 9-6](#).

Figure 9-6 *RSVP Signaling for an Unsuccessful Reservation*

The following considerations apply to Figure 9-6:

- The RSVP protocol attempts to establish an end-to-end reservation by checking for available bandwidth resources on all RSVP-enabled routers along the path from Device 3 to Device 4. As the RSVP messages progress through the network, the available RSVP bandwidth gets decremented by 24 kbps on the involved router interfaces, as shown in Figure 9-6.
- In this example, the available bandwidth on the link between routers R5 and R6 is not sufficient to accept the new data stream, so the reservation fails and the application is notified. The available RSVP bandwidth on each interface along the path is then restored to its previous value.
- The application can then decide what to do. It could abandon the data transfer or decide to send it anyway with no QoS guarantees as best-effort traffic.

RSVP Operation Models

By issuing the following Cisco IOS command in interface configuration mode on each Cisco IOS router interface, you can enable RSVP and define the maximum amount of bandwidth that it can control:

```
ip rsvp bandwidth [interface-kbps] [single-flow-kbps]
```

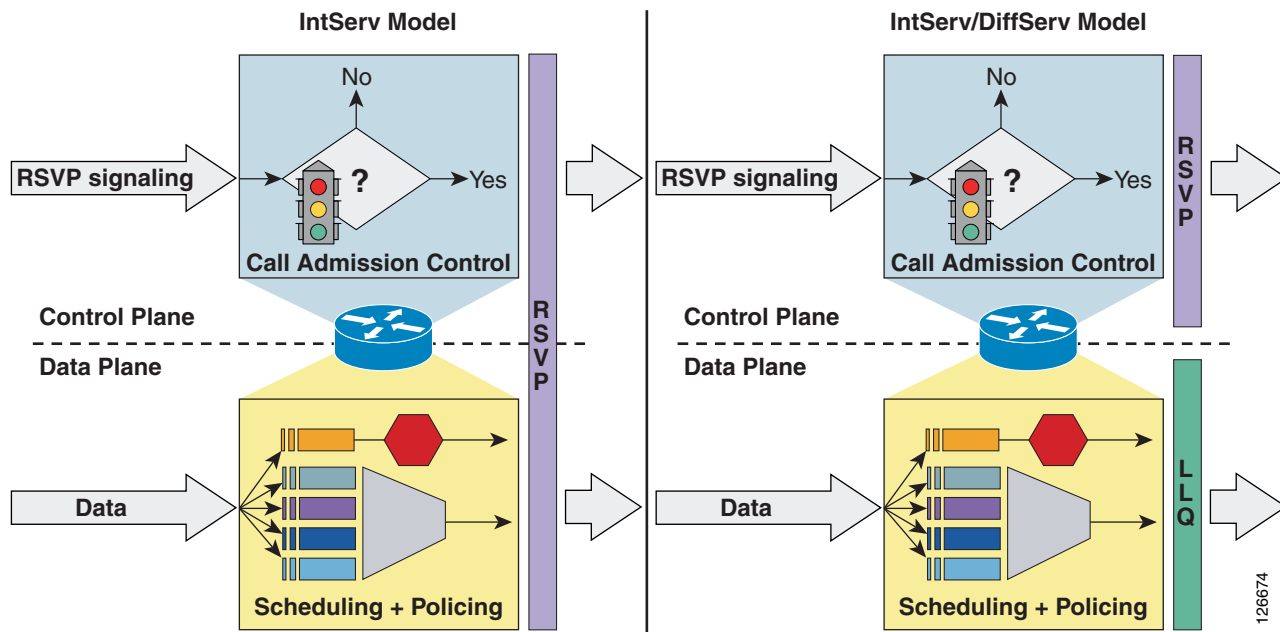
The *interface-kbps* parameter specifies the upper limit of bandwidth that RSVP can reserve on the given interface, while the *single-flow-kbps* parameter provides an upper bandwidth limit for each individual reservation (so that flows with higher bandwidth requests will be rejected even if there is bandwidth available on the interface).

Within Cisco IOS, RSVP can be configured to operate according to two different models: the Integrated Services (IntServ) model, described in RFC 2210, or the Integrated Services/Differentiated Services (IntServ/DiffServ) model, described in RFC 2998. Both RFC documents are available on the IETF website at

<http://www.ietf.org>

Figure 9-7 shows the difference between these two approaches from the perspective of a Cisco IOS router.

Figure 9-7 The Two RSVP Operation Models: IntServ and IntServ/DiffServ



The IntServ Model

As shown on the left side of Figure 9-7, RSVP in the IntServ model involves both the control plane and the data plane. In the control plane, RSVP admits or denies the reservation request. In the data plane, it classifies the data packets, polices them based on the traffic description contained in the RSVP messages, and queues them in the appropriate queue. The classification that RSVP performs is based on the 5-tuple consisting of the source IP address, source port, destination IP address, destination port, and protocol number.

In this model, all data packets transiting through the router must be intercepted by RSVP so that RSVP can inspect the 5-tuple and look for a match among the established reservations. If a match is found, the packets are scheduled and policed by RSVP according to the reservation's traffic specification.

To use the IntServ operation model on a Cisco IOS router, use the following commands in interface configuration mode:

```
ip rsvp resource-provider wfq [interface | pvc]
no ip rsvp data-packet classification
```

When these commands are active, RSVP admits or rejects new reservations, not only based on the upper bandwidth limit defined within the **ip rsvp bandwidth** command, but also based on the actual bandwidth resources available. For example, if there are LLQ classes with bandwidth statements, these amounts are deducted from the bandwidth pool that can be assigned to RSVP reservations. While LLQ classes

statically allocate bandwidth at configuration time, RSVP does not allocate any amount until a reservation request is received. Therefore, it is important to ensure that an appropriate percentage of the available interface bandwidth is *not* allocated to LLQ classes, so that it can be used by RSVP as reservation requests are received.

Because RSVP is in control of assigning packets to the various queues within this model, it is possible to define a mechanism for RSVP to know whether or not to place flows in the Priority Queue (PQ) by using the following Cisco IOS command in interface configuration mode:

```
ip rsvp pq-profile [r [b [p-to-r]]]
```

RSVP uses the parameters *r*, *b*, and *p-to-r* to determine if the flow being signaled for is a voice flow that needs PQ treatment. These parameters represent the following values:

- *r* = the average traffic rate in bytes per second
- *b* = the maximum burst of a flow in bytes
- *p-to-r* = the ratio of peak rate to average rate, expressed as a percentage

If the traffic characteristics specified by the RSVP messages for a certain flow are less than or equal to the parameters in the command, then RSVP will direct the flow into the PQ. If no parameters are provided with the command, the following values, representing the largest of the commonly used voice codecs (G.711), are used as default:

- *r* = 12288 bytes per second
- *b* = 592 bytes
- *p-to-r* = 110%

The IntServ/DiffServ Model

As shown on the right side of [Figure 9-7](#), RSVP in the IntServ/DiffServ model involves only the control plane performing admission control but does not involve the data plane. This means that the call admission control function is separate from the scheduling and policing functions, which can be performed by the Low Latency Queuing (LLQ) algorithm according to predefined class maps, policy maps, and service policies.

With the IntServ/DiffServ model, it is therefore possible to add RSVP call admission control to a network that is already using a Differentiated Services approach to QoS. RSVP admits or rejects calls based on a preconfigured bandwidth amount, but the actual scheduling is based on the pre-existing LLQ criteria such as the DSCP value of each packet.

To use the IntServ/DiffServ operation model on a Cisco IOS router, use the following commands in interface configuration mode:

```
ip rsvp resource-provider none
ip rsvp data-packet classification none
```

When these commands are active, RSVP admits or rejects new reservations uniquely based on the upper bandwidth limits defined within the **ip rsvp bandwidth** command, independently from the actual bandwidth resources available on the interface. Once admitted, the RSVP flows are subject to the same scheduling rules as all other non-RSVP traffic (for example, LLQ class and policy maps). Therefore, it is important to ensure that the RSVP-enabled traffic is marked with the appropriate DSCP value and that the bandwidth of the corresponding PQ or CBWFQ queues is provisioned to accommodate both RSVP-enabled traffic and all other traffic.

In this operating model, the **ip rsvp pq-profile** command is inactive because RSVP does not control the scheduling function.

Design Best Practices

In a Cisco IP Communications network, because all devices mark their bearer and signaling traffic with the appropriate DSCP values and the QoS approach is currently based on the Differentiated Services model, two design choices are possible when adding RSVP to the IP WAN: the IntServ model or the IntServ/DiffServ model.

Cisco recommends that you use the IntServ/DiffServ model if one of the following statements is true:

- The only traffic destined for the Priority Queue in your IP WAN interfaces is RSVP-enabled traffic.
- All the non-RSVP traffic destined for the Priority Queue can be deterministically limited to a certain amount by an out-of-band call admission control mechanism (such as Cisco CallManager locations or a Cisco IOS gatekeeper).

Under such conditions, you can configure your router so that the bandwidth value specified in the LLQ priority command is never exceeded and the Priority Queue is never oversubscribed.

If all the PQ traffic is RSVP-enabled, you can simply ensure that the value specified in the **ip rsvp bandwidth** command and in the **priority** command match. If some PQ traffic is not RSVP-enabled, you must ensure that the sum of the values specified in the **ip rsvp bandwidth** command and in the out-of-band call admission control mechanism do not exceed the bandwidth value specified in the **priority** command.

Meanwhile, other types of traffic can keep accessing their Class-Based Weighted Fair Queues (CBWFQ) based on their DSCP values. With this approach, the processing overhead introduced by RSVP is minimal because it needs to analyze only RSVP signaling messages and is not involved in the scheduling of data packets. However, it is very important to ensure that the non-RSVP traffic destined for the PQ does not exceed a preconfigured bandwidth value because any excess traffic would affect both the misbehaving application and the RSVP-enabled applications.

If the Priority Queue in your IP WAN interfaces must accommodate both RSVP and non-RSVP traffic but you cannot guarantee an upper limit for the non-RSVP traffic, then Cisco recommends that you use the IntServ model instead.

In this case, the non-RSVP traffic is granted a certain amount of PQ bandwidth through an LLQ class with the **priority** keyword, and you have to ensure that sufficient bandwidth is left unallocated on the interface so that RSVP can use it for its flows. You also must use the **ip rsvp pq-profile** command to define which RSVP flows should be assigned to the PQ. With this approach, RSVP introduces more processing overhead on the IP WAN routers because it needs to analyze every single data packet. However, the advantage is that, if the non-RSVP PQ traffic misbehaves and exceeds the bandwidth value configured in LLQ, the RSVP traffic will not be affected by the policing performed on the LLQ priority class.

Regardless of the approach chosen, use the following command to ensure that the RSVP signaling packets are marked according to the Cisco baseline QoS recommendations and that they are treated like the other types of signaling traffic:

```
ip rsvp signalling dscp 24
```

Refer to the chapter on [Network Infrastructure, page 3-1](#), for more details about the recommended DSCP values for the various types of traffic.

IP-to-IP Gateway

The Cisco Multiservice IP-to-IP Gateway (also referred to as IP-IP gateway or IPIPGW) can be used to ease the restriction of hub-and-spoke topologies for the IP WAN connections between Cisco CallManager clusters and/or H.323 gateways.

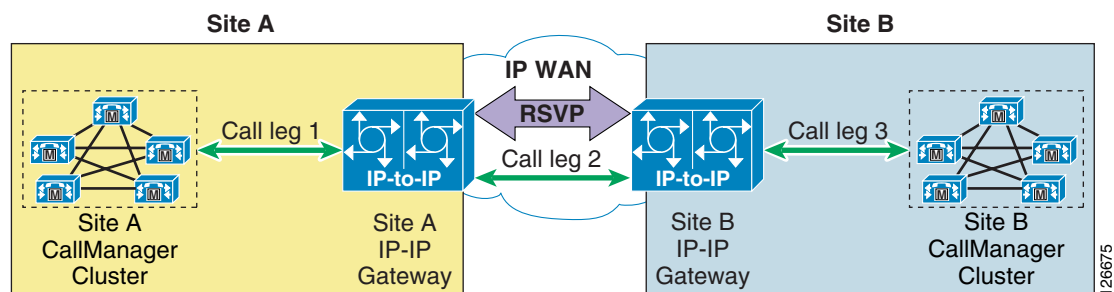
This Cisco IOS feature provides a mechanism to enable H.323 Voice over IP (VoIP) and videoconferencing calls from one IP network to another. The main purpose of the IP-IP gateway is to provide a control point and a demarcation for VoIP and video calls traversing administrative domains. This gateway performs most of the same functions of a PSTN-to-IP gateway, but typically joins two IP call legs rather than a PSTN and an IP call leg.

The most interesting feature of the IP-IP gateway in an enterprise IP Communications environment is that it can generate an RSVP reservation for each call that traverses it. As described in the section on [RSVP, page 9-7](#), RSVP is a network-based signaling protocol that provides a topology-aware call admission control mechanism that does not require a hub-and-spoke topology but works with any network topology.

As a consequence, you can perform call admission control over any IP WAN topology by inserting two IP-IP gateways in the call flow and enabling RSVP between them. [Figure 9-8](#) shows a basic example where two sites, A and B, each have a Cisco CallManager cluster and are connected via an IP WAN that has an arbitrary topology. An IP-IP gateway is also located at each site, and the two Cisco CallManager clusters are configured so that all inter-site calls are routed via a trunk that points to the local IP-IP gateway. When a call is set up between Site A and Site B, the following events occur:

- Cisco CallManager at Site A sets up a call through an H.323 trunk to Site A's IP-IP gateway. (This is call leg 1 in the figure.)
- Site A's IP-IP gateway attempts to establish another call to Site B's IP-IP gateway, but first it uses RSVP to allocate bandwidth resources along the IP WAN path.
- If the RSVP reservation is successful, call leg 2 is established between the two IP-IP gateways.
- Site B's IP-IP gateway generates another call to Site B's Cisco CallManager cluster. (This is call leg 3 in the figure.)

Figure 9-8 Simple Example of the IP-to-IP Gateway for RSVP Call Admission Control



The example in [Figure 9-8](#) is a simple scenario in which all calls between Cisco CallManager clusters are routed via a pair of IP-IP gateways. However, in many real-world cases this approach might not prove scalable or flexible enough. For these cases, Cisco IOS gatekeepers can be used to provide a wider range of communication options between Cisco CallManager clusters, H.323 gateways, H.323 videoconferencing endpoints, and IP-IP gateways.

The following sections provide more details on how to use IP-IP gateways in conjunction with via-zone gatekeepers and also present design best practices, scalability and redundancy considerations, and configuration guidelines.

**Note**

All scenarios involving IP-IP gateways described in this section apply to calls between different Cisco CallManager clusters. Cisco does not recommend inserting IP-IP gateways for calls between endpoints registered to the same Cisco CallManager cluster.

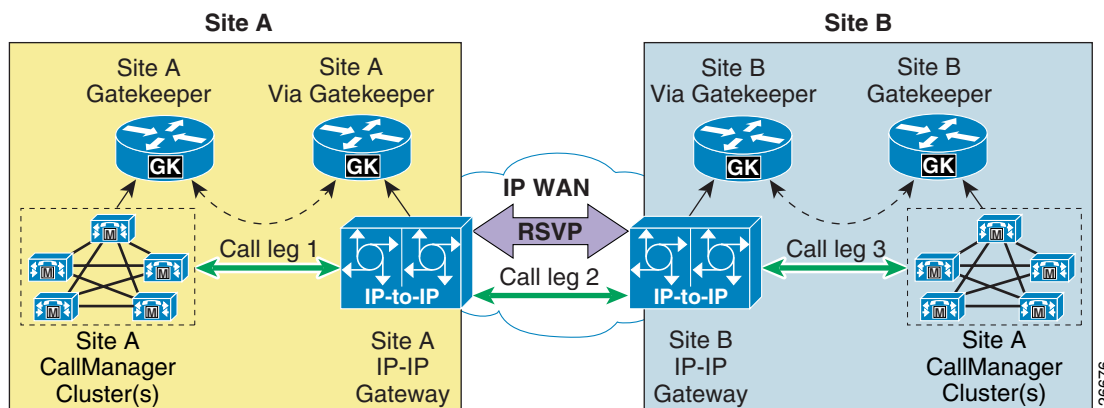
Via-Zone Gatekeeper

Traditional Cisco IOS gatekeeper functionality has been extended to accommodate for IP-IP gateways through the concept of a *via-zone gatekeeper*. A via-zone gatekeeper differs from legacy gatekeepers in how it uses LRQ and ARQ messages for call routing. Using via-zone gatekeepers will maintain normal clusters and functionality. Legacy gatekeepers examine incoming LRQs based on the called number, and more specifically the dialedDigits field in the destinationInfo portion of the LRQ. Via-zone gatekeepers look at the origination point of the LRQ before looking at the called number. If an LRQ comes from a gatekeeper listed in the via-zone gatekeeper's remote zone configurations, the gatekeeper checks to see that the zone remote configuration contains an **invia** or **outvia** keyword. If the configuration contains these keywords, the gatekeeper uses the new via-zone behavior; if not, it uses legacy behavior.

For ARQ messages, the gatekeeper determines if an **outvia** keyword is configured on the destination zone. If the **outvia** keyword is configured and the zone named with the **outvia** keyword is local to the gatekeeper, the call is directed to an IP-IP gateway in that zone by returning an ACF pointing to the IP-IP gateway. If the zone named with the **outvia** keyword is remote, the gatekeeper sends a location request to the outvia gatekeeper rather than the remote zone gatekeeper. The **invia** keyword is not used in processing the ARQ.

Figure 9-9 shows an example of how IP-IP gateways and via-zone gatekeepers can be used in conjunction with Cisco CallManager clusters and legacy gatekeepers to provide call routing and call admission control. The following considerations apply to this scenario:

- Site A's Cisco CallManager clusters use the Site A gatekeeper to route calls directly between them.
- The Site A gatekeeper sends all calls directed to Site B's E.164 numbers to the Site A via-zone gatekeeper.
- The Site A via-zone gatekeeper inserts an IP-IP gateway for all calls coming from or destined to the Site A gatekeeper.
- The Site A IP-IP gateway attempts RSVP reservations before sending calls toward Site B's IP-IP gateway.
- The Cisco CallManager clusters, gatekeepers, and the IP-IP gateway at Site B are configured in a similar fashion to their counterparts at Site A.

Figure 9-9 IP-to-IP Gateway for RSVP Using Via-Zone Gatekeepers

Design Best Practices

When deploying IP-IP gateways in conjunction with Cisco CallManager in order to enable RSVP call admission control in the IP WAN, observe these design best practices:

- When configuring trunks in Cisco CallManager for voice-only communication with other Cisco CallManager clusters through one or more IP-IP gateways, use intercluster trunks and check "Media Termination Point required" in the trunk configuration page so that supplementary services can be used across the IP-IP gateway(s).
- Cisco recommends that you place the MTP resources in the same site as the IP-IP gateway to avoid further IP WAN bandwidth usage when calling across the intercluster trunks. These MTP resources may be software-based (for example, on a Cisco MCS server or on a Cisco IOS router) or hardware-based (for example, on a Catalyst 6500 with a Cisco Communications Media Module or on a Cisco IOS router with a NM-HDV network module). Refer to the chapter on [Media Resources](#), page 6-1, for a complete list of available MTP resources.
- Cisco CallManager supplementary services such as hold/resume, transfer, and conference are supported across IP-IP gateways. However, because MTPs are used, media packets will transit through the initial MTP resources for the entire duration of the call, with the potential for hairpinning in the case of subsequent call transfers.
- If video calls need to be established across clusters through IP-IP gateways, define a dedicated intercluster trunk for video calls without selecting the "Media Termination Point required" option (because MTPs do not support video calls). No supplementary services will be possible across these trunks.
- Configure direct intercluster trunks (non-gatekeeper controlled) from Cisco CallManager to the IP-IP gateway only when you want all intercluster calls to use the IP-IP gateway. In this case, the IP-IP gateway can still use a gatekeeper to resolve the remote destination.
- Configure gatekeeper-controlled intercluster trunks in Cisco CallManager when you want to use a gatekeeper to resolve intercluster calls and decide whether they need to go through an IP-IP gateway or be routed directly. This approach is more flexible and allows for greater scalability.
- Compatibility with Cisco CallManager on the IP-IP gateway is available on Cisco IOS Release 12.3(1) or later. Cisco recommends that you use Cisco IOS Release 12.3(7)T or later.
- Keep the gatekeeper and via-zone gatekeeper functions separate by running them on different router platforms. Each IP-IP gateway should have a dedicated via-zone gatekeeper.

- Configure the IP-IP gateway to request a guaranteed-delay RSVP reservation, and make the reservation mandatory for the success of the call. (For details, see the section on [Configuration Guidelines](#), page 9-17.)
- You can run the via-zone gatekeeper function and the IP-IP gateway function on the same router platform (coresident); however, be aware of the scalability requirements described in the section on [Redundancy](#), page 9-17.
- Do not use IP-IP gateways for calls between endpoints controlled by the same Cisco CallManager cluster.

Redundancy

Redundancy and scalability can be provided by registering multiple IP-IP gateways with the same via-zone gatekeeper and in the same via-zone. The via-zone gatekeeper will automatically distribute the incoming calls between all the IP-IP gateways registered in the same via-zone, using a round-robin algorithm.

When an IP-IP gateway fails, it loses its registration to the via-zone gatekeeper, and the gatekeeper removes it from the list of available resources.

It is also possible to manually configure maximum-load thresholds within the IP-IP gateway so that a certain IP-IP gateway stops being selected for new calls when more than a certain percentage of its circuits are in use, and it becomes available again when the circuits in use drop below a certain percentage. The Cisco IOS commands used for this configuration are:

- On the IP-IP gateway:

```
ip circuit max-calls max-call-number
```

- On the gatekeeper:

```
endpoint resource-threshold onset onset-threshold abatement abatement-threshold
```

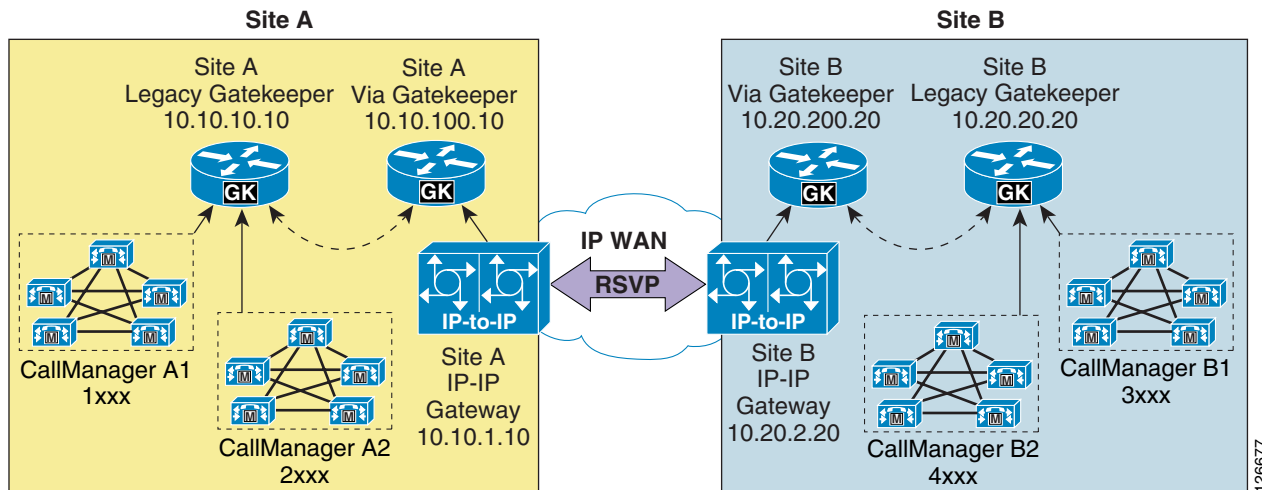
For more details on these commands, refer to the Cisco IOS command reference documentation available at

<http://www.cisco.com>

Configuration Guidelines

This section presents a simple configuration example based on the network diagram shown in [Figure 9-10](#). This section is not intended as an exhaustive command reference guide, but rather as a collection of guidelines that can prove helpful in the most common deployment scenarios. Complete information on how to configure IP-IP gateways and via-zone gatekeepers is provided in the online documentation for the Cisco Multiservice IP-to-IP Gateway, available at

<http://www.cisco.com>

Figure 9-10 Configuration Example for IP-IP Gateway with Via-Zone Gatekeeper

For the network shown in [Figure 9-10](#), assume that there are two Cisco CallManager clusters at Site A: cluster A1, with phone extensions 1xxx, and cluster A2, with phone extensions 2xxx. There are also two Cisco CallManager clusters at Site B: cluster B1, with phone extensions 3xxx, and cluster B2, with phone extensions 4xxx.

The following subsections show the relevant configurations for devices located at Site A, so that calls within Site A are routed directly between Cisco CallManager clusters (using Site A's legacy gatekeeper) while calls to Site B are routed through the two IP-IP gateways (using the respective legacy gatekeepers and via-zone gatekeepers).

Cisco CallManager

Both cluster A1 and cluster A2 use two gatekeeper-controlled intercluster trunks. The first one, T1, is configured with MTP required and points to Site A's legacy gatekeeper. The second one, T2, also points to Site A's legacy gatekeeper, but it has no MTP requirement.

A [34]XXX route pattern points to trunk T1 through a route list/route group construct, so as to reach Site B's clusters through the gatekeepers and the IP-IP gateways.

Another route pattern (2XXX for cluster A1 and 1XXX for cluster A2) allows cluster A1 and A2 to communicate with each other through the gatekeeper, by pointing to trunk T2 through a route list/route group construct.

Legacy Gatekeeper

The legacy gatekeeper at Site A routes calls between clusters A1 and A2 directly, while it sends all calls for Site B (extensions 3xxx and 4xxx) to Site A's via-zone gatekeeper. [Example 9-1](#) shows the relevant configuration.

Example 9-1 Legacy Gatekeeper Configuration for Site A

```
gatekeeper
zone local CCM-A1 customer.com 10.10.10.10
zone local CCM-A2 customer.com
zone remote A-VIAGK customer.com 10.10.100.10
zone prefix CCM-A1 1...
zone prefix CCM-A2 2...
```

```

zone prefix A-VIAGK 3...
zone prefix A-VIAGK 4...
gw-type-prefix 1#* default-technology
arq reject-unknown-prefix
no shutdown

```

Via-zone Gatekeeper

The via-zone gatekeeper at Site A sends calls directed to Site B's Cisco CallManager clusters (extensions 3xxx and 4xxx) to Site B's via-zone gatekeeper and invokes an IP-IP gateway for calls to or from Site B. Calls directed to Site A's clusters are routed to Site A's legacy gatekeeper without invoking an IP-IP gateway. [Example 9-2](#) shows the relevant configuration.

Example 9-2 Via-Zone Gatekeeper Configuration for Site A

```

gatekeeper
zone local A-VIAGK customer.com 10.10.100.10
zone remote CCM-A1 customer.com 10.10.10.10
zone remote CCM-A2 customer.com 10.10.10.10
zone remote B-VIAGK customer.com 10.20.200.20 invia A-VIAGK outvia A-VIAGK
zone prefix B-VIAGK 3...
zone prefix B-VIAGK 4...
zone prefix CCM-A1 1...
zone prefix CCM-A2 2...
arq reject-unknown-prefix
no shutdown

```

The following considerations apply to the configuration shown in [Example 9-2](#):

- The presence of the **invia** and **outvia** keywords in the command line related to the B-VIAGK remote zone triggers the via-zone gatekeeper behavior for that zone. This means that, for all calls destined to or coming from the B-VIAGK remote zone, the via-zone gatekeeper will invoke an IP-IP gateway resource registered in the A-VIAGK local zone.
- The absence of **invia** or **outvia** keywords in the command line related to the CCM-A1 and CCM-A2 remote zones means that the standard gatekeeper behavior is applied, and no IP-IP gateway is invoked for calls to or from these zones.

IP-IP Gateway

The IP-IP gateway at Site A requests RSVP reservations for voice and video calls directed toward Site B's Cisco CallManager clusters (extensions 3xxx and 4xxx) but not for calls directed toward Site A's Cisco CallManager clusters (extensions 1xxx and 2xxx). [Example 9-3](#) shows the relevant configuration.

Example 9-3 IP-IP Gateway Configuration for Site A

```

voice service voip
allow-connections h323 to h323
h323
h225 h245-address
ccm-compatible
call sync-rsvp slow-start
!
gateway
!
interface FastEthernet0/1
ip address 10.10.1.10 255.255.255.0
h323-gateway voip interface

```

```

h323-gateway voip id A-VIAGK ipaddr 10.10.100.10
h323-gateway voip h323-id A-IPIPGW
h323-gateway voip bind srcaddr 10.10.1.10
!
dial-peer voice 10 voip
destination-pattern [3-4]...
session target ras
req-qos guaranteed-delay audio
req-qos guaranteed-delay video
acc-qos guaranteed-delay audio
acc-qos guaranteed-delay video
codec transparent
!
dial-peer voice 11 voip
destination-pattern [1-2]...
session target ras
codec transparent

```

The following considerations apply to the configuration shown in [Example 9-3](#):

- In Cisco IOS Release 12.3(7)T or later, Cisco CallManager compatibility mode is not enabled by default on the IP-IP gateway, hence the need for the **h225 h245-address** and **ccm-compatible** commands.
- The **call sync-rsvp slow-start** command enables synchronization between H.323 slow-start voice and video calls and RSVP signaling. (This feature is normally enabled by default.) Ensure that this command is not disabled if you want to make RSVP reservations mandatory for voice and video calls.
- The **req-qos guaranteed-delay [audio | video]** commands specify that the IP-IP gateway should request a guaranteed-delay RSVP reservation for voice and video calls that use dial-peer 10, which is the dial-peer used to reach Site B's extensions.
- The **acc-qos guaranteed-delay [audio | video]** commands specify that the minimally acceptable QoS level for voice and video calls is also a guaranteed-delay RSVP reservation. This means that, if the RSVP request fails, the call will also fail, and equates to making the RSVP reservation mandatory. To configure the IP-IP gateway so that the RSVP reservation is optional (so that the call can go through even if the reservation fails), use the commands **acc-qos best-effort [audio | video]** instead.

Call Admission Control Design

This section describes how to apply the call admission control mechanisms to the various Cisco CallManager deployment models and to the following IP WAN topologies:

- [Simple Hub-and-Spoke Topologies, page 9-21](#)
- [Two-Tier Hub-and-Spoke Topologies, page 9-26](#)
- [MPLS-Based Topologies, page 9-34](#)
- [Complex Topologies, page 9-40](#)

For each topology, different sets of design considerations are given based on the Cisco CallManager deployment model adopted.

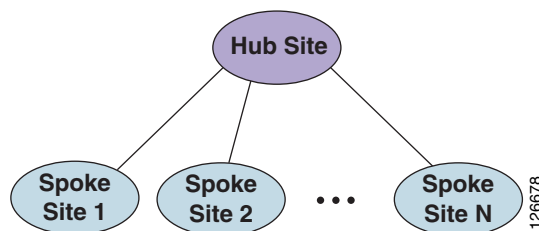
**Note**

All the call admission control solutions described in this section, except for those based on RSVP, rely on static configuration of the available bandwidth for each site. The call admission control mechanism cannot dynamically react to topology changes and link failures. If a site has dual connectivity to the IP WAN, you must account for the worst-case scenario by choosing the link with the smallest bandwidth for your calculations.

Simple Hub-and-Spoke Topologies

Figure 9-11 depicts a simple hub-and-spoke topology, also known as a star topology. In this type of network topology, all sites (called *spoke sites*) are connected to a single, central site (called the *hub site*). There are no direct links between the spoke sites, and every communication between them must transit through the hub site.

Figure 9-11 A Simple Hub-and-Spoke Topology



Design considerations for simple hub-and-spoke topologies apply to traditional Layer 2 IP WAN technologies such as:

- Frame Relay
- ATM
- Frame Relay/ATM Service Interworking
- Leased Lines

For IP WAN deployments based on the MPLS technology, refer to the section on [MPLS-Based Topologies](#), page 9-34.

The remainder of this section contains design best practices for simple hub-and-spoke topologies according to the Cisco CallManager deployment model adopted:

- [Purely Centralized Deployments](#), page 9-22

One or more Cisco CallManager clusters are located at the hub site, but only phones and gateways are located at the spoke sites.

- [Purely Distributed Deployments](#), page 9-22

A Cisco CallManager cluster is located at each site.

- [Hybrid Centralized/Distributed Deployments](#), page 9-25

Cisco CallManager clusters are located at the hub site and at some of the spoke sites, while only phones and gateways are located at the remaining spoke sites.

Purely Centralized Deployments

In multi-site WAN deployments with centralized call processing on a simple hub-and-spoke topology, use Cisco CallManager *locations* for implementing call admission control. Follow these guidelines when using locations for call admission control:

- Configure a separate location in Cisco CallManager for each spoke site.
- Configure the appropriate bandwidth limit for each site according to the type of codec used at that site. (See [Table 9-1](#) for recommended bandwidth settings.)
- Assign all devices at each spoke site to the appropriate location.
- Leave devices at the hub site in the <None> location.
- If you move a device to another location, change its location configuration as well.
- Cisco CallManager supports up to 500 locations.
- If you require automatic rerouting over the PSTN when the WAN bandwidth is not sufficient, configure the automated alternate routing (AAR) feature on Cisco CallManager. (See [Automated Alternate Routing](#), page 10-19.)
- If multiple Cisco CallManager clusters are located at the central site, leave the intercluster trunk devices in the <None> location. You may use a gatekeeper for dial plan resolution. However, gatekeeper call admission control is not necessary in this case because all IP WAN links are controlled by the locations algorithm.

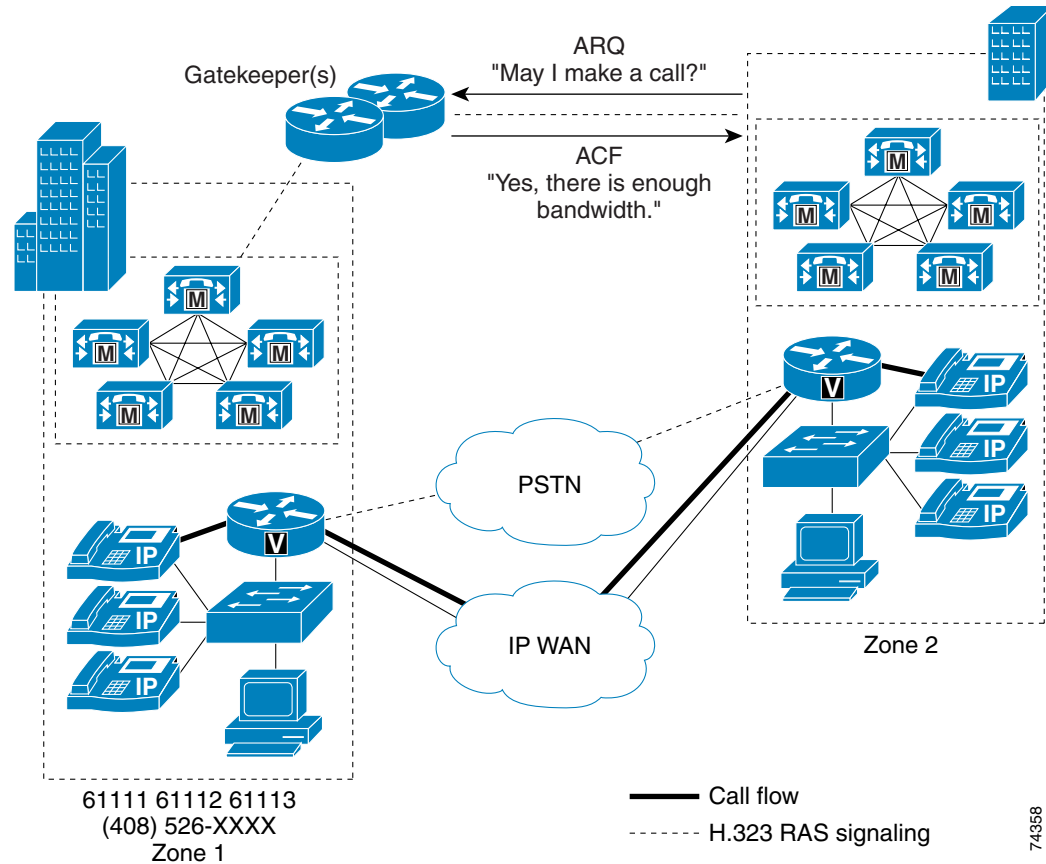
**Note**

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.

Purely Distributed Deployments

For distributed call processing deployments on a simple hub-and-spoke topology, you can implement call admission control with a Cisco IOS gatekeeper. In this design, the call processing agent registers with the Cisco IOS gatekeeper and queries it each time the agent wants to place an IP WAN call. The Cisco IOS gatekeeper associates each call processing agent with a zone that has specific bandwidth limitations. Thus, the Cisco IOS gatekeeper can limit the maximum amount of bandwidth consumed by IP WAN voice calls into or out of a zone.

[Figure 9-12](#) illustrates call admission control with a gatekeeper. In brief, when the call processing agent wants to place an IP WAN call, it first requests permission from the gatekeeper. If the gatekeeper grants permission, the call processing agent places the call across the IP WAN. If the gatekeeper denies the request, the call processing agent can try a secondary path (the PSTN, for example) or can simply fail the call. This design essentially consists of a call accounting method for providing admission control, in which the gatekeeper keeps track of the bandwidth consumed by the IP WAN calls. When you set the maximum bandwidth for a zone, take into account the limitation that voice traffic should not normally consume more than 33% of the WAN link.

Figure 9-12 Call Admission Control Using a Gatekeeper

In summary, the major design implications for gatekeeper call admission control are as follows:

- The gatekeeper provides support for hundreds of sites in a multi-site, distributed call processing environment.
- The gatekeeper tracks the used bandwidth into and out of a zone. The amount of bandwidth consumed by each call depends on the amount requested by the call processing agent and on the type of codec used for the call.
- Bandwidth calculations for the call Admission Request (ARQ) do not include compressed Real-time Transport Protocol (cRTP) or any other transport overhead.
- The topology is a logical hub-and-spoke, based on the gatekeeper zone concept.

The H.225 gatekeeper-controlled trunk enables Cisco CallManager to communicate with Cisco CallManager clusters and other H.323 devices registered to the H.323 gatekeeper. The H.225 gatekeeper-controlled trunk is not recommended in a pure Cisco CallManager environment, but it is required in a mixed environment with Cisco CallManager and Cisco CallManager Express or other H.323 gateway. The H.225 trunk attempts to discover the other H.323 device on a call-by-call basis. If it discovers a device that understands intercluster trunk protocol, it will automatically use that protocol. If it cannot discover the other device, Cisco CallManager will use the standard H.225 protocol.

Intercluster trunk protocol discovery was added to Cisco CallManager in Release 3.2. When interworking with older releases, you must use an intercluster trunk to communicate with the Cisco CallManager clusters or else upgrade all clusters to a minimum of Release 3.2 for correct operation with H.225/H.323 trunks. Without the use of both intercluster and H.225 trunks, it is not possible to have both H.323 devices and Cisco CallManagers all register with a gatekeeper and work correctly.

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, matching the zone to the correct gatekeeper zone for the site.
- Each Cisco CallManager subscriber listed in the device pool's CallManager Redundancy Group registers an H.225 gatekeeper-controlled trunk with the gatekeeper. (Maximum of three.)
- Calls are load-balanced across the registered trunks in the Cisco CallManager cluster.
- Cisco CallManager supports multiple gatekeepers and trunks.
- You can place the trunk in a route group/route list construct to provide automatic PSTN failover. (See [Dial Plan, page 10-1](#), for more details.)
- Configure a separate zone in the gatekeeper for each site supporting Cisco CallManagers, Cisco CallManager Express, or voice gateways.
- Use the **bandwidth interzone** command on the gatekeeper to control bandwidth between Cisco CallManager clusters, Cisco CallManager Express servers, and H.323 devices registered directly with the gatekeeper. (See [Table 9-2](#) for bandwidth settings by codec type.)
- A single Cisco IOS gatekeeper can support up to 100 zones or sites. [Example 9-4](#) illustrates a typical gatekeeper configuration.
- 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 9-4 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
arg reject-unknown-prefix
no shutdown
```

[Example 9-4](#) 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.

**Note**

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 only in deployments based entirely on Cisco CallManager.

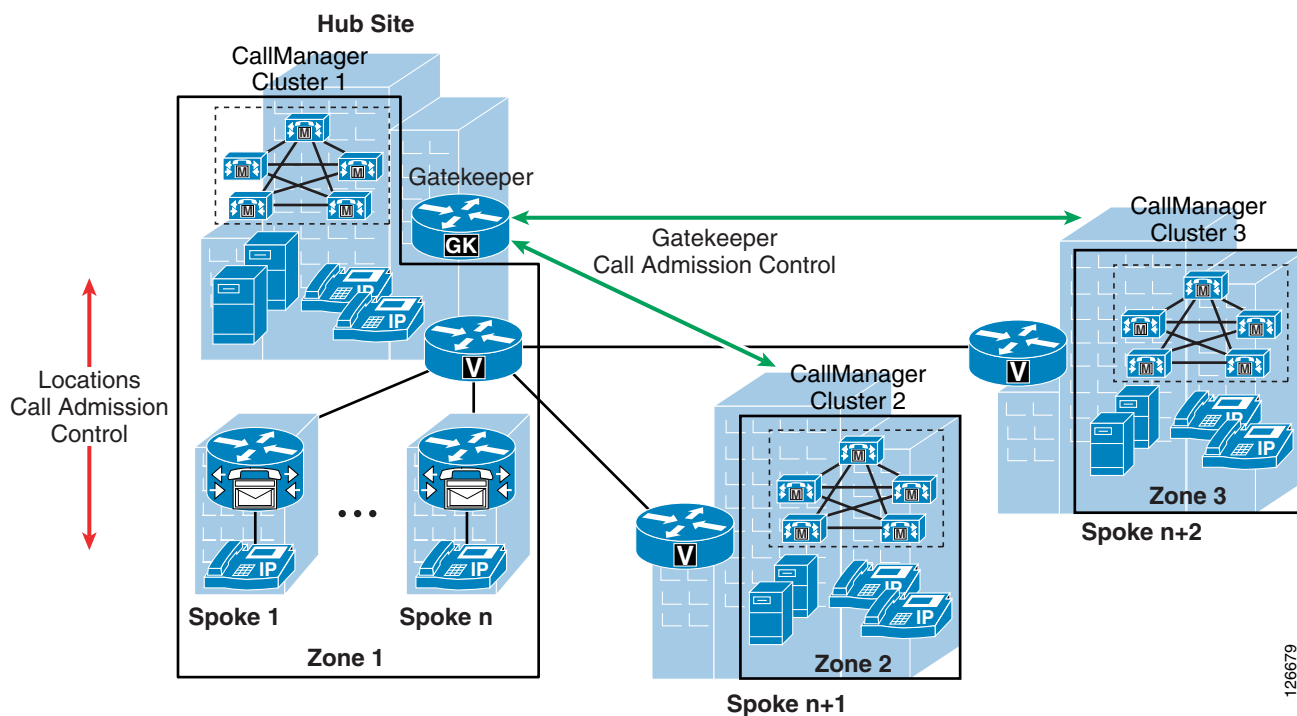
Hybrid Centralized/Distributed Deployments

In hybrid deployments that combine centralized and distributed call processing on a simple hub-and-spoke topology, you can use both a Cisco IOS gatekeeper and Cisco CallManager locations to provide call admission control. [Figure 9-13](#) shows a topology where spoke sites 1 through n are centrally controlled by Cisco CallManager cluster 1, located at the hub site, while spoke sites (n + 1) and (n + 2) have their own on-site Cisco CallManager cluster (clusters 2 and 3).

The following guidelines apply to this type of deployment:

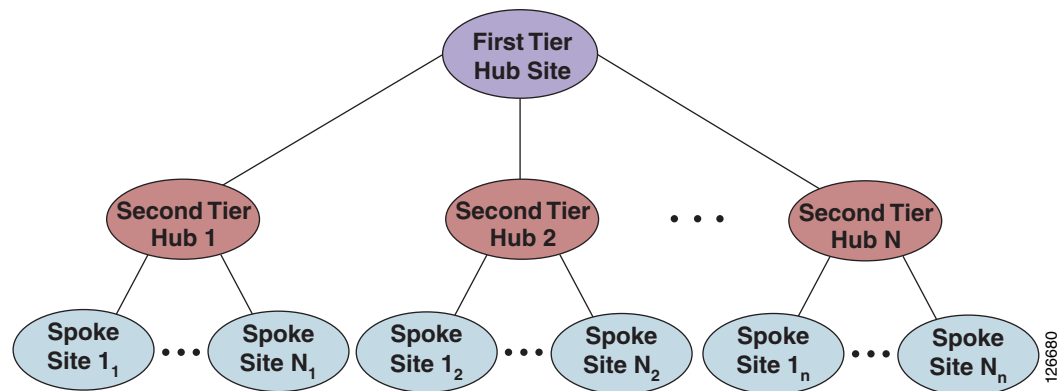
- Use Cisco CallManager locations to provide call admission control to spoke sites 1 through n. Assign each of these spoke sites to a different location, and refer to the section on [Purely Centralized Deployments, page 9-22](#), for additional considerations.
- Use a Cisco IOS gatekeeper located at the hub site to provide call admission control to spoke sites (n + 1) and (n + 2). Assign each Cisco CallManager cluster to a different gatekeeper zone, and refer to the section on [Purely Distributed Deployments, page 9-22](#), for additional considerations.
- Leave the gatekeeper-controlled intercluster (or H.225) trunks in the <None> location because call admission control between clusters is handled by the gatekeeper.

Figure 9-13 Simple Hub-and-Spoke, Hybrid Centralized/Distributed Deployments



Two-Tier Hub-and-Spoke Topologies

Figure 9-14 depicts a two-tier hub-and-spoke topology. This type of network topology can be seen as a hub-and-spoke topology "squared," and it consists of sites at three hierarchical levels: the first-tier hub site, the second-tier hub sites, and the spoke sites. A group of spoke sites are connected to a single second-tier hub site, and each second-tier hub site is in turn connected to the single first-tier hub site. As in the simple hub-and-spoke topology, there are no direct links between the spoke sites, and every communication between them must transit through the second-tier hub site. Similarly, there are no direct links between the second-tier hub sites, and all communications between them must transit through the first-tier hub site.

Figure 9-14 A Two-Tier Hub-and-Spoke Topology

Design considerations for two-tier hub-and-spoke topologies apply to traditional Layer 2 IP WAN technologies such as:

- Frame Relay
- ATM
- Frame Relay/ATM Service Interworking
- Leased Lines

For IP WAN deployments based on the MPLS technology, refer to the section on [MPLS-Based Topologies](#), page 9-34.

The remainder of this section contains design best practices for two-tier hub-and-spoke topologies according to the Cisco CallManager deployment model adopted:

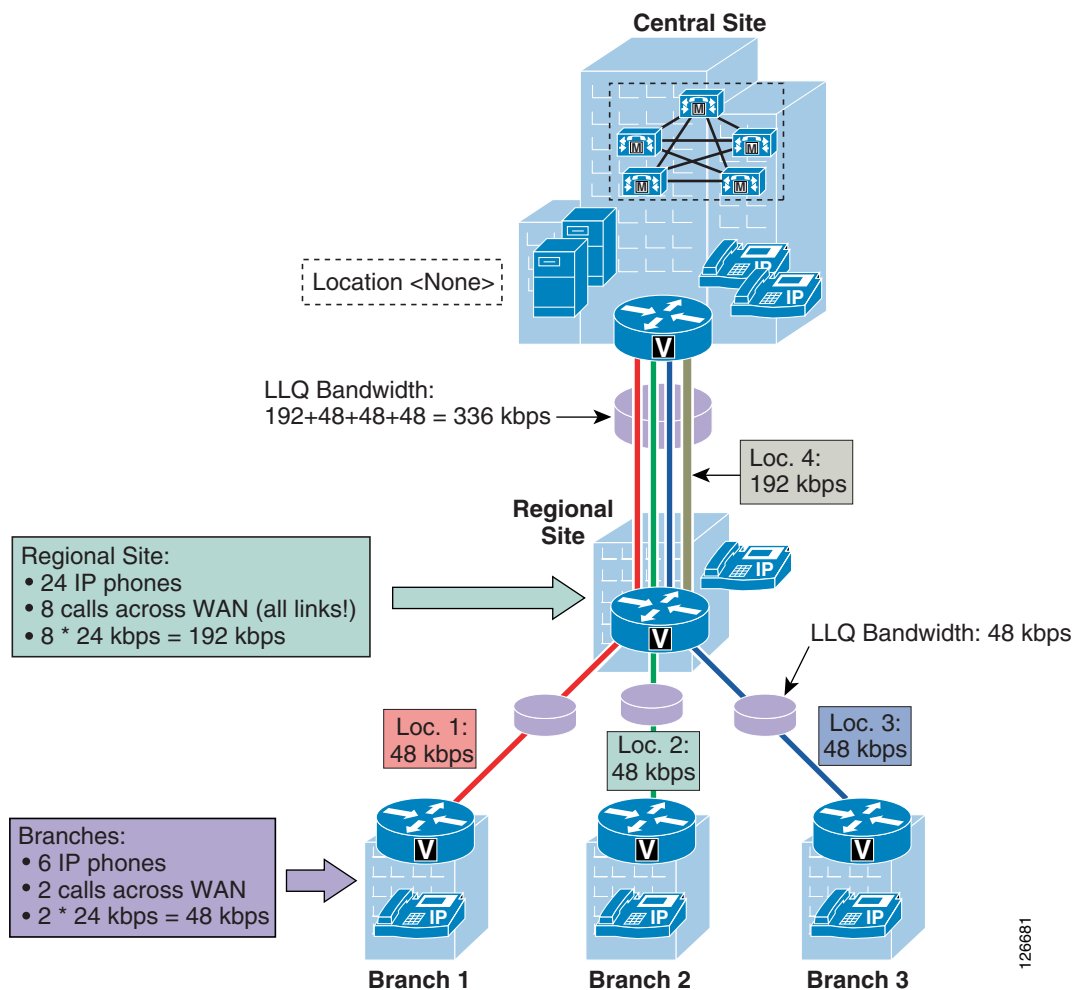
- [Purely Centralized Deployments](#), page 9-27
One or more Cisco CallManager clusters are located at the first-tier hub site, but only phones and gateways are located at the second-tier hub sites and the spoke sites.
- [Purely Distributed Deployments](#), page 9-31
A Cisco CallManager cluster is located at each site.
- [Hybrid Centralized/Distributed Deployments](#), page 9-33
Cisco CallManager clusters are located at the first-tier hub site and at the second-tier hub sites, while only phones and gateways are located at the spoke sites.

Purely Centralized Deployments

Centralized call processing deployments that use a two-tier hub-and-spoke topology pose a problem in terms of call admission control. The Cisco CallManager *locations* algorithm is the only available mechanism for call admission control in a centralized call processing deployment, and it is designed to work for simple hub-and-spoke topologies.

This section describes an attempt to mitigate the issues generated by such topologies, but you should keep in mind that this workaround has limitations, and the recommended approaches to solve this problem are either to change the Cisco CallManager cluster distribution or to change the topology to a simple hub-and-spoke, as described at the end of this section.

[Figure 9-15](#) shows a possible approach to allow a Cisco CallManager cluster in a centralized call processing deployment to support a two-tier hub-and-spoke topology.

Figure 9-15 Two-Tier Hub-and-Spoke in a Purely Centralized Deployment

When using this approach, adhere to the following guidelines:

- Leave the devices at the first-tier hub site (Central Site) in the <None> location.
- Place devices at each spoke site in their own location, and define the location bandwidth according to how many calls you want to allow in or out of each site. In the example of [Figure 9-15](#), the location bandwidth for the three spoke sites (branches 1, 2 and 3) is set to 48 kbps, or enough for two calls using the G.729 codec.
- Provision the priority queue bandwidth in the LLQ configuration for the links between the second-tier hub site and the spoke sites so that it matches the spoke site location bandwidth. In this example, this bandwidth setting is 48 kbps.
- Place devices at each second-tier hub site in their own location, and define the location bandwidth according to how many calls you want to allow in or out of each site. Keep in mind that this bandwidth accounts for all calls entering or leaving the second-tier hub site, so it includes calls toward the underlying spoke sites as well as calls toward the first-tier hub site. In the example of [Figure 9-15](#), the location bandwidth for the second-tier hub site shown (Regional Site) is set to 192 kbps, or enough for eight calls using the G.729 codec.

- Provision the priority queue bandwidth in the LLQ configuration for the links between the first-tier hub site and each second-tier hub site so that it matches the sum of the location bandwidth of the second-tier hub site itself plus all the location bandwidths of the spoke sites connected to that second-tier hub site. In this example, this bandwidth setting is $192 + 48 + 48 + 48 = 336$ kbps.

**Note**

For simplicity, this example shows LLQ bandwidth provisioning based on Layer 3 values (that is, 24 kbps for a G.729 call). In reality, Layer 2 overhead also must be taken into account when provisioning the LLQ queues. Refer to [Table 3-5](#) in the section on [Bandwidth Provisioning, page 3-28](#), for a complete list of bandwidth values for the various Layer 2 WAN technologies.

As shown by the example in [Figure 9-15](#), this approach relies on over-provisioning the priority queue bandwidth between the first-tier and the second-tier hubs to compensate for the unawareness of Cisco CallManager locations about the fact that groups of spoke sites are actually connected via a certain second-tier hub site. As a consequence, the configuration must account for bandwidth as if communications between any two sites has to transit through the first-tier hub site.

While this approach protects voice quality under all circumstances, it also involves the following design considerations and caveats:

- As the number of spoke sites increases per second-tier hub site, the amount of priority queue bandwidth that must be provisioned between the second-tier sites and the first-tier hub site also increases significantly.
- The bandwidth utilization is suboptimal because the calculations are based on the worst-case scenario. For example, in [Figure 9-15](#), if each of the three branches has two active calls to the Regional Site, Cisco CallManager will allow only two additional calls from the Regional Site to the Central Site, even if the actual bandwidth available in the priority queue would allow for 14 additional calls.
- Call completion across the IP WAN cannot be guaranteed even if the actual bandwidth resources would allow for it. For example, in [Figure 9-15](#), if the Regional Site has eight active calls to the Central Site, Cisco CallManager will not allow any of the branches to call the Regional Site, even if the actual bandwidth available would allow for two calls from each branch. On the other hand, in the same situation the branches will be allowed to call the Central Site.

**Note**

Cisco recommends that you do not place all spoke sites connected to a given second-tier hub site in the same location as the second-tier hub site. This solution cannot guarantee voice quality in all scenarios and is *not* supported by Cisco.

As mentioned previously in this section, the workaround illustrated in this section has several limitations. For a better solution, Cisco recommends that you alter the deployment characteristics by implementing one of the following changes:

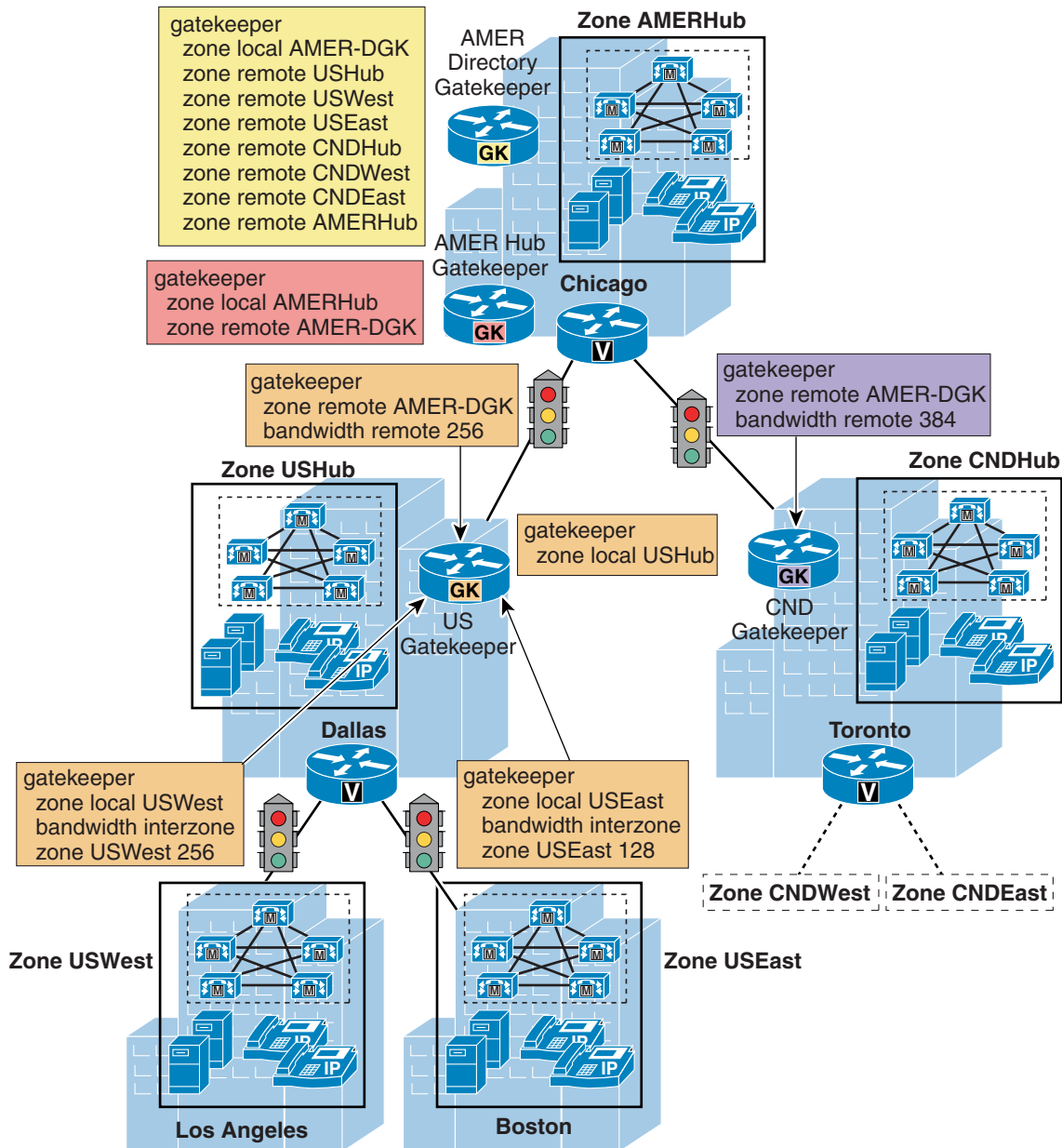
- Based on the number of phones located at the second-tier hub sites and at the spoke sites, increase the link speed and priority queue bandwidth between the second-tier hub sites and the first-tier hub site to values high enough that they cannot possibly be oversubscribed. Under these conditions, the second-tier hub sites can be considered logically co-located with the first-tier hub site, and their devices can be left in the <None> location. Locations can then be used uniquely to control bandwidth usage on the links between the spoke sites and the second-tier hub site.
- Change the network topology to a simple hub-and-spoke by connecting all spoke sites directly to the first-tier hub site. Use Cisco CallManager locations to perform call admission control, as described in the section on [Simple Hub-and-Spoke Topologies, page 9-21](#).

- Change the network topology to an MPLS-based network. Use Cisco CallManager locations to perform call admission control, as described in the section on [MPLS-Based Topologies](#), page 9-34.
- Opt for a hybrid centralized/distributed Cisco CallManager deployment by placing Cisco CallManager clusters at the second-tier hub sites as well as at the first-tier hub site. Use a Cisco IOS gatekeeper to perform call admission control on the links between the first-tier and second-tier hub sites, and use Cisco CallManager locations for the links between the second-tier hub sites and the spoke sites. Refer to the section on [Hybrid Centralized/Distributed Deployments](#), page 9-33, for more details.

Purely Distributed Deployments

In distributed call processing deployments that use a two-tier hub-and-spoke topology, you can use a Cisco IOS gatekeeper to perform call admission control on all IP WAN links. [Figure 9-16](#) shows an example of such a deployment, where a Cisco CallManager cluster is located at each site, Cisco IOS gatekeepers are located at the first- and second-tier hub sites, and a directory gatekeeper is located at the first-tier hub site.

Figure 9-16 Two-Tier Hub-and-Spoke in a Purely Distributed Deployment



126682

The following design recommendations apply to the example shown in [Figure 9-16](#):

- On the gatekeeper located at the second-tier hub site, define a local zone for each spoke site as well as a local zone for the second-tier hub site itself. Use the **bandwidth interzone** command on this gatekeeper to perform call admission control on the links between the spoke sites and the second-tier hub site. In the example, the US Gatekeeper located in Dallas has three local zones configured: USHub for the Dallas site, USWest for the Los Angeles site, and USEast for the Boston site. The **bandwidth interzone** command is then applied to the zones assigned to the spoke sites (USWest and USEast).
- On the gatekeeper located at the second-tier hub site, define a remote zone that points to the directory gatekeeper located at the first-tier hub site. Configure call routing so that all calls that cannot be handled by any local zone are sent to this remote zone. Use the **bandwidth remote** command on each second-tier hub site gatekeeper to perform call admission control on the link to the first-tier hub site. In the example, the US Gatekeeper located in Dallas has a remote zone configured to connect to the directory gatekeeper (AMER-DGK), and it uses the **bandwidth remote** command to limit the amount of bandwidth used on the link to Chicago.
- On the gatekeeper located at the first-tier hub site, define a local zone for the co-located Cisco CallManager cluster and a remote zone for the co-located directory gatekeeper. No call admission control is needed on this gatekeeper.
- On the directory gatekeeper located at the first-tier hub site, define a local zone for the directory gatekeeper itself and a remote zone for each Cisco CallManager cluster (that is, for each site in this topology). Configure call routing so that the calls are dispatched to the appropriate gatekeeper according to the called number.

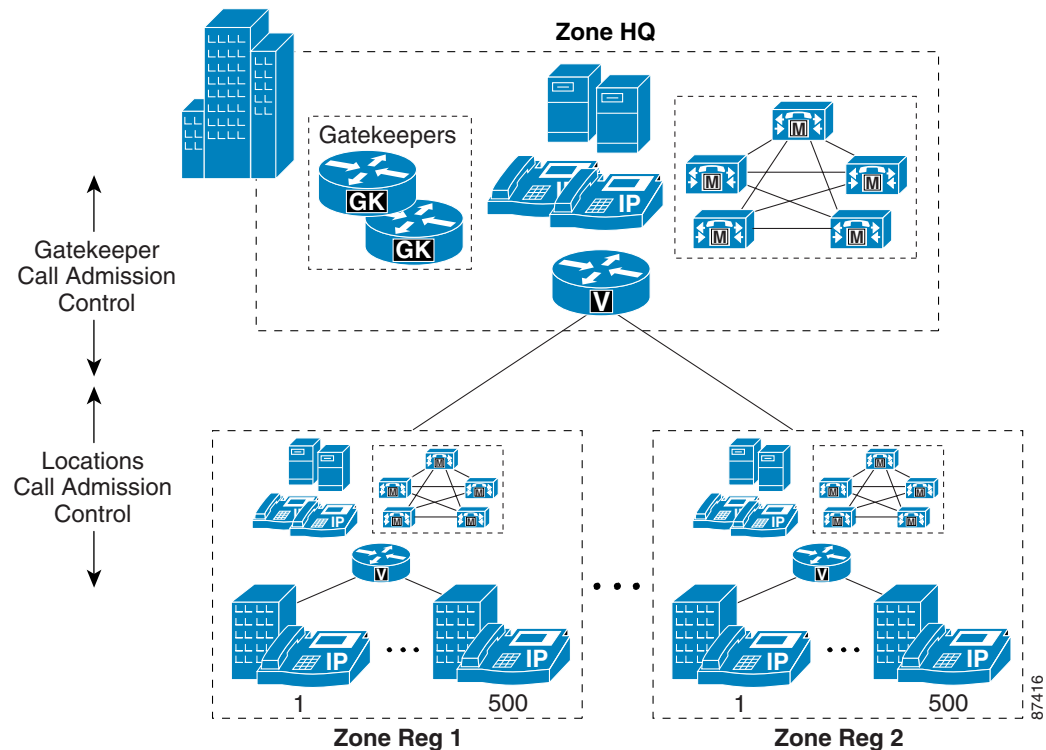
**Note**

The **bandwidth remote** command can be used in this deployment model because, for each second-tier hub site gatekeeper, the topology guarantees that all calls not destined for one of the local zones must transit through the link toward the first-tier hub site. If there were "peer-to-peer" links between the second-tier hub sites, it would not be possible to control bandwidth on those links as well as the link to the first-tier hub site because the **bandwidth remote** command accepts only a single bandwidth value.

Hybrid Centralized/Distributed Deployments

To provide call admission control in deployments that use a two-tier hub-and-spoke topology, with Cisco CallManagers at the first-tier and second-tier hub sites, you can combine the locations and gatekeeper mechanisms as illustrated in [Figure 9-17](#).

Figure 9-17 *Combining the Locations and Gatekeeper Mechanisms for Call Admission Control*



Follow these recommendations when combining H.323 trunks with locations for call admission control:

- Use locations-based call admission control for sites with no local Cisco CallManager (that is, the spoke sites).
- Use gatekeeper-based call admission control between Cisco CallManager clusters (that is, between the first-tier hub site and the second-tier hub sites).
- 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 9-1](#) and [Table 9-2](#) 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 cluster registers an H.323 gatekeeper-controlled trunk with the gatekeeper.
- Configure interzone bandwidth to control the number of calls between clusters.

Example 9-5 illustrates a typical gatekeeper configuration.

Example 9-5 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
```

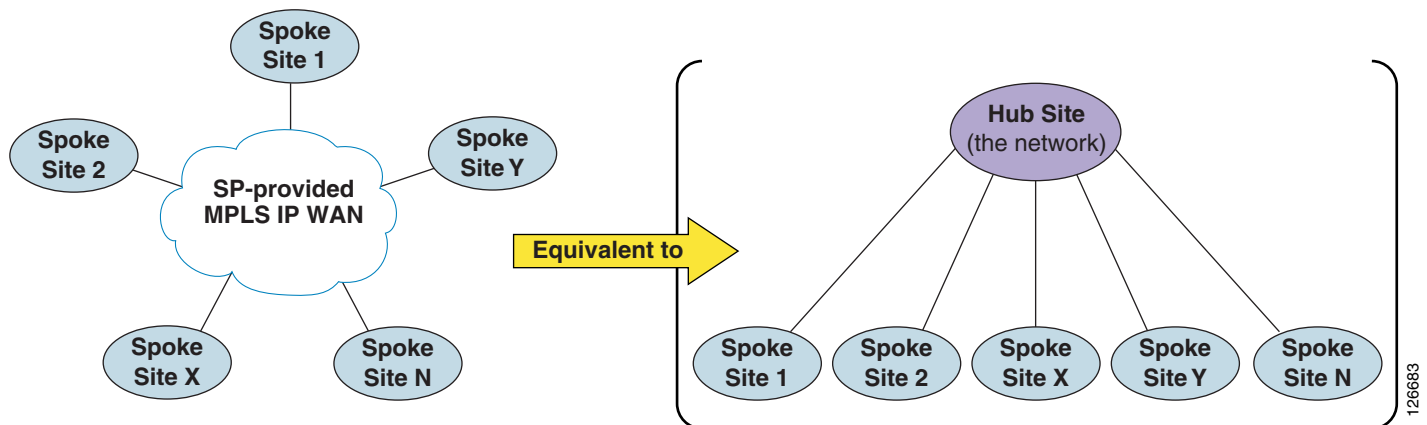
Example 9-5 illustrates the following points:

- 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.

MPLS-Based Topologies

Figure 9-18 shows an IP WAN (from a service provider) based on the Multiprotocol Label Switching (MPLS) technology. The main design difference between traditional Layer 2 WAN services offered by service providers and services based on MPLS is that, with MPLS, the IP WAN topology does not conform to a hub-and-spoke but instead provides "full-mesh" connectivity between all sites.

This topology difference means that, from an IP routing perspective on the enterprise side of the network, each site is one IP hop away from all other sites. Thus, there is no need to transit through a hub site to reach any other site. In fact, there is no concept of a "hub site." All sites are considered equal, and the only difference between them is the amount of bandwidth that they are allowed to use across the IP WAN.

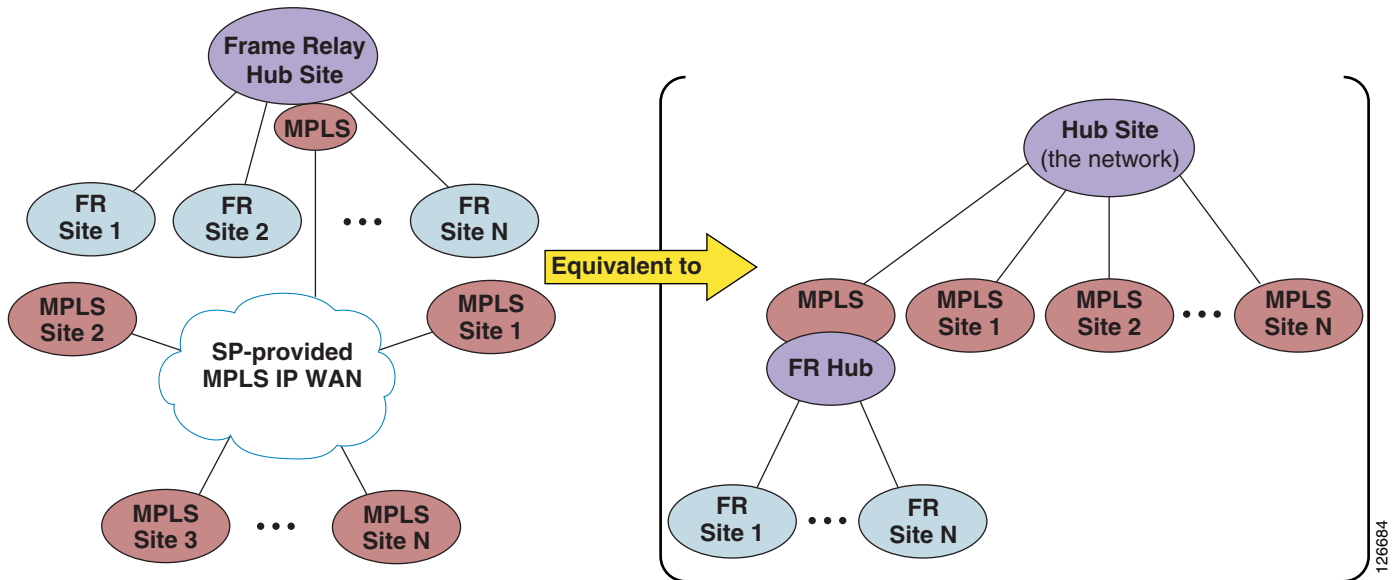
Figure 9-18 *MPLS IP WAN from a Service Provider, and Its Topology Equivalent*

Based on these considerations, it is easy to see that, from a call admission control perspective, a service-provider IP WAN service based on MPLS is in reality equivalent to a hub-and-spoke topology without a hub site. (See [Figure 9-18](#).) In fact, the network itself could be considered as the hub site, while all the enterprise sites (including the headquarters, or central site) are equivalent to spoke sites. These differences have implications on how to perform call admission control, which are described in the remainder of this section.

An exception to the above considerations that is worth mentioning here is represented by multi-site deployments where an MPLS-based WAN co-exists with an IP WAN based on a traditional Layer 2 technology, such as Frame Relay or ATM. Such a scenario could occur, for example, in the case of network migration phases, company mergers, or various other situations.

As shown in [Figure 9-19](#), integrating a hub-and-spoke IP WAN based on a traditional Layer 2 technology (such as Frame Relay) with an MPLS-based IP WAN results in a network topology that is neither a simple hub-and-spoke nor a full-mesh, but rather is equivalent to a two-tier hub-and-spoke.

In this case the first-tier hub site is represented by the MPLS network, the second-tier hub sites are represented by the MPLS-based sites as well as the MPLS-enabled Frame Relay hub site, and the spoke sites are represented by the Frame Relay spoke sites. Therefore, for design considerations on such deployments, refer to the section on [Two-Tier Hub-and-Spoke Topologies](#), [page 9-26](#).

Figure 9-19 Co-existence of MPLS Sites and Frame Relay Sites, and the Topology Equivalent

The remainder of this section contains design best practices for MPLS-based topologies according to the Cisco CallManager deployment model adopted:

- [Purely Centralized Deployments, page 9-36](#)

A single Cisco CallManager cluster is located at one site, but only phones and gateways are located at all other sites.

- [Purely Distributed Deployments, page 9-39](#)

A Cisco CallManager cluster is located at each site.

- [Hybrid Centralized/Distributed Deployments, page 9-39](#)

Cisco CallManager clusters are located at some sites, while only phones and gateways are located at the remaining sites.



Note

This section focuses on enterprise deployments where the MPLS WAN service is provided by a service provider. In cases where the MPLS network is deployed by the enterprise itself, call admission control can be performed effectively if one of the following two conditions is satisfied: (1) routing in the MPLS network is configured so that it is equivalent to a hub-and-spoke, or (2) bandwidth in the core of the MPLS network is heavily over-provisioned so that congestion can occur only at the edge.

Purely Centralized Deployments

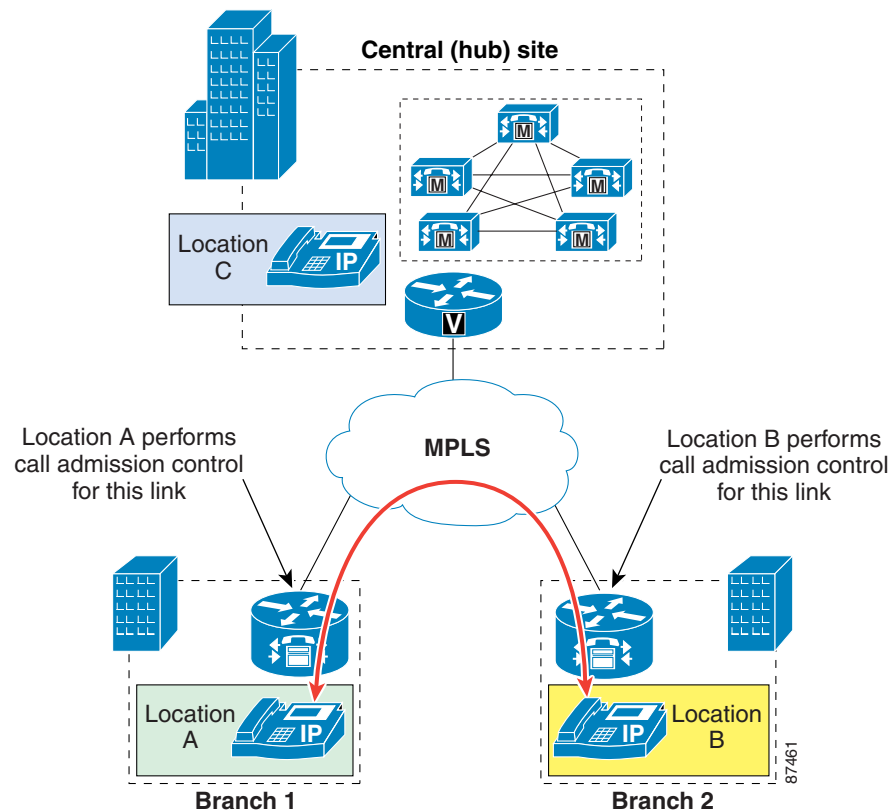
In single-cluster centralized call processing deployments, the call admission control function is performed by the *locations* construct within Cisco CallManager.

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

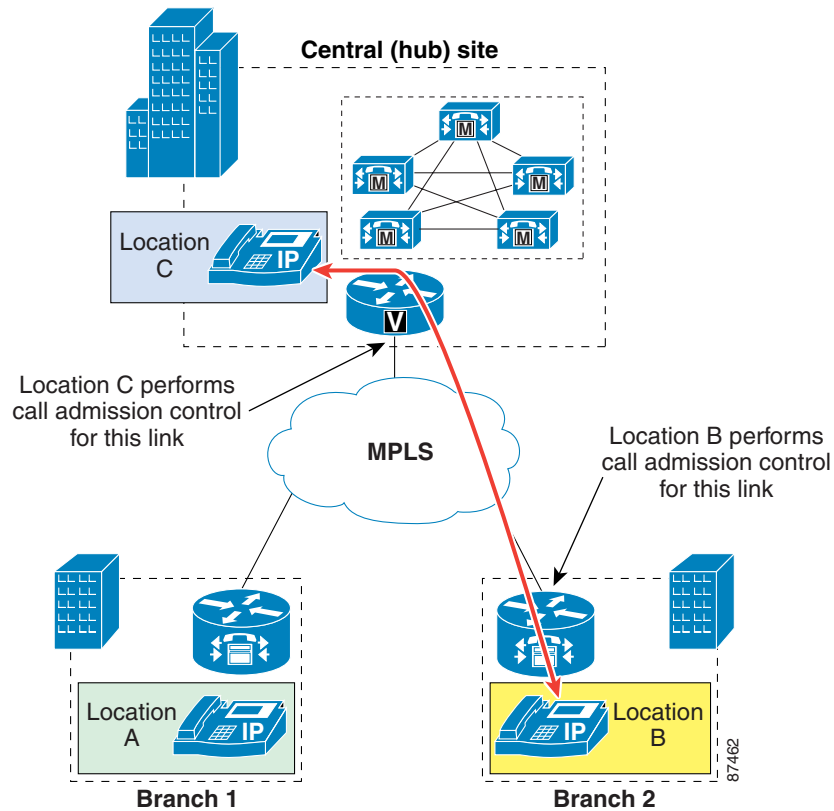
accounting occurs at the branch ends of the WAN links. Therefore, within the Cisco CallManager locations configuration, devices at the central site are left in the <None> location, while devices at each branch are placed in their appropriate location to ensure proper call admission control.

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 9-20](#) illustrates a spoke-to-spoke call between two branch sites in this type of deployment.

Figure 9-20 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 9-21](#).)

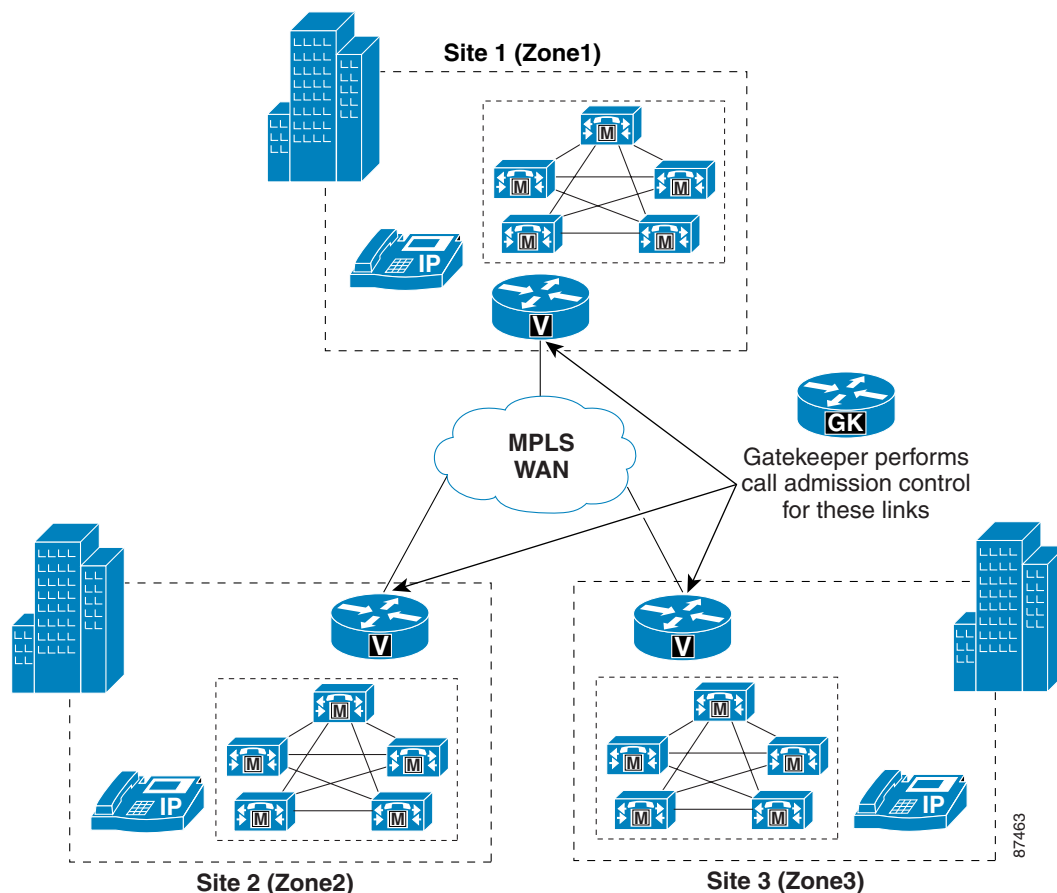
Figure 9-21 Calls to and from the Hub in an MPLS Deployment

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 [Automated Alternate Routing, page 10-19](#).)

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 9-22](#).)

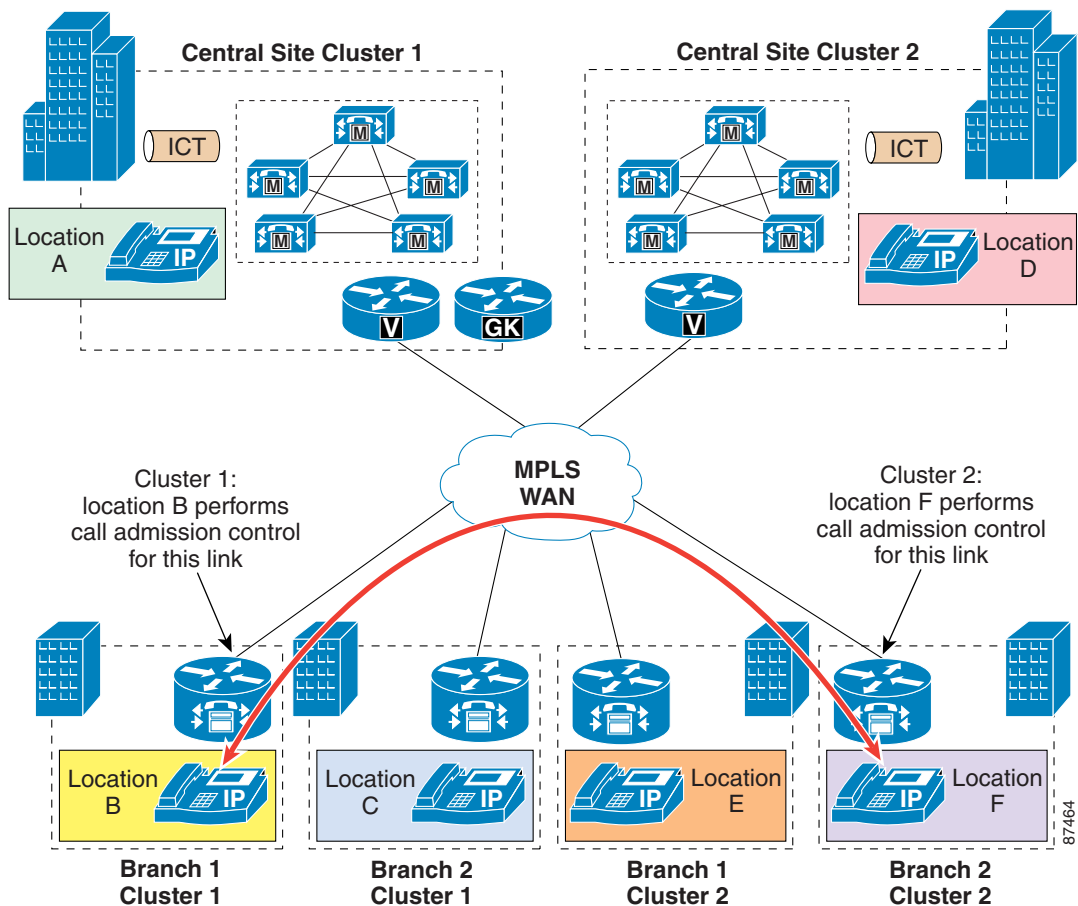
Figure 9-22 Gatekeeper Call Admission Control in a Distributed Deployment with MPLS



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 9-23](#).)

Figure 9-23 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.

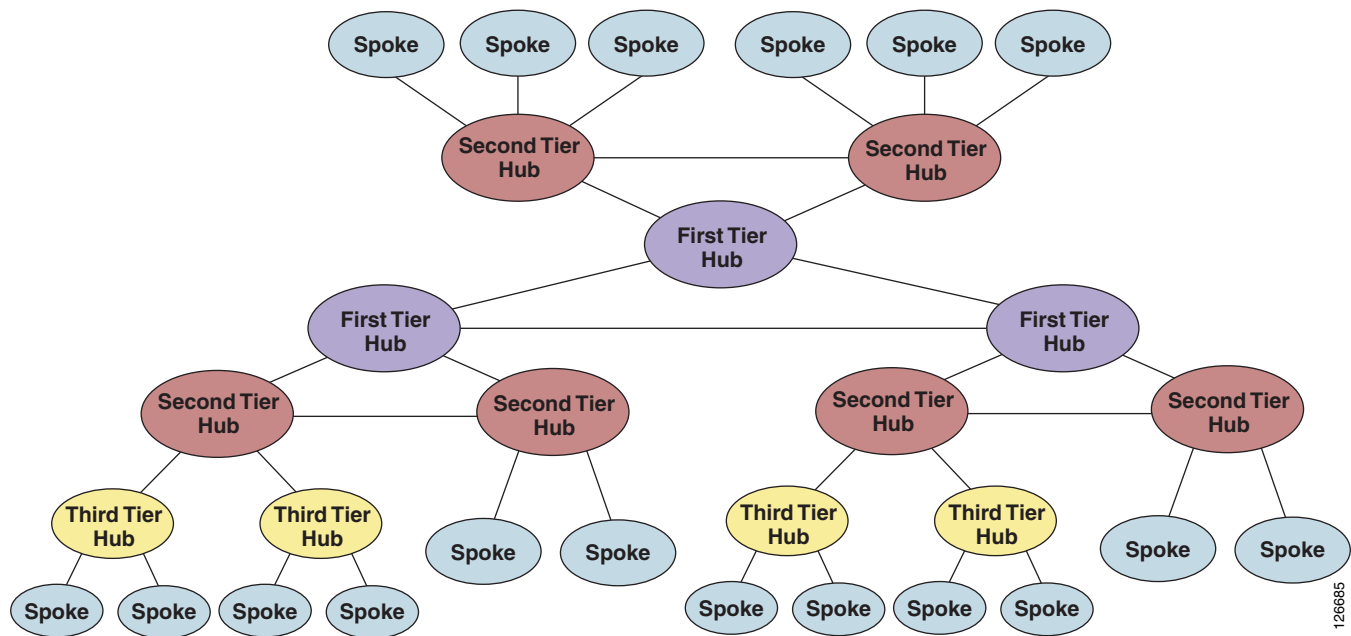
When all the available bandwidth for a particular site has been used, you can provide automatic failover to the PSTN by using a combination of the following two methods:

- 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 [Automated Alternate Routing](#), page 10-19.)

Complex Topologies

In the context of this document, a *complex topology* is any network topology that cannot be reduced to one of the topology types described in the previous sections of this chapter.

As the example in [Figure 9-24](#) illustrates, complex network topologies can present full-mesh features, hub-and-spoke features, partial mesh features, or possibly all combined in a single network.

Figure 9-24 Example of a Complex Topology

Generally speaking, the only call admission control mechanism that can support arbitrary topologies is RSVP. RSVP is a dynamic, distributed signaling protocol that is completely independent of the network topology, so it is ideally suited to work with topologies that cannot be logically simplified into a hub-and-spoke.

In order to add support for RSVP-based call admission control to a Cisco IP Communications network, you can use the via-zone gatekeeper and the Cisco IP-to-IP Gateway, as described in the section on [IP-to-IP Gateway](#), page 9-14.

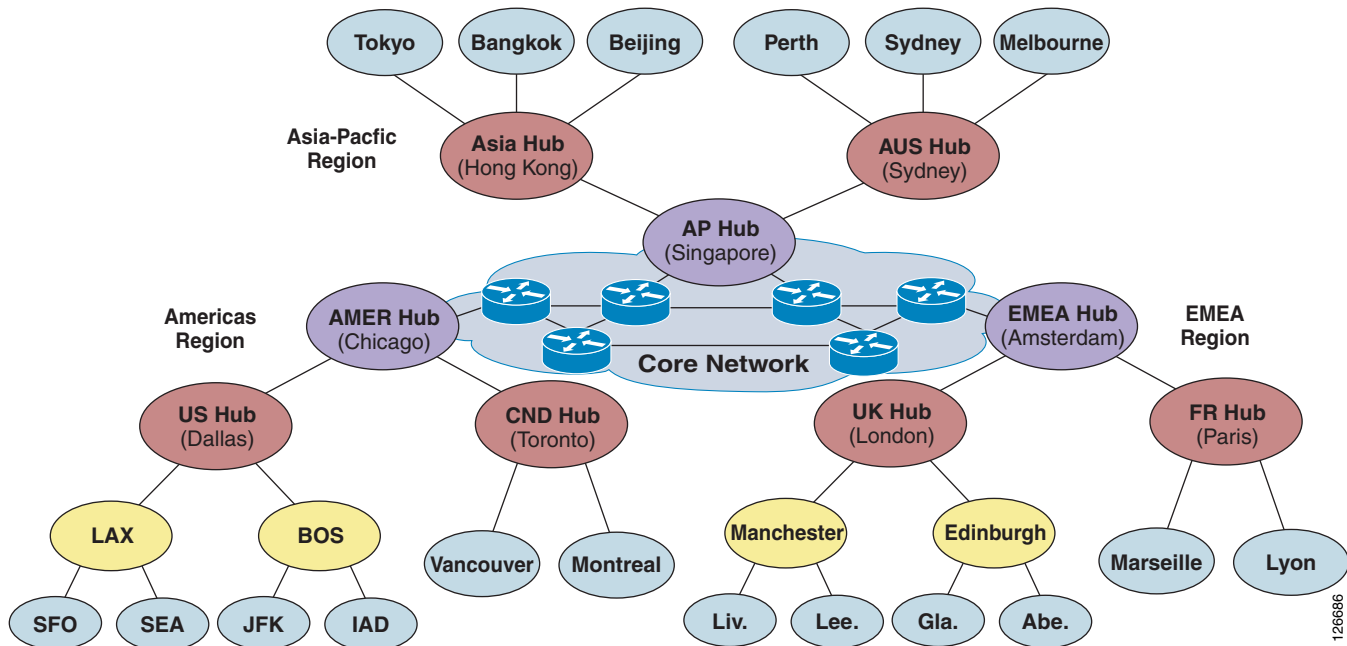
However, it might prove unfeasible to add IP-IP gateway functionality to all sites within a large network. Assuming that certain parts of the network can be simplified to a hub-and-spoke topology (either by altering the network connectivity or by making assumptions based on the available bandwidth), it is possible to combine the "static" call admission control mechanisms such as Cisco CallManager locations and Cisco IOS gatekeepers with dynamic call admission control provided by RSVP in the same network.

The remainder of this section presents a detailed case study based on a hypothetical large customer network whose topology cannot be simplified to a hub-and-spoke. This example combines the following call admission control mechanisms to provide an end-to-end solution:

- Cisco CallManager locations
- Cisco IOS Gatekeeper
- RSVP (using the Cisco IP-to-IP Gateway)

Case Study

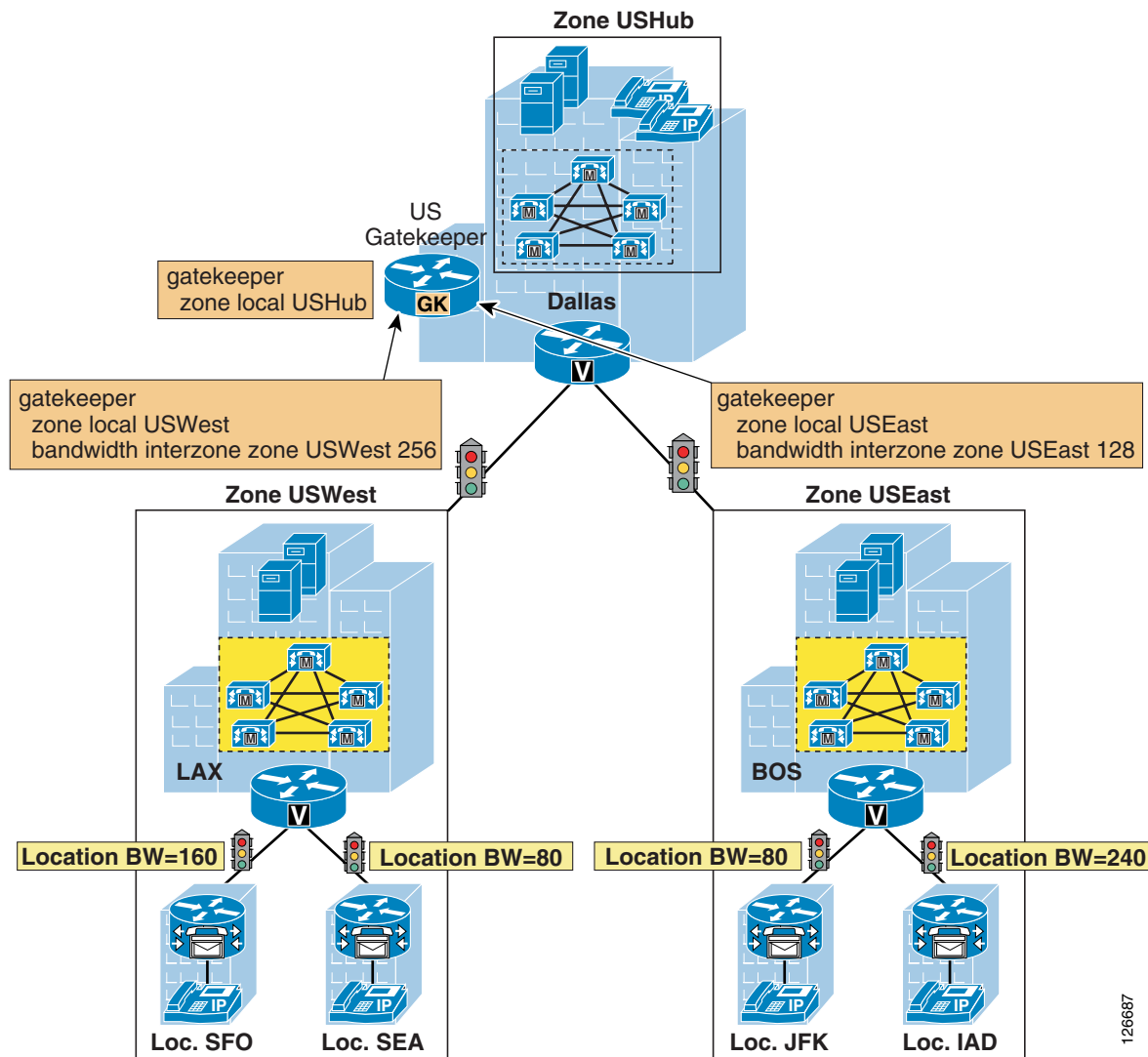
[Figure 9-25](#) shows the high-level topology of a hypothetical large customer network. The customer sites are divided into three main regions: the Americas (AMER), Europe-Middle East-Africa (EMEA), and Asia-Pacific (AP). Within each of these regions, the network topology is a three-tiered hub-and-spoke, with a regional hub and several country hubs, each of which can have one or two lower levels of smaller sites. The core network interconnects the three regions across an arbitrary network topology, possibly containing multiple equal-cost paths between any two destinations.

Figure 9-25 Hypothetical Customer Network

This complex topology illustrates where to apply the various call admission control mechanisms and how to combine them. To provide a better understanding of the overall solution, this section analyzes different portions of the network separately.

First, consider the AMER region. As illustrated in [Figure 9-25](#), there are two country hubs that connect to the region hub: the US hub site, located in Dallas, and the Canada hub site, located in Toronto. [Figure 9-26](#) shows how to implement call admission control for the US portion of the network.

Figure 9-26 Call Admission Control for the First Two Levels of the Customer Network in the US



A Cisco IOS gatekeeper is placed at the US hub site in Dallas. A number of first-level spoke sites (up to 100 per gatekeeper) can connect to this hub site. In the example shown in [Figure 9-26](#), these sites are Los Angeles (LAX) and Boston (BOS). By defining a local zone for each site in the gatekeeper (USHub, USWest, and USEast in the example), you can perform call admission control on the links between the first-level spoke sites and the US hub site using the Cisco IOS gatekeeper command **bandwidth interzone zone zone-name bandwidth-value**. This command ensures that the bandwidth used by voice and video calls in and out of each zone does not exceed the configured value.

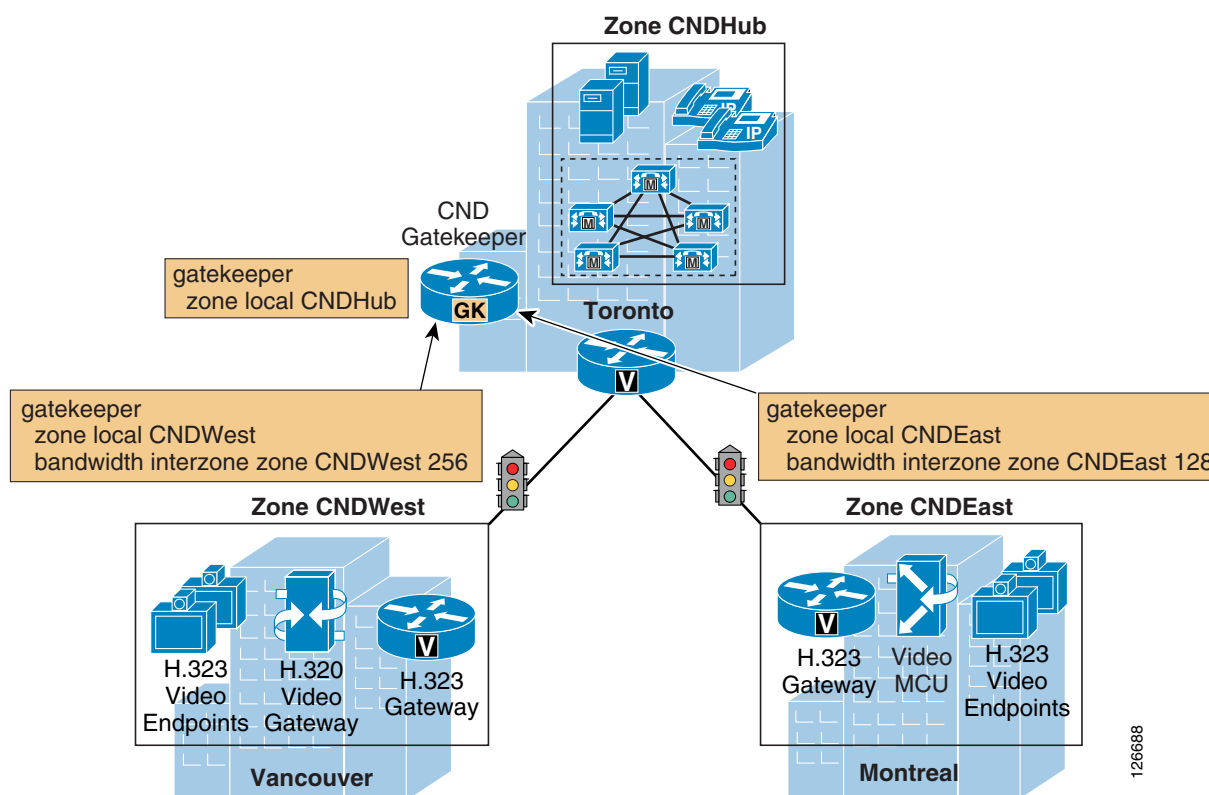
In this figure, the first-level spoke sites (LAX and BOS) each host a Cisco CallManager cluster. Because Cisco CallManager uses its own call admission control mechanism (the locations construct), it is possible to add a number of second-level spoke sites (up to 500 per cluster) to each of the first-level spoke sites. In this example, two additional sites are controlled by each Cisco CallManager cluster: San Francisco (SFO) and Seattle (SEA) for the Los Angeles cluster and New York (JFK) and Washington, DC (IAD), for the Boston cluster. The bandwidth on the links between these second-level spoke sites and their respective first-level site is therefore controlled by the Cisco CallManager locations.

**Note**

As far as the gatekeeper is concerned, the second-level spoke sites are considered to belong to the same zone as their "parent" first-level spoke site because they are controlled by the same Cisco CallManager cluster. This analysis is correct because calls from any of the second-level spoke sites directed to an on-net destination not controlled by the Cisco CallManager cluster will have to traverse the link between the "parent" first-level spoke site and the country hub site.

Figure 9-27 shows the Canada (CND) sites, which provide another example of how call admission control is implemented at this level in the network.

Figure 9-27 Call Admission Control for the First Level of the Customer Network in Canada



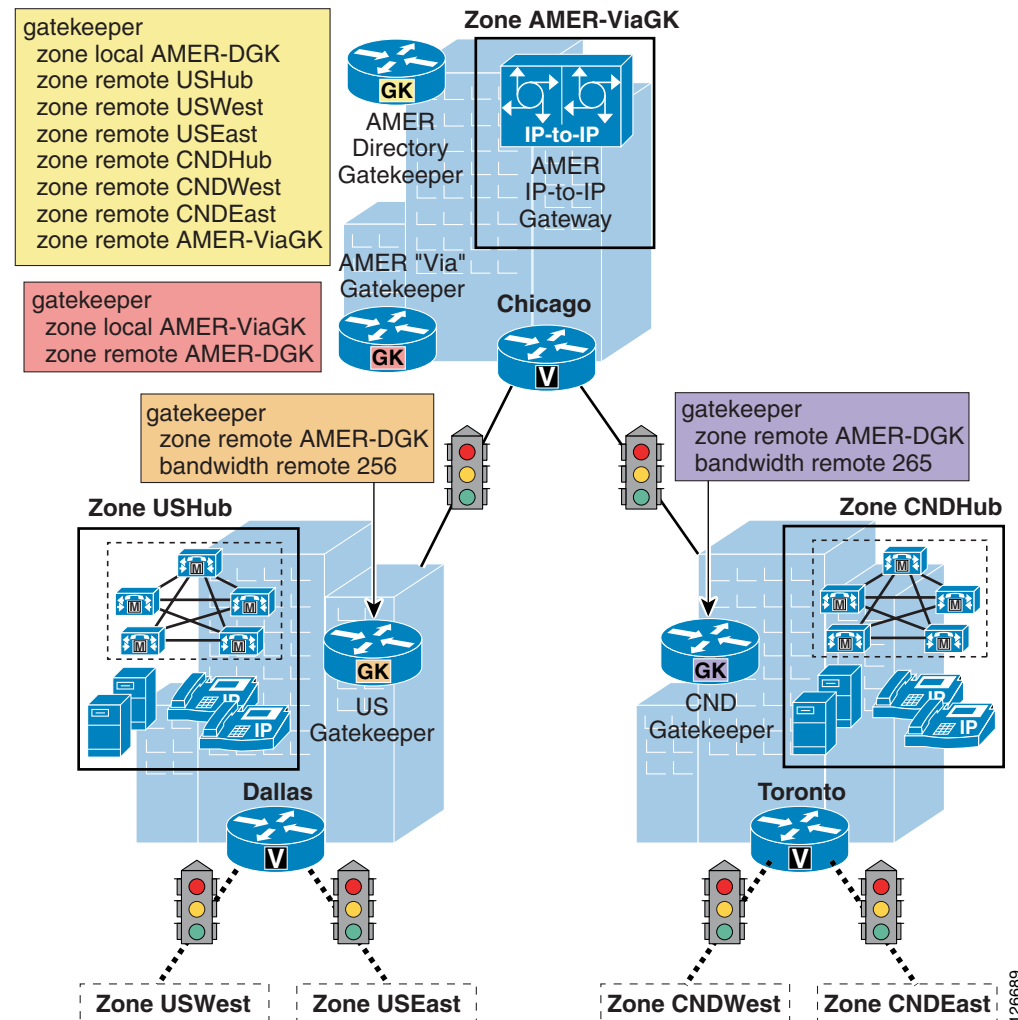
Another Cisco IOS gatekeeper is placed at the Canada hub site in Toronto, and a number of spoke sites connect to the hub site. In the example, the two spoke sites are Vancouver and Montreal. Note that these spoke sites do not contain a Cisco CallManager cluster but do contain a variety of H.323-based devices such as voice gateways, video endpoints, Multipoint Control Units (MCUs), and H.320 video gateways. For each site, a local zone (CNDHub, CNDWest, and CNDEast in the example) is defined within the gatekeeper. The bandwidth on the links between the Canada hub site and the spoke sites can again be controlled using the Cisco IOS gatekeeper command **bandwidth interzone zone zone-name bandwidth-value**. This command ensures that the bandwidth used by voice and video calls in and out of each zone does not exceed the configured value.

Because the H.323 devices at the spoke sites cannot perform call admission control, it is not possible to add a second level of spoke sites in this case. However, because all these devices can also be controlled by Cisco CallManager, it would be possible to add a Cisco CallManager cluster to the first-level spoke sites and distribute these H.323 devices across multiple second-level spoke sites.

Also note that, within the sites controlled by the same gatekeeper, it is possible to mix and match Cisco CallManager spoke sites and H.323 spoke sites.

Now consider the next level up in the network hierarchy. Figure 9-28 illustrates the call admission control mechanisms used between the AMER region hub site in Chicago and the two country hub sites for the US and Canada, located in Dallas and Toronto respectively.

Figure 9-28 Call Admission Control for the Second Level of the Customer Network (Region Hubs)



This part of the network is also a hub-and-spoke, but in this case the approach used to perform call admission control is different from that used between the country hub sites and the first-level spoke sites.

As discussed earlier, the country hub sites contain a Cisco IOS gatekeeper that controls bandwidth to the first-level spoke sites using local gatekeeper zones and the **bandwidth interzone** Cisco IOS command. The AMER region hub site contains a gatekeeper used as a directory gatekeeper, whose main purpose is to handle call routing between the country gatekeepers.

To control bandwidth usage on the links between the country hub sites and the region hub site, you can use the **bandwidth remote bandwidth-value** Cisco IOS command on the two gatekeepers located at the country hub sites. This command ensures that the bandwidth used by voice and video calls between all the local zones and all non-local zones does not exceed the configured value. Because all calls to

destinations that do not belong to a local zone must traverse the link between the country hub site and the region hub site, you can use the **bandwidth remote** command to perform call admission control effectively on this link.

In Figure 9-28, also notice the presence of a via-zone gatekeeper at the AMER region hub site. The directory gatekeeper is configured (configuration not shown in Figure 9-28) to route calls to destinations in other regions (such as AP and EMEA) to the AMER via-zone gatekeeper.

Next consider the core section of the network, which connects the three region hub sites. Figure 9-29 and Figure-35 illustrate how call routing and call admission control are performed across the core network.

Figure 9-29 Call Routing Across the Core of the Customer Network Using Via-Zone Gatekeepers

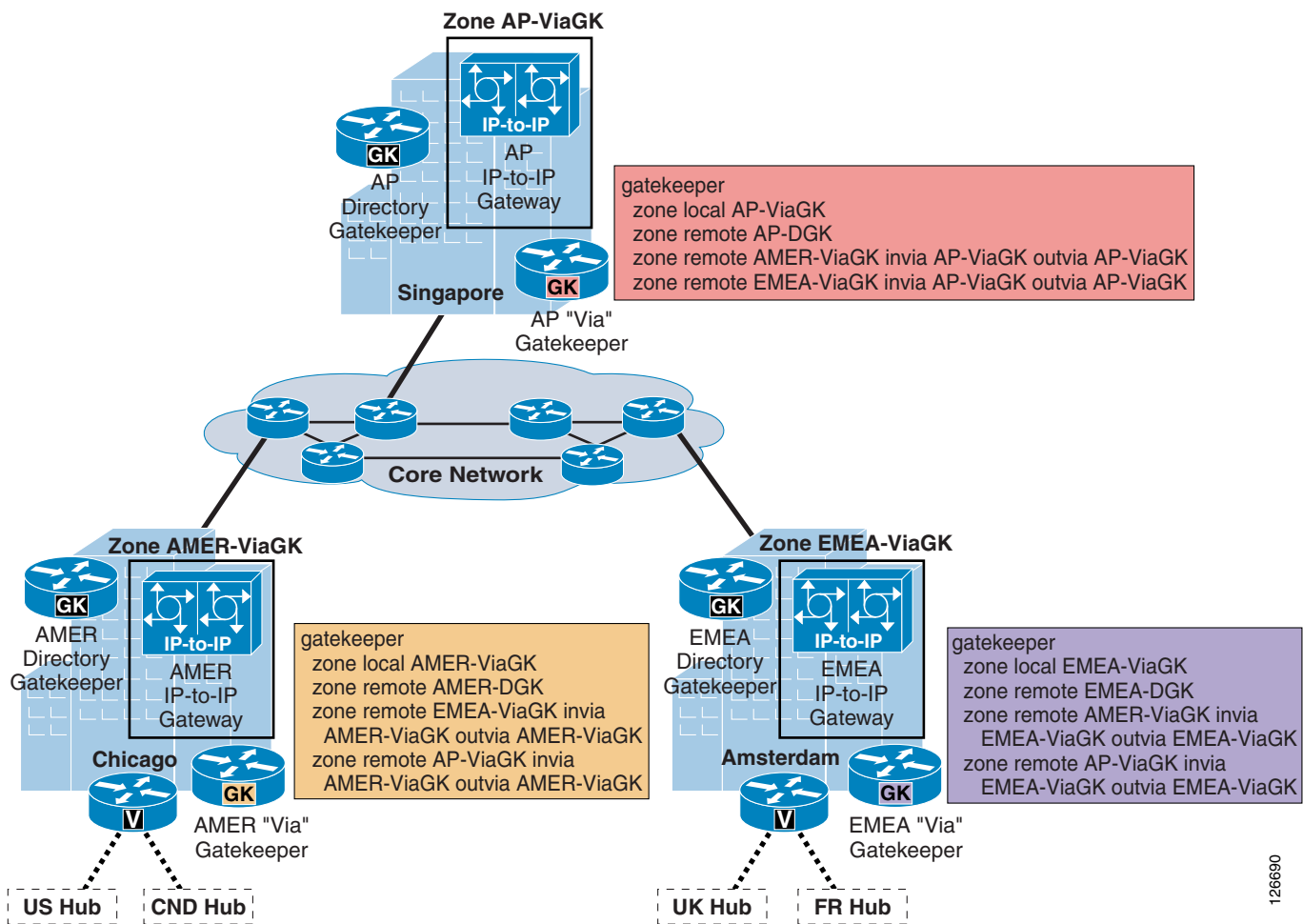


Figure 9-29 shows partial configurations of the three via-zone gatekeepers located in each of the three region hub sites. Each region's via-zone gatekeeper is configured so that all calls destined to or coming from the other two regions will cause the gatekeeper to insert an IP-to-IP gateway into the call.

For example, the configuration for the AMER region's via-zone gatekeeper defines the following zones:

- A single local zone called AMER-ViaGK
- The remote zone AMER-DGK, used to reach the AMER directory gatekeeper and, hence, all destinations within the AMER region

- The remote zone EMEA-DGK, used to reach destinations in the EMEA region
- The remote zone AP-DGK, used to reach destinations in the AP region

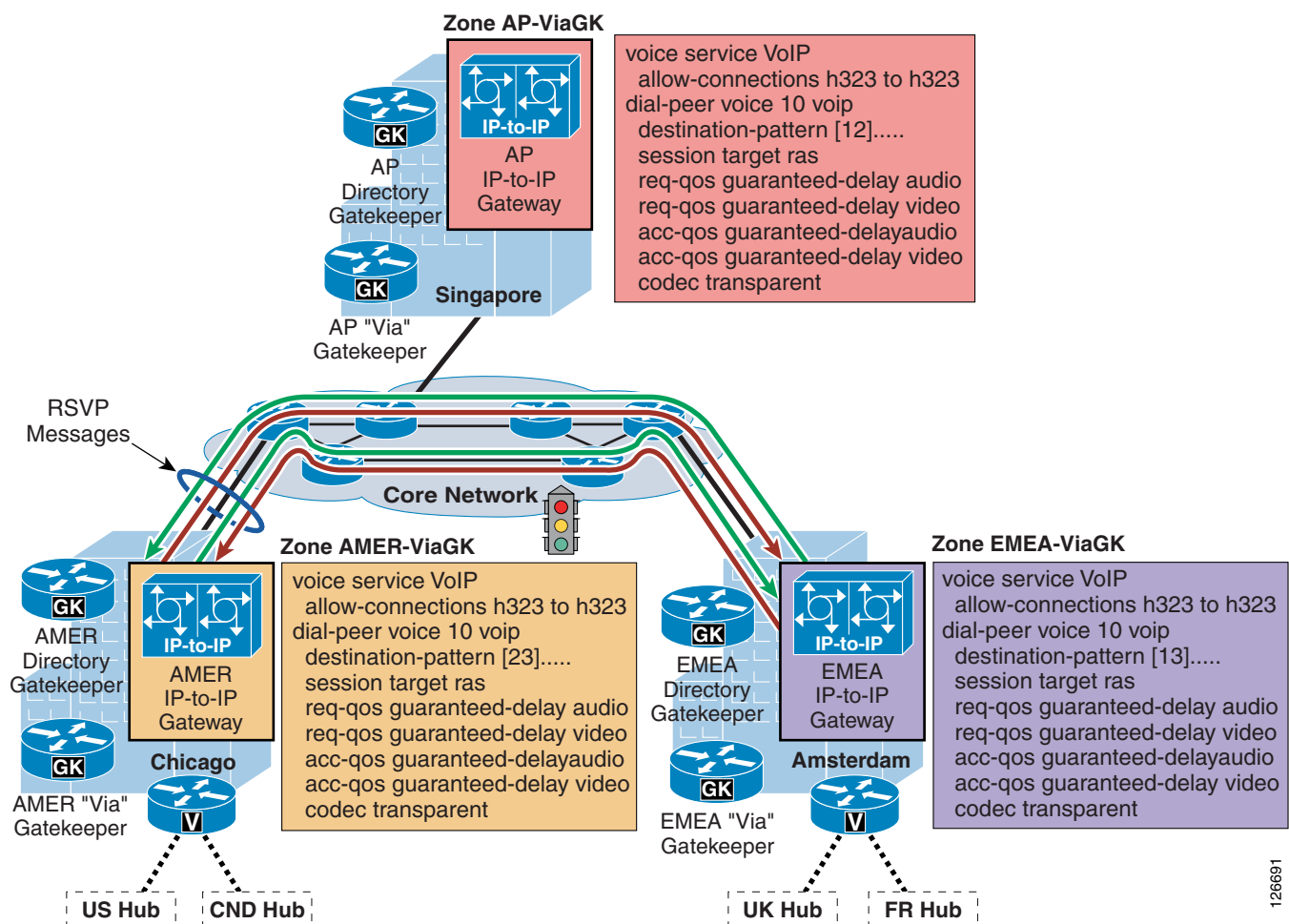
The definitions of the EMEA-DGK and AP-DGK remote zones also specify that the local zone AMER-ViaGK should be used as the **invia** and **outvia** zone to reach these destinations. This means that all calls destined to or coming from these remote zones will cause the AMER via-zone gatekeeper to insert an IP-to-IP gateway into the call. The specific IP-to-IP gateway resource will be selected from among all those registered with the gatekeeper in the local zone AMER-ViaGK.

**Note**

Inserting an IP-to-IP gateway into a call effectively splits the call into two call legs. In this example, because an IP-to-IP gateway is inserted at each region hub site, a call between an endpoint in the AMER region and one in the EMEA region would consist of three call legs: the first leg from the AMER endpoint to the AMER IP-to-IP gateway, the second leg between the AMER IP-to-IP gateway and the EMEA IP-to-IP gateway, and the third leg between the EMEA IP-to-IP gateway and the EMEA endpoint.

Figure 9-30 illustrates the call admission control mechanism used in the core network.

Figure 9-30 RSVB-Based Call Admission Control Across the Core of the Customer Network



126691

The IP-to-IP gateways are inserted into the call flow by the via-zone gatekeepers whenever the call needs to traverse the core network. Each IP-to-IP gateway is in turn configured to use the via-zone gatekeeper to route its calls and to use RSVP call admission control for the calls it originates and terminates across the core network. Because RSVP reservations are unidirectional, each voice call generates two RSVP reservations, one for each direction.

Figure 9-30 shows a call between the AMER IP-to-IP gateway and the EMEA IP-to-IP gateway, assuming that multiple paths exist between the two regional hubs across the core network. If the routes taken are asymmetrical, the two RSVP reservations will travel along different paths. However, for a given reservation, the RSVP messages will always travel along the same path because RSVP uses reverse hop-by-hop routing to establish the reservation. RSVP ensures that the bandwidth used by voice and video calls across the core network does not exceed the configured values on each router's interface.

In summary, end-to-end call admission control is achieved by combining different techniques in different portions of the network.

Referring to the network depicted in Figure 9-25, for example, a call between two IP phones located in San Francisco in the US and Liverpool in the UK would use the following mechanisms:

- Cisco CallManager locations for the link between San Francisco and Los Angeles
- Cisco IOS gatekeeper **bandwidth inter-zone** command for the link between Los Angeles and Dallas
- Cisco IOS gatekeeper **bandwidth remote** command for the link between Dallas and Chicago
- RSVP between Chicago and Amsterdam (the two regional hub sites)
- Cisco IOS gatekeeper **bandwidth remote** command for the link between Amsterdam and London (the country hub site for the UK)
- Cisco IOS gatekeeper **bandwidth inter-zone** command for the link between London and Manchester (the hub site for England)
- Cisco CallManager locations for the link between Manchester and Liverpool



Dial Plan

The dial plan is one of the key elements of an IP Telephony system, and an integral part of all call processing agents. Generally, the dial plan is responsible for instructing the call processing agent on how to route calls. Specifically, the dial plan performs the following main functions:

- Endpoint addressing

Reachability of internal destinations is provided by assigning directory numbers (DNs) to all endpoints (such as IP phones, fax machines, and analog phones) and applications (such as voicemail systems, auto attendants, and conferencing systems)

- Path selection

Depending on the calling device, different paths can be selected to reach the same destination. Moreover, a secondary path can be used when the primary path is not available (for example, a call can be transparently rerouted over the PSTN during an IP WAN failure).

- Calling privileges

Different groups of devices can be assigned to different classes of service, by granting or denying access to certain destinations. For example, lobby phones might be allowed to reach only internal and local PSTN destinations, while executive phones could have unrestricted PSTN access.

- Digit manipulation

In some cases, it is necessary to manipulate the dialed string before routing the call; for example, when rerouting over the PSTN a call originally dialed using the on-net access code, or when expanding an abbreviated code (such as 0 for the operator) to an extension.

- Call coverage

Special groups of devices can be created to handle incoming calls for a certain service according to different rules (top-down, circular hunt, longest idle, or broadcast).

This chapter examines the following main aspects of dial plan:

- [Planning Considerations, page 10-2](#)

This section analyzes the thought process involved in planning an IP Telephony dial plan, ranging from the number of digits used for internal extensions to the overall architecture of a company's internal dial plan. (Prerequisite: Some familiarity with dial plans in general.)

- [Dial Plan Elements, page 10-7](#)

This section provides detailed explanations of the elements of a Cisco IP Telephony dial plan. Covered topics include call routing logic, calling privileges, and digit manipulation techniques for various Cisco products. (Prerequisite: A working knowledge of Cisco CallManager and Cisco IOS is recommended.)

- [Design Considerations, page 10-43](#)

This section contains design and deployment guidelines related to multi-site IP Telephony networks, endpoint addressing methods, approaches to building classes of service, and call coverage functionality. (Prerequisite: A working knowledge of Cisco CallManager and Cisco IOS is recommended.)

For more details, refer to the *Cisco CallManager System Guide*, the *Cisco IOS Voice, Video, and Fax Configuration Guide, Release 12.2*, and other product documentation available at

<http://www.cisco.com>

Planning Considerations

The dial plan is the most fundamental attribute of a telephony system. It is at the very core of the user experience because it defines the rules that govern how a user reaches any destination. These rules include:

- Extension dialing — how many digits must be dialed to reach an extension on the system
- Extension addressing — how many digits are used to identify extensions
- Dialing privileges — allowing or not allowing certain types of calls
- Path selection — for example, using the IP network for on-net calls, or using one carrier for local PSTN calls and another for international calls
- Automated selection of alternate paths in case of network congestion — for example, using the local carrier for international calls if the preferred international carrier cannot handle the call
- Blocking of certain numbers— for example, pay-per-minute calls
- Transformation of the called number — for example, retaining only the last five digits of a call dialed as a ten-digit number
- Transformation of the calling number — for example, replacing a caller's extension with the office's main number when calling the PSTN

A dial plan suitable for an IP telephony system is not fundamentally different from a dial plan designed for a traditional TDM telephony system; however, an IP-based system presents the dial plan architect with some new possibilities. For example, because of the flexibility of IP-based technology, telephony users in separate sites who used to be served by different, independent TDM systems can now be included in one, unified IP-based system. These new possibilities afforded by IP-based systems require some rethinking of the way we look at dial plans. This section examines some of the elements that the system planner must consider to properly establish the requirements that drive the design of the dial plan.

On-Net versus Off-Net Dialing

Calls that originate and terminate on the same telephony network are considered to be on-network (or on-net). By contrast, if a call originates in company A and terminates at company B, it probably has to be routed through different telephony networks: first company A's network, followed by the PSTN, and finally into company B's network. From the caller's perspective, the call was routed off-network (or off-net); from the called party's perspective, the call originated off-net.

In TDM systems, the on-net boundaries of a telephony system are established by the PBX or Centrex system, and they typically do not extend outside of a single site. When they do, they typically do not include sites not immediately on the periphery of a large system hub.

One of the key attributes of IP telephony is its ability to expand the boundaries of calls that can be considered on-net. For instance, telephony users in an enterprise with six dispersed branch offices might be used to reaching one colleague with abbreviated dialing (for example, four-digit dialing) if the called party is located in the same site but dialing a full PSTN number to reach another colleague located in a different site. With an IP-based system where all users are served by the same IP network, it now becomes economically feasible to unite all six branches under a four-digit abbreviated dialing plan, with the IP network as the preferred path and automated overflow to the PSTN as a secondary path if the IP network is congested.

Abbreviated Dialing

Consider an extension with direct inward dial (DID) capability, which can be reached directly from the PSTN. An off-net PSTN caller has to dial the fully qualified PSTN number (for example, 1 415 555 1234) to reach a DID extension. An on-net caller might, however, prefer the ability to reach that same extension by simply dialing the last few digits of the DID number. In a four-digit abbreviated dial plan, the on-net caller would dial only 1234 in this example to reach the same extension.

Avoiding Overlap of Extension Dialing

A telephony system must be configured so that any extension can be reached in an unambiguous manner. To accomplish this goal, the dial plan must satisfy the following requirements:

- All on-net extension dialing must be globally unique. For instance, in a system using an abbreviated four-digit on-net dial plan, there cannot be an extension 1000 in site A and another extension 1000 in site B if the requirement is to reach either of them by dialing only four digits from site C.
- There cannot be any partial overlap between different dial strings.
 - For instance, if 9 is used as an off-net access code in a four-digit abbreviated dial plan (for example, to make PSTN calls), there cannot be any extensions in the 9XXX range. Attempting to do so would create situations where calls are not routed immediately. For example, if a user dialed 9141, the system would have to wait for either more digits (if the user were dialing 9 1 415 555 1234, for example) or the expiration of the interdigit timeout before routing the call to extension 9141. Likewise, if an operator code is used (for example, 0), the entire 0XXX extension range would have to be excluded from a four-digit uniform dial plan.
 - There cannot be overlapping strings of different length. For example, a system with extensions 1000 and 10000 would force users to wait for the interdigit timeout when they dial 1000.

Dialing String Length

The number of dialable extensions determines the quantity of digits needed to dial extensions. For example, a four-digit abbreviated dial plan cannot accommodate more than 10000 extensions (from 0000 to 9999). If 0 and 9 are reserved as operator code and off-net access code, respectively, the number range is further reduced to 8000 (1000 to 8999).

Uniform On-Net Dial Plan

A dial plan can be designed so that all extensions within the system are reached in a uniform way; that is, a fixed quantity of digits are used to reach a given extension from any on-net origination point. Uniform dialing is desirable because of the simplicity it presents to users. A user does not have to remember different ways to dial a number when calling from various on-net locations.

For example, if phone A is reached by dialing 1234 from any on-net location, whether the calling phone is in the same office or at a different site, the enterprise's dial plan is deemed uniform.

When the enterprise consists of few sites, this approach can be used with few complications. The larger the enterprise, in terms of number of extensions and sites, the more of the following challenges it faces in designing a uniform dial plan:

- The number of extensions can exceed the range afforded by the quantity of digits being considered for the dial plan. For instance, if more than 8000 extensions are required (considering the exclusions of the 0XXX and 9XXX ranges), the system may require that an abbreviated dial plan use more than four digits.
- Matching on-net abbreviated extensions to DID numbers means that, when a new DID range is obtained from a local exchange carrier, it cannot conflict with the pre-existing on-net abbreviated dial ranges. For example, if the DID range of 415 555 1XXX exists in a system using a four-digit uniform abbreviated dial plan, and DID range 650 556 1XXX is also being considered, it might be desirable to increase the quantity of digits for on-net dialing to five. In this example, the five-digit on-net ranges 51XXX and 61XXX would not overlap.
- Most systems require the exclusion of certain ranges due to off-net access codes and operator dialing. In such a system where 9 and 0 are reserved codes, no dial plan (uniform or not) could accommodate on-net extension dialing that begins with 9 or 0. This means that DID ranges could not be used if they would force the use of 9 or 0 as the first digit in the dial plan. For instance, in a five-digit abbreviated dial plan, the DID range 415 559 XXXX (or any subset thereof) could not be used. In this example, alternatives include increasing the length of the abbreviated dialing to six or more digits, or avoiding any DID range whose last five digits start with 9.

Once a given quantity of digits has been selected and the requisite ranges have been excluded (for example, ranges beginning with 9 or 0), the remaining dialing space has to be divided between all sites.

Most systems require that two ranges be excluded, thus leaving eight different possibilities for the leading digit of the dial range. [Table 10-1](#) exemplifies the distribution of the dialing space for a typical four-digit uniform dial plan.

Table 10-1 Distribution of Digits in a Typical Four-Digit Uniform Dial Plan

Range	Use	DID Ranges	Non-DID Ranges
0XXX	Excluded; 0 is used as off-net access code		
1XXX	Site A extensions	418 555 1XXX	N/A
2XXX	Site B extensions	919 555 2XXX	N/A
3XXX	Site C extensions	415 555 30XX	3[1-9]XX
4[0-4]XX	Site D extensions	613 555 4[0-4]XX	N/A
4[5-9]XX	Site E extensions	450 555 4[5-9]XX	N/A
5XXX	Site A extensions	418 555 5XXX	N/A
6XXX	Site F extensions	514 555 6[0-8]XX	69XX

Table 10-1 Distribution of Digits in a Typical Four-Digit Uniform Dial Plan (continued)

Range	Use	DID Ranges	Non-DID Ranges
7XXX	Future	XXX XXX 7XXX	7XXX
8XXX	Future	XXX XXX 8XXX	8XXX
9XXX	Excluded; 9 is used as off-net access code		

For the example in [Table 10-1](#), the various sites were assigned numbers in the following ways:

- Site A, the company headquarters, requires more than 1000 extensions, so two entire ranges of numbers have been retained (1XXX and 5XXX). Note that the corresponding DID ranges must also be retained from the site's local exchange carrier.
- Site B has been assigned an entire range (2XXX), allowing for up to 1000 extensions.
- Site C was also assigned an entire range, but it has been split between 100 DID extensions (415 555 30XX) and up to 900 non-DID extensions. If growth requires more extensions for DID, any unassigned numbers from the non-DID range could be used.
- Sites D and E were each assigned 500 numbers from the 4XXX range. Note that their corresponding DID ranges must match each of the site's respective portions of the 4XXX range. Because the DID ranges are for different sites (probably from different PSTN service providers), more coordination effort is required to split ranges between sites. As the quantity of sites assigned within a given range increases, it becomes increasingly difficult (sometimes impossible) to make full use of an entire range.
- Site F's range is split between 900 DID numbers (6[0-8]XX) and 100 non-DID numbers (69XX).
- The ranges 7XXX and 8XXX are reserved for future use.

When implementing a new dial plan, one of the main desires of any planner is to avoid having to change phone numbers. In addition, the extension ranges of any existing phone systems may have overlapped without any problems in the past, but they could be incompatible with a uniform dial plan.

Variable Length On-Net Dial Plan

Systems with many sites or overlapping site extension ranges can benefit from the use of a variable-length dial plan with the following characteristics:

- Within a site, the system retains the use of abbreviated dialing for calls to on-net extensions (for example, four-digit dialing).
- Between sites, users dial an access code followed by a site code and the destination's on-net extension.
- Off-net calls require an access code followed by a PSTN number.

The use of access and site codes (see [Table 10-2](#)) enables the on-net dial plan to differentiate between extensions that would overlap if a uniform abbreviated dial plan were implemented.

Table 10-2 Typical Use of Site Codes

Site Code	Range	Use	DID Ranges	Non-DID Ranges
1	1XXX	Site A extensions	418 555 10XX	1[1-9]XX
2	1XXX	Site B extensions	919 555 1XXX	N/A
3	1XXX	Site C extensions	907 555 1XXX	N/A

In [Table 10-2](#), sites A, B, and C are independently assigned the four-digit range 1XXX. For calls from site A to site B under the old telephony system, the calls had to be routed as off-net calls. With the new system, these calls can be dialed as on-net calls.

From site A, users simply dial 1234 to reach extension 1234. But from site B, the dial plan must accommodate the ability to reach extension 1234 at site A without conflicting with site B's own extension 1234. Therefore, each site is assigned a site code.

From site B, merely dialing the combination of site A's code with the desired extension is not feasible: in this case because 11234 would partially overlap with site B's extension 1123, thus causing interdigit timeout issues. If, instead, we assign 8 as an inter-site on-net access code, this would allow site B to dial 81234 to reach site A's extension 1234.

The following factors determine the quantity of digits required to dial an on-net, off-site extension:

- One digit for the inter-site access code
- N digits for the site code, where N is chosen to satisfy the quantity of site codes required (For example, if a system has 13 sites, then a minimum of two digits are required for the site code.)
- The quantity of digits required by the destination site's local dial plan

For example, a system with 75 sites which each use four-digit abbreviated dialing would require a format of 8 + SS + XXXX, where 8 is the on-net access code, SS is a two-digit site code, and XXXX is a four-digit extension number, giving a total of seven digits.

On-Net and Off-Net Access Codes

It is common practice in most enterprise telephony systems to dedicate a digit (for example, 9) as an off-net access code to steer calls to an off-net destination. In the variable-length on-net dial plan, another steering digit (for example, 8) is also required as an on-net access code to dial calls to on-net extensions at other sites. These two access codes, along with the use of an operator access code (for example, 0), implicitly exclude three of the ten possible leading digits of any dialed string. This restriction might not prove convenient for either of the following reasons:

- The users would be required to know the difference between on-net and off-net destinations, and to select the proper access code.
- The exclusion of three entire dialing ranges can become too restrictive or can conflict with some pre-assigned extension ranges. For instance, if a site already uses an abbreviated dialing range beginning with 8, the use of that same digit as an access code would require a change.

In systems where a uniform off-net access code (for example, 9) is already in use by all sites, it might be desirable to use the same code for both off-net and on-net off-site destinations. This approach has two main implications:

- To avoid partial overlap and interdigit timeout situations, the quantity of digits expected after the access code should be uniform.
- The telephony system must be able to recognize all on-net numbers dialed as off-net numbers and to route them over the IP network. This task can be simple for small systems with only one Cisco CallManager cluster but complex for large systems with multiple Cisco CallManager clusters.

Plan Ahead

Implementing an IP-based system might require changing certain existing user practices. Although it is preferable to plan a new system so that the implementation is as transparent as possible to users, dialing habits might have to be adapted to accommodate the integration of multiple sites that used to be on separate telephony systems. For instance, adapting to a new global, enterprise-wide dial plan might require changing the way a user reaches another user at a different site, the quantity of digits used to make intra-site calls and, in some cases, the extension numbers. To avoid exposing users to multiple generations of dial plan changes, try to anticipate expansion of the enterprise, which could result in the addition of sites in different dialing regions, an increase in the required number of on-net extensions, PSTN renumbering (for example, an area code split), or business expansion into different countries.

Dial Plan Elements

This section provides design and configuration guidelines for the following dial plan elements within a Cisco IP Telephony system:

- [Call Routing in Cisco CallManager, page 10-8](#)
- [Calling Privileges in Cisco CallManager, page 10-14](#)
- [Digit Manipulation in Cisco CallManager, page 10-18](#)
- [Automated Alternate Routing, page 10-19](#)
- [Extension Mobility, page 10-22](#)
- [Hunt Lists and Line Groups, page 10-23](#)
- [Call Routing in Cisco IOS with H.323 Dial Peers, page 10-27](#)
- [Calling Privileges in Cisco IOS with H.323 Dial Peers, page 10-39](#)
- [Digit Manipulation in Cisco IOS with H.323 Dial Peers, page 10-41](#)

Call Routing in Cisco CallManager

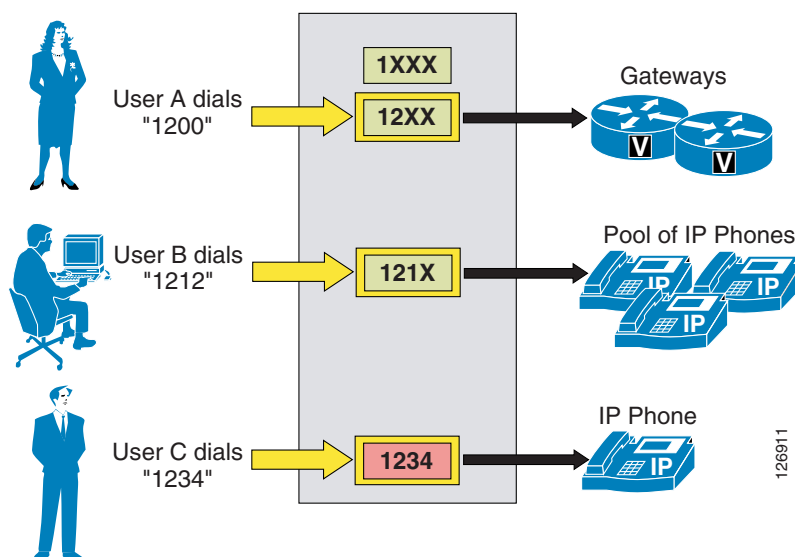
All dialing destinations configured in Cisco CallManager are added to its internal call routing table as patterns. These destinations include IP phone lines, voicemail ports, route patterns, translation patterns, and CTI route points.

When a number is dialed, Cisco CallManager uses closest-match logic to select which pattern to match from among all the patterns in its call routing table. In practice, when multiple potentially matching patterns are present, the destination pattern is chosen based on the following criteria:

- It matches the dialed string, and
- Among all the potentially matching patterns, it matches the fewest strings other than the dialed string.

For example, consider the case shown in [Figure 10-1](#), where the call routing table includes the patterns 1XXX, 12XX, and 1234.

Figure 10-1 Cisco CallManager Call Routing Logic Example



When user A dials the string 1200, Cisco CallManager compares it with the patterns in its call routing table. In this case, there are two potentially matching patterns, 1XXX and 12XX. Both of them match the dialed string, but 1XXX matches a total of 1000 strings (from 1000 to 1999) while 12XX matches only 100 strings (from 1200 to 1299). Therefore, 12XX is selected as the destination of this call.

When user B dials the string 1212, there are three potentially matching patterns, 1XXX, 12XX and 121X. As mentioned above, 1XXX matches 1000 strings and 12XX matches 100 strings. However, 121X matches only 10 strings; therefore it is selected as the destination of the call.

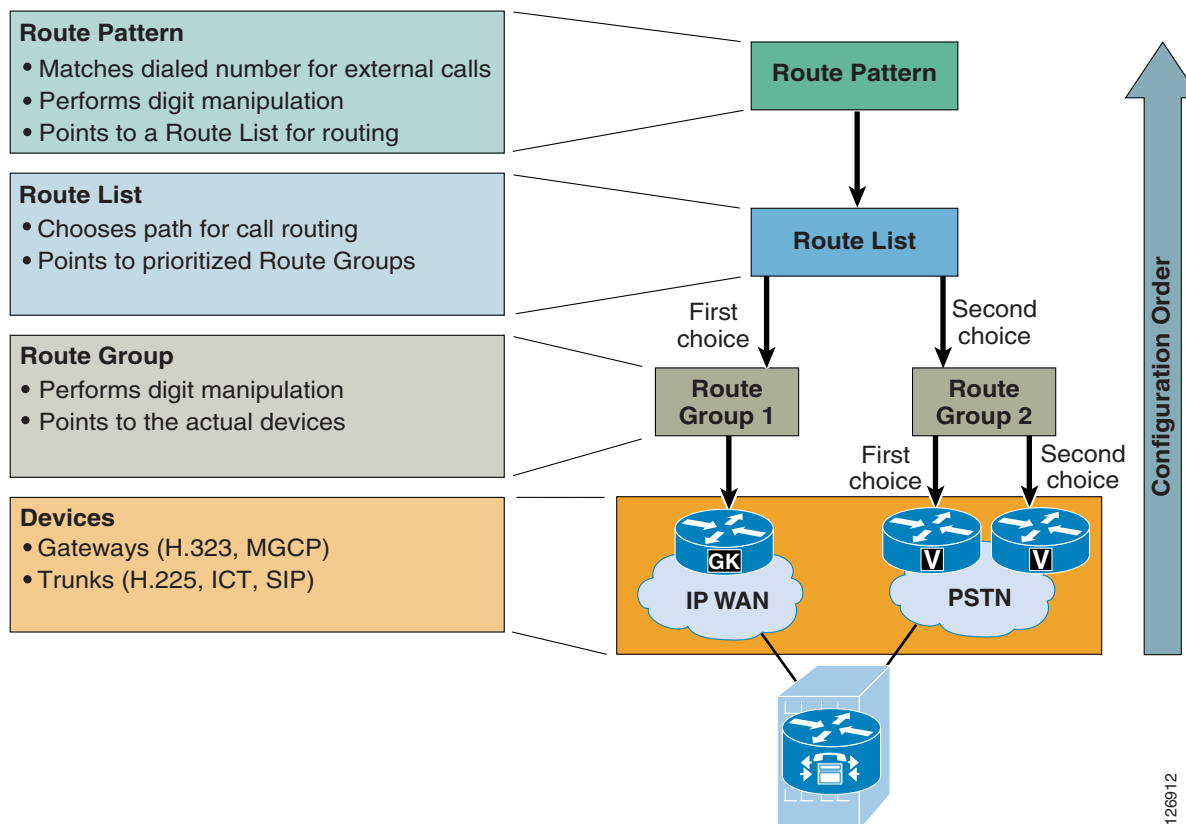
When user C dials the string 1234, there are three potentially matching patterns, 1XXX, 12XX, and 1234. As mentioned above, 1XXX matches 1000 strings and 12XX matches 100 strings. However, 1234 matches only a single string (the dialed string); therefore it is selected as the destination of this call.

**Note**

Whenever a directory number (DN) is configured in Cisco CallManager Release 4.0 and later, it is placed in the call routing table, regardless of whether the respective device (for example, an IP phone) is registered or not. This behavior differs from previous Cisco CallManager versions, where patterns were added to the call routing table only if the respective devices were registered. An implication of this modified behavior is that it is no longer possible to rely on secondary, identical patterns to provide failover capabilities to applications when the application (and hence the primary pattern) is unregistered. Because the primary pattern is now permanently in the call routing table, the secondary pattern will never be matched. It is, however, possible to achieve the same effect by using the Forward Busy field on primary patterns such as CTI route points because this field accepts wildcards in Cisco CallManager Release 4.0 and later.

Cisco CallManager automatically "knows" how to route calls to destinations within the same cluster. For external destinations such as PSTN gateways, H.323 gatekeepers, or other Cisco CallManager clusters, you have to use the external route construct (described in the following sections) to configure explicit routing. This construct is based upon a three-tiered architecture that allows for multiple layers of call routing as well as digit manipulation. Cisco CallManager searches for a configured route pattern that matches the external dialed string and uses it to select a corresponding route list, which is a prioritized list of the available paths for the call. These paths are known as route groups and are very similar to trunk groups in traditional PBX terminology. Figure 10-2 depicts the three-tiered architecture of the Cisco CallManager external route construct.

Figure 10-2 External Route Pattern Architecture



128912

The following sections describe the individual elements of the external route construct in Cisco CallManager:

- [Route Patterns, page 10-10](#)
- [Route Lists, page 10-13](#)
- [Route Groups, page 10-13](#)
- [Route Group Devices, page 10-13](#)

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 route pattern can point directly to a gateway for routing calls or point to a route list, which in turn points to a route group and finally to a gateway.

Cisco strongly recommends that you use the complete route pattern, route list, and route group construct because it provides the greatest flexibility for call routing, digit manipulation, and future dial plan growth.

The @ Wildcard

- The @ wildcard is a special macro function that expands into a series of patterns representing the entire national numbering plan for a certain country. For example, configuring a single unfiltered route pattern such as 9.@ with the North American Numbering Plan really adds 166 individual route patterns to the Cisco CallManager internal dial plan database.
- It is possible to extend Cisco CallManager to accept other national numbering plans by using the files created with the International Dial Plan Tool published on Cisco.com. Once this is done, the @ wildcard can be used for different numbering plans even within the same Cisco CallManager cluster, depending on the value selected in the Numbering Plan field on the Route Pattern configuration page. Contact your local Cisco account team for more information on which numbering plans are currently available.
- The @ wildcard can be practical in several small and medium deployments, but it can become harder to manage and troubleshoot in large deployments because adopting the @ wildcard forces the administrator to use route filters to block certain patterns (see [Route Filters, page 10-10](#)).

Route Filters

- Use route filters only with the @ route pattern 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. In most European countries, the same result is accomplished with the 0.00! route pattern.
- The ! wildcard is also used for deployments in countries where the dialed numbers can be of varying lengths. 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 or 0.00!# for Europe), and instruct the users to dial # to indicate end of dialing. This action is analogous to hitting the "send" button on a cell phone.

Overlap Sending and Overlap Receiving

In countries whose national numbering plan is not easily defined with static route patterns, you can configure Cisco CallManager for overlap sending and overlap receiving.

Overlap sending means that Cisco CallManager keeps collecting digits as they are dialed by the end users, and passes them on to the PSTN as they are dialed. To enable overlap sending in Cisco CallManager Release 4.0 or later, check the Allow Overlap Sending box on the Route Pattern Configuration page. (In earlier Cisco CallManager releases, overlap sending is enabled by setting the SendingCompleteIndicator service parameter to False.) The route pattern needs only to include the PSTN access code (for example, "9." in North America or "0." in many European countries).

Overlap receiving means that Cisco CallManager receives the dialed digits one-by-one from a PRI PSTN gateway, and it then waits for completion of the dialed string before attempting to route the call to an internal destination. To enable overlap receiving in Cisco CallManager Release 3.3(3) or later, set the OverlapReceivingFlagForPRI service parameter to True. (In earlier Cisco CallManager releases, the parameter name was OverlapReceivingForPriFlag.)

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.
- Similarly, if you configure digit manipulation in the route pattern, the IP phone display of the calling party and the Placed Calls register will show the manipulated number. If you configure digit manipulation only in the route group, the manipulations will be transparent to the end user.

Calling Line ID

- The calling line ID presentation can be enabled or disabled on the gateway and also can be manipulated in the route pattern, based on site requirements.
- If you select the option Use Calling Party's External Phone Number Mask, 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.

Urgent Priority

- The Urgent Priority checkbox is often used to force immediate routing of certain calls as soon as a match is detected, without waiting for the T302 timer to expire. For example, in North America, if the patterns 9.911 and 9.[2-9]XXXXXX are configured and a user dials 9911, Cisco CallManager has to wait for the T302 timer before routing the call because further digits may cause the 9.[2-9]XXXXXX to match. If Urgent Priority is enabled for the 9.911 route pattern, Cisco CallManager makes its routing decision as soon as the user has finished dialing 9911, without waiting for the T302 timer.
- It is important to note that the Urgent Priority checkbox only forces the T302 to expire as soon as the configured pattern is a potential match for the dialed number. This does not mean that the urgent pattern has a higher priority than other patterns; the closest-match logic described in the section on [Call Routing in Cisco CallManager, page 10-8](#), still applies. For example, assume the route pattern 1XX is configured as urgent and the pattern 12! is configured as a regular route pattern. If a user dials 123, Cisco CallManager will make its routing decision as soon as it receives the third digit because 1XX is an urgent pattern, but in this case it will select the pattern 12! because it is a better match (10 total patterns matched by 12! versus 100 patterns matched by 1XX).

Call Classification

- Calls using this route pattern can be classified as on-net or off-net calls. This route pattern can be used to prevent toll fraud by prohibiting off-net to off-net call transfers or by tearing down a conference bridge when no on-net parties are present. (Both of these features are controlled by Service Parameters within Cisco CallManager Administration.)
- When the "Allow device override" box is enabled, the calls are classified based on the call classification settings on the associated gateway or trunk.

Forced Account Codes (FAC)

- The Forced Account Codes (FAC) checkbox is used to restrict the outgoing calls when using a particular route pattern. If you enable FAC through route patterns, users must enter an authorization code to reach the intended recipient of the call.
- When a user dials a number that is routed through a FAC-enabled route pattern, the system plays a tone that prompts for the authorization code. To complete the call, the user authorization code must meet or exceed the level of authorization that is specified to route the dialed number.
- Only the authorization name appears in the call detail records (CDR); the authorization code does not appear in the CDR.
- The FAC feature is not available if the "Allow overlap sending" checkbox is enabled.

Client Matter Codes (CMC)

- The Client Matter Code (CMC) checkbox is used to track calls to certain numbers when using a particular route pattern. (For example, a company can use it to track calls to certain clients.)
- If you enable CMC for a route pattern, users must enter a code to reach the intended destination.
- When a user dials a number that is routed through a CMC-enabled route pattern, the system plays a tone that prompts for the code. The user must enter the correct code in order to complete the call.
- Client Matter Codes appear in the call detail records so that they can be used by the CDR analysis and reporting tool to generate reports for client billing and accounting.
- The CMC feature is not available if the "Allow overlap sending" checkbox is enabled.
- If both CMC and FAC are enabled, the user dials a number, enters the FAC when prompted to do so, and then enters the CMC at the next prompt.

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), H.323 trunks to a gatekeeper or remote Cisco CallManager cluster, or SIP trunks to a SIP proxy. (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.)

Cisco CallManager sends calls to the devices according to the distribution algorithm assigned. Cisco CallManager supports top-down and circular algorithms.

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:

- Media Gateway Control Protocol (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
- SIP trunk — trunk to a SIP proxy (available with Cisco CallManager Release 4.0 or later)

**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 Privileges in Cisco CallManager

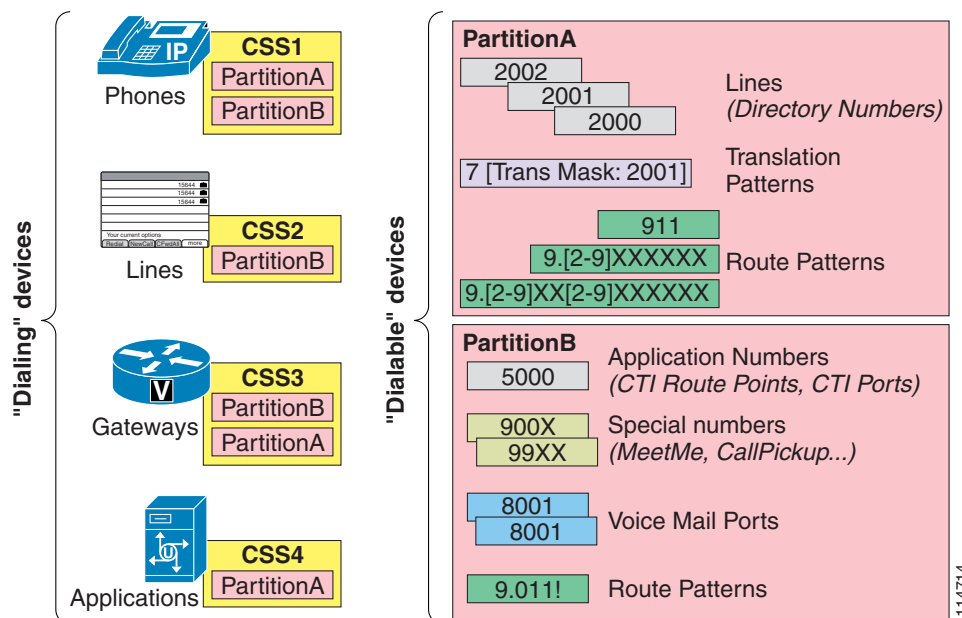
To implement calling privileges, configure the following elements in Cisco CallManager:

- [Partitions, page 10-15](#)
- [Calling Search Spaces, page 10-16](#)

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

As illustrated in [Figure 10-3](#), items that can be placed in partitions all have a dialable pattern, and they include phone lines, route patterns, translation patterns, CTI route group lines, CTI port lines, voicemail ports, and Meet-Me conference numbers. Conversely, items that have a calling search space are all devices capable of dialing a call, such as phones, phone lines, gateways, and applications (via their CTI route groups or voicemail ports).

Figure 10-3 Partitions and Calling Search Spaces

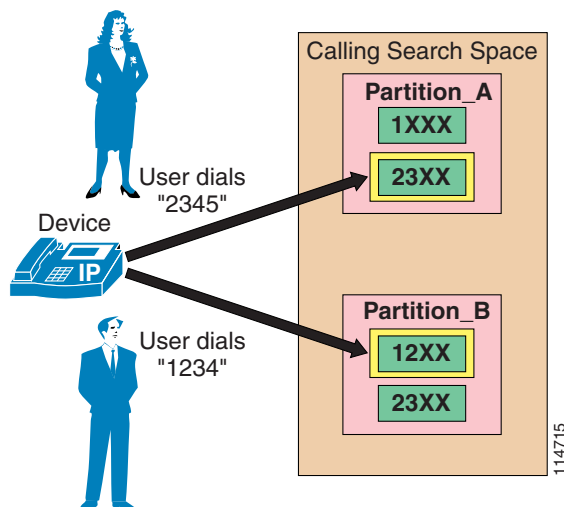


Partitions

The dial plan entries that you may place in a partition include IP phone directory numbers, translation patterns, route patterns, CTI route points, and voicemail ports. As described in the section on [Call Routing in Cisco CallManager, page 10-8](#), 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. In cases where two dial plan entries match the dialed pattern equally, Cisco CallManager selects the dial plan entry that appears first in the calling search space of the device making the call.

For example, consider [Figure 10-4](#), where route patterns 1XXX and 23XX are part of Partition_A and route patterns 12XX and 23XX are part of Partition_B. The calling search space of the calling device lists the partitions in the order Partition_A:Partition_B. If the user of this device dials 2345, Cisco CallManager selects route pattern 23XX in Partition_A as the matching entry because it appears first in the calling device's calling search space. However, if the user dials 1234, Cisco CallManager selects route pattern 12XX in Partition_B as the matching entry because it is a closer match than 1XXX in Partition_A. Remember that the partition order in a calling search space is used exclusively as a tie-breaker in case of equal matches based on the closest-match logic.

Figure 10-4 Impact of Partition Order on the Matching Logic



Note

When multiple equal-precision matches occur in the same partition, Cisco CallManager selects the entry that is listed first in its local dial plan database. Since you cannot configure the order in which the dial plan database lists dial plan entries, Cisco strongly recommends that you avoid any possibility of equal-precision matches coexisting within the same partition because the resulting dial plan logic is not predictable in such cases.

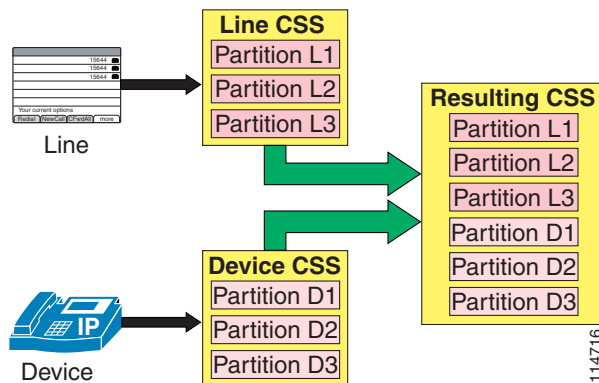
Beginning with Cisco CallManager Release 4.1, partitions can be activated or deactivated based on the time and date. You can activate or deactivate partitions by first configuring time periods and schedules within Cisco CallManager Administration and then assigning a specific time schedule to each partition. Outside of the times and days specified by the schedule, the partition is inactive, and all patterns contained within it are ignored by the Cisco CallManager call routing engine. For more information on this feature, see [Time-of-Day Routing, page 10-26](#).

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, as shown in Figure 10-5.

Figure 10-5 Concatenation of Line and Device Calling Search Spaces for IP Phones



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 according to the rules described in the section on [Partitions, page 10-15](#), 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).

For recommendations on how to set the line and device calling search spaces, refer to the sections on [Building Classes of Service for Cisco CallManager with the Traditional Approach, page 10-63](#), and [Building Classes of Service for Cisco CallManager with the Line/Device Approach, page 10-67](#).



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. This reverse behavior is still adopted by CTI ports and CTI route groups.

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>

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.

Call-Forward Calling Search Spaces

While the main calling search space configured on the line is concatenated with the device calling search space, the calling search spaces configured for the three types of call forward (Forward All, Forward Busy, and Forward No Answer) are standalone values not concatenated with any other calling search space. When the Forward All calling search space is left as <None>, the results are difficult to predict and depend on the Cisco CallManager release. Therefore, Cisco recommends the following best practices when configuring call-forward calling search spaces:

- Always provision the call-forward calling search spaces with a value other than <None>. This practice avoids confusion and facilitates troubleshooting because it enables the network administrator to know exactly which calling search space is being used for forwarded calls.
- Configure the Call Forward Busy and Call Forward No Answer calling search spaces with values that allow them to reach the DNs for the voicemail pilot and voicemail ports but not external PSTN numbers.
- Configure the Call Forward All calling search space according to your company's policy. Many companies choose to restrict forwarded calls to internal numbers only, to prevent users from forwarding their IP phone lines to a long-distance number and dialing their local IP phone number from the PSTN to bypass long-distance toll charges on personal calls.

When the Call Forward All calling search space is left as <None>, the following behavior applies:

- If Call Forward All is invoked from the IP phone, the call-forward calling search space becomes the concatenation of the line and device calling search spaces of that IP phone.
- If Call Forward All is invoked from the User Options page, the behavior depends on the Cisco CallManager release:
 - In Cisco CallManager Release 3.3(3) and earlier, the call-forward calling search space is copied from the line calling search space only.
 - In Cisco CallManager Releases 3.3(4) and 4.0(2) or later, the call-forward calling search space becomes the concatenation of the line calling search space and the calling search space of the "best-guess" device (that is, the last device from which the user invoked Call Forward All).

**Note**

In Cisco CallManager Release 4.0(1), when the Call Forward All calling search space is left as <None>, Cisco CallManager does not use the line or device calling search space of the line that invoked the Forward All; instead, it uses the calling search space of the calling device that is being redirected.

Digit Manipulation in Cisco CallManager

The following tools provide digit manipulation capability in Cisco CallManager:

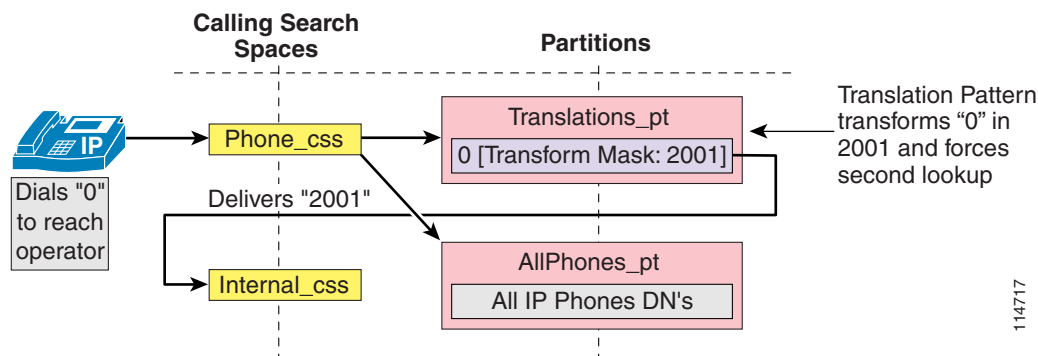
- The external route construct (route patterns, route lists, and route groups)
- Translation patterns

The external route construct can be used to introduce some digit manipulation while routing calls to external devices, and it is described in the section on [Call Routing in Cisco CallManager, page 10-8](#).

Translation patterns are the most powerful tool in Cisco CallManager to manipulate digits for any type of call. They follow the same general rules and use the same wildcards as route patterns. 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 first and then routes the call again, this time using the calling search space configured within the translation pattern.

Translation patterns can be used for a variety of applications, as shown by the example in [Figure 10-6](#).

Figure 10-6 Application Example for Translation Patterns



In this example, the administrator wishes to provide users with an operator service that is reached by dialing 0, while also maintaining a fixed-length internal numbering plan. The IP phones are configured with the Phone_css calling search space, which contains the Translations_pt partition (among others). A translation pattern 0 is defined in this partition, and the configured Called Party Transform Mask instructs Cisco CallManager to replace the dialed string (0) with the new string 2001, which corresponds to the DN of the operator phone. A second lookup (of 2001 this time) is forced through the call routing engine, using the Internal_css calling search space, and the call can now be extended to the real operator DN of 2001, which resides in the AllPhones_pt partition.



Note

When a dialed number is manipulated using a translation pattern, the translated number is recorded in the call detail record (CDR) and in the Placed Calls directory on the IP phone. However, when the digit manipulation occurs within a route list, the CDR and the Placed Calls directory on the IP phone show the originally dialed number, not the translated one.

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.

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, four-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 is incompatible with the Extension Mobility feature when users roam across different sites. Refer to [Extension Mobility, page 10-22](#), for more details.

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

- [Establish the PSTN Number of the Destination, page 10-19](#)
- [Prefix the Required Access Codes, page 10-20](#)
- [Select the Proper Dial Plan and Route, page 10-20](#)

**Note**

Beginning with Cisco CallManager Release 4.1.3, automated alternate routing (AAR) can be applied to voicemail hunt group members.

Establish the PSTN Number of the Destination

The rerouting of calls requires using a destination directory number (DN) that can be routed 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 can be routed 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 (212555XXXX) and uses it to derive the fully qualified number (2125551234) that can be routed 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.

For deployments that cover multiple countries within the same Cisco CallManager cluster, keep in mind that the external phone number mask needs to be configured in such a way that the destination phone can be reached from the same country or from a different country by simply prefixing digits. This means that any in-country prefixes (such as 0 in many countries) should not be included in the external phone number mask if they are not part of the E.164 address.

In order to better understand this, consider an example where a Cisco CallManager cluster has a site in London (United Kingdom), one in Paris (France), and one in Nice (France). The E.164 address of the DID range in Paris is +33145678XXX, but these extensions are usually reached as 0145678XXX when calling from within the French PSTN.

When somebody in the London office wishes to dial the Paris office via the PSTN, the dialed string is 90033145678XXX. However, when somebody in the Nice office wishes to dial the Paris office via the PSTN, the dialed string is 00145678XXX. Therefore, in this case the external phone number mask for the Paris office phones must be set to 145678XXX and not to the usual French national number of 0145678XXX because, if the 0 were included in the mask, it would not be possible to obtain the string 90033145678XXX by simply prefixing additional digits.

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 area code (also known as Numbering Plan Area, or NPA), then 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.

For deployments that cover multiple countries, you will typically need at least one AAR group per country. Considering the example introduced in the previous section, two AAR groups could be defined: the UK AAR group (assigned to all DNs in London) and the France AAR group (assigned to all DNs in Paris and Nice). The UK AAR group configuration prefixes 90033 for calls to the France AAR group, while the France AAR group simply prefixes 00 for calls within the same AAR group.

Select the Proper Dial Plan and Route

AAR calls should egress through a gateway within the same location as the calling phone, thus causing the completed dial string to 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.

**Note**

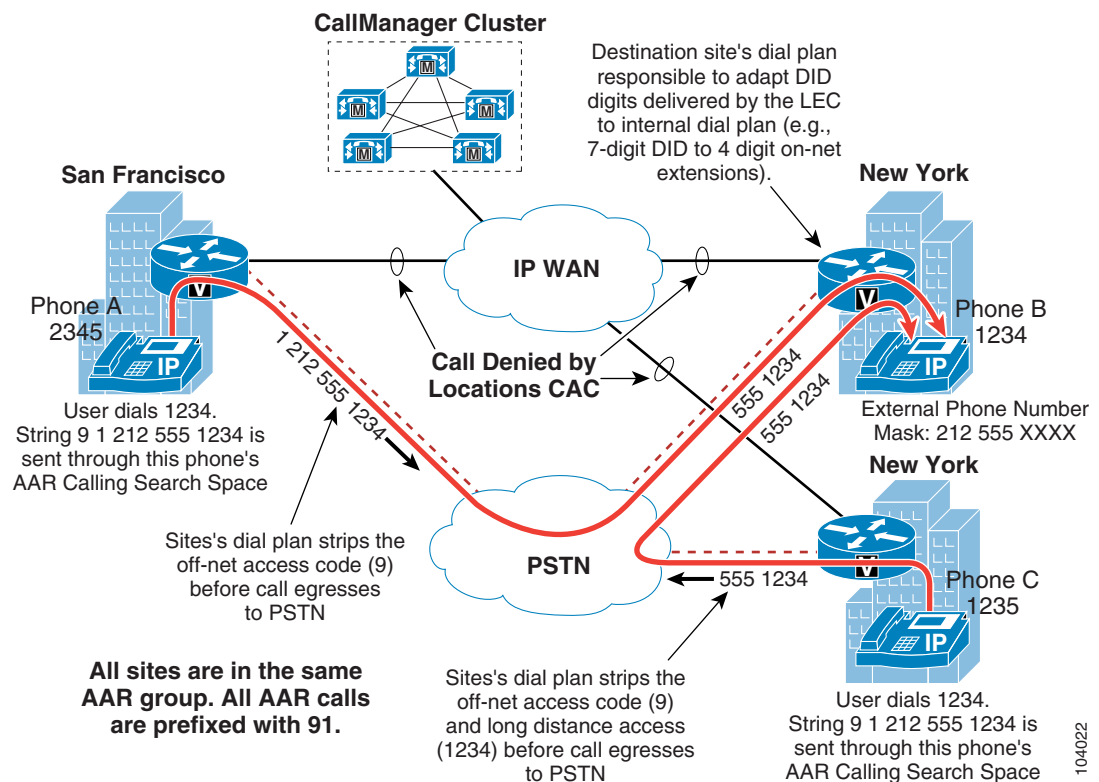
If you have configured additional route patterns to force on-net internal calls dialed as PSTN calls, ensure that these patterns are not matched by the AAR feature. Refer to [Design Guidelines for Multi-Site Deployments](#), page 10-44, for more details.

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 area code of 212. (See [Figure 10-7](#).) 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 10-7 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.

Extension Mobility

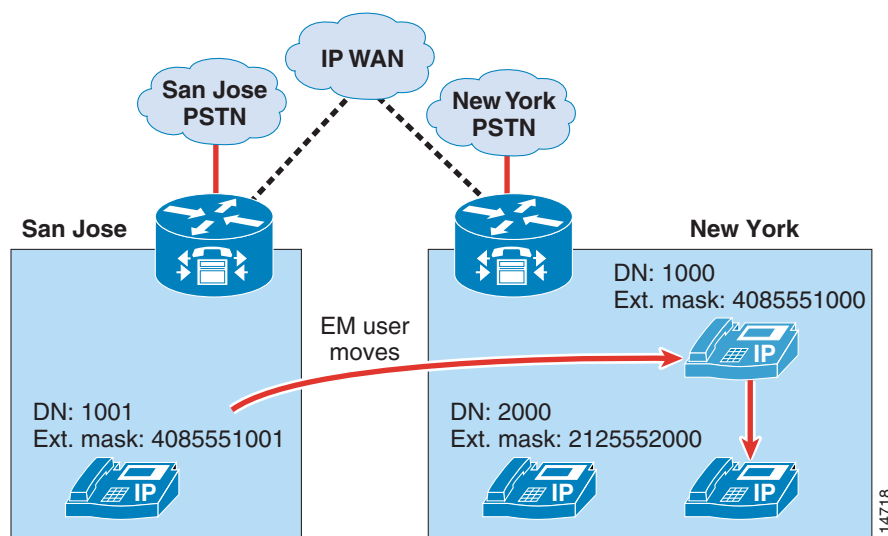
The Extension Mobility feature enables a user to log in to an IP phone and automatically apply his or her profile to that phone, including extension, speed dials, and calling privileges. This mechanism relies on the creation of a device profile associated with each Extension Mobility user. The device profile is effectively a virtual IP phone on which you can configure one or more lines and define calling privileges, speed dials, and so on.

When an IP phone is in the logged-out state, (that is, no Extension Mobility user has logged into it), the phone characteristics are determined by the device configuration page and the line configuration page(s). When a user logs in to an IP phone, the device configuration does not change, but the existing line configuration is saved in the Cisco CallManager database and is replaced by the line configuration of the user's device profile.

One of the key benefits of Extension Mobility is that users can be reached at their own extensions regardless of where they are located, provided that they can log in to an IP phone controlled by the same Cisco CallManager cluster. When Extension Mobility is applied to multi-site deployments with centralized call processing, this capability is extended to multiple sites geographically separated from each other.

However, if you combine the Extension Mobility feature with the AAR feature described in the section on [Automated Alternate Routing](#), [page 10-19](#), some limitations exist. Consider the example shown in [Figure 10-8](#), where Extension Mobility and AAR are deployed in a centralized call processing Cisco CallManager cluster with one site in San Jose and one in New York.

Figure 10-8 Extension Mobility and AAR



In this example, assume that an Extension Mobility user who is normally based in San Jose has a DN of 1000 and a DID number of (408) 555-1000. That user's external phone number mask is therefore configured as 4085551000. The user now moves to the New York site and logs in. Also, assume that the IP WAN bandwidth between San Jose and New York has been entirely utilized.

When the user in San Jose with extension 1001 tries to call 1000, AAR is triggered and, based on the AAR calling search space of the calling party and the AAR group of the called party, a new call to 914085551000 is attempted by the San Jose phone. This call uses the San Jose gateway to access the PSTN, but because the DID (408) 555-1000 is owned by that same gateway, the PSTN sends the call back to it. The San Jose gateway tries to complete the call to the phone with extension 1000, which is now in New York. Because no bandwidth is available to New York, the AAR feature is invoked again, and one of the following two scenarios will occur:

- If the gateway's AAR calling search space contains external PSTN route patterns, this is the beginning of a loop that eventually uses all the PSTN trunks at the San Jose site.
- If, on the other hand, the gateway's AAR calling search space contains only internal numbers, the call fails and the caller hears a fast-busy tone. In this case, one PSTN call is placed and one is received, so two PSTN trunks are utilized on the San Jose gateway for the duration of the call setup.

**Tip**

To prevent routing loops such as the one described here, always configure all calling search spaces on the gateway configuration pages to include only internal destinations and no external route patterns.

This example highlights the fact that Extension Mobility leverages the dynamic aspect of Cisco IP Communications and, therefore, requires that the call routing between sites use the IP network. Because the E.164 numbers defined in the PSTN are static and the PSTN network is unaware of the movements of the Extension Mobility users, the AAR feature, which relies on the PSTN for call routing, cannot be used to reach Extension Mobility users who move to a site other than their home site.

**Note**

However, if the Extension Mobility user moves to a remote site that belongs to the same AAR group as his or her home site, he or she can use the AAR feature to place calls to other sites when the available IP WAN bandwidth is not sufficient.

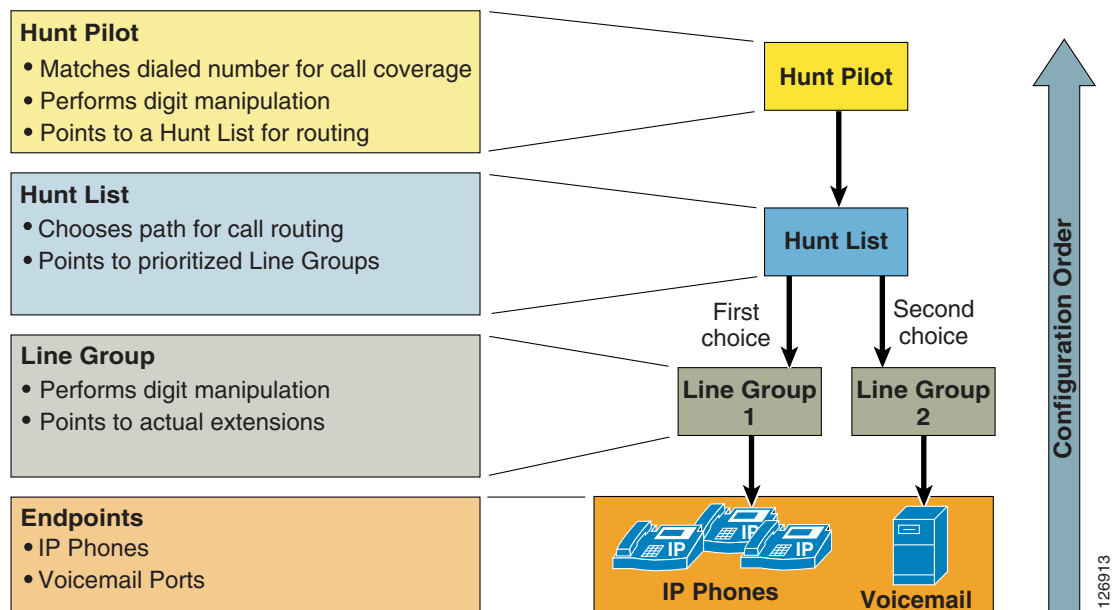
Hunt Lists and Line Groups

The *hunt pilot* is typically used for call coverage, or distributing a call through a list of Skinny Client Control Protocol (SCCP) endpoints. For call distribution, you can use a hunt construct. This hunt construct is based upon a three-tiered architecture, similar to that used to route external calls, that allows for multiple layers of call routing as well as digit manipulation.

Cisco CallManager searches for a configured hunt pilot that matches an incoming called number and uses it to select a corresponding hunt list, which is a prioritized list of the available paths for the call. These paths are known as *line groups*. [Figure 10-9](#) depicts the three-tiered architecture of the Cisco CallManager hunt construct.

**Note**

In Cisco CallManager Release 3.3 and earlier, call coverage functionality was provided by hunt groups, which are controlled by the Telephony Call Dispatcher (TCD) service, also used by the Cisco Attendant Console. Cisco CallManager Release 4.0 introduced hunt pilots, hunt lists, and line groups, but in that release the hunt pilot structure was combined with the route pattern structure and the hunt list was combined with the route list. In Cisco CallManager release 4.1, these structures are now independent.

Figure 10-9 Three-Tiered Architecture for the Hunt Construct in Cisco CallManager Release 4.1

Hunt Pilot

Hunt pilots are strings of digits and wildcards similar to route patterns, such as 9.[2-9]XXXXXX, configured in Cisco CallManager to route calls to directory numbers. The hunt pilot points directly to a hunt list. Hunt lists point to line groups, which finally point to SCCP endpoints.

Beginning with Cisco CallManager Release 4.1, calls can be redirected to a final destination when the hunting fails because of one or both of the following reasons:

- All hunting options have been exhausted and the call still is not answered.
- A time-out period has expired.

This call redirection is configured in the Hunt Forward Settings section of the Hunt Pilot configuration page, and the destination for this redirect can be either of the following options:

- A specific pattern in the internal call routing table of Cisco CallManager
- Personal preferences, which point to the Call Forward No Coverage settings for the originally called number when hunting on behalf of that number fails

For example, you can implement the personal preferences option by configuring a user's phone so that the Forward No Answer field redirects the call to a hunt pilot, in order to search for someone else who can answer the call. If the call hunting fails, either because all the hunting options were exhausted or because a time-out period expired, the call can be sent to a destination personalized for the person who was originally called. For example, if you set the Forward No Coverage field within the person's DN configuration page to the voicemail number, the call will be sent to that person's voicemail box if hunting fails.

The following considerations apply to calls handled by hunt pilots:

- Call Pickup and Group Call Pickup are not supported on calls distributed by a hunt pilot. A member of the line group cannot pick up the hunt pilot call offered to another member in the line group, even if they belong to the same call pickup group.

- Calls distributed by the hunt pilot number do not support separate call-forward treatment configured on a directory number in a line group. Thus, the Immediate Divert (iDivert) softkey or any call-forward configurations on the directory number do not work for calls distributed by the hunt pilot. Only the call-forward conditions available as hunt options in a line group configuration apply to the hunt pilot call. However, the iDivert softkey or call-forward configurations work for all incoming calls other than those distributed by the hunt pilot.
- The hunt pilot can distribute calls to any of its line group and route group members, even if the members and the hunt pilot reside in different partitions. A call distributed by the hunt pilot overrides all the partitions and calling search space restrictions.

Hunt List

A hunt list is a prioritized list of eligible paths (line groups) for call coverage. Hunt lists have the following characteristics:

- Multiple hunt pilots may point to the same hunt list.
- A hunt list is a prioritized list of line groups that function as alternate paths to a given destination. For example, you can use a hunt list to distribute a call within a particular site and use a hunt list to distribute a call at some other remote site.
- Hunt lists do not do any digit manipulation.
- Multiple hunt lists can contain the same line group.

Line Group

Line group members are user extension numbers that are controlled by Cisco CallManager. Thus, when the call is being distributed through the line group members, Cisco CallManager is in control of the call. Hunt options can be applied to the call when it is not answered or if the extension is busy or unregistered.

Line groups control the order in which the call is distributed, and they have the following characteristics:

- Line groups point to specific extensions, which are typically IP phone extensions or voicemail ports.
- A single extension may be present in multiple line groups.
- Computer Telephony Integration (CTI) ports and CTI route points cannot be added within a line group. Therefore, calls cannot be distributed to endpoints controlled through CTI applications such as Cisco Customer Response Solution (CRS), IP Interactive Voice Response (IP IVR), and so forth.
- Cisco CallManager distributes calls to the devices according to the distribution algorithm assigned. Cisco CallManager supports the following algorithms:
 - Top-down
 - Circular
 - Longest idle time
 - Broadcast
- In the event of No-Answer, Busy, or Not-Available, line groups redirect a call distributed to an extension based on the hunt options. Cisco CallManager supports the following hunt options:
 - Try next member, then try next group in hunt list.
 - Try next member, but do not go to next group.
 - Skip remaining members and go directly to next group.
 - Stop hunting.

For more information about hunt algorithms and hunt options, refer to the *Cisco CallManager Administration Guide*, available at

<http://www.cisco.com>

Line Group Devices

The line group devices are the endpoints accessed by line groups, and they can be of any of the following types:

- Any Skinny Client Control Protocol (SCCP) endpoints, such as Cisco IP Phones, VG248, or ATA 188
- Voicemail ports (for Cisco Unity)
- H.323 clients
- FXS extensions attached to an MGCP gateway

Time-of-Day Routing

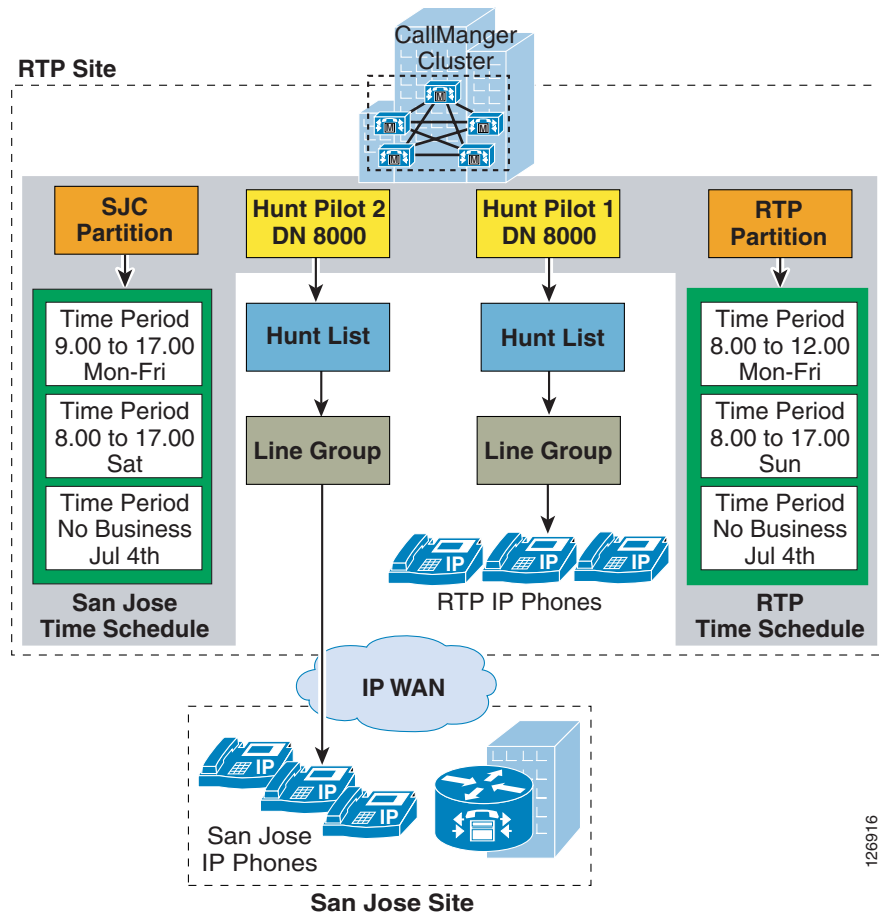
Cisco CallManager Release 4.1 introduces the time-of-day (ToD) routing feature. To use this feature, configure the following elements:

- Time period
- Time schedule

The time period allows you to configure start and end times for business hours. The start and end times indicate the times during which the calls can be routed. In addition to these times, you can set the event to repeat itself on a weekly or yearly basis. Moreover, you can also configure non-business hours by selecting "No business hours" from the Start Time and End Time options. All incoming calls will be blocked when this option is selected.

A time schedule is a group of specific time periods assigned to the partition. It determines whether the partition is active or inactive during the specified time periods. A matching/dialing pattern can be reached only if the partition in which the dialing pattern resides is active.

As illustrated in [Figure 10-10](#), two hunt pilots with the same calling pattern (8000) are configured in two partitions (namely, RTP_Partition and SJC_Partition). Each of these partitions is assigned a time schedule, which contains a list of defined time periods. For example, RTP phones can be reached using Hunt Pilot 1 from 8:00 AM to 12:00 PM EST (GMT - 5.00) Monday through Friday as well as 8:00 AM to 5:00 PM on Sundays. In the same way, SJC phones can be reached using Hunt Pilot 2 from 8:00 AM to 5:00 PM PST (GMT - 8.00) Monday through Friday and 8:00 AM to 5:00 PM on Saturdays. Both of the hunt pilots in this example are inactive on July 4th.

Figure 10-10 Time-of-Day Routing

For the example in Figure 10-10, an incoming call to the hunt pilot (8000) on Wednesday at 3:00 PM will be forwarded to the SJC phones, while a person calling the hunt pilot on July 4th will get a fast busy tone unless there is another pattern that matches 8000.

Call Routing in Cisco IOS with H.323 Dial Peers

The call routing logic on Cisco IOS routers using the H.323 protocol relies on the dial peer construct. Dial peers are similar to static routes; they define where calls originate and terminate and what path the calls take through the network. Dial peers are used to identify call source and destination endpoints and to define the characteristics applied to each call leg in the call connection. Attributes within the dial peer determine which dialed digits the router collects and forwards to telephony devices.

For a detailed description of dial peers and their configuration, refer to *Configuring Dial Plans, Dial Peers, and Digit Manipulation*, part of the *Cisco IOS Voice, Video, and Fax Configuration Guide, Release 12.2*, available at

<http://www.cisco.com>

One of the keys to understanding call routing with dial peers is the concept of incoming versus outgoing call legs and, consequently, of incoming versus outgoing dial peers. Each call passing through the Cisco IOS router is considered to have two call legs, one entering the router and one exiting the router. The call leg entering the router is the *incoming call leg*, while the call leg exiting the router is the *outgoing call leg*.

Call legs can be of two main types:

- Traditional TDM telephony call legs, connecting the router to the PSTN, analog phones, or PBXs
- IP call legs, connecting the router to other gateways, gatekeepers, or Cisco CallManagers

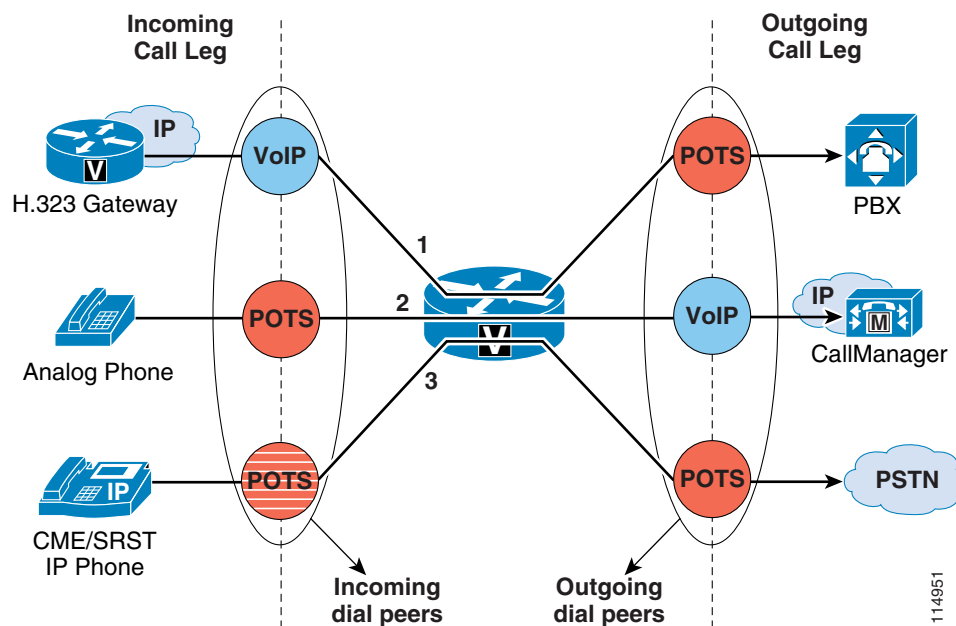
For all calls going through the router, Cisco IOS associates one dial peer to each call leg. Dial peers are also of two main types, according to the type of call leg with which they are associated:

- POTS dial peers, associated with traditional TDM telephony call legs
- VoIP dial peers, associated with IP call legs

Figure 10-11 shows the following examples of different types of calls going through a Cisco IOS router:

- Call 1 is from another H.323 gateway across the IP network to a traditional PBX connected to the router (for example, via a PRI interface). For this call, an incoming VoIP dial peer and an outgoing POTS dial peer are selected.
- Call 2 is from an analog phone connected to an FXS port on the router to a Cisco CallManager cluster across the IP network. For this call, an incoming POTS dial peer and an outgoing VoIP dial peer are selected by the router.
- Call 3 is from an IP phone controlled by Cisco CallManager Express or SRST to a PSTN interface on the router (for example, a PRI interface). For this call, an automatically generated POTS dial peer (corresponding to the **ephone** configured on the router) and an outgoing POTS dial peer are selected.

Figure 10-11 Incoming and Outgoing Dial Peers



To match incoming call legs to incoming dial peers, the router selects a dial peer by matching the information elements in the setup message (called number/DNIS and calling number/ANI) with four configurable dial peer attributes. The router attempts to match these items in the following order:

1. Called number with **incoming called-number**
2. Calling number with **answer-address**
3. Called number with **destination-pattern**
4. Incoming voice port with configured voice port

The router must match only one of these conditions. It is not necessary for all the attributes to be configured in the dial peer or that every attribute match the call setup information; only one condition must be met for the router to select a dial peer. The router stops searching as soon as one dial peer is matched, and the call is routed according to the configured dial peer attributes. Even if there are other dial peers that would match, only the first match is used.

How the router selects an outbound dial peer depends on whether **direct-inward-dial** (DID) is configured in the inbound POTS dial peer:

- If DID is not configured in the inbound POTS dial peer, the router performs two-stage dialing and collects the incoming dialed string digit-by-digit. As soon as one dial peer matches the destination pattern, the router immediately places the call using the configured attributes in the matching dial peer.
- If DID is configured in the inbound POTS dial peer, the router uses the full incoming dial string to match the destination pattern in the outbound dial peer. With DID, the setup message contains all the digits necessary to route the call, so no additional digit collection is required. If more than one dial peer matches the dial string, all of the matching dial peers are used to form a *hunt group*. The router attempts to place the outbound call leg using all of the dial peers in the hunt group until one is successful.

By default, dial peers in a hunt group are selected according to the following criteria, in the order listed:

1. Longest match in phone number

This method selects the destination pattern that matches the greatest number of dialed digits. For example, if one dial peer is configured with a dial string of 345.... and a second dial peer is configured with 3456789, the router would first select 3456789 because it has the longest explicit match of the two dial peers.

2. Explicit preference

This method uses the priority configured with the **preference** dial peer command. The lower the preference number, the higher the priority. The highest priority is given to the dial peer with preference order 0. If the same preference is defined in multiple dial peers with the same destination pattern, a dial peer is selected randomly.

3. Random selection

In this method, all destination patterns are weighted equally.

You can change this default selection order or choose different methods for hunting dial peers by using the **dial-peer hunt** global configuration command. An additional selection criteria is *least recent use*, which selects the destination pattern that has waited the longest since being selected (equivalent to *longest idle* for Cisco CallManager line groups).

Observe the following best practices when configuring H.323 dial peers on a Cisco IOS router:

- To ensure that incoming PSTN calls are directly routed to their destination based on the DNIS information, create a default POTS dial peer with the **direct-inward-dial** attribute, as follows:

```
dial-peer voice 999 pots
  incoming called-number .
  direct-inward-dial
  port 1/0:23
```
- When using the router as an H.323 gateway connected to a Cisco CallManager cluster, provide redundancy by configuring at least two VoIP dial peers with the same destination pattern pointing to two different Cisco CallManager servers. Use the **preference** attribute to select the priority order between primary and secondary Cisco CallManager servers.

Call Routing in Cisco IOS with a Gatekeeper

An H.323 gatekeeper is an optional node that manages endpoints (such as H.323 terminals, gateways, and Multipoint Control Units (MCUs), as well as Cisco CallManager Express and Cisco CallManager clusters) in an H.323 network, providing them with call routing and call admission control functions. The endpoints communicate with the gatekeeper using the H.323 Registration Admission Status (RAS) protocol.

Endpoints attempt to register with a gatekeeper on startup. When they want to communicate with another endpoint, they request admission to initiate a call using a symbolic alias for the endpoint, such as an E.164 address or an email address. If the gatekeeper decides that the call can proceed, it returns a destination IP address to the originating endpoint. This IP address might not be the actual address of the destination endpoint, but instead it might be an intermediate address, such as the address of an IP-to-IP gateway or a gatekeeper that routes call signaling.

For more details about the H.323 protocol and the message exchange between H.323 endpoints and gatekeepers, refer to the *Cisco IOS H.323 Configuration Guide*, available at

<http://www.cisco.com>

The Cisco 2600, 3600, 3700, 2800, 3800, and 7200 Series routers all support the gatekeeper feature. You can configure Cisco IOS gatekeepers in a number of different ways for redundancy, load balancing, and hierarchical call routing. This section focuses on the call routing capabilities of the gatekeeper feature. For redundancy and scalability considerations, refer to [Gatekeeper Redundancy, page 8-16](#); for call admission control considerations, refer to [Gatekeeper, page 9-6](#).

Call routing in Cisco IOS gatekeeper is based on the following types of information:

- Statically configured information, such as zone prefixes and default technology prefixes
- Dynamic information, such as E.164 addresses and technology prefixes provided by H.323 devices during the registration phase
- Per-call information, such as called number and technology prefix

A zone is a collection of H.323 devices (such as endpoints, gateways, or MCUs) that register with a gatekeeper. There can be only one active gatekeeper per zone, and you can define up to 100 local zones on a single gatekeeper.

When an H.323 endpoint registers with the gatekeeper, it is assigned to a zone and it can optionally register one or more E.164 addresses for which it is responsible, as well as a technology prefix that specifies which kinds of calls it can handle (for example, voice, video, fax, and so on).

For each zone, you can configure one or more *zone prefixes* on the gatekeeper. Zone prefixes are strings that contain digits and wildcards and are used by the gatekeeper to facilitate call routing decisions. The following characters are allowed in a zone prefix string:

- All numbers between 0 and 9, which match a single specific digit
- The dot (.) wildcard, which matches one digit between 0 and 9
- The * wildcard, which matches one or more digits between 0 and 9

To understand the gatekeeper call routing behavior, it is helpful to consider the message parsing logic. [Figure 10-12](#) illustrates the parsing logic for an Admission Request (ARQ). 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.

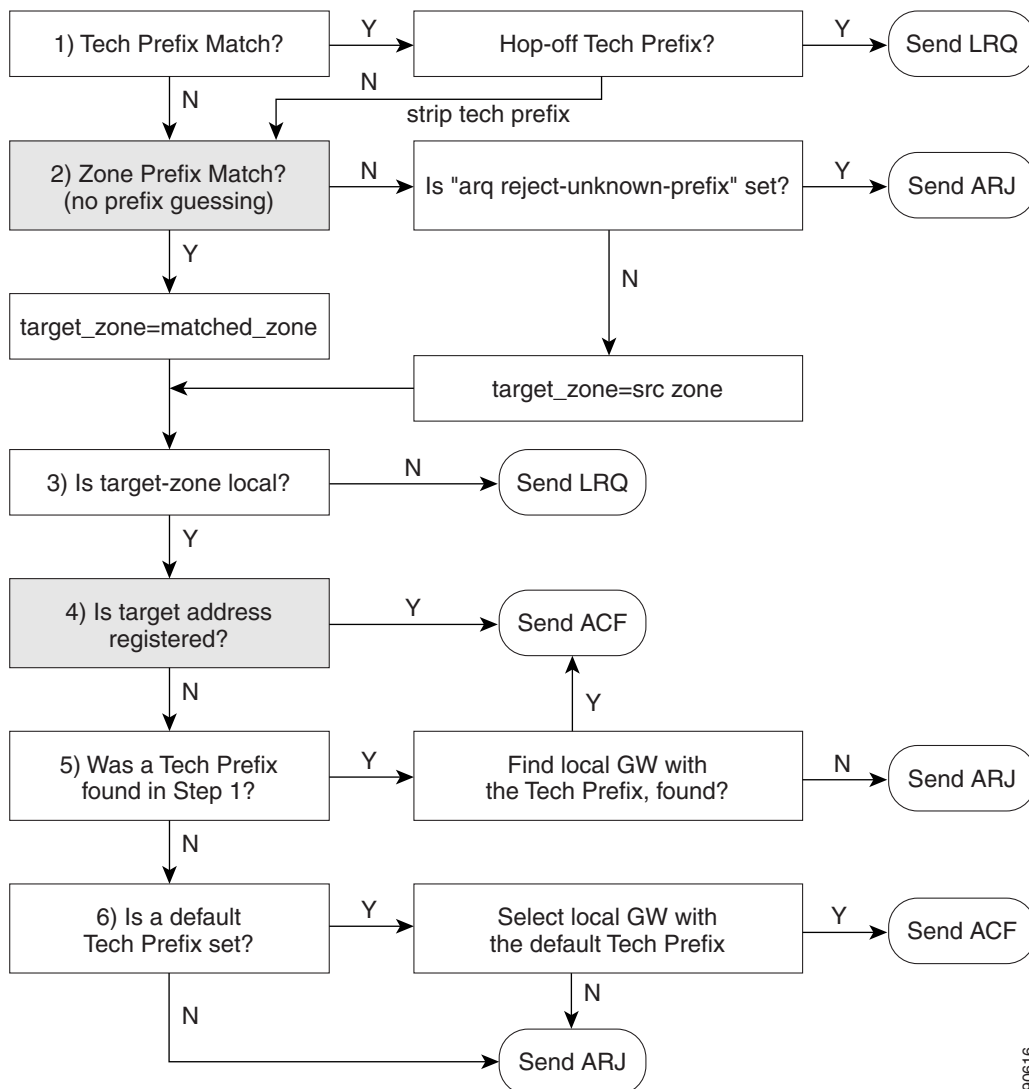
If the ARQ contains the E.164 address (with Cisco CallManager, the ARQ always contains an E.164 address), the ARQ may or may not contain a technology prefix. If the ARQ contains a technology prefix, the gatekeeper strips it from the called number. If the ARQ does not contain a technology prefix, the gatekeeper uses the default technology prefix if one is configured (see the **gw-type-prefix** command in the section on [Centralized Gatekeeper Configuration, page 10-34](#)). The technology prefix thus obtained is stored in memory, and the gatekeeper continues with the call routing algorithm.

Next, the gatekeeper tries to match the called number with one of the configured zone prefixes. Longest-match is used if multiple potential matches exist. If no zone prefix can be matched, and if the gatekeeper is configured to accept calls with an unknown prefix, the gatekeeper then assumes that the destination zone is equal to the source zone.

At this point, the gatekeeper searches in the chosen destination zone for a registered E.164 address that matches the called number. If there is a match, the gatekeeper can send an Admission Confirm (ACF), provided that the requested bandwidth for the call is available and that 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.

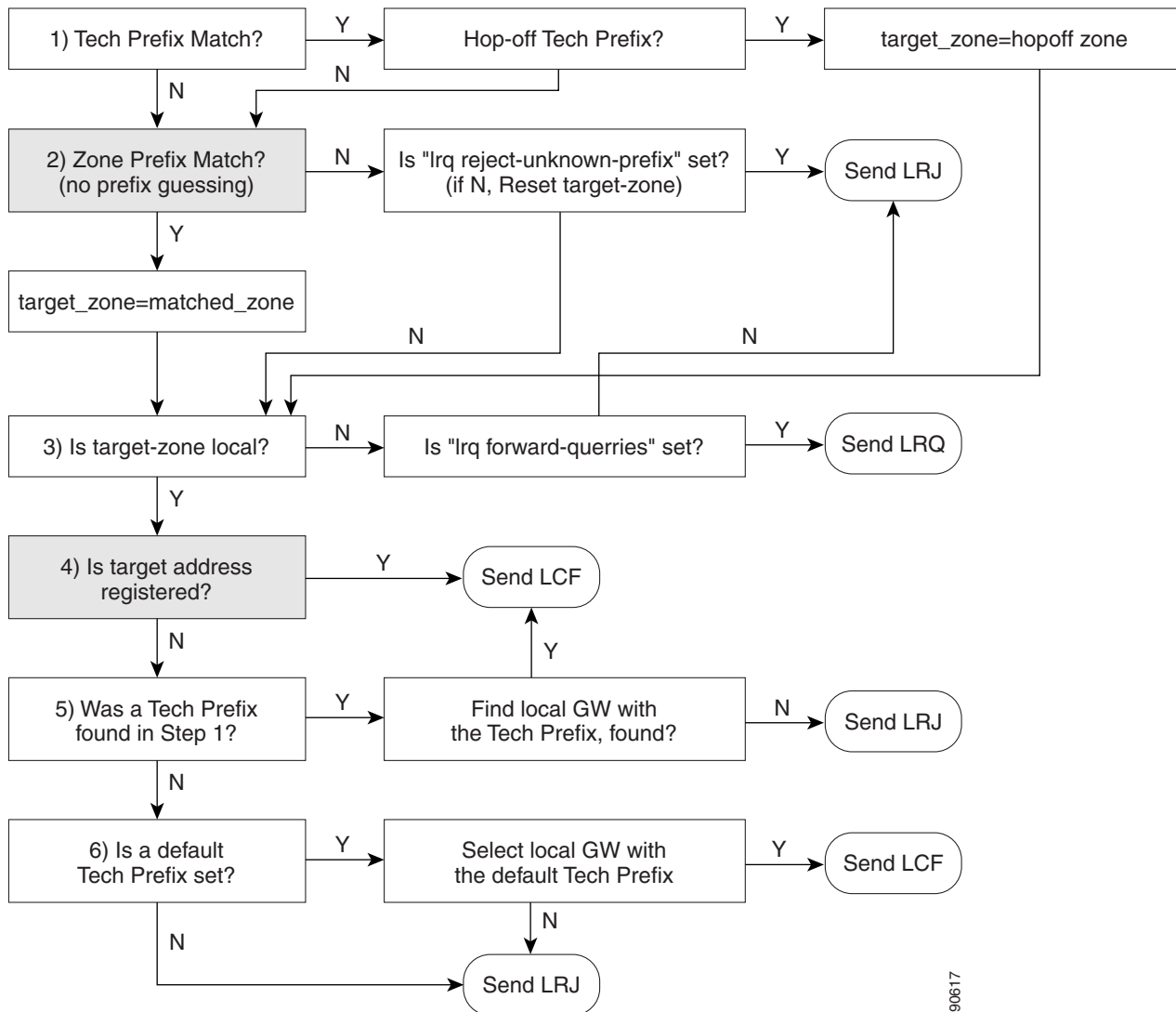
If there is no matching E.164 address registered in the destination zone, the gatekeeper will use the previously stored technology prefix to choose a gateway registered in that zone as the destination for the call. The same considerations regarding bandwidth availability and endpoint registration dictate whether the gatekeeper will send an ACF or an 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 10-12 Gatekeeper Address Resolution for an ARQ

90616

Figure 10-13 illustrates the parsing logic for a Location Request (LRQ). LRQ messages are exchanged between gatekeepers and are used for inter-zone (remote zone) calls. For example, gatekeeper A receives an ARQ from a local zone gateway requesting call admission for a remote zone device. Gatekeeper A then sends an LRQ message to gatekeeper B. Gatekeeper B replies to the LRQ message with either a Location Confirm (LCF) or Location Reject (LRJ) message, depending on whether it is configured to admit or reject the inter-zone call request and whether the requested resource is registered.

Figure 10-13 Gatekeeper Address Resolution for an LRQ

90617

Traditional Cisco IOS gatekeeper functionality has been extended to accommodate for IP-to-IP gateways through the concept of a *via-zone gatekeeper*. (Refer to [IP-to-IP Gateway](#), page 9-14, for some deployment examples.)

A *via-zone* gatekeeper differs from legacy gatekeepers in how it uses LRQ and ARQ messages for call routing. Using *via-zone* gatekeepers will maintain normal clusters and functionality. Legacy gatekeepers examine incoming LRQs based on the called number, and more specifically the `dialedDigits` field in the `destinationInfo` portion of the LRQ. *Via-zone* gatekeepers look at the origination point of the LRQ before looking at the called number. If an LRQ comes from a gatekeeper listed in the *via-zone* gatekeeper's remote zone configurations, the gatekeeper checks to see that the zone remote configuration contains an **invia** or **outvia** keyword. If the configuration contains either of these keywords, the gatekeeper uses the new *via-zone* behavior; if not, it uses legacy behavior.

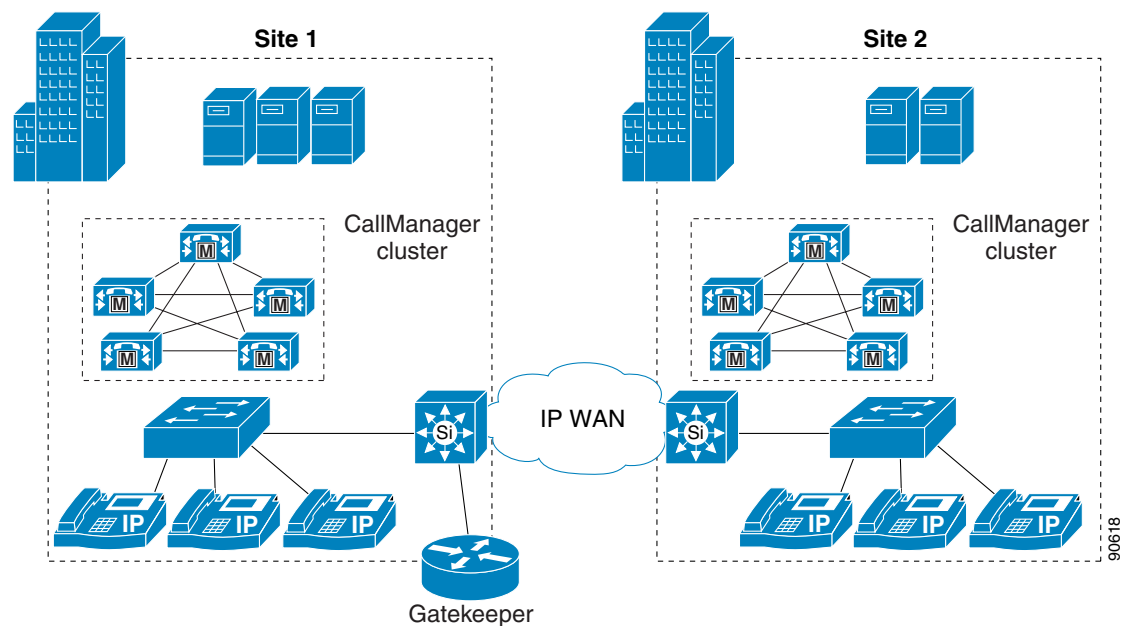
For ARQ messages, the gatekeeper determines if an **outvia** keyword is configured on the destination zone. If the **outvia** keyword is configured and the zone named with the **outvia** keyword is local to the gatekeeper, the call is directed to an IP-IP gateway in that zone by returning an ACF pointing to the IP-IP

gateway. If the zone named with the **outvia** keyword is remote, the gatekeeper sends a location request to the **outvia** gatekeeper rather than the remote zone gatekeeper. The **invia** keyword is not used in processing the ARQ.

Centralized Gatekeeper Configuration

A single gatekeeper can support call routing between clusters and call admission control for up to 100 Cisco CallManager clusters. [Figure 10-14](#) illustrates a distributed call processing environment with two Cisco CallManager clusters and a single, centralized gatekeeper.

Figure 10-14 Centralized Gatekeeper Supporting Two Clusters



[Example 10-1](#) shows the gatekeeper configuration for the example in [Figure 10-14](#).

Example 10-1 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 10-14](#):

- 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.

- Bandwidth statements are configured for each site. Cisco recommends that you use the **bandwidth interzone** command because 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 entire deployment.

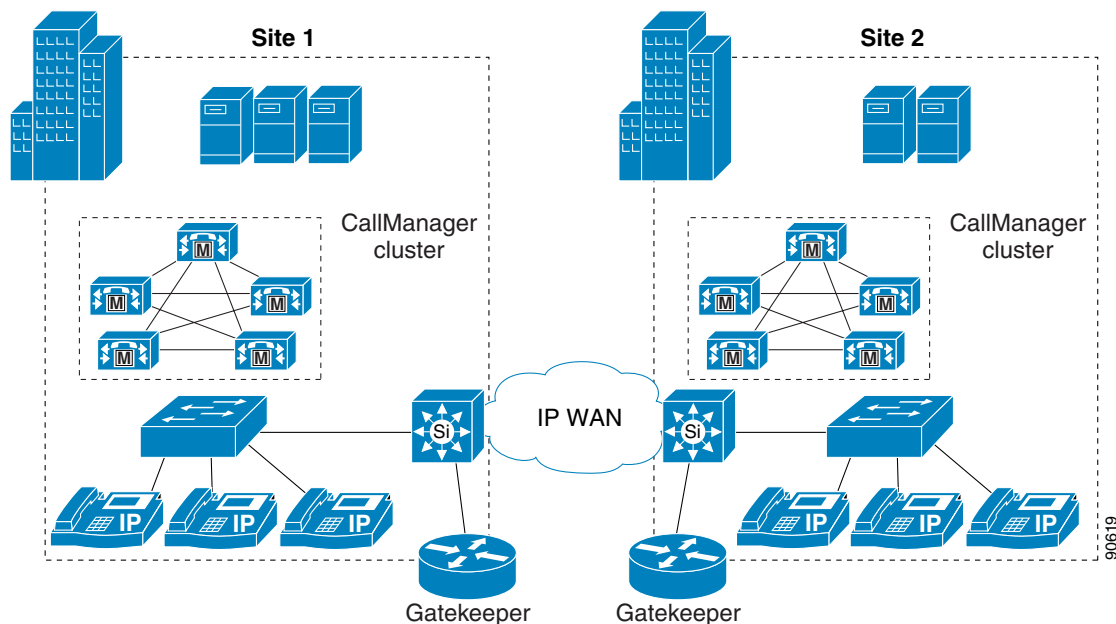
Cisco IOS gateways automatically add a technology prefix to outbound calls if the gateway has a prefix configured. The gateway also automatically strips the prefix from incoming H.323 calls. Cisco CallManager can register with the gatekeeper using a technology prefix, as specified on the configuration page for gatekeeper-controlled H.323 trunks. However, this technology prefix is not automatically added to outgoing calls to the gatekeeper, and is not automatically stripped from inbound calls to Cisco CallManager. You can use translation patterns and significant digits to manipulate the called number so as to add or strip the technology prefix as needed.

- 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 10-15 illustrates a distributed call processing environment with two clusters and two gatekeepers.

Figure 10-15 *Distributed Gatekeepers Supporting Two Clusters*



Example 10-2 shows the gatekeeper configuration for Site 1 in Figure 10-15.

Example 10-2 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 10-2:

- 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 10-3](#) shows the gatekeeper configuration for Site 2 in [Figure 10-15](#).

Example 10-3 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 10-3](#):

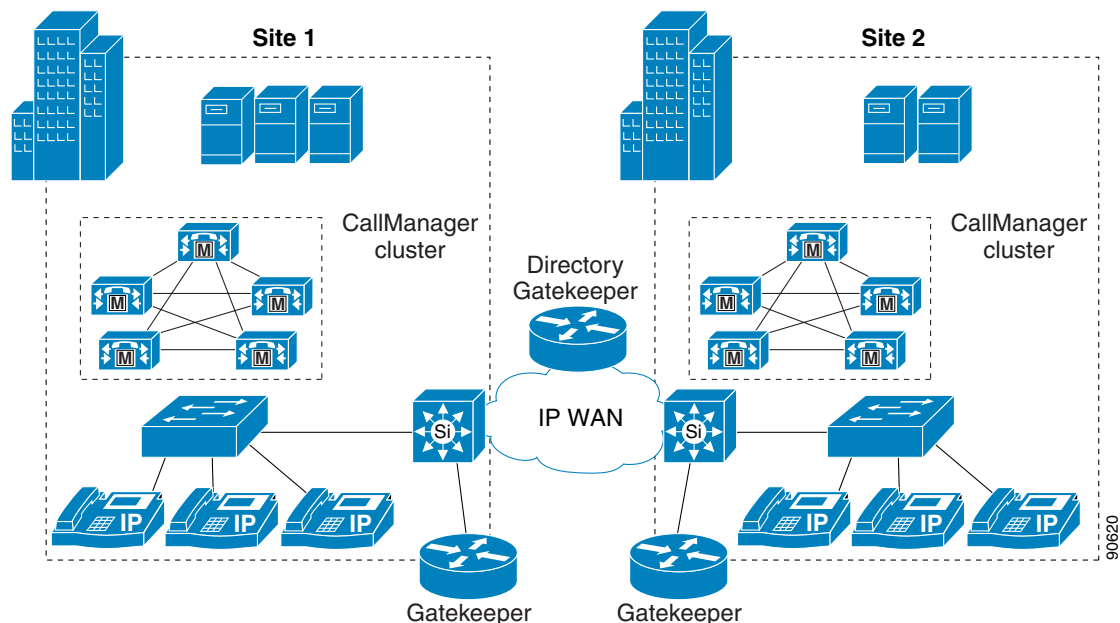
- A local zone is configured for registration of local Cisco CallManager cluster trunks.
- A remote zone is configured for routing calls to the Site 1 gatekeeper.
- Zone prefixes are configured for both zones for inter-zone call routing.
- The **bandwidth remote** command is used to limit bandwidth between 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 10-16](#) 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 10-16 Distributed Gatekeepers with a Directory Gatekeeper

Example 10-4 shows the gatekeeper configuration for Site 1 in Figure 10-16. Because the Site 1 and Site 2 gatekeeper configurations are almost identical in this example, only Site 1 is covered here.

Example 10-4 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
arg reject-unknown-prefix
no shutdown
```

The following notes also apply to Example 10-4:

- 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 10-5 shows the directory gatekeeper configuration for the example in Figure 10-16.

Example 10-5 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 10-5:

- 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 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.

Calling Privileges in Cisco IOS with H.323 Dial Peers

Use the class of restriction (COR) functionality to implement calling privileges with Cisco IOS-based systems using H.323, including H.323 gateways, SRST, and Cisco CallManager Express. This functionality provides flexibility in network design, allows administrators to block calls for all users (for example, calls to 900 numbers), and applies different calling privileges to call attempts from different originators (for example, it allows some users but not others to call international numbers).

The fundamental mechanism at the center of the COR functionality relies on the definition of incoming and outgoing *corlists* that are associated to existing dial peers, where the concepts of incoming and outgoing are relative to the Cisco IOS router (as in the case of dial peers). Each corlist is defined to include a number of members, which are simply tags previously defined within Cisco IOS.

When a call goes through the router, an incoming dial peer and an outgoing dial peer are selected based on the Cisco IOS dial peer routing logic. If corlists are associated with the selected dial peers, the following additional check is performed before extending the call:

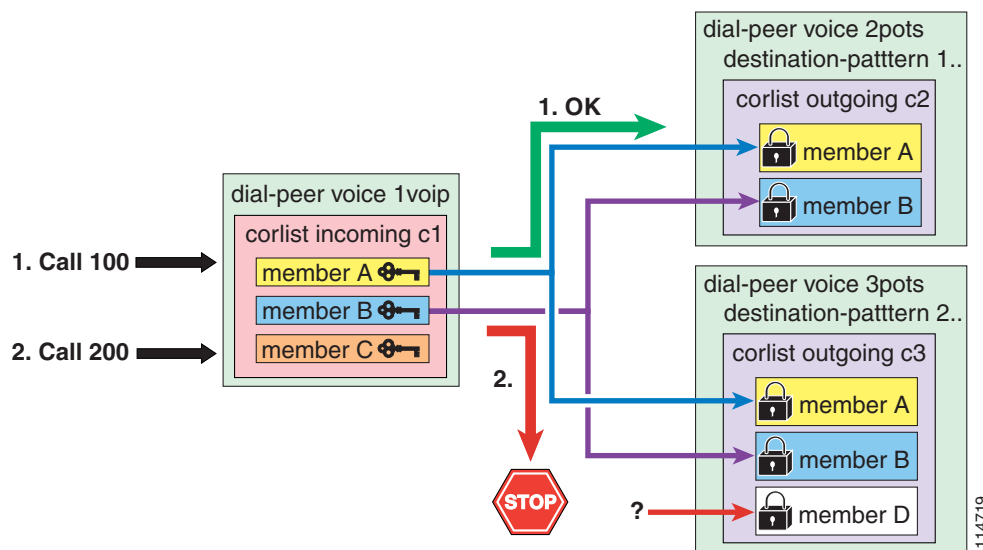
- If the members of the outgoing corlist associated with the outgoing dial peer are a subset of the members of the incoming corlist associated with the incoming dial peer, the call is permitted.
- If the members of the outgoing corlist associated with the outgoing dial peer are *not* a subset of the members of the incoming corlist associated with the incoming dial peer, the call is rejected.

If no corlist statements are applied to some dial peers, the following properties apply:

- When no incoming corlist is configured on a dial-peer, the default incoming corlist is used. The default incoming corlist has the highest possible priority, and it therefore allows this dial-peer to access all other dial-peers, regardless of their outgoing corlist.
- When no outgoing corlist is configured on a dial-peer, the default outgoing corlist is used. The default outgoing corlist has the lowest possible priority, and it therefore allows all other dial-peers to access this dial-peer, regardless of their incoming corlist.

This behavior is best illustrated with an example as shown in [Figure 10-17](#), where one VoIP dial-peer and two POTS dial-peers are defined.

Figure 10-17 Example of COR Behavior



The VoIP dial-peer is associated with the c1 incoming corlist, with members A, B, and C. You can think of members of the incoming corlist as "keys."

The first POTS dial-peer has a destination-pattern of 1.. and is associated with the c2 outgoing corlist, with members A and B. The second POTS dial-peer has a destination-pattern of 2.. and is associated with the c3 outgoing corlist, with members A, B, and D. You can think of members of the outgoing corlists as "locks."

For the call to succeed, the incoming corlist of the incoming dial-peer must have all the "keys" needed to open all the "locks" of the outgoing corlist of the outgoing dial-peer.

In the example shown in [Figure 10-17](#), a first VoIP call with destination 100 is received by the router. The Cisco IOS call routing logic matches the incoming call leg with the VoIP dial-peer and the outgoing call leg with the first POTS dial-peer. The COR logic is then applied; because the c1 incoming corlist has all the keys needed for the c2 outgoing corlist locks (A and B), the call succeeds.

A second VoIP call with destination 200 is then received by the router. The Cisco IOS call routing logic matches the incoming call leg with the VoIP dial-peer and the outgoing call leg with the second POTS dial-peer. The COR logic is then applied; because the c1 incoming corlist is missing one "key" for the c3 outgoing corlist (D), the call is rejected.

Follow these steps when configuring the COR functionality in Cisco IOS:

-
Step 1	Define "tags" to be used as corlist members with the command dial-peer cor custom .
Step 2	Define corlists with the command dial-peer cor list <i>corlist-name</i> .
Step 3	Associate corlists with existing VoIP or POTS dial-peers by using the command corlist {incoming outgoing} <i>corlist-name</i> within the dial-peer configuration.
-

With Cisco IOS Release 12.2(8)T and later, you can apply the COR functionality to SRST-controlled IP phones. Because IP phones register with the SRST router dynamically, SRST has no prior knowledge of the individual phones until they lose connectivity to the Cisco CallManager cluster. Therefore, the COR feature configuration for SRST is based on the phone DN's instead. When the IP phones register with the SRST router, they communicate their DN to it, thus allowing the SRST router to assign them to the appropriate corlist.

Configure COR for IP phones controlled by SRST by using the command **cor {incoming | outgoing}** *corlist-name* {*corlist-number starting-number – ending-number* | **default**} within the **call-manager-fallback** configuration mode.

The following limitations apply to this command:

- The maximum number of **cor {incoming | outgoing}** statements under **call-manager-fallback** is 5 (plus the default statement) in SRST version 2.0, available on Cisco IOS Release 12.2(8)T or later.
- The maximum number of **cor {incoming | outgoing}** statements under **call-manager-fallback** is 20 (plus the default statement) in SRST version 3.0, available on Cisco IOS Release 12.3(4)T or later.

The COR functionality can also be deployed with Cisco CallManager Express, using Cisco IOS Release 12.2(8)T and later. Because the individual IP phones are specifically configured within Cisco CallManager Express, you can apply corlists directly to the IP phones themselves by using the command **cor {incoming | outgoing}** *corlist-name* within the **ephone-dn dn-tag** configuration mode of each IP phone.

Refer to the section on [Building Classes of Service in Cisco IOS with H.323, page 10-75](#), for an example of how to apply all these concepts in practice.

For more details on configuration of Cisco SRST and Cisco CallManager Express, refer to the *Cisco SRST 3.0 System Administrator Guide* and the *Cisco CallManager Express 3.1 System Administrator Guide* available at

<http://www.cisco.com>

Digit Manipulation in Cisco IOS with H.323 Dial Peers

In Cisco IOS routers running H.323, digit manipulation is performed via voice translation profiles, which are used to manipulate the calling number (ANI) or called number (DNIS) digits for a voice call or to change the numbering type of a call.

Voice translation profiles are available starting with Cisco IOS Release 12.2(11)T, and they are used to convert a telephone number into a different number before the call is matched to an incoming dial peer or before the call is forwarded by the outgoing dial peer. For example, within your company you might dial a five-digit extension to reach an employee at another site. If the call is routed through the PSTN to reach the other site, the originating gateway must use voice translation profiles to convert the five-digit extension into the ten-digit format that is recognized by the central office switch.

You configure voice translation profiles by using the **voice translation-rule** and **voice translation-profile** Cisco IOS commands, which use regular expressions to define the digit strings to be modified. You then specify if the manipulations should be associated to calling numbers, called numbers, or redirected called numbers. After you define a voice translation profile, you can apply it to any of the following elements:

- All incoming POTS call legs that terminate on a specific voice port
- All incoming VoIP call legs into the router
- Outgoing call legs associated with a specific VoIP or POTS dial peer
- All incoming or outgoing call legs that terminate on the IP phones controlled by SRST
- Incoming call legs for calls originated by all IP phones controlled by SRST


Note

Voice translation profiles using the **voice translation-rule** command replace and enhance the functionality previously provided by the **translation-rule** command. The syntax of the new command differs from that used by the old command. For more details, refer to **voice translation-rule** in the *Cisco IOS Voice Command Reference*, Release 12.2(11)T or later, available at <http://www.cisco.com>.

A typical application of voice translation profiles is to enable the preservation of on-net inter-site dialing habits from a branch site even when the IP WAN is unavailable and the router is running in SRST mode. For example, consider a simple deployment with a central site in San Jose and three remote sites in San Francisco, New York, and Dallas. [Table 10-3](#) shows the DID ranges and the internal site codes for this example.

Table 10-3 Example of DID Ranges and Site Codes for Translation Rule Application

	San Jose	San Francisco	New York	Dallas
DID Range	(408) 555-1XXX	(415) 555-1XXX	(212) 555-1XXX	(972) 555-1XXX
Site Code	1	2	3	4

Inter-site calls are normally placed across the IP WAN by dialing the on-net access code 8 followed by the one-digit site code and the four-digit extension of the called party. To preserve these dialing habits even when the IP WAN is down and Cisco SRST is active, the internal numbers must be converted back into the E.164 numbers before being sent to the PSTN. A sample configuration for the San Francisco router is as follows:

```
voice translation-rule 1
  rule 1 /^81/ /91408555/
  rule 2 /^83/ /91212555/
  rule 3 /^84/ /91972555/

voice translation-profile on-net-xlate
  translate called 1

call-manager-fallback
  translation-profile outgoing on-net-xlate

dial-peer voice 2 pots
  destination-pattern 91[2-9]..[2-9].....
  port 1/0:0
  direct-inward-dial
  forward-digits 11
```

With this configuration, when the San Francisco site is in SRST mode and a user dials 831000, the router will match **rule 2** of **voice translation-rule 1** and translate the called number to 912125551000. This new number will then be used to match the outgoing dial peer (**dial-peer voice 2**).

For a detailed description of dial peers and their configuration, refer to *Configuring Dial Plans, Dial Peers, and Digit Manipulation*, part of the *Cisco IOS Voice, Video, and Fax Configuration Guide, Release 12.2*, available at

<http://www.cisco.com>

For a thorough explanation of regular expression syntax in Cisco IOS, refer to the document on *Regular Expressions*, available at

http://www.cisco.com/en/US/partner/products/sw/iosswrel/ps1835/products_configuration_guide_chapter09186a00800ca7e6.html

Design Considerations

This section presents the following dial plan design considerations for multi-site deployments:

- [Design Guidelines for Multi-Site Deployments, page 10-44](#), covers guidelines and best practices that apply to all multi-site deployment models.
- [Choosing a Dial Plan Approach, page 10-46](#), presents the various approaches to organizing a dial plan for uniform versus variable-length on-net dialing and, for this second option, partitioned versus flat addressing.
- The following sections analyze in detail three dial plan approaches and provide configuration guidelines for each:
 - [Deploying Uniform On-Net Dial Plans, page 10-48](#)
 - [Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing, page 10-50](#)
 - [Deploying Variable-Length On-Net Dial Plans with Flat Addressing, page 10-54](#)
- The following sections present two alternative ways of configuring classes of service within Cisco CallManager:
 - [Building Classes of Service for Cisco CallManager with the Traditional Approach, page 10-63](#)
 - [Building Classes of Service for Cisco CallManager with the Line/Device Approach, page 10-67](#)
- [Building Classes of Service in Cisco IOS with H.323, page 10-75](#), explains how to implement classes of service within a Cisco IOS router running the H.323 protocol.
- [Deploying Call Coverage, page 10-78](#), provides guidelines and best practices for implementing call coverage functionality using hunt lists and line groups with Cisco CallManager.

Design Guidelines for Multi-Site Deployments

The following guidelines and best practices apply in common to all multi-site IP Telephony deployments. For deployments that involve more than one Cisco CallManager cluster, also refer to the section on [Additional Considerations for Distributed Call Processing Deployments](#), page 10-45.

- To prevent routing loops, make sure the calling search spaces of all PSTN gateways do not include any partitions that contain external route patterns.
- When choosing DID ranges with your Local Exchange Carrier (LEC), try to select them so that no overlap occurs within a site. For example, if you use four-digit dialing within a site and you need two blocks of 1000 DIDs, the blocks (408)555-1XXX and (408)999-1XXX would overlap when reduced to four-digit numbers and would introduce additional complexity due to inbound and outbound translations.
- Allow for multiple ways of dialing emergency numbers. For example, in North America, configure both 911 and 9.911 as emergency route patterns within Cisco CallManager.
- When automated alternate routing (AAR) is deployed, ensure that the external phone number mask configured on the IP phones is compatible with all the prefixes added by the various AAR groups. For example, in a multi-national deployment, do not include national access codes such as 0 in the mask unless they are part of the global E.164 address.
- You can force on-net calls dialed as PSTN for destinations within the cluster by adding translation patterns that match the E.164 DID ranges for each site and that manipulate the digits so they match the destination's internal extension. However, remember to provision AAR accordingly, using one of the following methods, so that PSTN failover is possible when the IP WAN is not available:
 - Configure the AAR calling search space to exclude the partition containing the "force on-net" translation patterns but to include a partition containing the regular route patterns pointing to the PSTN.
 - Configure the AAR group to prefix digits with a special character such as *, and configure additional route patterns such as *9.!, *9.!! (or *0.!, *0.!!) in the regular partition.

The second approach is preferable because it does not require additional calling search spaces to be defined for AAR, and the additional * route patterns provide a good troubleshooting and testing tool that allows for overriding the "force on-net" configuration and placing calls over the PSTN even when AAR is not invoked.

- Within a centralized call processing cluster with N sites, you can implement Tail-End Hop-Off (TEHO) using one of the following methods:
 - TEHO with centralized failover
This method involves configuring a set of N route patterns in a global partition, with each pattern pointing to a route list that has the appropriate remote site route group as the first choice and the central site route group as the second choice.
 - TEHO with local failover
This method involves configuring N sets of N route patterns in site-specific partitions, with each pattern pointing to a route list that has the appropriate remote site route group as the first choice and the local site route group as the second choice.

While the second approach allows for an optimal failover scenario when the remote gateway or the IP WAN is unavailable, it also introduces a high level of complexity into the dial plan because it requires a minimum of N^2 route patterns and N^2 route lists, as opposed to the N route patterns and N route lists needed with the first approach.

- When appropriate for your national numbering plan, you may configure an additional translation pattern at each site to catch local PSTN calls dialed as long-distance calls and to translate them into the proper abbreviated form. This translation pattern should be accessible only from phones located within the site. Such a configuration also helps simplify the AAR configuration (see [Special Considerations for Sites Located Within the Same Local Dialing Area, page 10-21](#)).
- Do not use the multilevel precedence and preemption (MLPP) feature to assign higher priority to emergency calls. An emergency-related call might not appear as such to the IP Telephony system, and you would risk terminating an existing emergency call to place another call to the main emergency service routing number.

**Note**

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 increase to allow additional time for the configuration to initialize. For details on the service parameters, refer to the online help for Service Parameters in Cisco CallManager Administration.

Additional Considerations for Distributed Call Processing Deployments

In addition to the considerations made in the previous section, observe the following best practices when designing a dial plan for a distributed call processing deployment (that is, any multi-site deployment involving multiple Cisco CallManager clusters):

- Avoid splitting DID ranges across multiple Cisco CallManager clusters. This practice would make intercluster routing very difficult because summarization would not be possible. Each DID range should belong to a single Cisco CallManager cluster.
- Avoid splitting devices within a remote site between multiple Cisco CallManager clusters. Locations-based call admission control is significant only within a cluster, and having devices belong to different clusters at the same remote site would lead to poor utilization of the IP WAN bandwidth because you would have to "partition" the available bandwidth between the clusters. Each remote site should belong to a single Cisco CallManager cluster.
- Use gatekeeper-controlled intercluster trunks to route calls between Cisco CallManager clusters. This practice enables you to add or modify clusters easily in your network without reconfiguring all other clusters.
- Implement redundancy in the connection between Cisco CallManager and the gatekeeper by using a gatekeeper cluster and by assigning the intercluster trunk to a device pool that uses a Cisco CallManager group with multiple servers configured.
- When sending calls to the gatekeeper, expand the called number to the full E.164 address. This practice simplifies PSTN failover when the IP WAN is not available because no additional digit manipulation is required to reroute the call via a PSTN gateway. Also, this practice eliminates the need to configure the local (calling) Cisco CallManager with dial length information for each remote site.
- Within the gatekeeper, configure one zone per Cisco CallManager cluster. For each cluster/zone, add zone prefix statements to match all DN ranges owned by that cluster.

- You can implement Tail-End Hop-Off (TEHO) across multiple Cisco CallManager clusters by following these guidelines:
 - Add specific route patterns for the relevant E.164 ranges to the source (originating) Cisco CallManager cluster, and point them to a route list that has the IP WAN route group as the first choice and a local PSTN route group as the second choice
 - Within the Cisco IOS gatekeeper configuration, add zone prefix statements for all the relevant E.164 ranges and point them to the appropriate Cisco CallManager cluster
 - Ensure that the intercluster trunk calling search space in the destination Cisco CallManager cluster includes partitions featuring route patterns that match the local PSTN numbers, and that digit manipulation is applied as needed (for example, stripping the area code before sending the call to the PSTN).

For details on how to configure a Cisco IOS gatekeeper for distributed call processing deployments, see [Gatekeeper Considerations, page 8-16](#).

Choosing a Dial Plan Approach

As introduced in [Planning Considerations, page 10-2](#), there are two main approaches to a dial plan for internal destinations within an IP Telephony system:

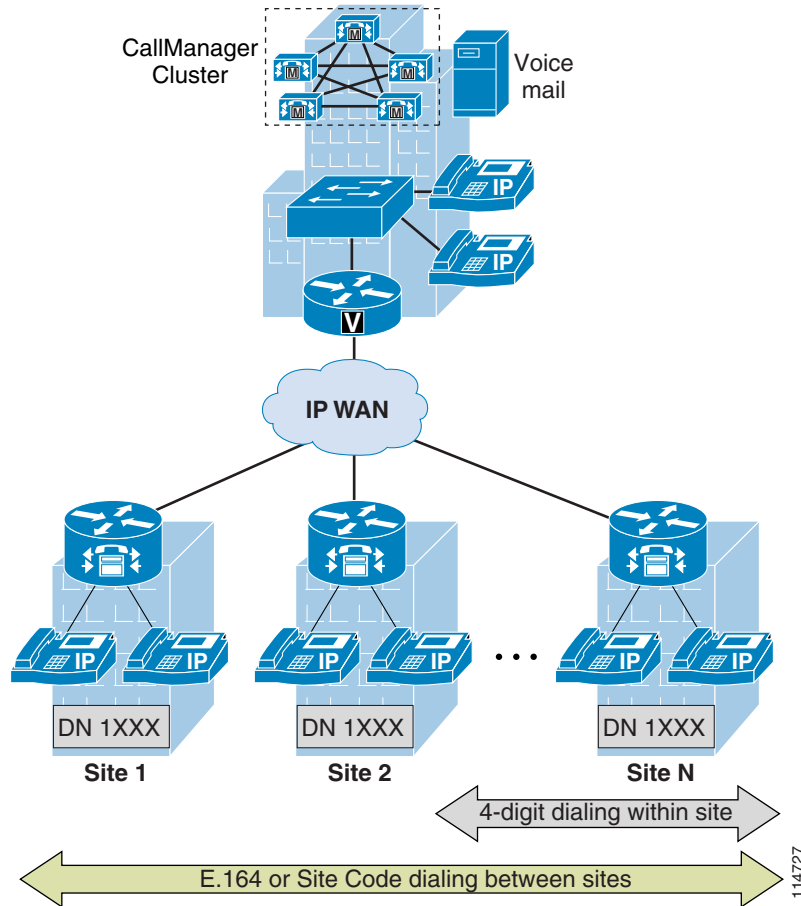
- Uniform on-net dial plan, where each internal destination is dialed in the same way regardless of whether the caller is located in the same site or in a different site.
- Variable-length on-net dial plan, where internal destinations are dialed differently within a site than across sites. Typically, this approach uses four or five-digit abbreviated dialing for calls within a site and either full E.164 addresses or an on-net access code followed by a site code and the extension for calls across sites.

To help you decide which approach is best suited for your needs, consider the following high-level design questions:

- How many sites will eventually be served by the IP Telephony system?
- What are the calling patterns between sites or branches?
- What do users dial within a site and to reach another site?
- Are there any calling restrictions applied to on-net inter-site calls?
- What transport network (PSTN or IP WAN) will be used for most inter-site calls?
- What (if any) CTI applications are being used?
- Is there a desire for a standardized on-net dialing structure using site codes?

Uniform on-net dial plans are the easiest to design and configure; however, they work best for small to medium deployments, and they become impractical when the number of sites and users increases. They are described and analyzed in detail in the section on [Deploying Uniform On-Net Dial Plans, page 10-48](#).

Variable-length on-net dial plans are more scalable but also more complex to design and configure. [Figure 10-18](#) shows the typical requirements for a large-scale deployment using the variable-length on-net dial plan approach.

Figure 10-18 Typical Dialing Requirements for Large Multi-Site Deployments

With Cisco CallManager, there are two main methods for implementing a variable-length on-net dial plan:

- Partitioned addressing

The internal extensions are placed in different partitions according to the site where they are located. This method is typically based on full E.164 addresses for inter-site calls and is analyzed in detail in the section on [Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing](#), page 10-50.

- Flat addressing

All internal extensions are placed in the same partition. This method is typically based on on-net site codes for inter-site calls and is analyzed in detail in the section on [Deploying Variable-Length On-Net Dial Plans with Flat Addressing](#), page 10-54. In some cases it is possible to use this approach even when using full E.164 addresses for inter-site calls, as described in the section on [Special Considerations for Deployments Without Site Codes](#), page 10-61.

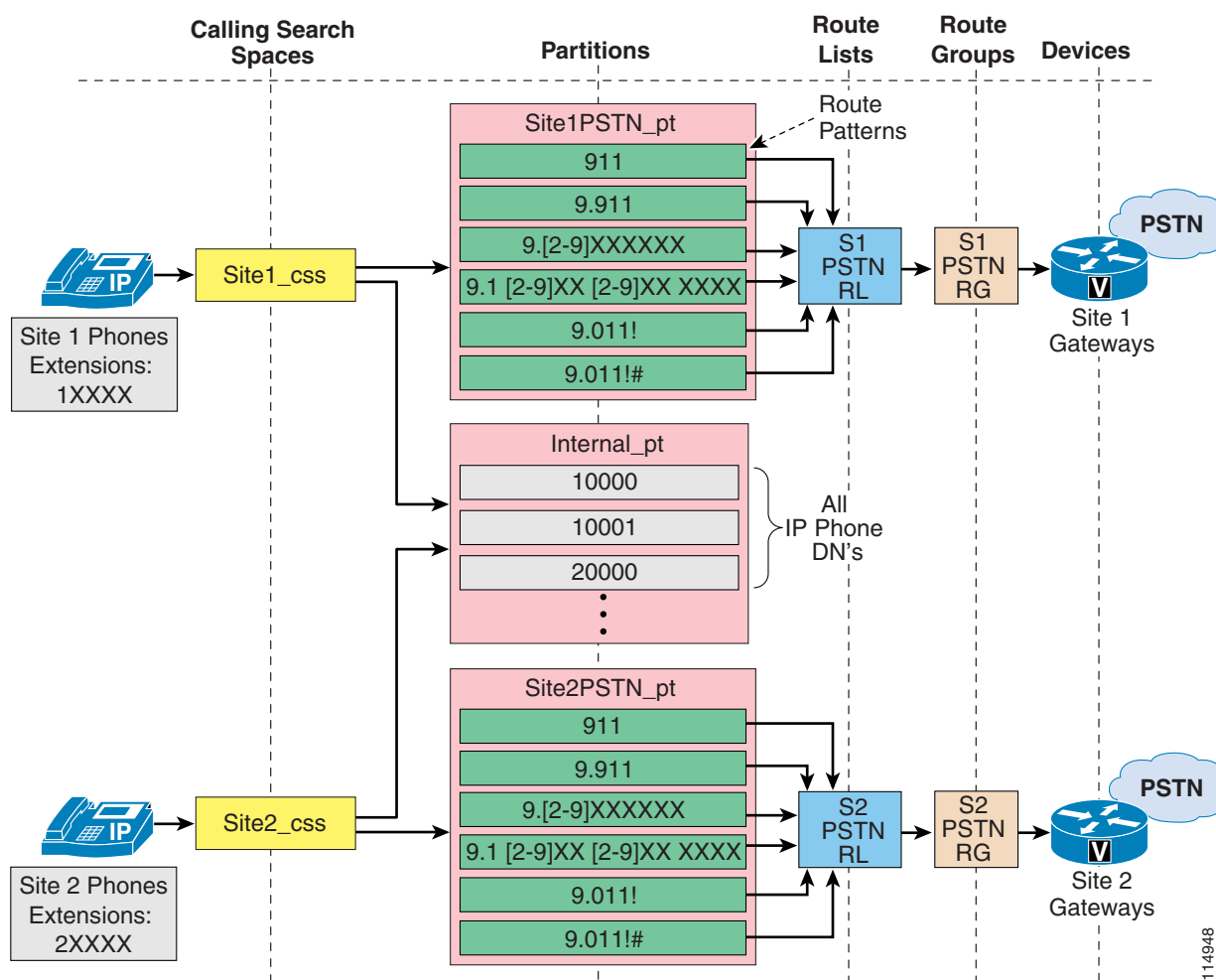
Deploying Uniform On-Net Dial Plans

You can implement a uniform on-net dial plan by following these guidelines:

- Uniquely identify all phones with an abbreviated extension.
- Place all the phone DN's in a single partition.
- At each site, place PSTN route patterns in one or more site-specific partitions, according to the chosen class-of-service approach.

Figure 10-19 shows an example implementation for a single Cisco CallManager cluster deployment.

Figure 10-19 Example of Uniform On-Net Dial Plan Deployment



Use this approach if both of the following conditions apply:

- The DID ranges available do not overlap when considering the number of digits chosen to identify internal extensions.
- The number of sites covered by the IP Telephony system is not expected to grow significantly over time.

The following sections analyze implementation details and best practices for different types of calls within the framework of uniform on-net dial plans:

- [Inter-Site Calls Within a Cluster, page 10-49](#)
- [Outgoing PSTN and IP WAN Calls, page 10-49](#)
- [Incoming Calls, page 10-49](#)
- [Voicemail Calls, page 10-49](#)

Inter-Site Calls Within a Cluster

Because all internal DNs are directly reachable from every device's calling search space, all on-net calls (intra-site and inter-site) are automatically enabled and need no additional configuration within Cisco CallManager.

Outgoing PSTN and IP WAN Calls

PSTN calls are enabled via the site-specific partitions and route patterns, so that emergency and local calls can be routed via the local branch gateway. Long-distance and international calls may be routed via the same branch gateway (as shown in [Figure 10-19](#)) or via a centralized gateway, depending on company policy. This second option would simply require an additional route list per site, with the first-choice route group pointing to the central site gateway and, optionally, a second-choice route group pointing to the local branch gateway.

Abbreviated dialing to another Cisco CallManager cluster or Cisco CallManager Express is also possible via a gatekeeper. For these IP WAN calls, Cisco recommends that you expand the abbreviated string to the full E.164 via a translation pattern before sending it to the gatekeeper.

Incoming Calls

Incoming PSTN calls simply require stripping the excess digits in order to match the extension length configured in Cisco CallManager. This can be done within the gateway configuration or, alternatively, via a translation pattern included in the gateway's calling search space.

Voicemail Calls

Because every extension is unique within the system, the extension itself can be used to configure voicemail boxes within the voicemail system. No translations are necessary to send calls to the voicemail system or to enable a Message Waiting Indicator (MWI) within Cisco CallManager.

However, when users access the voicemail system from the PSTN, they need to be trained to enter their eight-digit extension to access their voicemail boxes.

Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing

You can implement a variable-length on-net dial plan with partitioned addressing by placing the abbreviated DN's for phones at each site in a different partition (one partition per site), and by relying on a set of translation patterns placed in a global partition (one translation pattern per site) to perform the inter-site call routing.

This method enables you to satisfy the key requirement of supporting abbreviated dialing (typically four or five digits) within a site and full E.164 dialing between sites, at the price of higher complexity in the dial plan.



Note

These deployments are also known as *overlapping dial plans* or *dial plans with overlapping extensions* because the abbreviated DN's defined at the various sites usually overlap with each other.

[Table 10-4](#) illustrates the basic relationship between calling search spaces and partitions at each site, without taking into account additional elements required to implement classes of service.

Table 10-4 Calling Search Spaces and Partitions for Variable-Length Dial Plans with Partitioned Addressing

Calling Search Space	Partition	Partition Contents
Site1_css	Site1Phones_pt	Site 1 phone DN's (abbreviated form)
	Site1PSTN_pt	PSTN route patterns for Site 1 (more partitions may be needed based on classes of service)
	Translations_pt	Translation patterns for inter-site calling within the cluster
...	...	...
SiteN_css	SiteNPhones_pt	Site N phone DN's (abbreviated form)
	SiteNPSTN_pt	PSTN route patterns for Site N (more partitions may be needed based on classes of service)
	Translations_pt	Translation patterns for inter-site calling within the cluster

Use this approach if one or more of the following conditions apply:

- A global on-net numbering plan using site codes is *not* desired



Note

While this approach can also be used in the presence of an on-net internal numbering plan using site codes, for this scenario Cisco recommends that you follow the flat addressing approach described in the section on [Deploying Variable-Length On-Net Dial Plans with Flat Addressing, page 10-54](#), because it enables you to greatly simplify the dial plan structure and to better manage and troubleshoot the system.

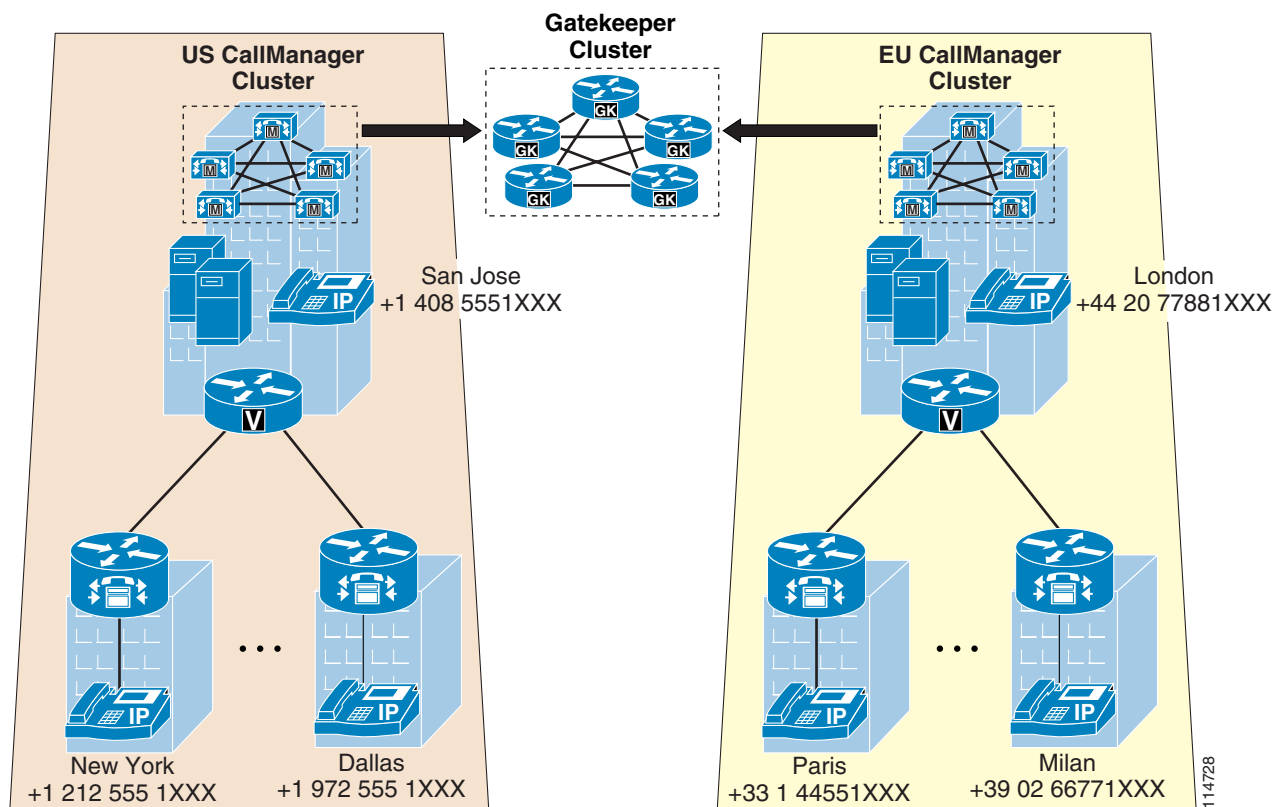
- Policy restrictions must be applied to on-net inter-site calls (that is, some or all users are not allowed to dial other sites on-net).
- Inter-site calls are always routed over the PSTN (as described in the section on [Voice Over the PSTN as a Variant of Centralized Call Processing, page 2-9](#)).
- CTI-based applications are not used across sites.

**Note**

Some CTI-based applications, as well as Cisco Attendant Console, do not support overlapping extensions and cannot be deployed with the partitioned addressing approach. Other applications, such as Cisco Personal Assistant and Cisco IP Manager Assistant (IPMA), require additional dial plan configuration that becomes more complex and might be impaired in presence of overlapping extensions. If you are planning to deploy these applications, Cisco recommends that you choose the flat addressing approach described in the next section.

To better understand how to deploy the partitioned addressing approach, consider the hypothetical customer network shown in [Figure 10-20](#). This network has a main site (San Jose) and a number of smaller branch sites (New York and Dallas) in the United States, and a similar topology with a main site (London) and several smaller branch sites (Paris and Milan) in Europe. Given the user count, the administration needs, and the network topology, two Cisco CallManager clusters with centralized call processing are deployed, one in the United States and one in Europe. A Cisco IOS gatekeeper cluster is used to provide E.164 address resolution and call admission control between the two clusters. It has also been decided that a variable-length on-net dial plan is required, with four-digit dialing within each site (the 1XXX extension range is chosen at each site) and full E.164 dialing between sites.

Figure 10-20 Example of a Large Multi-Site Deployment



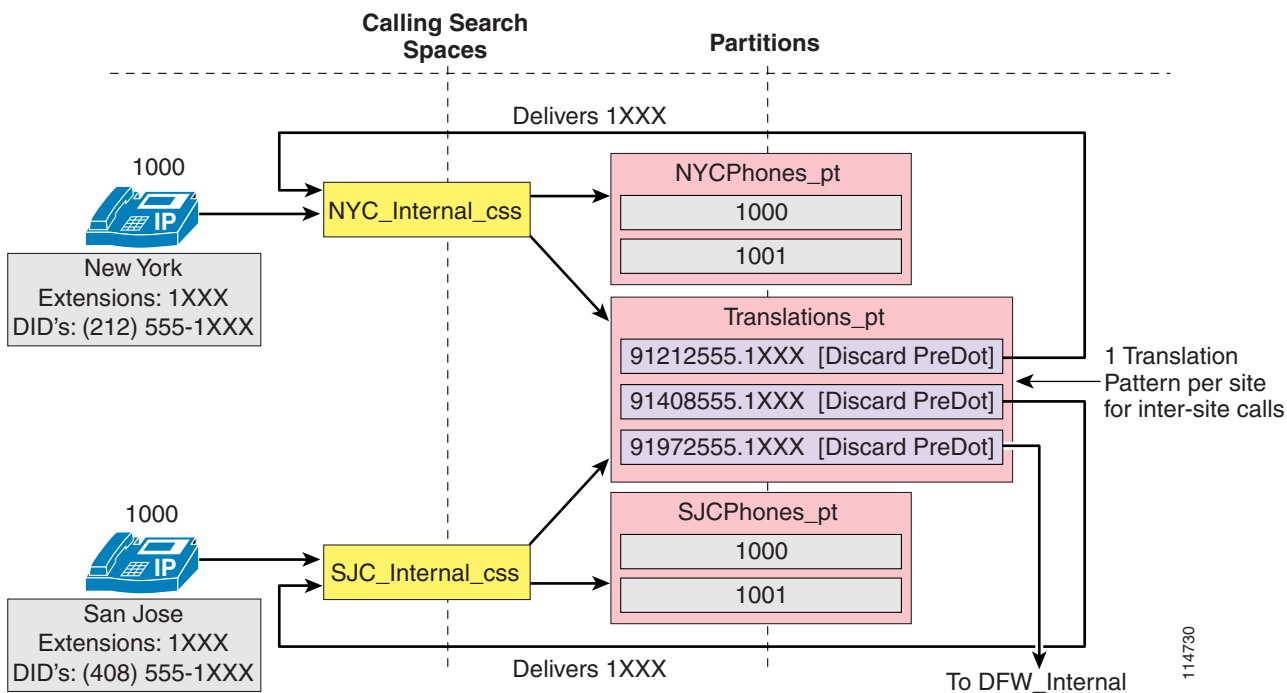
Using the United States (US) cluster from [Figure 10-20](#) as an example, the following sections analyze implementation details and best practices for different types of calls within the framework of the partitioned addressing approach:

- [Inter-Site Calls Within a Cluster, page 10-52](#)
- [Outgoing PSTN and IP WAN Calls, page 10-53](#)
- [Incoming Calls, page 10-54](#)
- [Voicemail Calls, page 10-54](#)

Inter-Site Calls Within a Cluster

[Figure 10-21](#) shows an example configuration for inter-site calls within the US cluster.

Figure 10-21 Inter-Site Calls Within a Cluster for the Partitioned Addressing Method



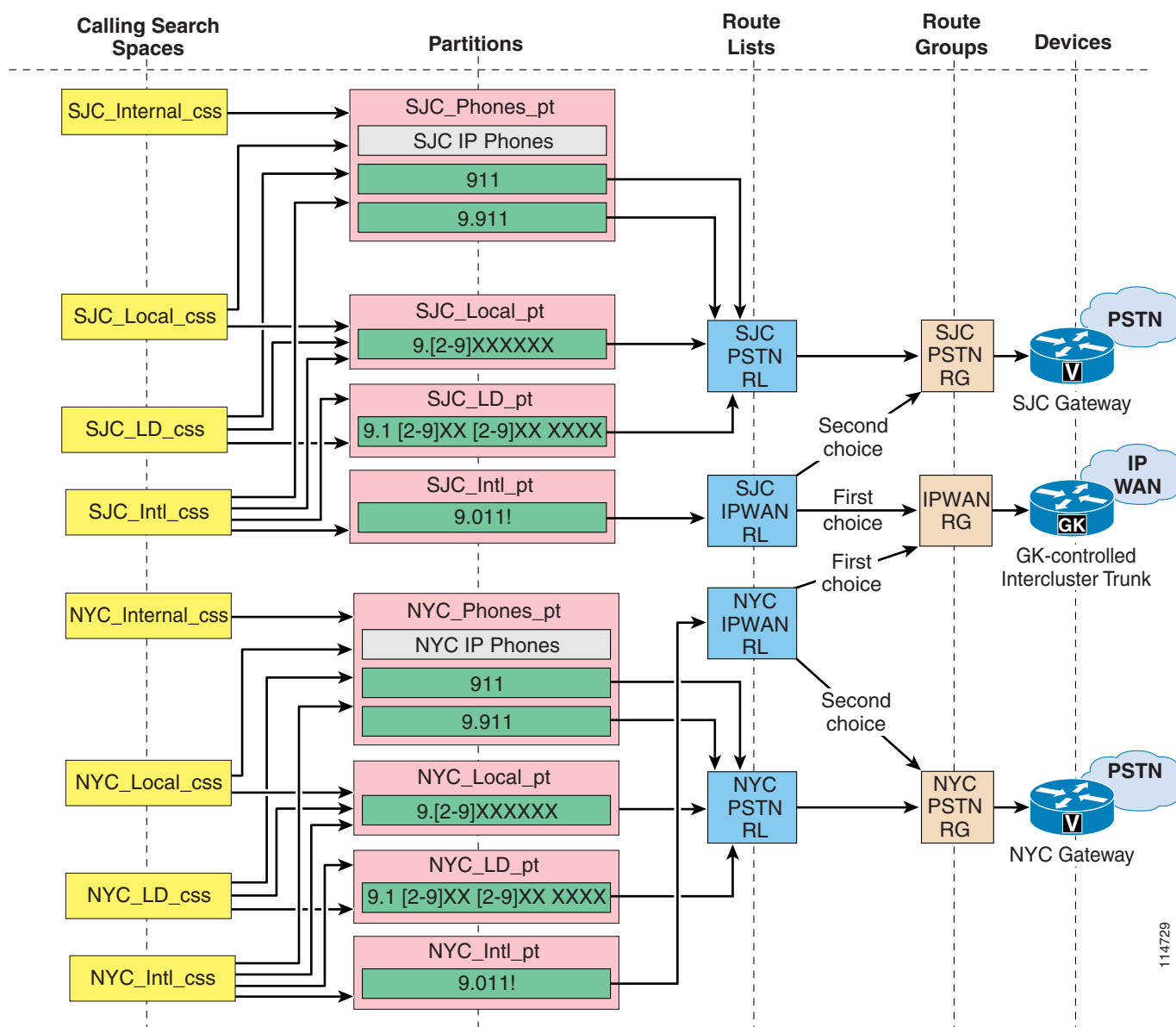
To provide connectivity between sites and partitions, use translation patterns according to the following guidelines:

- Define one translation pattern for each site, and place them all in the Translations_pt partition.
- Each pattern should match a site's E.164 address range, including the PSTN access code (9 in the example).
- The resulting called numbers after translation should match the site's extensions (1XXX in the example).
- The calling search space where the call is sent after translation should include the partition that contains the destination site's IP phone DNs.

Outgoing PSTN and IP WAN Calls

Depending on how the various types of PSTN calls need to be routed (centralized gateways versus distributed gateways), the configuration may vary. In the example shown in Figure 10-22, all domestic PSTN calls are routed via the local branch gateway. International calls are first routed through the gatekeeper to intercept calls toward the sites controlled by the Europe cluster, and then through the local PSTN gateway.

Figure 10-22 Outgoing PSTN and IP WAN Calls for the Partitioned Addressing Method.



As Figure 10-22 illustrates, for outgoing PSTN and IP WAN calls, there are no special considerations due to the partitioned nature of the internal addressing.

**Note**

Figure 10-22 uses the traditional approach for building classes of service. However, the line/device approach can also be combined with the partitioned addressing approach. Class-of-service approaches and endpoint addressing approaches are independent and interchangeable.

Incoming Calls

To dispatch incoming PSTN or IP WAN calls to the appropriate extension, you can reuse the translation patterns in the Translations_pt partition mentioned previously. It is sufficient to assign all PSTN gateways a calling search space that contains only the Translations_pt partition, making sure that the gateways also prefix a 9 or a 91 to the dialed number to match the already defined translation patterns.

Voicemail Calls

Voicemail integration requires special attention to the following requirements with the partitioned addressing approach:

- Voicemail boxes must have unique identifiers. This means that the IP phone extension cannot be used as a voicemail box, and some sort of digit manipulation is needed to obtain a unique number.
- Message waiting indicator (MWI) messages from the voicemail 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 voicemail system and to 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 reusing the translation patterns in the on-cluster partition. If the voicemail 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 voicemail system with the full E.164 number will be translated to the appropriate extension in the specific partition. For example, the voicemail ports can be configured with a calling search space that contains a VM_Translations_pt partition with a single translation pattern that prefixes 9 to the E.164 numbers dialed by the voicemail system. The calling search space of this translation pattern contains the Translations_pt partition, which then provides access to all the extensions via the previously defined translation patterns.

**Note**

This approach 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.

Deploying Variable-Length On-Net Dial Plans with Flat Addressing

You can implement a variable-length on-net dial plan with flat addressing by defining phone DNs as unique strings containing an on-net access code, a site code, and the extension (for example, 8-123-1000). You can place all these DNs in the same global partition, thus enabling inter-site calls using the site code, and you can define translation patterns in site-specific partitions (one translation pattern and one partition per site) to enable abbreviated dialing within a site.

This internal structure can be hidden from the end users by configuring the Line Text Label parameter within the Directory Number configuration page with the four or five-digit number that users are accustomed to dialing within the site. The external phone number mask should also be provisioned with the corresponding PSTN number in order to enable AAR and to give users a visual indication of their own DID number on the IP phone display.

[Table 10-5](#) illustrates the basic relationship between calling search spaces and partitions at each site, without taking into account additional elements required to implement classes of service:

Table 10-5 Calling Search Spaces and Partitions for Variable-Length Dial Plans with Flat Addressing

Calling Search Space	Partition	Partition Contents
Site1_css	Site1Translations_pt	Translation patterns for Site 1's abbreviated dialing
	Site1PSTN_pt	PSTN route patterns for Site 1 (more partitions may be needed based on classes of service)
	Internal_pt	All IP phone DNs (unique form)
...	...	...
SiteN_css	SiteNTranslations_pt	Translation patterns for Site N's abbreviated dialing
	SiteNPSTN_pt	PSTN route patterns for Site N (more partitions may be needed based on classes of service)
	Internal_pt	All IP phone DNs (unique form)

Use this approach if one or more of the following conditions apply:

- No dialing restrictions are needed for on-net inter-site calls.
- A global on-net numbering plan using site codes is desired.
- Inter-site calls are normally routed over the IP WAN.
- CTI-based applications are used across sites.



Note

If dialing restrictions need to be applied to on-net inter-site calls, or an on-net numbering plan using site codes is not desired, refer to the section on [Special Considerations for Deployments Without Site Codes, page 10-61](#), for a variant of this approach that can accommodate these needs.

The following considerations apply to this approach:

- The destination numbers of local four-digit calls get expanded to the unique internal DN on the IP phone display and in the Placed Calls directory.
- Calling number, and numbers in the Missed Calls and Received Calls directories, appear as the unique internal DN.
- If you wish to preserve the four-digit dialing feature when the IP WAN is unavailable and the branch phones are in SRST mode, you need to apply a translation rule to the **call-manager-fallback** configuration within the SRST router.
- When the branch phones are in SRST mode, the Line Text Label that masks the unique internal DN as a four-digit number on the IP phone display is not available, so the users will see their full internal DN appear instead.

To better understand how to deploy the partitioned addressing approach, consider again the hypothetical customer network shown in [Figure 10-20](#). In this case, it has been decided that a variable-length on-net dial plan is required, with four-digit dialing within each site (the 1XXX extension range is chosen at each site) and inter-site dialing with an eight-digit string consisting of an on-net access code (8 in this example), a three-digit site code, and the four-digit extension. The three-digit site code is derived from the NANP area code for the sites located in the United States, and from the E.164 country code followed by a site identifier for the sites located in Europe. [Table 10-6](#) summarizes the site codes chosen.

Table 10-6 Site Codes for the Customer Network in [Figure 10-20](#)

	San Jose	New York	Dallas	London	Paris	Milan
Site Code	408	212	972	442	331	392

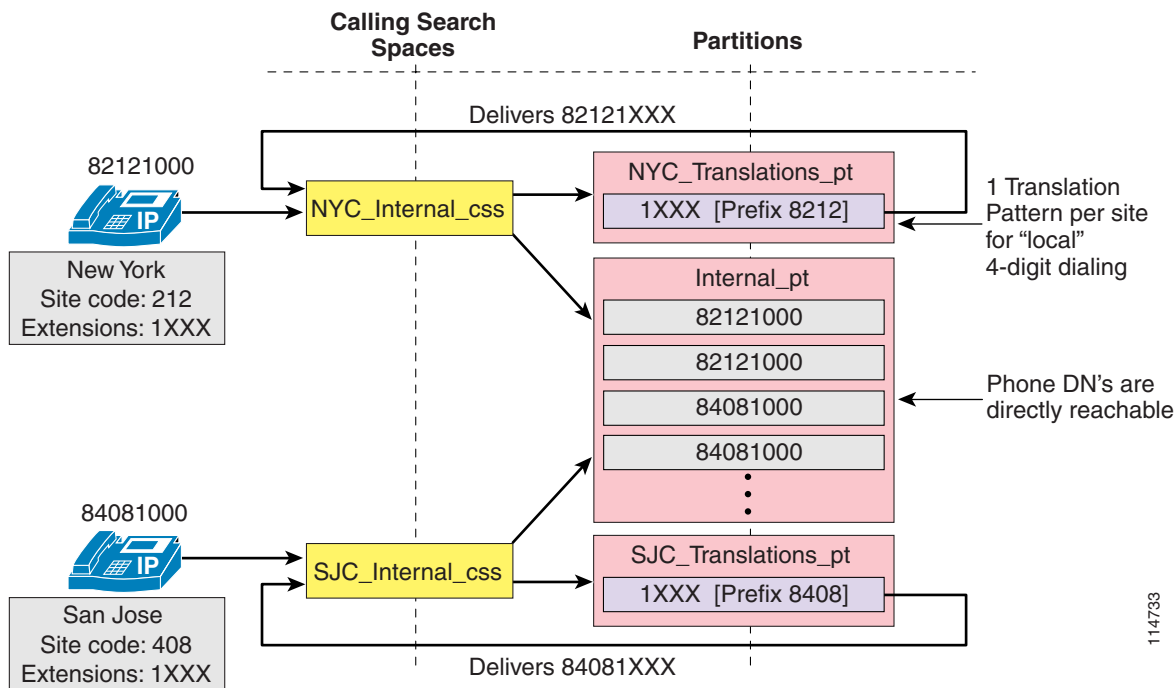
Using the US cluster from this example, the following sections analyze implementation details and best practices for different types of calls within the framework of the flat addressing approach:

- [Inter-Site Calls Within a Cluster, page 10-57](#)
- [Outgoing PSTN and IP WAN Calls, page 10-58](#)
- [Incoming Calls, page 10-61](#)
- [Voicemail Calls, page 10-61](#)
- [Special Considerations for Deployments Without Site Codes, page 10-61](#)

Inter-Site Calls Within a Cluster

Figure 10-23 shows an example configuration for inter-site calls within the US cluster.

Figure 10-23 Inter-Site Calls Within a Cluster for the Flat Addressing Method



To provide connectivity between sites and partitions, use the following guidelines:

- Place all unique DNs, including the on-net access code 8, in a global partition (named `Internal_pt` in this example).
- Create one partition per site, each containing a translation pattern that expands four-digit numbers into the fully qualified eight-digit number for that site, thus enabling abbreviated dialing within the site.
- For each site, include both the `Internal_pt` partition and the local translation partition in the phone's calling search space.

The inclusion of the on-net access code in the DN configured in Cisco CallManager enables you to place all internal extensions in a partition directly accessible by all phones, and at the same time ensures that all call directories on the IP phones are populated with numbers that can be directly redialed.

Outgoing PSTN and IP WAN Calls

Depending on how the various types of PSTN calls need to be routed (centralized gateways versus distributed gateways), the configuration may vary.

To provide on-net connectivity for inter-site calls to the Europe (EU) cluster, the following options are possible:

Option 1: Eight-Digit Numbers Only

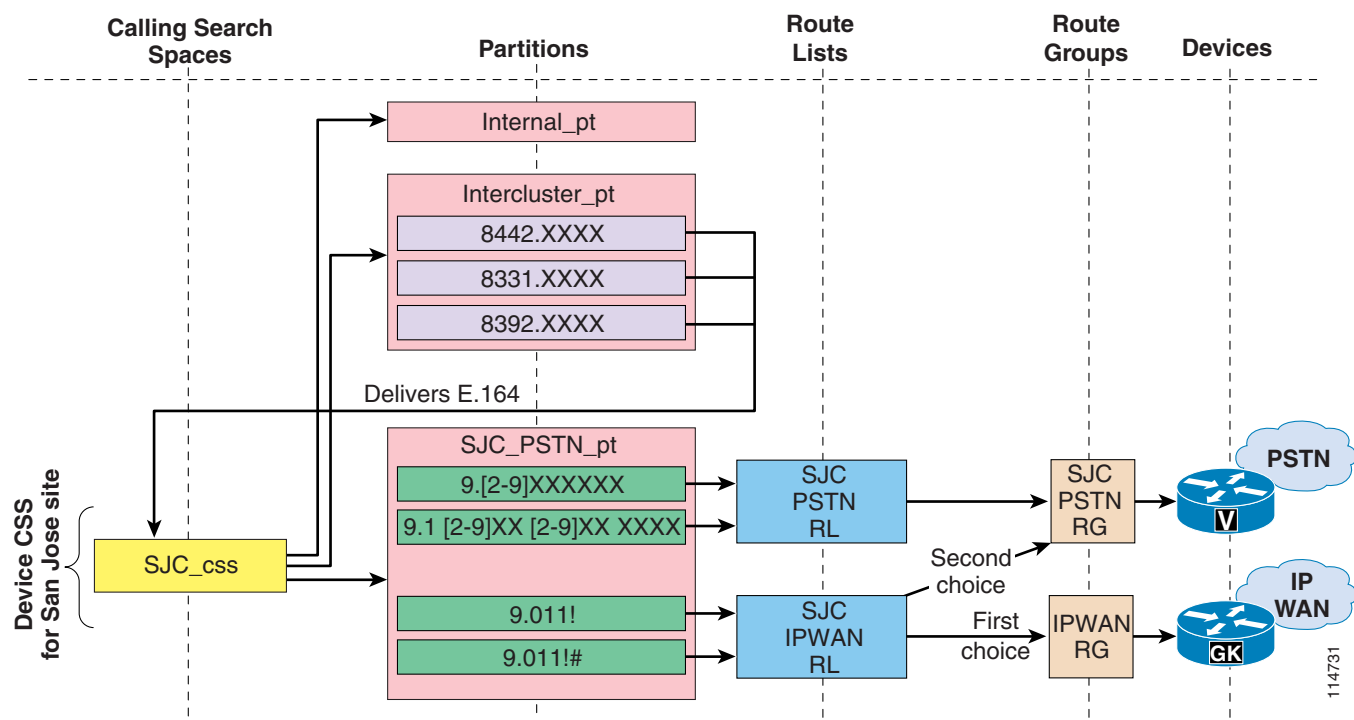
This option relies on a single route pattern that matches all eight-digit ranges (8XXXXXXX) and points to a route list or route group that contains only a gatekeeper-controlled intercluster trunk. The gatekeeper is configured to use the site codes as zone prefixes.

This solution is simple and easy to maintain because no information is needed about other clusters' site codes or E.164 ranges. However, no automatic PSTN failover is provided when the IP WAN is unavailable, and users are expected to redial manually using the PSTN access code and the E.164 address of the destination.

Option 2: Eight-Digit Numbers and E.164 Addresses with Centralized PSTN Failover

This option, illustrated in [Figure 10-24](#), uses a global set of translation patterns that match the Europe eight-digit ranges and translate them into the corresponding E.164 numbers. The translation patterns use the central site's calling search space (in this case, San Jose), and the call then matches the international PSTN route pattern within the central site's PSTN partition. At each site, the international PSTN route pattern points to a route list with the IP WAN route group as a first choice and the local PSTN route group as a second choice. The gatekeeper is configured to use the E.164 ranges as zone prefixes.

Figure 10-24 Outgoing PSTN and IP WAN Calls for the Flat Addressing Method with Centralized PSTN Failover for IP WAN Calls



**Note**

The configuration example in [Figure 10-24](#) assumes that the line/device approach to building classes of service is being used, but the same considerations apply when using the traditional approach.

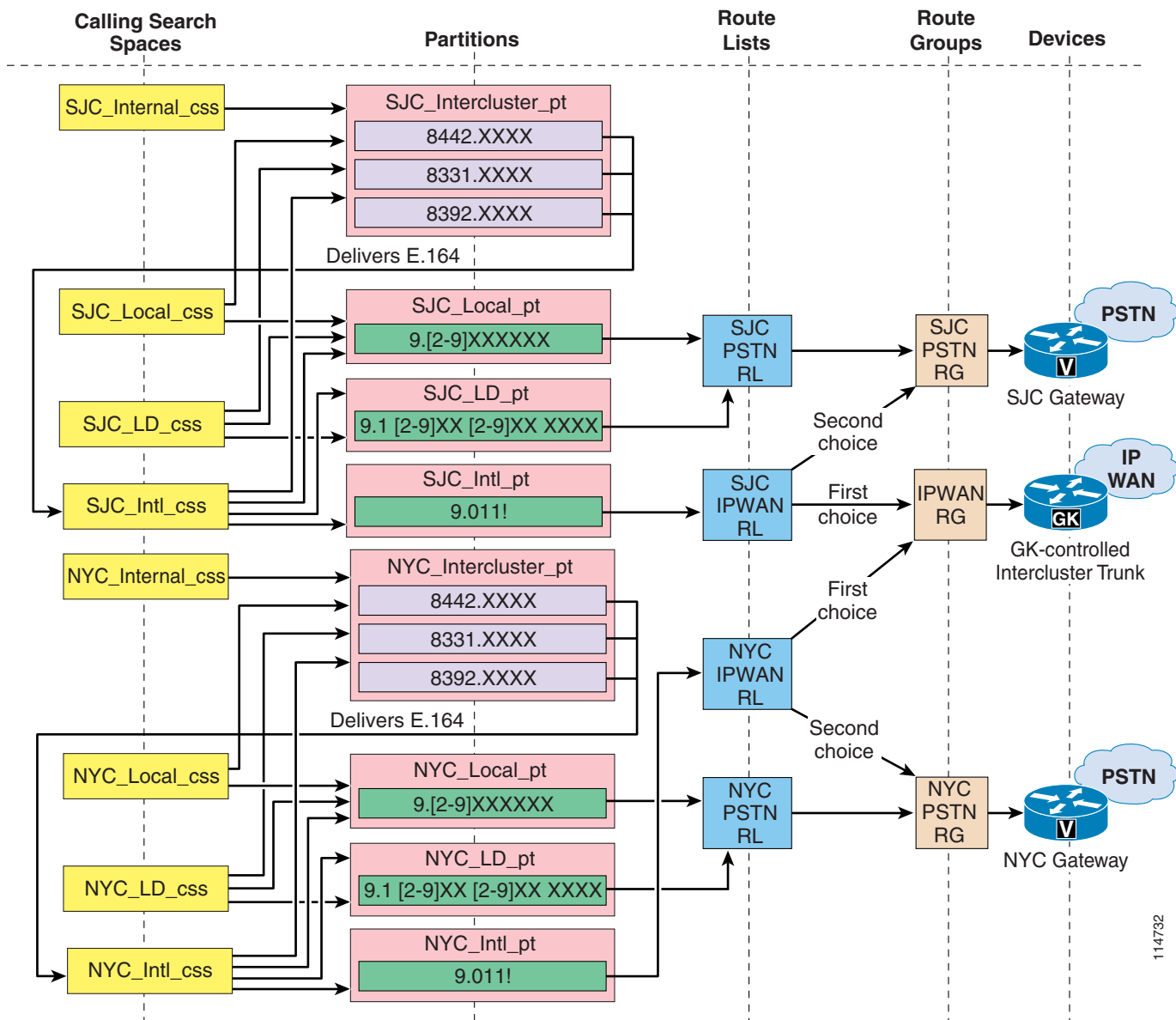
This solution requires a little more configuration and maintenance than that outlined in Option 1 because it requires that you configure and maintain information about other clusters' site codes and E.164 ranges. On the other hand, it provides automatic PSTN failover when the IP WAN is unavailable. PSTN failover is provided using the central site's gateway only, so the IP WAN bandwidth usage is not optimal.

Also observe that calls to the Europe sites dialed as PSTN calls will automatically be placed on-net if the IP WAN is available, with an automatic PSTN failover that in this case uses the local gateway.

Option 3: Eight-Digit Numbers and E.164 Addresses with Distributed PSTN Failover

This option, illustrated in [Figure 10-25](#), uses a set of translation patterns per site. Each set matches the Europe eight-digit ranges and translates them into the corresponding E.164 numbers. The translation patterns use the originating site's calling search space, and the call then matches the international PSTN route pattern within the originating site's PSTN partition. At each site, the international PSTN route pattern points to a route list with the IP WAN route group as a first choice and the local PSTN route group as a second choice. The gatekeeper is configured to use the E.164 ranges as zone prefixes.

Figure 10-25 Outgoing PSTN and IP WAN Calls for the Flat Addressing Method with Distributed PSTN Failover for IP WAN Calls



114732

This solution requires a lot more configuration and maintenance than that outlined in Option 2 because it requires that you configure and maintain information about other clusters' site codes and E.164 ranges, and that you repeat the configuration for each remote site within the cluster. On the other hand, it provides automatic PSTN failover when the IP WAN is unavailable, using the local site's gateway so that the IP WAN bandwidth usage is optimal.

Also in this case, as for Option 2, calls to the Europe sites dialed as PSTN calls will automatically be placed on-net if the IP WAN is available, with an automatic PSTN failover using the local gateway.

Incoming Calls

Incoming PSTN calls require that the E.164 number be manipulated to obtain the eight-digit internal number in order to reach the destination phone. You can implement this requirement in any of the following ways:

- Configure the Num Digits and Prefix Digits fields within the Gateway Configuration page in Cisco CallManager to strip and then prefix the needed digits.
- If you have configured translation patterns to force on-net inter-site calls within the cluster, you can reuse these patterns by simply prefixing the PSTN access code to the incoming called number on the gateway.
- If you are using an H.323 gateway, you can use translation rules within the gateway to manipulate the digits before sending the call to Cisco CallManager.

The third approach has the advantage that the translation rules you configured can be reused to provide incoming PSTN connectivity to the IP phones when the branch is in SRST mode.

Voicemail Calls

Because every eight-digit extension is unique within the system, the extension itself can be used to configure voicemail boxes within the voicemail system. No translations are necessary to send calls to the voicemail system or to enable Message Waiting Indicators (MWIs) within Cisco CallManager.

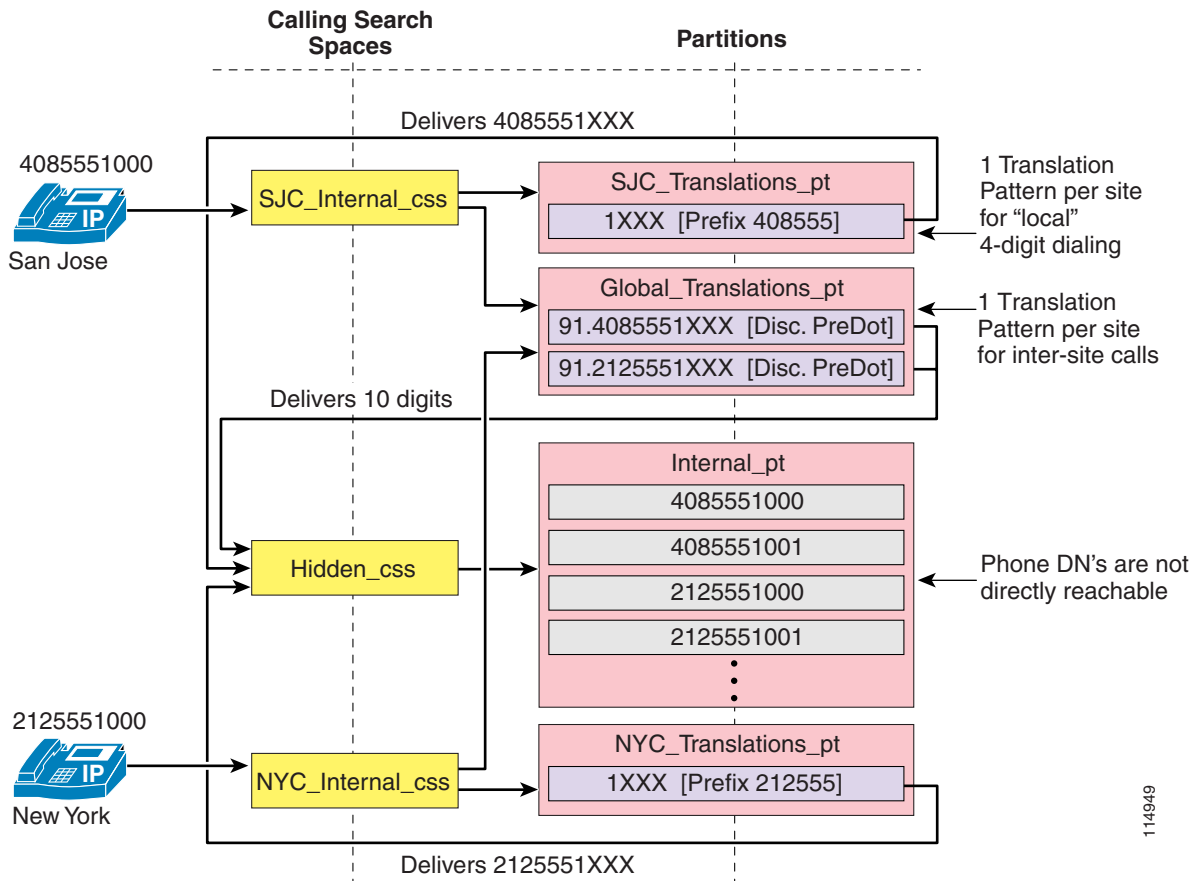
Special Considerations for Deployments Without Site Codes

This scenario is a variant of the flat addressing approach that does not rely on the definition of an on-net numbering plan based on site codes. In this scenario, intra-site calls are still dialed as four-digit numbers, while inter-site calls are dialed as regular PSTN calls and are then intercepted and routed across the IP WAN by Cisco CallManager.

To implement this mechanism, follow these guidelines, illustrated in [Figure 10-26](#):

- Define the phone DNs as the full E.164 addresses and place them all in the same partition (named Internal_pt in this example).
- Configure the phones' calling search spaces so that they do not have direct access to the Internal_pt partition. Instead, they point to a global partition named Global_Translations_pt, which contains one translation pattern per site or DID range.
- Configure the translation patterns to strip the PSTN access code (for example, 9) and send the call to the Internal_pt partition via the Hidden_css calling search space.
- Abbreviated four-digit dialing within each site can still be provided via site-specific partitions containing translation patterns (one partition and one translation pattern per site or DID range).

Figure 10-26 Variable-Length Dial Plans with Flat Addressing Without Site Codes



This configuration variant allows you to establish dialing restrictions for inter-site calls. For example, if you must prevent a certain group of users from calling other sites, configure their calling search space so that it does not contain the Global_Translations_pt partition. However, keep in mind the following factors when choosing this approach:

- Additional complexity is introduced in the form of translation patterns.
- Because you are effectively forcing on-net PSTN calls, remember to configure AAR so that calls can still be placed across the PSTN when the IP WAN bandwidth is not sufficient. Refer to [Design Guidelines for Multi-Site Deployments, page 10-44](#), for details.
- The placed calls, missed calls, and received calls directories display the unique phone DNs. Because these DNs are not directly reachable, users will not be able to redial numbers directly from these directories.
- The configuration shown in [Figure 10-26](#) assumes that PSTN calls are dialed in the same way in all sites, which is usually true for deployments contained within a single country. International deployments require one additional set of translation patterns per country in order to intercept inter-site calls.
- International deployments also introduce the complexity of having to deal with variable-length internal DNs because E.164 addresses have different lengths in different countries (and even within a single country in some cases).

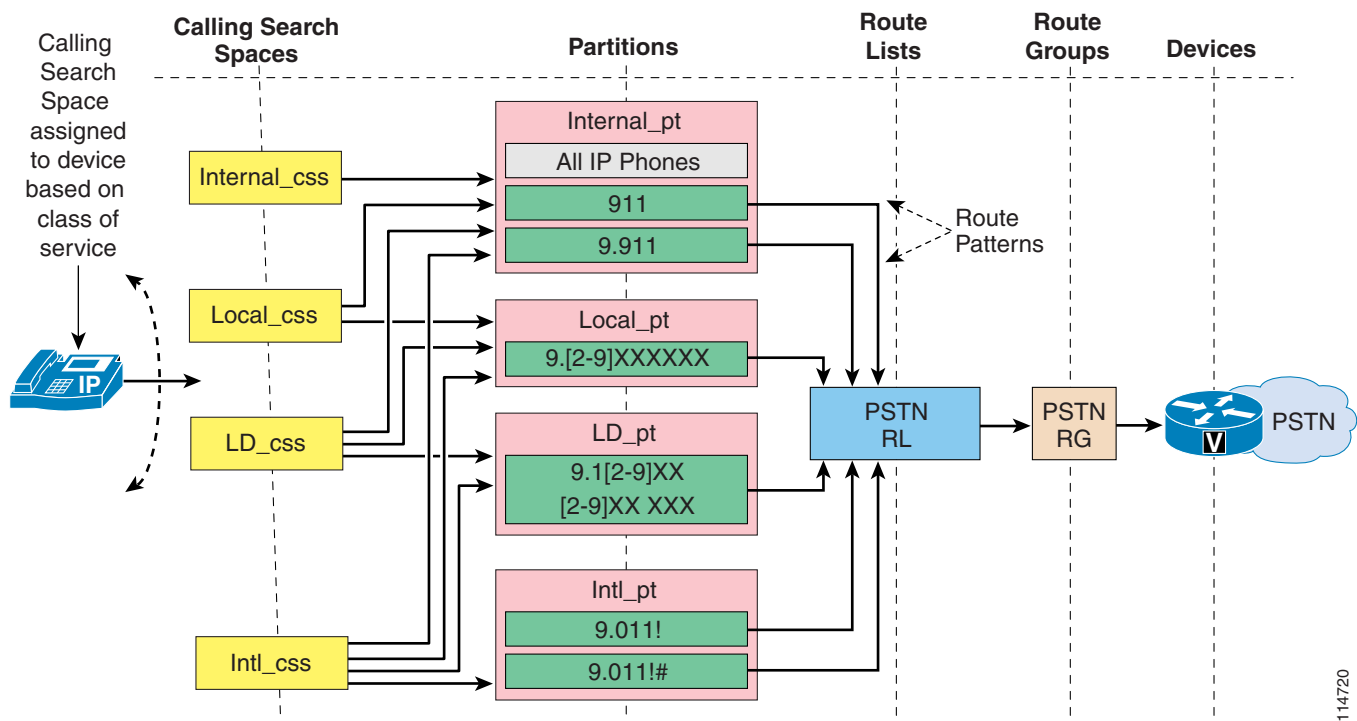
Building Classes of Service for Cisco CallManager with the Traditional Approach

With Cisco CallManager, you can define classes of service for IP Telephony users by combining partitions and device 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. Cisco recommends that you group route patterns in partitions according to the reachability policies for the various classes of service.
- 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 internal and local calls.
- Assign these calling search spaces to the phones by configuring them on the Cisco CallManager device pages. In this way, all lines configured on the device automatically receive the same class of service.

Figure 10-27 shows a simple example for a single-site deployment.

Figure 10-27 Basic Example of Classes of Service Using the Traditional Approach



114720

With this approach, the device calling search space performs two distinct logical functions:

- Path selection

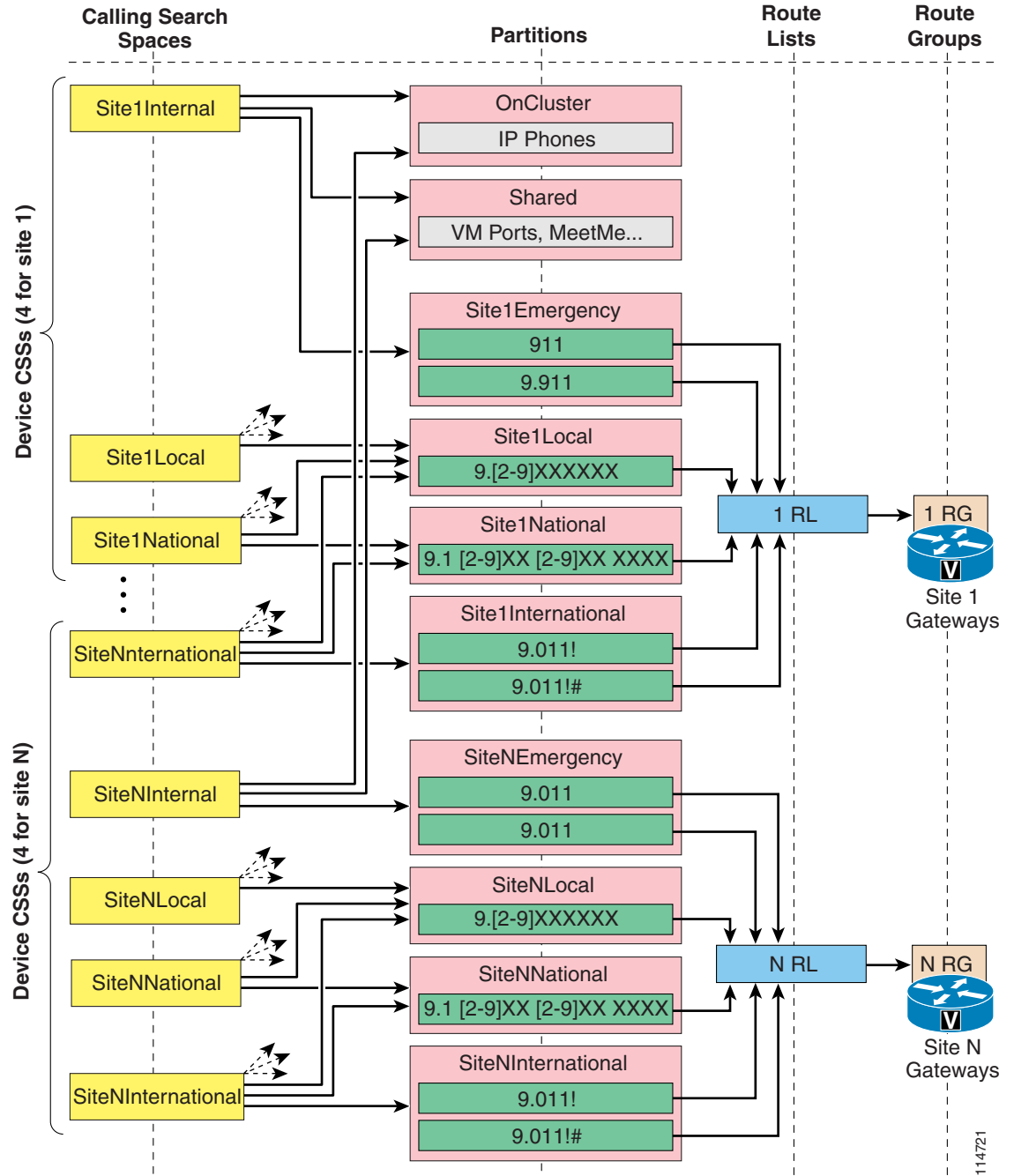
The calling search space contains specific partitions, which in turn contain specific route patterns that point to specific PSTN gateways.

- Class of service

By selectively including certain partitions and not others in the device calling search space, you effectively apply calling restrictions to certain groups of users.

As a consequence, when you apply this approach to a multi-site deployment with centralized call processing, you have to replicate partitions and calling search spaces for each site because for each site you have to create classes of service and, at the same time, route the PSTN calls out of the local branch gateways, as illustrated in [Figure 10-28](#).

Figure 10-28 Calling Search Spaces and Partitions Needed with the Traditional Approach



114721

Follow these additional guidelines when applying this dial plan approach to a multi-site deployment with centralized call processing:

- To facilitate on-net dialing between sites, place all IP phone DNs in an on-cluster or internal partition that can be reached from the calling search spaces of all sites. Note that this is not possible if the IP phone DNs overlap. For more details on dial plans that use overlapping extensions, refer to [Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing, page 10-50](#).
- 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.
- Depending on your company's call forwarding restriction policy, you can reuse one of the site-specific device calling search spaces for the Forward All calling search space.

As a general rule, use the following formulas to calculate the total number of calling search spaces and partitions needed:

Total Partitions = (Number of classes of service) * (Number of sites) + (1 Partition for all IP phone DNs)

Total Calling Search Spaces = (Number of classes of service) * (Number of sites)

**Note**

These values represent the minimum number of partitions and calling search spaces required. You may need additional partitions and calling search spaces for special devices and applications, as well as for on-net patterns for other call processing agents.

Extension Mobility Considerations with the Traditional Approach

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).

With the traditional approach for building classes of service, consider the following guidelines to apply calling restrictions when using Extension Mobility:

- At each site, configure the device calling search space for all IP phones to point to only PSTN emergency services (using the local gateway).
- Configure the line calling search spaces for IP phones used for Extension Mobility in a logged-out state to point to internal numbers only.
- For each Extension Mobility user, configure the line calling search space within the device profile to point to internal numbers and the additional PSTN route patterns allowed for their specific class of service (again, using an appropriate gateway according to company policy).

Note that, when an Extension Mobility user who is normally based in Site 1 logs into an IP phone in Site 2, the path selection for PSTN calls will change as follows:

- Emergency calls will be correctly routed using Site 2's PSTN gateway because the emergency services are provided by the device calling search space of the IP phone at Site 2.
- All other PSTN calls will be routed according to the Extension Mobility user's profile (more specifically, the line calling search space configured in the device profile). Typically, this means that these PSTN calls will traverse two WAN links and use Site 1's gateway to access the PSTN.

You can use one of the following methods to modify this behavior and ensure that PSTN calls are always routed via the local PSTN gateway even when Extension Mobility users roam across different sites:

- Include local PSTN route patterns in the device calling search space and remove them from the line calling search space within the device profile. This method ensures that local PSTN calls will be routed via the co-located branch gateway, but it also means that users will be able to dial these calls even without logging into the IP phones. Long-distance and international calls will still be routed according to the Extension Mobility user's device profile, so this solution is suitable only if these calls are usually routed via a centralized gateway.
- Define multiple device profiles for each user, one for each site to which they usually roam. Each device profile is configured so that its line calling search space points to PSTN route patterns that use the local gateway for that site. This method might prove burdensome to configure and manage if a significant number of users roam to a significant number of sites.
- Implement the line/device approach described in the next section on [Building Classes of Service for Cisco CallManager with the Line/Device Approach](#), page 10-67.

Building Classes of Service for Cisco CallManager with the Line/Device Approach

The traditional approach outlined in the preceding section can result in a large number of partitions and calling search spaces when applied to large multi-site deployments with centralized call processing. This configuration is required because the device calling search space is used to determine both the path selection (which PSTN gateway to use for external calls) and the class of service.

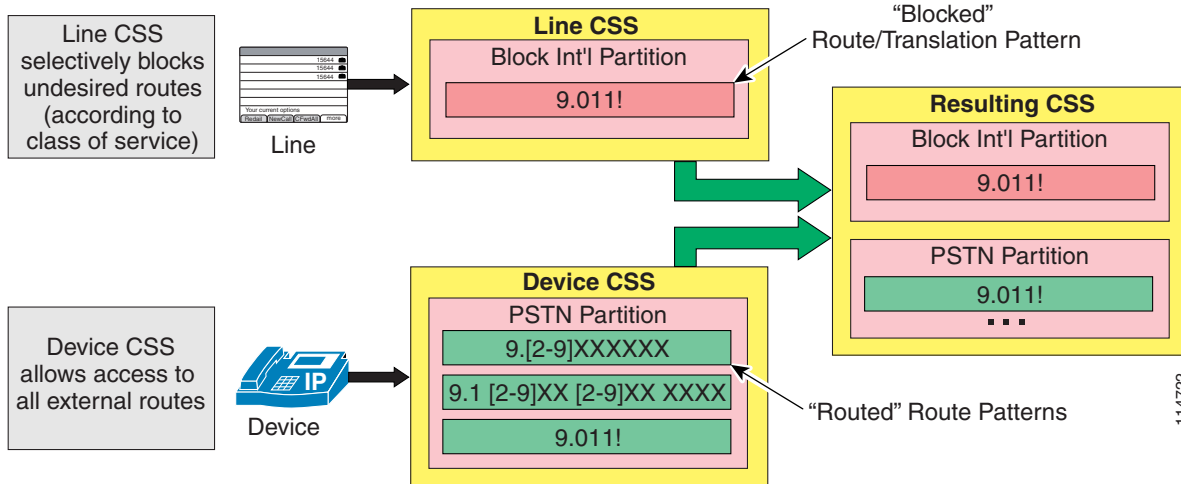
It is possible to significantly decrease the total number of partitions and calling search spaces needed by dividing these two functions between the line calling search space and the device calling search space, in what is called the *line/device approach*.

Keeping in mind how the line calling search space and the device calling search space for each given IP phone are combined within Cisco CallManager, and how the line calling search space partitions appear first in the resulting calling search space (see [Calling Privileges in Cisco CallManager](#), page 10-14), you can apply the following general rules to the line/device approach:

- Use the device calling search space to provide call routing information (for example, which gateway to select for PSTN calls).
- Use the line calling search space to provide class-of-service information (for example, which calls to allow).

To better understand how to apply these rules, consider the example shown in [Figure 10-29](#), where the device calling search space contains a partition with route patterns to all PSTN numbers, including international numbers. The route patterns point to a PSTN gateway via the route list and route group construct.

Figure 10-29 Key Concepts in the Line/Device Approach



At the same time, the line calling search space contains a partition with a single translation pattern that matches international numbers and that has been configured as a blocked pattern.

The resulting calling search space therefore contains two identical patterns matching international numbers, with the blocked pattern in the line calling search space appearing first. The result is that international calls from this line will be blocked.

It is possible to use route patterns instead of translation patterns to block calls within the line calling search space. To configure a blocked route pattern, first create a "dummy" gateway with an unused IP address and place it into a "dummy" route list and route group construct. Then point the route pattern to the dummy route list. The main difference between using a route pattern and a translation pattern to block calls is the end-user experience when trying to dial a blocked number, as follows:

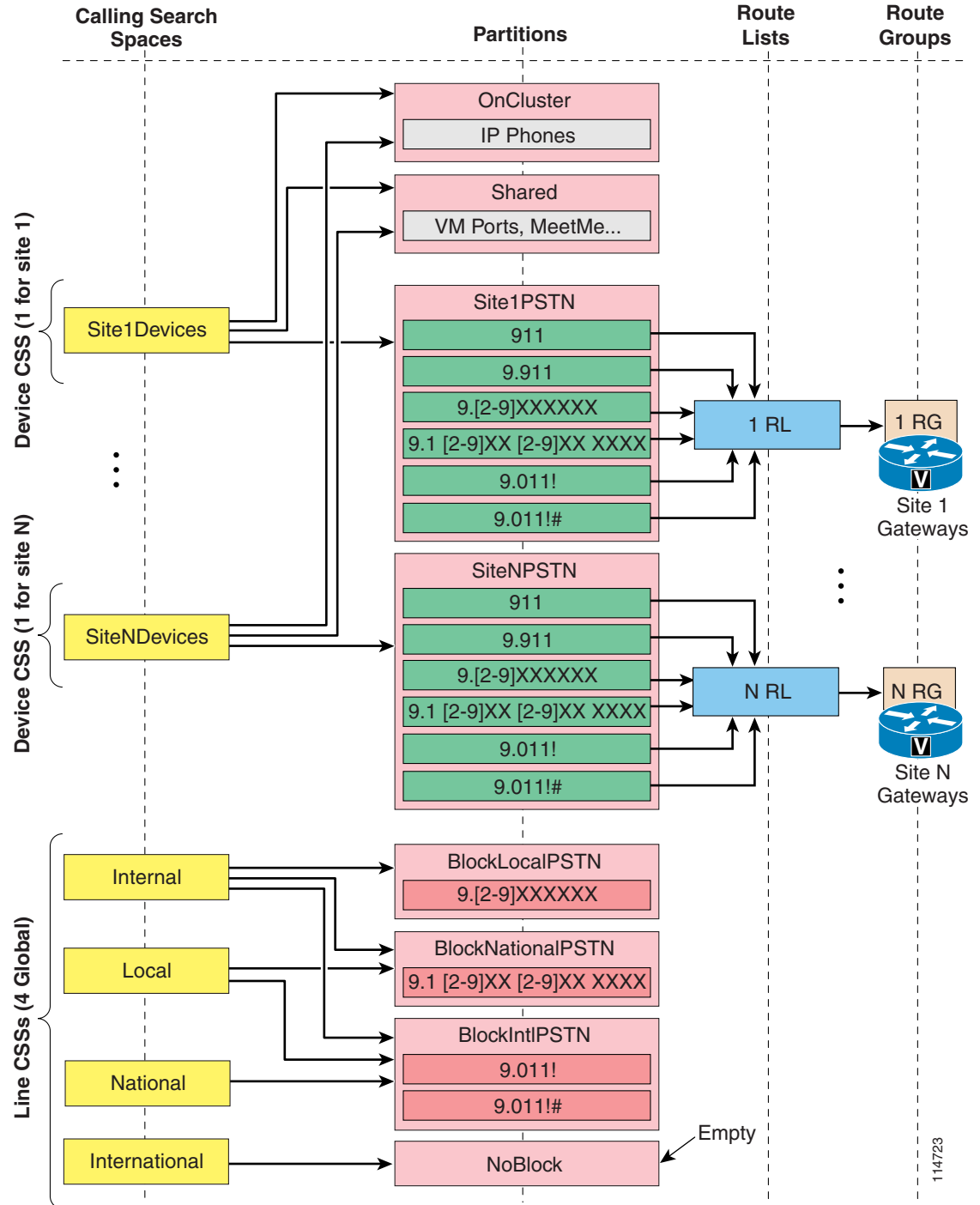
- When using a translation pattern, the end users will be able to dial the entire number and only then will they hear a fast-busy tone.
- When using a route pattern, the end users will hear a fast-busy tone as soon as the number they are dialing can no longer match any allowed pattern.

Follow these guidelines to implement the line/device approach in a multi-site deployment with centralized call processing:

- Create an unrestricted calling search space for each site and assign it to the phone's device calling search space. This calling search space should contain a partition 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 translation/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 that blocks the 9.011! route pattern.

Figure 10-30 shows an example of how these guidelines can apply to a multi-site deployment with N sites.

Figure 10-30 Calling Search Spaces and Partitions Needed with the Line/Device Approach



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

Consequently, you can use the following formulas to calculate the total number of calling search spaces and partitions needed:

Total Partitions = (Number of classes of service) + (Number of sites) + (1 Partition for all IP phone DNs)

Total Calling Search Spaces = (Number of classes of service) + (Number of sites)



Note

These values represent the minimum numbers of partitions and calling search spaces required. You may need additional partitions and calling search spaces for special devices and applications, as well as for on-net patterns for other call processing agents.

When applied to centralized call processing deployments with large numbers of sites, the line/device approach drastically reduces the number of partitions and calling search spaces needed. For example, a deployment with 100 remote sites and 4 classes of service requires at least 401 partitions and 400 calling search spaces with the traditional approach but only 105 partitions and 104 calling search spaces with the line/device approach.

However, the line/device approach relies on the fact that you can globally identify the types of PSTN calls that must be restricted for certain classes of service (for example, local, long-distance, and international calls). If the national numbering plan of your country does not allow this global identification of the different types of calls, the efficiency of this approach (in terms of configuration savings) is lower than that indicated in the formulas above.

For example, in France the numbering plan is based on five area codes, from 01 to 05 (plus the 06 area code for cellular phones), followed by eight digits for the subscriber number. The key characteristic is that each PSTN destination is reached by dialing exactly the same number (for example, 01XXXXXXXX for Paris numbers, 04XXXXXXXX for Nice numbers, and so on), whether calling from the same local area or from a different area. This means that it is not possible to block access to long-distance calls with a single partition and a single route pattern because the concept of "long-distance call" changes depending on the area where the calling party is located. (For example, a call to 014455667788 is a local call if the caller is in Paris but a long-distance call if the caller is in Nice or Lyon.)

In such cases, you will have to configure additional sets of blocking calling search spaces and partitions, one for each area within which local and long distance calls are dialed the same way. In the example of France, you would have to define five additional blocking calling search spaces and partitions, one for each area code, as shown in [Table 10-7](#):

Table 10-7 The Line/Device Approach Applied to the French National Numbering Plan

Calling Search Space	Partition	Blocked Route Patterns
Internal_css	BlockAllNational_pt	0.0[1-6]XXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
Local01_css	BlockLD01_pt	0.0[2-6]XXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
Local02_css	BlockLD02_pt	0.0[13-6]XXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
Local03_css	BlockLD03_pt	0.0[124-6]XXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#

Table 10-7 The Line/Device Approach Applied to the French National Numbering Plan (continued)

Calling Search Space	Partition	Blocked Route Patterns
Local04_css	BlockLD04_pt	0.0[1-356]XXXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
Local05_css	BlockLD05_pt	0.0[1-46]XXXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
LD_css	BlockIntl_pt	0.00!, 0.00!#
Intl_css	NoBlock_pt	none

Guidelines for the Line/Device Approach

Consider the following guidelines when using the line/device approach:

- Call Forward calling search spaces (Forward Busy, Forward No Answer, and Forward All) are not concatenated with the line or device calling search space.
- Set the Forward Busy and Forward No Answer calling search spaces to a global calling search space that can reach the voicemail pilot number and ports.
- Set the Forward All calling search space according to your company's policy:
 - If forwarded calls must have unrestricted privileges, set the Forward All calling search space to match the site-specific device calling search space. (For additional considerations with Extension Mobility, see [Extension Mobility Considerations for the Line/Device Approach, page 10-72](#).)
 - If forwarded calls must be restricted to internal numbers only, set the Forward All calling search space to a global calling search space that can reach only internal numbers.
 - If forwarded calls must have some intermediate restriction (such as access to local PSTN numbers but not international numbers), the line/device approach might become less efficient as additional site-specific calling search spaces are required. In such cases, the traditional approach may be preferable.
- Remember that, for this approach to work, the blocked patterns configured within the line calling search space must be at least as specific as the route patterns configured within the device calling search space. Wherever possible, Cisco recommends that you configure the blocked patterns as more specific than the routed ones, to avoid any possibility of error. Use extra care when dealing with the @ wildcard because the patterns defined within it are very specific.
- AAR is triggered when on-net DN's are dialed. Access to these DN's can be controlled by the same processes described previously. AAR uses a different calling search space for rerouted calls. In most cases, the AAR calling search space can be the same as the site-specific, unrestricted device calling search space because it can never be dialed directly by end users.



Note

The priority order between line and device is reversed for CTI devices (CTI route points and CTI ports). For these devices, the partitions in the device calling search space are placed before the line calling search space in the resulting calling search space. Therefore, the line/device approach cannot be applied to CTI devices such as Cisco IP SoftPhone.

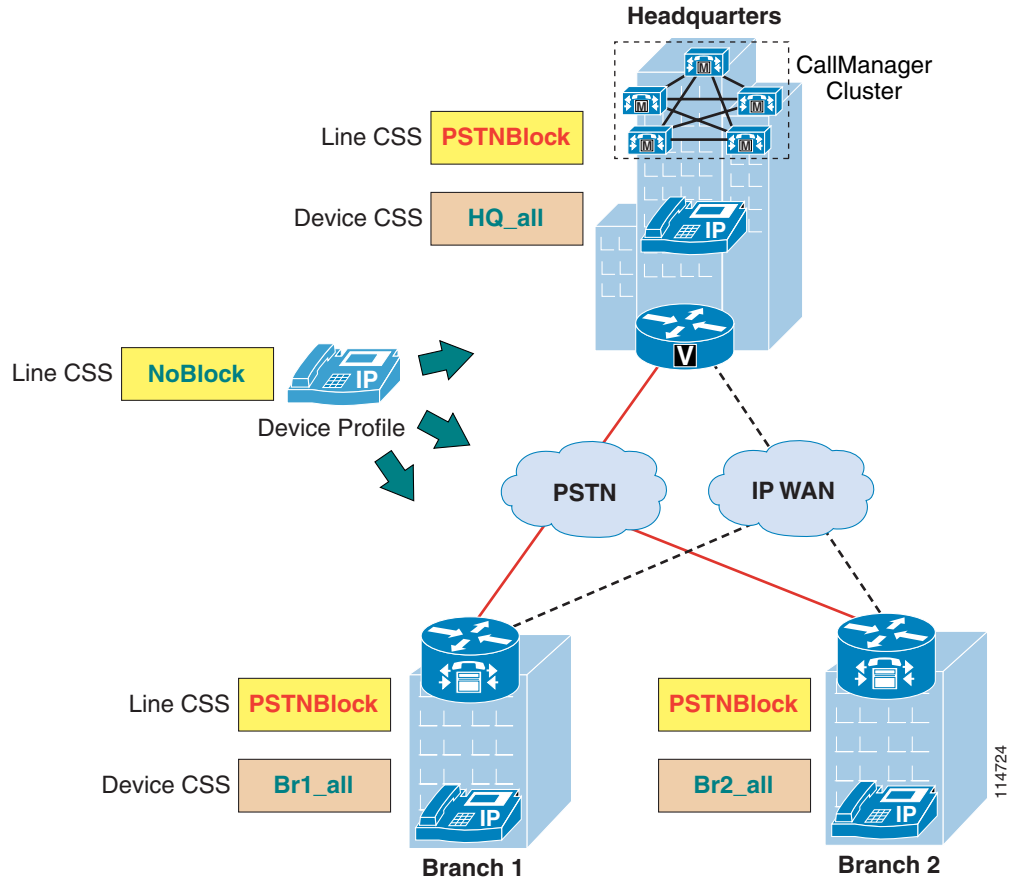
Extension Mobility Considerations for the Line/Device Approach

When using the Extension Mobility feature, the line/device approach provides a natural way to implement the dialing restrictions of a phone as 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).

With the line/device approach for building classes of service, you can simply apply the same rules described in the previous section to the Extension Mobility device profile construct. Consider the following guidelines to apply calling restrictions when using Extension Mobility:

- At each site, configure the device calling search space for all IP phones to point to a site-specific partition that contains all possible PSTN route patterns and that routes them appropriately (for example, using the local gateway for emergency and local calls and a centralized gateway for long distance calls).
- Configure the line calling search space for all IP phones (or the line calling search space for the default logout device profile) to point to a global calling search space featuring blocked translation/route patterns that block all calls except those to be allowed when no user is logged in (for example, internal extensions and emergency services).
- For each Extension Mobility user, configure the line calling search space within the device profile to point to a global calling search space featuring blocked translation/route patterns to selectively block the PSTN calls that are not to be allowed for their specific class of service (for example, block only international calls). If some users must have unrestricted calling privileges, assign them to a line calling search space featuring an empty partition.

The key advantage of using the line/device approach for extension mobility is that, in a multi-site deployment with centralized call processing, appropriate call routing is ensured even when users log in to IP phones located at branch sites different from their home site, as illustrated in [Figure 10-31](#).

Figure 10-31 Extension Mobility with the Line/Device Approach

As described previously in this chapter, the line calling search space configured within the device profile replaces the line calling search space configured on the physical IP phone when a user logs in through Extension Mobility. Because the call routing is correctly determined by the device calling search space, the login operation is used merely to "unlock" the phone by removing the phone's line calling search space (which contains blocked patterns) and replacing it with the device profile's line calling search space (which does not contain blocked patterns in this simplified example).

Because all the call routing is done within the device calling search space while the line calling search space only introduces blocked patterns, whenever a user logs in at a different site from their home site, they will automatically inherit all the local dialing habits for that site. For example, assume that phone DN's are defined as eight-digit numbers, but four-digit dialing is provided within each site by local translation patterns. In this case, a user roaming to a different site will not be able to dial a colleague at the home site by using only four digits because the four-digit number will now be translated according to the rules of the host site where the user logged in.

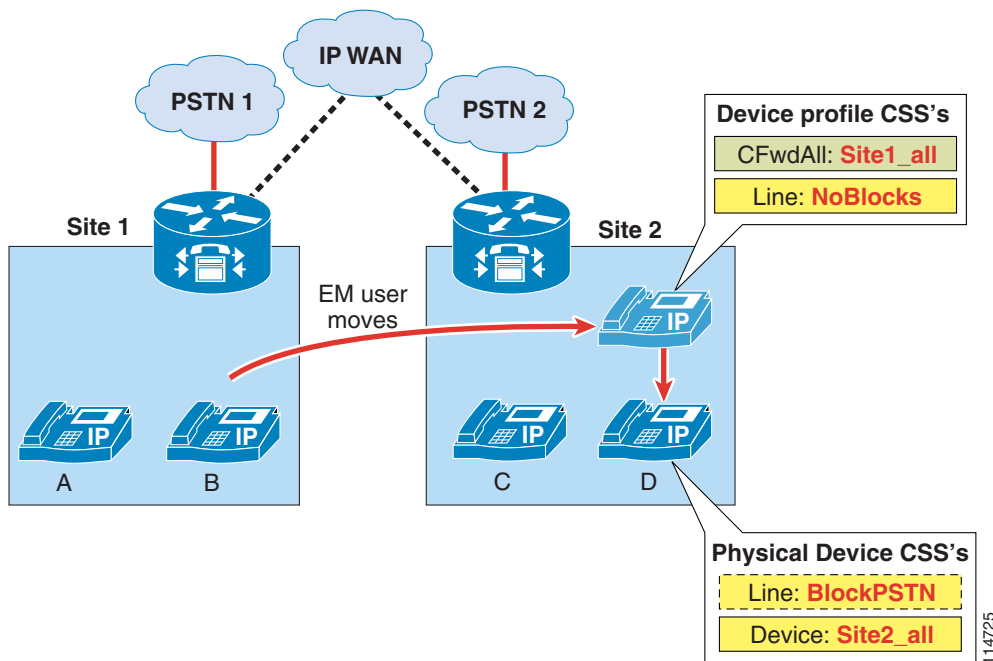
In summary, when you use the line/device approach for Extension Mobility, end-users have to adopt the dialing behavior of the site where they logged in.

Call Forwarding Considerations

When applying the Line/Device calling search space approach to a centralized call processing environment with Extension Mobility, you should be aware of the call forwarding behavior if users need to be allowed to forward all their calls to external PSTN numbers.

In [Figure 10-32](#), an Extension Mobility user is normally based in Site 1, with a device profile allows the user to place unrestricted PSTN calls and to forward all incoming calls to any PSTN number.

Figure 10-32 Call Forwarding Considerations for Extension Mobility with the Line/Device Approach



As described in the section on [Call-Forward Calling Search Spaces](#), [page 10-17](#), the Forward All calling search space is not concatenated with the line or the device calling search spaces and therefore needs to be set to Site1_all, which includes all PSTN routes using the Site 1 gateway.

When this user moves to Site 2 and logs into phone D, the user's device profile applies its line calling search space and Forward All calling search space to the physical device. For direct PSTN calls, the calling search space used is the concatenation of the line and device calling search space, and phone D's device calling search space (Site2_all) correctly points to the Site 2 gateway.

If the user now configures the phone to forward all calls to a PSTN number, any forwarded call will use the Site1_all calling search space, which still points to the Site 1 gateway. This condition results in the following behavior:

- Incoming PSTN calls enter the IP network at the Site 1 gateway and are hairpinned back into the PSTN within the same gateway.
- Calls originating from Site 1 phones (such as phone A) are correctly forwarded to the PSTN via the Site 1 gateway.
- Calls originating from Site 2 phones (such as phone C) traverse the WAN to Site 1 and access the PSTN via the Site 1 gateway. The same behavior applies to calls originating from other sites within the same Cisco CallManager cluster.

Keep this behavior in mind when designing the network and training the users.

Building Classes of Service in Cisco IOS with H.323

The following scenarios require you to define classes of service within Cisco IOS routers running the H.323 protocol:

- Cisco CallManager multi-site deployments with centralized call processing
- Cisco CallManager Express deployments

Under normal conditions in Cisco CallManager multi-site deployments with centralized call processing, classes of service are implemented using partitions and calling search spaces within Cisco CallManager. However, when IP WAN connectivity is lost between a branch site and the central site, Cisco SRST takes control of the branch IP phones, and all the configuration related to partitions and calling search spaces is unavailable until IP WAN connectivity is restored. Therefore, it is desirable to implement classes of service within the branch router when running in SRST mode.

Similarly, in Cisco CallManager Express deployments, the router needs a mechanism to implement classes of service for the IP phones.

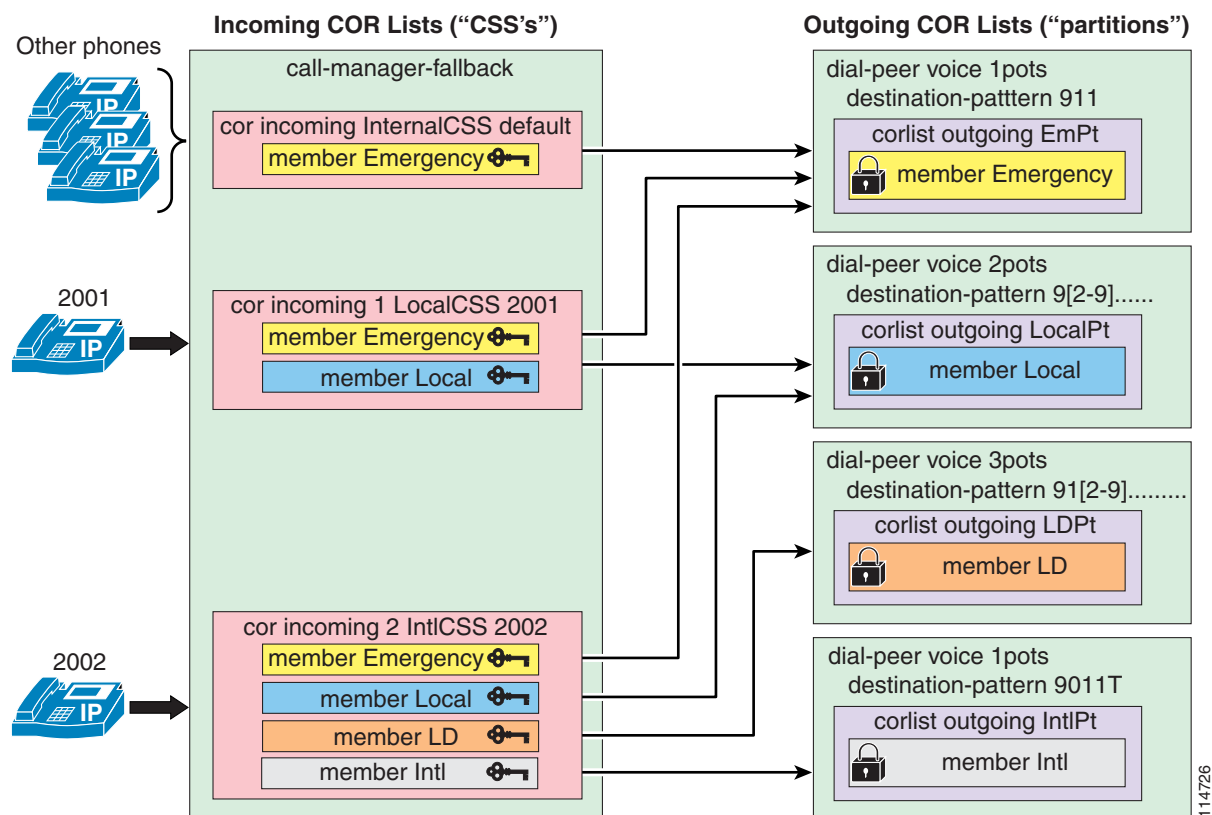
For both of these applications, define classes of service in Cisco IOS routers by using the class of restriction (COR) functionality (refer to [Calling Privileges in Cisco IOS with H.323 Dial Peers](#), page 10-39, for details on COR).

You can adapt the COR functionality to replicate the Cisco CallManager concepts of partitions and calling search spaces by following these main guidelines:

- Define tags for each type of call that you want to distinguish.
- Assign "basic" outgoing corlists (equivalent to partitions), containing a single member tag, to the respective POTS dial peers that route each type of call.
- Assign "complex" incoming corlists (equivalent to calling search spaces), containing subsets of the member tags, to the IP phones that belong to the various classes of service.

[Figure 10-33](#) illustrates an implementation example based on SRST, where the IP phone with DN 2002 is configured to have unrestricted PSTN access, the IP phone with DN 2001 is configured to have only local PSTN access, and all other IP phones are configured to have access only to internal numbers and emergency services.

Figure 10-33 Building Classes of Service for Cisco SRST using COR



The following steps provide examples and guidelines for implementing a Cisco IOS solution like the one shown in Figure 10-33.

- Step 1** Using the **dial-peer cor custom** command, define meaningful tags for the various types of calls (Emergency, VMail, Local, LD, Intl):

```
dial-peer cor custom
  name Emergency
  name VMail
  name Local
  name LD
  name Intl
```

- Step 2** Using the **dial-peer cor list** command, define basic corlists to be used as partitions, each containing a single tag as a member:

```
dial-peer cor list EmPt
  member Emergency

dial-peer cor list VMailPt
  member VMail

dial-peer cor list LocalPt
  member Local

dial-peer cor list LDPT
  member LD
```

```
dial-peer cor list IntlPt
member Intl
```

- Step 3** Using the **dial-peer cor list** command, define more complex corlists to be used as calling search spaces, each containing a subset of the tags as members according to the classes of service needed:

```
dial-peer cor list InternalCSS
member Emergency
member VMail
```

```
dial-peer cor list LocalCSS
member Emergency
member VMail
member Local
```

```
dial-peer cor list LDCSS
member Emergency
member VMail
member Local
member LD
```

```
dial-peer cor list IntlCSS
member Emergency
member VMail
member Local
member LD
member Intl
```

- Step 4** Using the command **corlist outgoing** *corlist-name*, configure the basic “partition” corlists as outgoing corlists assigned to the corresponding POTS dial peers:

```
dial-peer voice 100 pots
corlist outgoing EmPt
destination-pattern 911
no digit-strip
direct-inward-dial
port 1/0:23
```

```
dial-peer voice 101 pots
corlist outgoing VMailPt
destination-pattern 914085551234
forward-digits 11
direct-inward-dial
port 1/0:23
```

```
dial-peer voice 102 pots
corlist outgoing LocalPt
destination-pattern 9[2-9].....
forward-digits 7
direct-inward-dial
port 1/0:23
```

```
dial-peer voice 103 pots
corlist outgoing LDPT
destination-pattern 91[2-9]..[2-9].....
forward-digits 11
direct-inward-dial
port 1/0:23
```

```
dial-peer voice 104 pots
corlist outgoing IntlPt
destination-pattern 9011T
prefix-digits 011
```

```
direct-inward-dial
port 1/0:23
```

- Step 5** Using the **cor incoming** command within the **call-manager-fallback** configuration mode, configure the complex corlists acting as "calling search spaces" to be incoming corlists assigned to the various phone DNs:

```
call-manager-fallback
cor incoming InternalCSS default
cor incoming LocalCSS 1 3001 - 3003
cor incoming LDCSS 2 3004
cor incoming IntlCSS 3 3010
```

When deploying COR for SRST, keep in mind the following limitations:

- In SRST version 2.0, available on Cisco IOS Release 12.2(8)T or later, the maximum number of **cor incoming** statements allowed under **call-manager-fallback** is 5 (plus the default statement).
- In SRST version 3.0, available on Cisco IOS Release 12.3(4)T or later, the maximum number of **cor incoming** statements allowed under **call-manager-fallback** is 20 (plus the default statement).

Therefore, if the phone DNs of users with non-default privileges are not consecutive and the SRST site is relatively large, you might have to reduce the number of classes of service in SRST mode to accommodate all the DNs without exceeding these limitations.

Although the preceding example is based on Cisco SRST, the same concepts can be applied to a Cisco CallManager Express deployment, except for the following considerations:

- With Cisco CallManager Express, the corlist expressing the class of service (equivalent to a calling search space) can be assigned directly to the individual IP phones by using the **cor {incoming | outgoing} corlist-name** command under the **ephone-dn dn-tag** configuration mode.
- According to COR general rules, all IP phones for which no corlist is configured have unrestricted access to all dial peers, regardless of their outgoing corlist. Cisco CallManager Express has no mechanism equivalent to the **cor incoming corlist-name default** command, which applies default restrictions to all phones.

Deploying Call Coverage

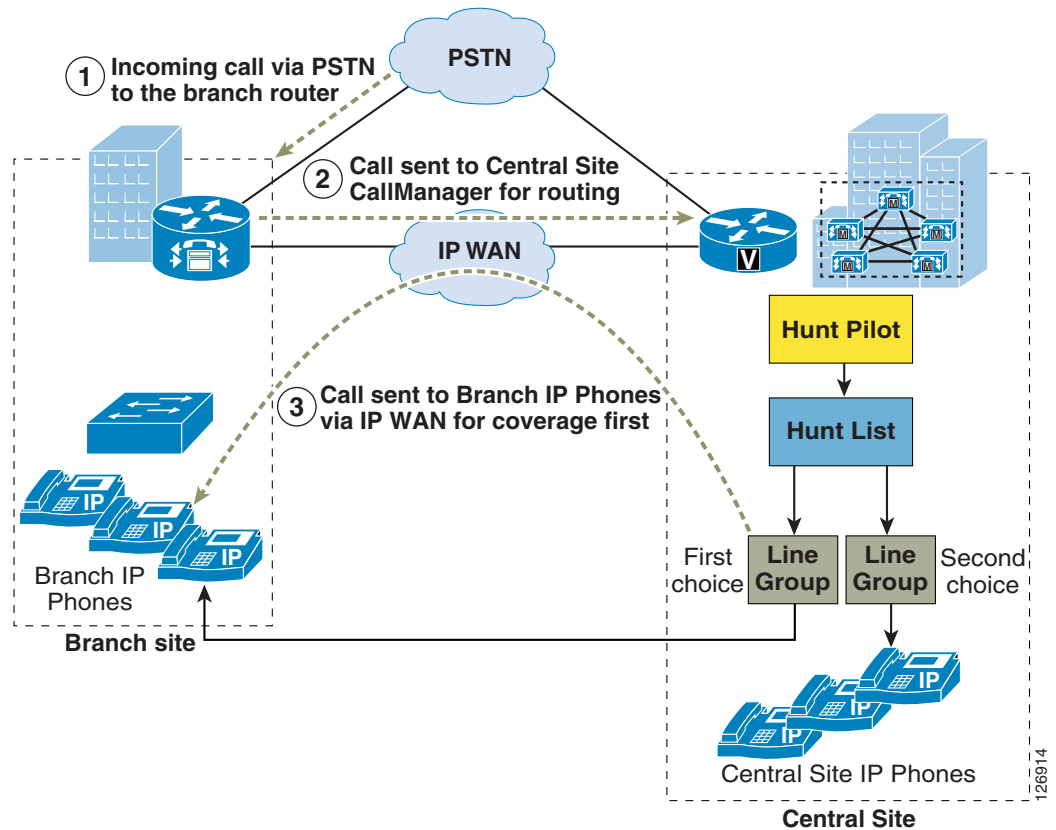
Call coverage functionality is a key feature in many IP telephony deployments. Many customer-focused service companies have to route customer calls to the appropriate service representatives expeditiously. This section focuses on design guidelines for using the hunting mechanism based on hunt pilots, hunt lists, and line groups in Cisco CallManager Release 4.1 to manage call distribution, and it covers the following main topics:

- [Deploying Call Coverage in a Multi-Site Centralized Call Processing Model, page 10-79](#)
- [Deploying Call Coverage in a Distributed Call Processing Model, page 10-80](#)

Deploying Call Coverage in a Multi-Site Centralized Call Processing Model

Figure 10-34 shows an example of configuring hunt lists for a multi-site centralized call processing deployment. This example assumes that the hunt pilot call will be distributed first through the remote office operators. If the call is not answered or is rejected due to call admission control, the call will then be routed to central-site operators or voicemail.

Figure 10-34 Call Coverage Between Multiple Sites in a Centralized Call Processing Deployment



In centralized IP telephony systems, features such as Automated Alternate Routing (AAR) and Survivable Remote Site Telephony (SRST) may be enabled for high availability. Consider the following guidelines when deploying call coverage functionality with AAR or SRST features enabled:

- Automated Alternate Routing (AAR)

The line and route group members can be assigned in different locations and regions. Call admission control implemented through locations works as expected. However, a call being distributed from a hunt pilot will not use AAR to reroute a call if Cisco CallManager blocks the call to one of its line or route group members due to insufficient location bandwidth. Instead, Cisco CallManager distributes the call to the next available member or next available line or route group.

- Survivable Remote Site Telephony (SRST)

- When Cisco CallManager receives a call for the hunt pilot, and if some of its line or route members are at the remote sites operating in SRST mode, then Cisco CallManager skips those members and distributes the call to the next available line or route group member. From the perspective of Cisco CallManager, the members operating in SRST mode are unregistered, and hunt pilot calls are not forwarded to unregistered members.

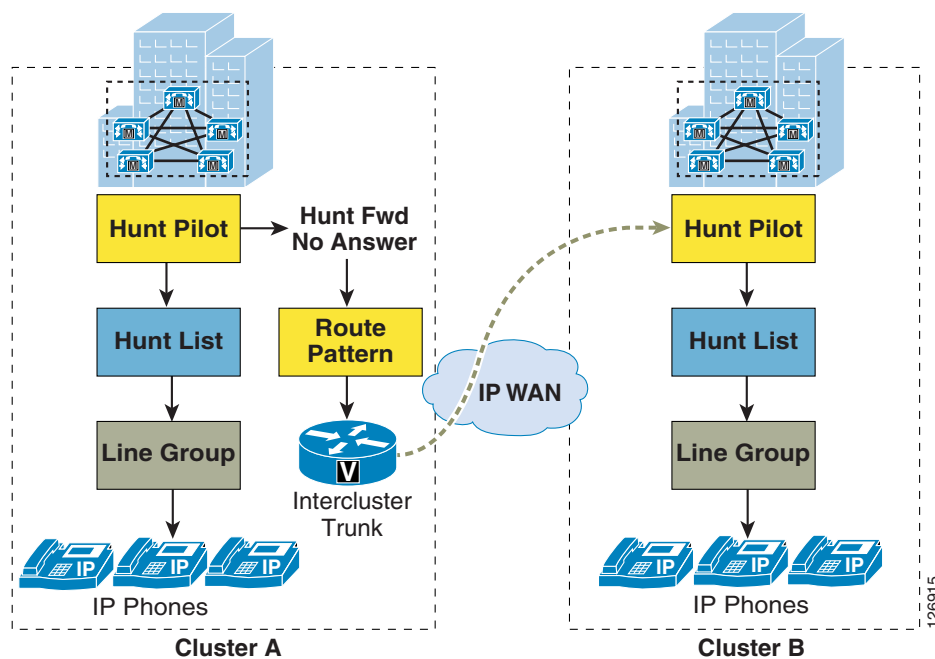
- When a router operating in SRST mode receives a call for the hunt pilot, call coverage functionality is unavailable. The call fails if no configuration is added to reroute the call to a registered and available extension. You can use the **alias** or the **default-destination** command under the **call-manager-fallback** mode in Cisco IOS to reroute the call destined for the hunt pilot to an operator extension or to voicemail.

Deploying Call Coverage in a Distributed Call Processing Model

Beginning with Cisco CallManager Release 4.1, route groups can no longer be added to hunt lists. Thus, a hunt list cannot be used to send the calls to other clusters or to a remote gateway. But the hunt option settings in Hunt Pilot, introduced in Cisco CallManager Release 4.1, can be used to match a route pattern that in turn points to gateways or trunks.

Figure 10-35 shows an example of configuring hunt lists for a distributed call processing deployment with an intercluster trunk. This example assumes that the hunt pilot call is first distributed within Cluster A. If the call is not answered, it is rerouted to Cluster B for call distribution using the Forward Hunt No Answer setting, which matches a route pattern. The route pattern, in turn, points to an intercluster trunk to Cluster B.

Figure 10-35 Call Coverage Between Clusters in a Distributed Call Processing Deployment



Tip

In distributed call processing deployments, load sharing of incoming hunt pilot calls can be managed using Cisco VoIP gateways and gatekeepers. In the event that the call is not answered within one cluster, it can overflow to another cluster for service. Calls can also be sent through gateways or trunks to IVR treatment. Tool Command Language (TCL) IVR applications can be implemented on Cisco IOS gateways.

Hunt Pilot Scalability

Cisco recommends using the following guidelines when deploying call coverage using top-down, circular, and longest-idle algorithms:

- The Cisco CallManager cluster supports a maximum of 15,000 hunt list devices.
- The hunt list devices may be a combination of 1500 hunt lists with 10 IP phones in each hunt list, or a combination of 750 hunt lists with 20 IP phones in each hunt list. However, an increase in a number of hunt lists can require increasing the dial plan initialization timer specified in the Cisco CallManager service parameters. Cisco recommends setting the dial plan initialization timer to 600 seconds if 1500 hunt lists are configured.



Note When using the broadcast algorithm for call coverage, the number of hunt list devices is limited by the number of busy hour call completions (BHCC).

- Cisco recommends having a maximum of 35 directory numbers in a single line group configured to send the calls simultaneously to all DNs. Additionally, the number of broadcast line groups depends on the BHCC. If there are multiple broadcast line groups in a Cisco CallManager system, the number of maximum directory numbers in a line group must be less than 35. The number of busy hour call attempts (BHCA) for all the broadcast line groups should not exceed 35 calls set up per second



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:

- [Planning for 911 Functionality, page 11-2](#)
- [Gateway Considerations, page 11-11](#)
- [Cisco Emergency Responder Considerations, page 11-13](#)

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/about/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.nena.org/florida/Directory/911Tutorial%20Study%20Guide.pdf>

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 serviced 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 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 Local Exchange Carrier (LEC) 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, contact the PSAP directly to verify the information.

Typical Situation

- For a given street address, the 911 network service provider is the incumbent Local Exchange Carrier (LEC). 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 a 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 a 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 WAN failure conditions where Survivable Remote Site Telephony (SRST) operation is activated.

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 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 of origin, 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 main 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 as 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 routing 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 11-7](#), 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 ANI 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, VIC-2FXO, and VIC-4FXO trunk cards.

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's 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 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 attempt 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, known 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 11-4](#).) 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 events 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 is one of the key advantages of IP telephony technology. To provide for moves, adds, and changes 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 11-13, 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 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 but there were only two ELINs for the ERL, the situation would allow for callback functionality to properly reach only the last two callers in the sequence.

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. Cisco ER's on-site notification capabilities can help in identifying the phone at the origin of such spurious 911 calls by providing detailed accounts of all calls made to 911, including calls made by mistake.

Gateway Considerations

Consider the following factors when selecting the gateways to handle emergency calls for your system:

- [Gateway Placement, page 11-11](#)
- [Gateway Blocking, page 11-12](#)
- [Answer Supervision, page 11-12](#)
- [Answer Supervision, page 11-12](#)

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 Cisco's IP-based enterprise telephony architecture, 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 choose 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. Be aware that, even if the 911 network in a given region lends itself to using a centralized gateway for 911 calls, it might be preferable to rely on gateways co-located with the calling phones to prevent 911 call routing from being impacted during WAN failures.

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 if it is acceptable to the applicable public safety authorities and only as a 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.

Be advised that, in situations where answer supervision is not provided, the call detail records (CDRs) will not accurately reflect the connect time or duration of 911 calls. This inaccuracy can impede the ability of a call reporting system to document the relevant statistics properly for 911 calls.

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 (Cisco 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 a call admission control location.

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 occurs, 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 Communicator 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 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 not to allow 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. An active call timer is an indication that answer supervision is working properly.

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.

Multi-Cluster Considerations

Enterprise telephony systems based on multiple Cisco CallManager clusters can benefit from the functionality of Cisco Emergency Responder (Cisco ER).

The *Cisco Emergency Responder Administration Guide* provides detailed descriptions of the terms used herein, as well as the background information required to support the following discussion. Of specific interest is the chapter on *Planning for Cisco Emergency Responder*. This documentation is available at

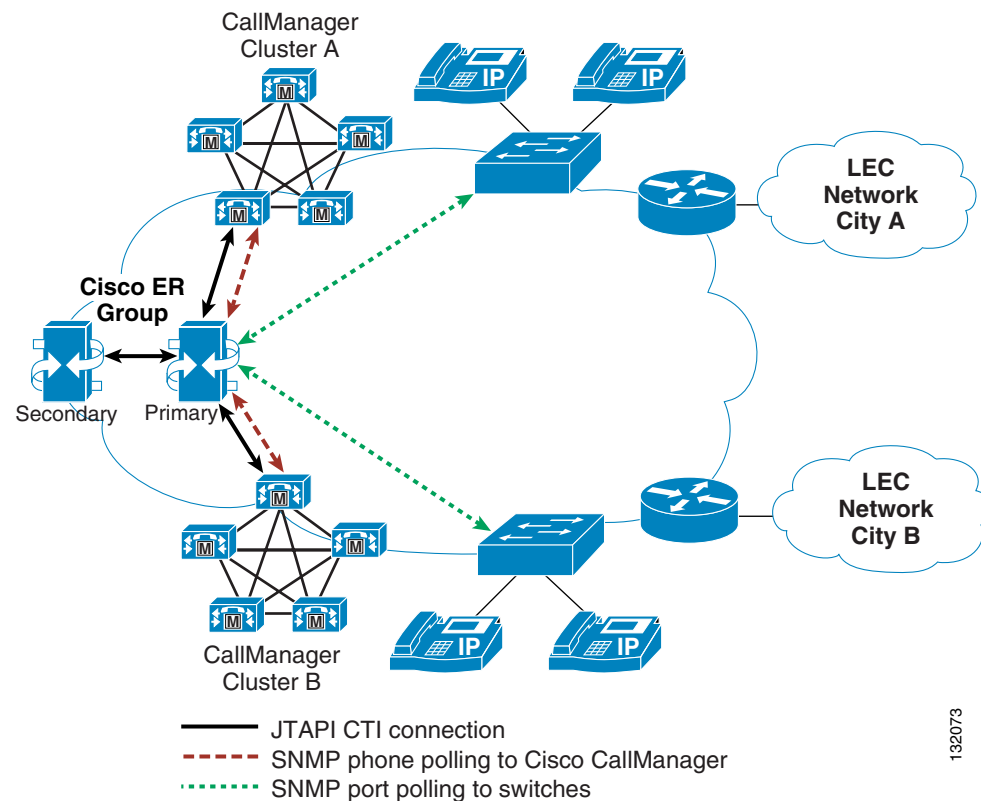
<http://www.cisco.com>

Single Cisco ER Group

A single Emergency Responder group can be deployed to handle emergency calls from two or more Cisco CallManager clusters. The design goal is to ensure that any phone's emergency call is routed to the Cisco ER group, which will assign an ELIN and route the call to the appropriate gateway based on the phone's location.

One advantage of using a single Cisco ER group is that all ERLs and ELINs are configured into a single system. A phone registered on any cluster will be located by the single Cisco ER group because that group is responsible for polling all of the system's access switches. [Figure 11-1](#) illustrates a single Cisco ER group interfaced with two Cisco CallManager clusters.

Figure 11-1 A Single Cisco ER Group Connected to Two Cisco CallManager Clusters



132073

The single Cisco ER group in [Figure 11-1](#) interfaces with the following components:

- Each Cisco CallManager cluster via SNMP, to collect information about their respective configured phones
- All of the enterprise's switches via SNMP, so that any cluster's phone, connected to any switch, can be located. This connection is not required if the phone locations are being identified based on IP subnets. For details on configuring IP subnet-based ERLs, refer to the chapter on Configuring Cisco Emergency Responder in the *Cisco Emergency Responder Administration Guide*, available at <http://www.cisco.com>
- Each Cisco CallManager cluster via JTAPI, to allow for the call processing required by any phone that dials 911 – for example, identification of the calling phone's ERL, assignment of the ELIN, redirection of the call to the proper gateway (based on the calling phone's location), and the handling of the PSAP callback functionality

The version of the JTAPI interface used by Cisco Emergency Responder is determined by the version of the Cisco CallManager software to which it is connected. At system initialization, Cisco ER interrogates the Cisco CallManager cluster and loads the appropriate JTAPI Telephony Service Provider (TSP). Because there can be only one version of JTAPI TSP on the Cisco ER server, all Cisco CallManager clusters to which a single Cisco ER group is interfaced *must* run the same version of Cisco CallManager software.

For some deployments, this software version requirement might present some difficulties. For instance, during a Cisco CallManager upgrade, different clusters will be running different versions of software, and some of the clusters will be running a version of JTAPI that is not compatible with the version running on the Cisco ER servers. When this situation occurs, emergency calls from the cluster running a version of JTAPI different than that of the Cisco ER group might receive the call treatment provided by the Call Forward Busy settings of the emergency number's CTI Route Point.

When considering if a single Cisco ER group is appropriate for multiple Cisco CallManager clusters, apply the following guidelines:

- Make Cisco CallManager upgrades during an acceptable maintenance window when emergency call volumes are as low as possible (for example, after hours, when system use is at a minimum).
- Use a single Cisco ER group only if the quantity and size of the clusters allow for minimizing the amount of time when dissimilar versions of JTAPI are in use during software upgrades.

For example, a deployment with one large eight-server cluster in parallel with a small two-server cluster could be considered for use with a single Cisco ER group. In this case, it would be best to upgrade the large cluster first, thus minimizing the number of users (those served by the small cluster) that might be without Cisco ER service during the maintenance window of the upgrade. Furthermore, the small cluster's users can more appropriately be served by the temporary static routing of emergency calls in effect while Cisco ER is not reachable because they can be identified by the single ERL/ELIN assigned to all non-ER calls made during that time.

Multiple Cisco ER Groups

Multiple Cisco ER groups can also be deployed to support multi-cluster systems. In this case, each ER group interfaces with the following components:

- A Cisco CallManager cluster via the following methods:
 - SNMP, to collect information about its configured phones
 - JTAPI, to allow for the call processing associated with redirection of the call to the proper gateway or, in the case of roaming phones, the proper Cisco CallManager cluster

- The access switches (via SNMP) to which most of the phones associated with the Cisco CallManager of the Cisco ER group are most likely to be connected

This approach allows Cisco CallManager clusters to run different versions of software because each is interfaced to a separate Cisco ER group.

To allow phones to roam between various parts of the network and still be tracked by Cisco ER, you might have to configure the Cisco ER groups into a Cisco ER cluster. For details on Cisco ER clusters and groups, refer to the chapter on *Planning for Cisco Emergency Responder* in the *Cisco Emergency Responder Administration Guide*, available at

<http://www.cisco.com>

Figure 11-2 presents a sample topology illustrating some of the basic concepts behind Cisco ER clustering.

Figure 11-2 Multiple Cisco ER Groups

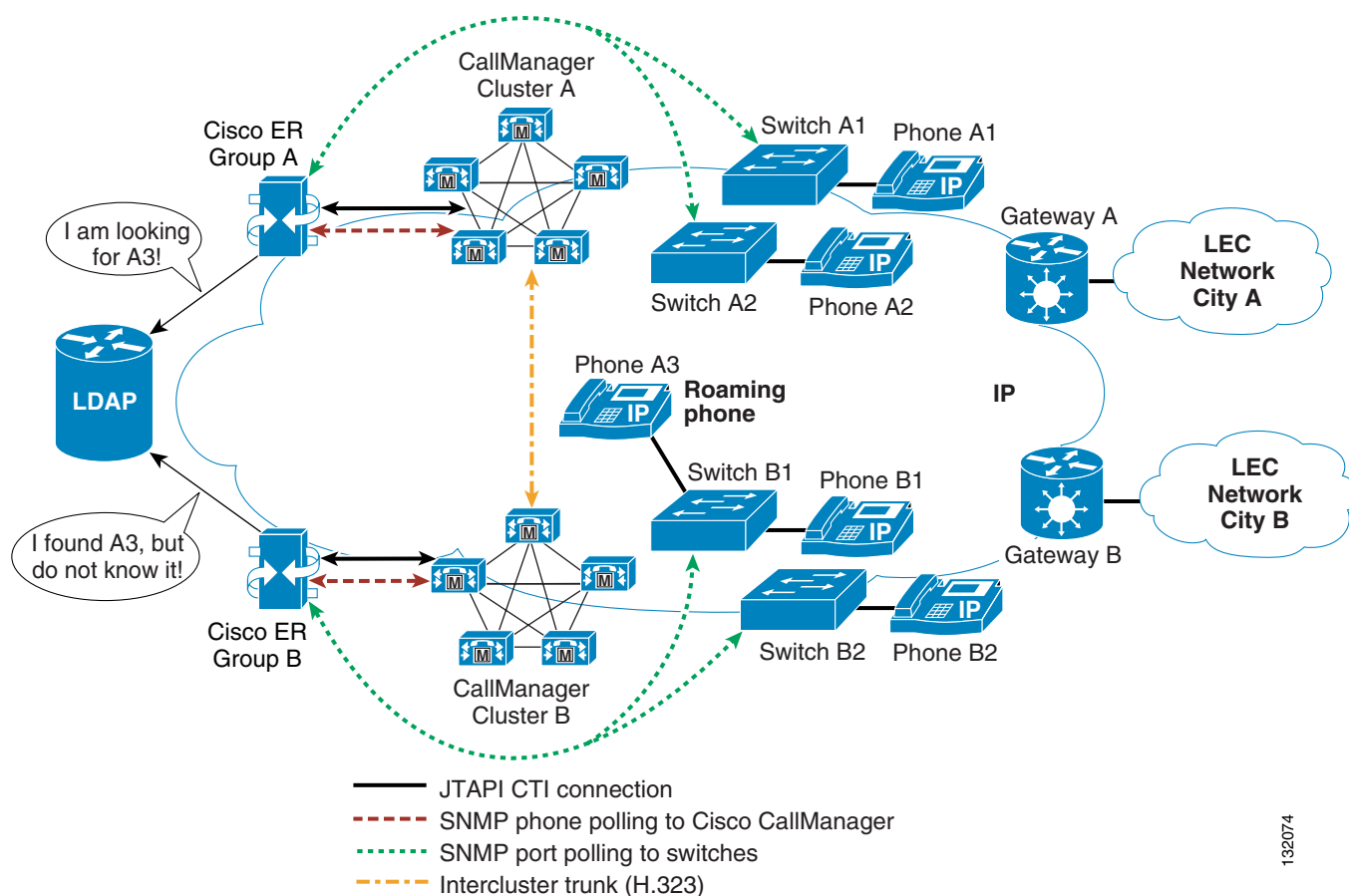


Figure 11-2 illustrates the following topology:

- Cisco ER group A is interfaced to Cisco CallManager cluster A to access switches A1 and A2, and it is deemed to be the home Cisco ER group of all phones registered to Cisco CallManager cluster A.
- Likewise, Cisco ER group B is interfaced to Cisco CallManager cluster B to access switches B1 and B2, and it is deemed to be the home Cisco ER group of all phones registered to Cisco CallManager cluster B.

132074

Phone Movements Within the Tracking Domain of a Cisco ER Group

The emergency call processing for phones moving between access switches controlled by the same home Cisco ER group is the same as the processing done for a deployment with a single Cisco CallManager cluster. For example, a phone moving between access switches A1 and A2 remains registered with Cisco CallManager cluster A, and its location is determined by Cisco ER group A both before and after the move. The phone is still under full control of Cisco ER group A, for both the discovery of the phone by Cisco CallManager cluster A and the determination of the phone's location by switch A2. The phone is therefore not considered to be an unlocated phone.

Phone Movements Between the Various Tracking Domains of a Cisco ER Cluster

A Cisco ER cluster is essentially a collection of Cisco ER groups that share location information through a Lightweight Directory Access Protocol (LDAP) database. Each group shares the location of any phone it finds on an access switch or in an IP subnet. However, any phone found in the Cisco ER group's own Cisco CallManager cluster is deemed to be *unknown*, and its information is not shared.

Cisco ER groups also share information about phones that cannot be located within a Cisco ER group's tracking domain (in switches or IP subnets) but which are known to be registered in the group's associated Cisco CallManager cluster. Such phones are deemed *unlocated*.

If a phone is roaming between access switches monitored by different Cisco ER groups, those groups must be configured in a Cisco ER cluster so they can exchange information about the phone's location. For example, phone A3 is registered with Cisco CallManager cluster A, but it is connected to an access switch controlled by Cisco ER group B. Cisco ER group A is aware that phone A3 is registered with Cisco CallManager cluster A, but group A cannot locate phone A3 in any of the site A switches. Therefore, phone A3 is deemed *unlocated* by Cisco ER group A.

Cisco ER group B, on the other hand, has detected the presence of phone A3 in one of the switches that it monitors. Because the phone is not registered with Cisco CallManager cluster B, phone A3 is advertised through the Cisco ER LDAP database as an *unknown* phone.

Because the two Cisco ER groups are communicating through an LDAP database, they can determine that Cisco ER group B's *unknown* phone A3 is the same as Cisco ER group A's *unlocated* phone A3.

The Unlocated Phone page in Cisco ER group A will display the phone's host name along with the remote Cisco ER group (in this, case Cisco ER group B).

Emergency Call Routing within a Cisco ER Cluster

Cisco ER clustering also relies on route patterns that allow emergency calls to be redirected between pairs consisting of a Cisco CallManager cluster and a Cisco ER. For more details, refer to the section on *Creating Route Patterns for Inter-Cisco Emergency Responder-Group Communications* in the *Cisco Emergency Responder Administration Guide*, available at

<http://www.cisco.com>

If phone A3 places an emergency call, the call signaling flow will be as follows:

1. Phone A3 sends the emergency call string to Cisco CallManager cluster A for processing.
2. Cisco CallManager cluster A sends the call to Cisco ER group A for redirection.
3. Cisco ER group A determines that phone A3 is located in Cisco ER group B's tracking domain, so it redirects the call to a route pattern that points to Cisco CallManager cluster B.
4. Cisco CallManager cluster A sends the call to Cisco CallManager cluster B.
5. Cisco CallManager cluster B sends the call to Cisco ER group B for redirection.

6. Cisco ER group B identifies the ERL and ELIN associated with phone A3's location and redirects the call to Cisco CallManager cluster B. The calling number is transformed into the ELIN associated with the ERL of phone A3, and the called number is modified to route the call to the proper gateway.
7. Cisco CallManager cluster B routes the call according to the new called number information obtained from Cisco ER group B.
8. Cisco CallManager cluster B sends the call out the gateway toward the Emergency PSTN network.

Scalability Considerations for Cisco ER Clustering

In a Cisco ER cluster, the quantity of phones roaming outside the tracking domain of their home Cisco ER group is a scalability factor that you must keep within the limits set forth in the section on *Network Hardware and Software Requirements* in the *Cisco Emergency Responder Administration Guide 1.2(3)*, available at:

http://www.cisco.com/en/US/products/sw/voicesw/ps842/prod_maintenance_guides_list.html

With the Cisco MCS 7845 server platform and version 1.2(3) of the Cisco ER software, a Cisco ER cluster can support a maximum of 3000 roaming phones. For deployments that have to exceed this limit (for instance, large campus deployments with multiple Cisco CallManager clusters), phone movement can be tracked by IP subnets. By defining the IP subnets in each of the Cisco ER groups and by assigning each ERL with one ELIN per Cisco ER group, you can virtually eliminate roaming phones because all phones in the campus will be part of the tracking domain of their respective Cisco ER group.

ALI Formats

In multi-cluster configurations, there might be instances where the physical locations of ERLs and ELINs defined in a single Cisco ER group span the territory of more than one phone company. This condition can lead to situations where records destined for different phone companies have to be extracted from a common file that contains records for multiple LECs.

Cisco ER exports this information in ALI records that conform to National Emergency Number Association (NENA) 2.0, 2.1, and 3.0 formats. However, many service providers do not use NENA standards. In such cases, you can use the ALI Formatting Tool (AFT) to modify the ALI records generated by Cisco ER so that they conform to the formats specified by your service provider. That service provider can then use the reformatted file to update their ALI database.

The ALI Formatting Tool (AFT) enables you to perform the following functions:

- Select a record and update the values of the ALI fields. AFT allows you to edit the ALI fields to customize them to meet the requirements of various service providers. Your service provider can then read the reformatted ALI files and use them to update their ELIN records.
- Perform bulk updates on multiple ALI records. Using the bulk update feature, you can apply common changes to all the records that you have selected, to one area code, or to one area code and one city code.
- Selectively export ALI records based on area code, city code, or a four-digit directory number. By selecting to export all the ALI records in an area code, for example, you can quickly access all the ELIN records for each service provider, thereby easily supporting multiple service providers.

Given the flexibility of the AFT, a single Cisco ER group can export ALI records in multiple ALI database formats. For a Cisco ER group serving a Cisco CallManager cluster with sites in the territories of two LECs, the basic approach is as follows:

1. Obtain an ALI record file output from Cisco Emergency Responder in standard NENA format. This file contains the records destined for multiple LECs.

2. Make a copy of the original file for each required ALI format (one copy per LEC).
3. Using the AFT of the first LEC (for example, LEC-A), load a copy of the NENA-formatted file and delete the records of all the ELINs associated with the other LECs. The information to delete can usually be identified by NPA (or area code).
4. Save the resulting file in the required ALI format for LEC-A, and name the file accordingly.
5. Repeat steps 3 and 4 for each LEC.

The ALI formatting tools are available at

http://cco/en/US/products/sw/voicesw/ps842/products_administration_guides_list.html

For LECs not listed at this URL, the output from Cisco CallManager can be formatted using standard text file editing tools, such as spreadsheet programs and standard text editors.



Voicemail Design

This chapter discusses the following options for deploying voicemail systems with Cisco CallManager:

- [Cisco Unity, page 12-1](#)
- [Cisco Unity Express, page 12-1](#)
- [Third-Party Voicemail Systems, page 12-7](#)



Note

This chapter does not discuss how to size a voicemail system for ports and/or storage. For this type of information, contact your voicemail vendor, who should be better able to discuss the individual requirements of their own system based upon specific traffic patterns.

Cisco Unity

Cisco Unity is a class-leading unified messaging solution that supports either Microsoft Exchange or IBM (Lotus) Domino as the back-end message store. Unity is also offered as a standalone voicemail system using only Microsoft Exchange as a local message store. For information on the various deployment options with Cisco Unity, see the chapter on [Cisco Unity, page 13-1](#).

For more information on Cisco Unity, refer to the Cisco Unity product documentation available at <http://www.cisco.com>

Cisco Unity Express

This section introduces Cisco Unity Express as a distributed voicemail and automated attendant (AA) solution for Cisco CallManager. Cisco Unity Express is an entry-level voicemail system for offices of up to 100 people (100 mailboxes), and it integrates physically as a blade within the branch office router at each site.

Use Cisco Unity Express as a distributed voicemail solution if any of the following conditions apply to your Cisco CallManager network deployment:

- Survivability of voicemail and AA access must be ensured regardless of WAN availability.
- Available WAN bandwidth is insufficient to support voicemail calls traversing the WAN to a central voicemail server.

- There is limited geographic coverage of the AA or branch site PSTN phone numbers published to the local community, and these numbers cannot be dialed to reach a central AA server without incurring toll charges.
- The likelihood is high that a PSTN call into a branch office will be transferred from the branch AA to a local extension in the same office.
- Management philosophy allows remote locations to select their own voicemail and AA technology.

**Note**

Interoperability with Cisco CallManager requires a minimum of Cisco Unity Express Release 1.1 and Cisco CallManager Release 3.3.3 or later 3.3-based release. Cisco Unity Express 2.0 provides interoperability with Cisco CallManager Release 4.0, and Cisco Unity Express 2.1 provides interoperability with Cisco CallManager Release 4.1.

For more information about Cisco Unity Express, refer to the product documentation available at

<http://www.cisco.com>

Deployment Models for Cisco Unity Express

Cisco Unity Express is supported with all of the Cisco CallManager deployment models, including single-site deployments, multi-site WAN with centralized call processing, and multi-site WAN with distributed call processing. [Figure 12-1](#) shows a centralized call processing deployment incorporating Cisco Unity Express, and [Figure 12-2](#) shows a distributed call processing deployment.

Figure 12-1 Cisco Unity Express in a Centralized Call Processing Deployment

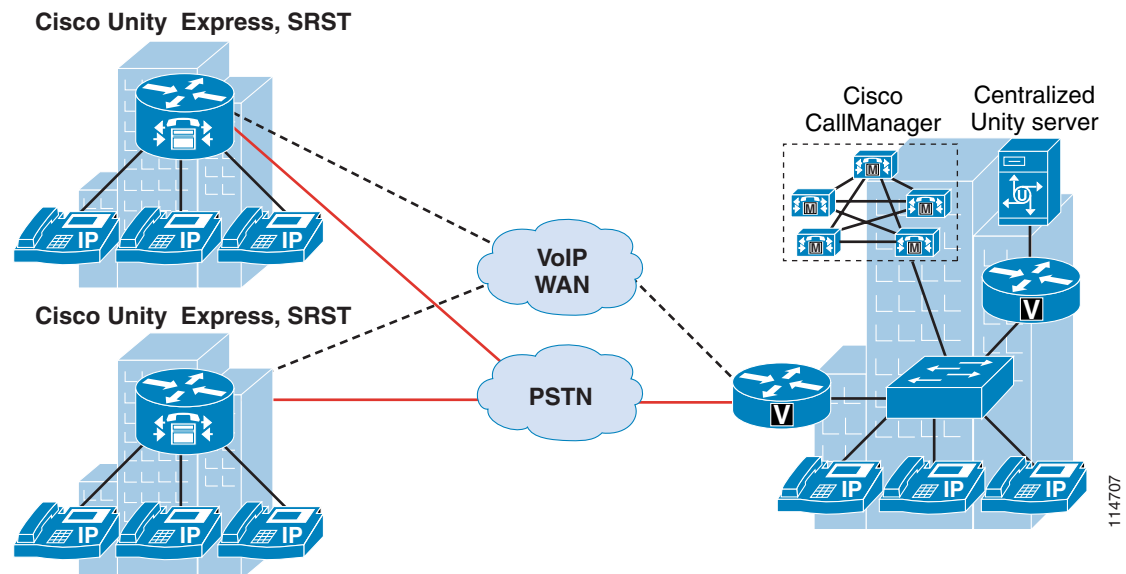
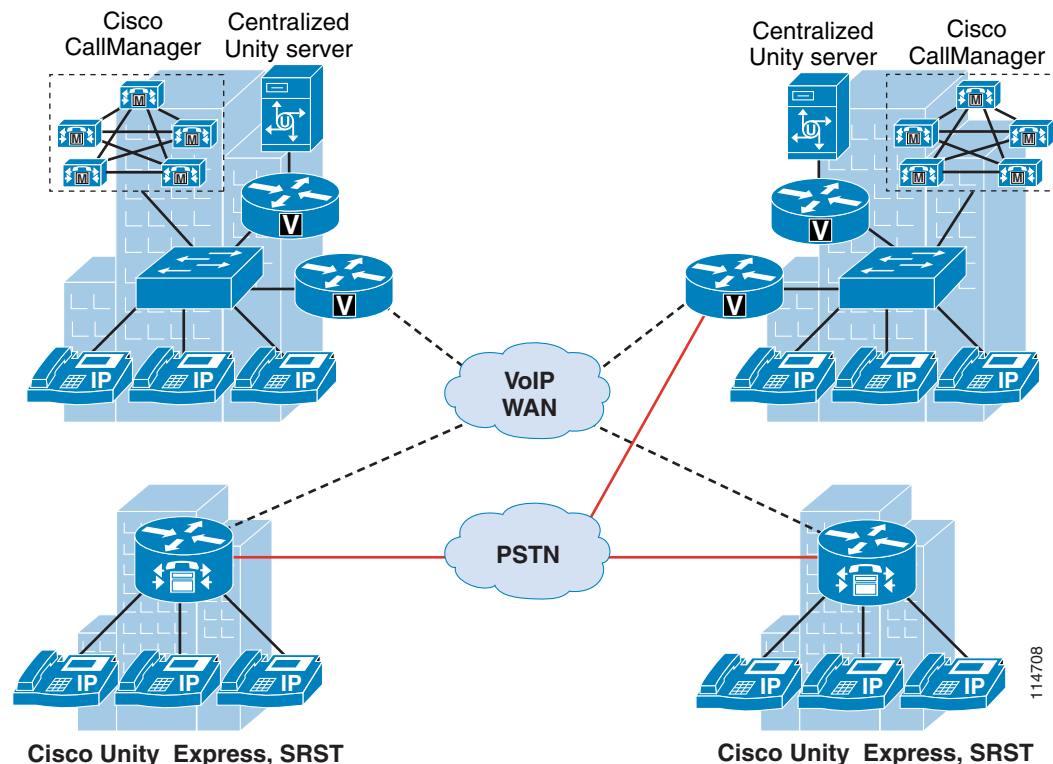


Figure 12-2 Cisco Unity Express in a Distributed Call Processing Deployment

The most likely deployment model to use Cisco Unity Express is the multi-site WAN with centralized call processing, where Cisco Unity Express provides distributed voicemail at the smaller remote offices and a central Cisco Unity system provides voicemail to the main campus and larger remote sites.

Cisco Unity Express can also be deployed as a co-resident voicemail solution with Cisco CallManager Express. Cisco Unity Express sites controlled by Cisco CallManager Express, as well as other sites controlled by Cisco CallManager, can be interconnected in the same network. Although Cisco Unity Express can integrate with either Cisco CallManager or Cisco CallManager Express, it cannot integrate with both simultaneously.

The following characteristics and guidelines apply to Cisco Unity Express in either a centralized or distributed Cisco CallManager deployment:

- A single Cisco Unity Express can be integrated with a single Cisco CallManager cluster.
- Cisco Unity Express integrates with Cisco Call Manager using a JTAPI application and Computer Telephony Integration (CTI) Quick Buffer Encoding (QBE) protocol. CTI ports and CTI route points control the Cisco Unity Express voicemail and automated attendant applications.
- Cisco Unity Express provides voicemail functionality to Cisco IP Phones running Skinny Client Control Protocol (SCCP). Future releases will provide support for Session Initiation Protocol (SIP) IP phones.

- The following CTI route points are defined on Cisco CallManager for Cisco Unity Express:
 - Automated attendant entry point (Cisco Unity Express can contain up to five distinct AAs and may therefore require up to five different route points.)
 - Voicemail pilot number
 - Greeting management system (GMS) pilot number (Optional; if the GMS is not used, then this route point need not be defined.)
- The following CTI ports are defined on Cisco CallManager for Cisco Unity Express:
 - A 12 or 25-mailbox system (4 ports)
 - A 50-mailbox AIM-CUE system (4 ports)
 - A 50-mailbox NM-CUE system (8 ports)
 - A 100-mailbox system (8 ports)
- Each Cisco Unity Express site has 100 mailboxes or less. For deployments that require more than 100 mailboxes, consider using Cisco Unity or other voicemail solutions.
- Each Cisco Unity Express mailbox can be associated with a maximum of two different extensions, if needed.
- The automated attendant function for any office deployed with Cisco Unity Express can be local to the office (using the AA application in Cisco Unity Express) or centralized (using Cisco Unity Express for voicemail only).
- Beginning with Cisco Unity Express Release 2.0, a single Cisco Unity Express can be networked only with other Cisco Unity Expresses or with Cisco Unity. Thus, a Cisco Unity Express subscriber can forward or send messages to another remote Cisco Unity Express or Cisco Unity subscriber.
- Cisco Unity Express allows you to specify up to three Cisco CallManagers for failover. If IP connectivity to all three Cisco CallManagers is lost, Cisco Unity Express switches to Survivable Remote Site Telephony (SRST) call control, thus providing AA call answering service as well as mailbox access to IP phones and PSTN calls coming into the branch office.
- Cisco Unity Express automated attendant supports dial-by-extension and dial-by-name functions. The dial-by-extension operation enables a caller to transfer a call to any user endpoint in the network. The dial-by-name operation uses the directory database internal to Cisco Unity Express and does not interact with external LDAP or Active Directory databases.

Figure 12-3 shows the protocols involved in the call flow between Cisco CallManager and Cisco Unity Express.

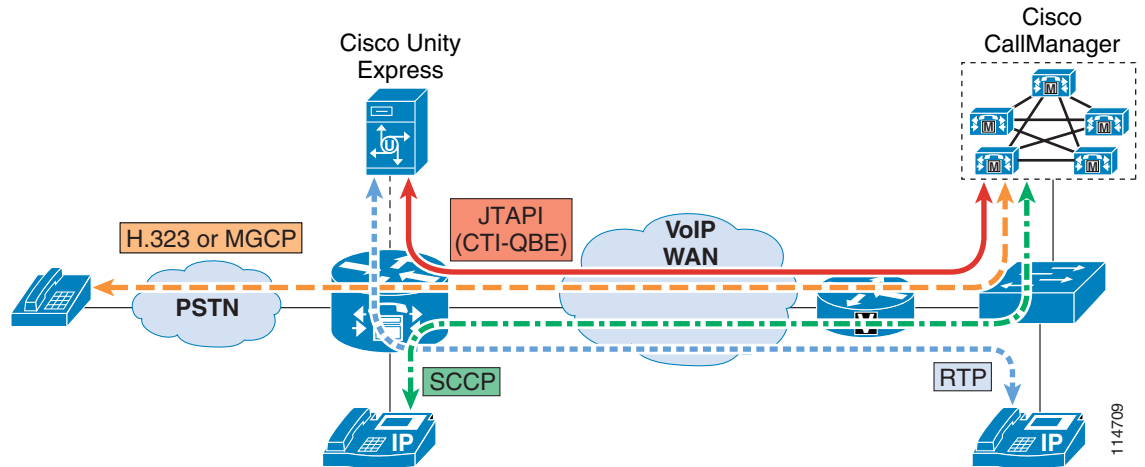
Figure 12-3 *Protocols Used Between Cisco Unity Express and Cisco CallManager*

Figure 12-3 illustrates the following signaling and media flows:

- Phones are controlled via SCCP from Cisco CallManager.
- Cisco Unity Express is controlled via JTAPI (CTI-QBE) from Cisco CallManager.
- The Message Waiting Indicator (MWI) on the phone is affected by Cisco Unity Express communicating a change of mailbox content to Cisco CallManager via CTI-QBE, and by Cisco CallManager in turn sending a MWI message to the phone to change the state of the lamp.
- The voice gateway communicates via either H.323 or MGCP to Cisco CallManager.
- Real-Time Transport Protocol (RTP) stream flows carry the voice traffic between endpoints.

Figure 12-4 shows the protocols involved in the call flow between the SRST router and Cisco Unity Express when the WAN link is down.

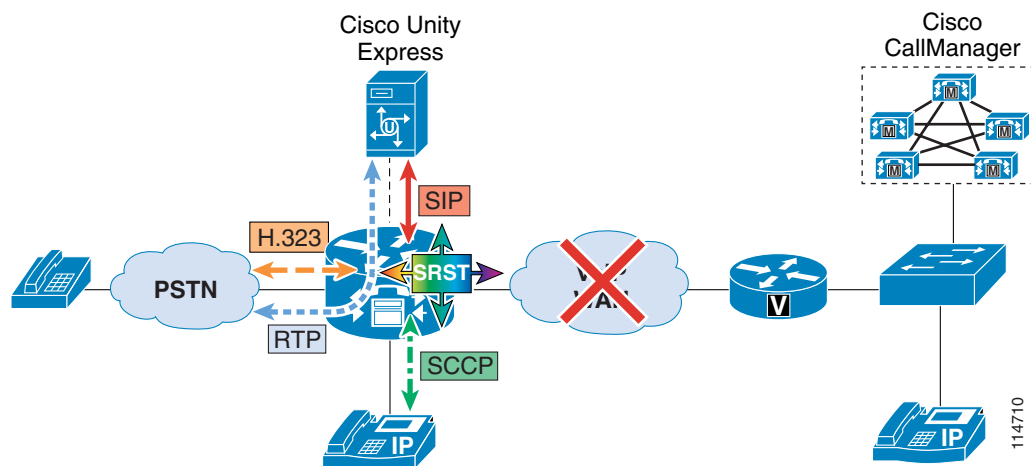
Figure 12-4 *Protocols Used Between Cisco Unity Express and the SRST Router*

Figure 12-4 illustrates the following signaling and media flows:

- Phones are controlled via SCCP from the SRST router.
- Cisco Unity Express communicates with the SRST router via an internal SIP interface.

- Currently, MWI changes are not supported in SRST mode. Voice messages can be sent and retrieved, as during normal operation, but MWI lamp state on the phone remains unchanged until the phone registers again with Cisco CallManager. At that time, all MWI lamp states are automatically resynchronized with the current state of the users' Cisco Unity Express voicemail boxes.
- The voice gateway communicates via H.323 to the SRST router.
- RTP stream flows carry the voice traffic between endpoints.

Best Practices for Deploying Cisco Unity Express

Use the following guidelines and best practices when deploying the Cisco Unity Express.

- Ensure that the IP phones having Cisco Unity Express as their voicemail destination are located on the same LAN segment as the router hosting Cisco Unity Express. Cisco Unity Express cannot provide mailboxes to phones that are not co-located with it.
- If uninterrupted AA and voicemail access is required for a site deployed with Cisco Unity Express, ensure that the Cisco Unity Express blade, SRST, and PSTN voice gateway interface are all integrated into the same physical router. Hot Standby Router Protocol (HSRP) or other redundant router configurations are not currently supported with Cisco Unity Express.
- Each mailbox can be associated with a primary extension number and a primary E.164 number. Typically, this number is the direct-inward-dial (DID) number that PSTN callers use. If the primary E.164 number is configured to any other number, use Cisco IOS translation patterns to match either the primary extension number or primary E.164 number so that the correct mailbox can be reached during SRST mode.
- Each Cisco Unity Express site must be associated with at least two CTI route points and four CTI ports. Ensure that the number of sites with Cisco Unity Express does not exceed the CTI scalability guidelines presented in the chapter on [Call Processing](#), page 8-1.
- Cisco Unity Express is associated with a JTAPI user on Cisco CallManager. Although a single JTAPI user can be associated with multiple Cisco Unity Expresses in a system, Cisco recommends associating each dedicated JTAPI user in Cisco CallManager with a single Cisco Unity Express.
- Calls into Cisco Unity Express use G.711 only. Cisco recommends using a local transcoder to convert the G.729 calls traversing the WAN into G.711 calls. You can configure Cisco CallManager regions with the G.711 voice codec for intra-region calls and the G.729 voice codec for inter-region calls.
- If transcoding facilities are not available at the Cisco Unity Express site, provision enough bandwidth for the required number of G.711 voicemail calls over the WAN. Configure the Cisco CallManager regions with the G.711 voice codec for calls between the IP phones and Cisco Unity Express devices (CTI ports and CTI route points).
- The CTI ports and CTI route points can be defined in specific locations. Cisco recommends using location-based call admission control between Cisco CallManager and Cisco Unity Express.
- Ensure proper Quality of Service (QoS) and bandwidth for signaling traffic that traverses the WAN between Cisco Unity Express and Cisco CallManager. Provision 20 kbps of bandwidth for CTI-QBE signaling for each Cisco Unity Express site. See the chapter on [Network Infrastructure](#), page 3-1, for more details.
- The CTI-QBE signaling packets from Cisco CallManager to Cisco Unity Express are marked with a DSCP value of AF31 (0x68). However, CTI-QBE signaling packets from Cisco Unity Express to Cisco CallManager are not marked. Cisco recommends using access control lists (ACLs) to mark

these CTI-QBE packets to ensure proper QoS. [Example 12-1](#) shows a sample configuration for achieving proper QoS for CTI-QBE signaling packets. Cisco CallManager uses TCP port 2748 for CTI-QBE signaling.

Example 12-1 QoS Configuration for CTI-QBE Signaling Packets

```
access-list 101 permit tcp host a.b.c.d any eq 2748
!
class-map match-all cti-qbe
  match access-group 101
!
policy-map cti-qbe
  class cti-qbe
    set dscp af31
    bandwidth 20
!
interface Serial0/1
  service-policy output cti-qbe
```

Third-Party Voicemail Systems

There are many voicemail vendors, and it is not uncommon for customers to want to continue to use an existing voicemail system when deploying Cisco CallManager. With this requirement in mind, Cisco provides support for the industry standard voicemail protocol known as Simplified Message Desk Interface (SMDI). SMDI is a serial protocol that provides all the necessary call information required for a voicemail system to answer calls appropriately.

There are also other options for integrating Cisco CallManager to voicemail systems, such as digital set emulation, Microsoft TAPI, QSIG, and so forth. Each method has its own pros and cons, and the method you employ will largely depend on how your voicemail system is integrated to your current PBX.

This section covers the following aspects of integrating third-party voicemail systems with Cisco CallManager:

- [SMDI, page 12-7](#)
- [Digital Set Emulation, page 12-9](#)
- [Centralized Voicemail, page 12-12](#)
- [Positive Disconnect Supervision, page 12-14](#)
- [Summary of Third-Party Voicemail Integration, page 12-15](#)

SMDI

Cisco CallManager supports use of SMDI through either of the following methods:

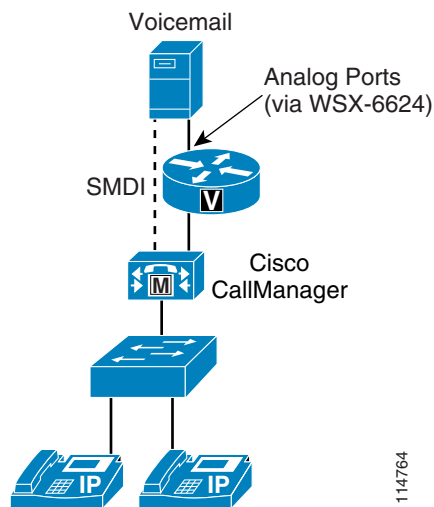
- [Cisco Messaging Interface, page 12-8](#)
- [Cisco VG248, page 12-8](#)

Cisco Messaging Interface

The Cisco Messaging Interface (CMI) is a Cisco CallManager service that should be run only on the publisher server. This service intercepts calls destined for voicemail and generates appropriate SMDI messages, which are then delivered to one of the server's Component Object Model (COM) ports. The CMI is compatible with any MGCP gateway that supports analog FXS ports or T1 CAS E&M; however, the WS-X6624 module is one of only two gateways that support positive disconnect supervision (see [Positive Disconnect Supervision, page 12-14](#)), and it is therefore the only gateway currently recommended for use with the CMI.

[Figure 12-5](#) illustrates the use of SMDI through the CMI in Cisco CallManager.

Figure 12-5 SMDI via Cisco CallManager



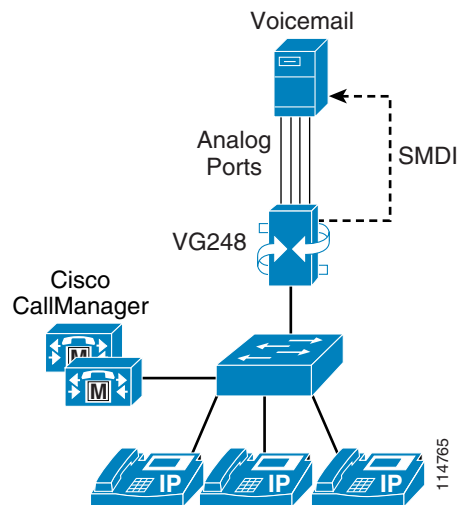
Through the CMI, Cisco CallManager supports integration with virtually any voicemail system that can provide SMDI with analog FXS ports, including (but not limited to) the following:

- Octel 100, 200/300, and 250/350
- Intuity Audix
- Siemens PhoneMail
- Centigram/BayPoint (OnePoint and NuPoint Messenger)
- Lyrinx ECS
- IBM Message Center

Cisco VG248

The Cisco VG248 is an SCCP gateway that supports 48 analog FXS ports and generates SMDI locally (that is, it runs independent of the CMI service). As with the WS-X6624 module, the VG248 also supports positive disconnect supervision.

[Figure 12-6](#) illustrates the use of SMDI with the VG248.

Figure 12-6 SMDI via the VG248

Voicemail integration through the VG248 provides the following features and advantages:

- Multiple SMDI links per Cisco CallManager
- SMDI failover capability
- Independence from the location of the voicemail system

The VG248 is also capable of supporting two other serial protocols that are sometimes used for voicemail integration: NEC Message Center Interface (MCI) and Ericsson MD110 proprietary protocol.

Considerations When Using FXS Ports

If your voicemail system is equipped with analog FXS ports, use the following Cisco gateways to integrate with the voicemail system:

- WS-X6624

Use this module when there is a slot available for it in the Catalyst 6500 chassis.

- VG248

Use this gateway when full failover is required for the serial port as well as voice ports, when serial protocols other than SMDI are required (such as NEC MCI or Ericsson MD110), or when no Catalyst 6500 chassis slot is available.

Digital Set Emulation

Digital set emulation (DSE) is another method of integrating PBXs to voicemail systems. In this mode, the voicemail ports look like proprietary digital handsets to the PBX. This integration method has the following advantages over the use of SMDI with analog FXS ports:

- The circuit is completely digital for both the voice path and signaling of call information.
- There is no out-of-band signaling because the voice and signaling are both transported over the same physical circuit.
- Overall quality of the call is generally better.

Digital PBX Adapter (DPA)

Cisco developed the Digital PBX Adapter (DPA) specifically to integrate Cisco CallManager with third-party voicemail systems via digital set emulation. The DPA essentially functions as a number of digital PBX extensions on one side, with IP connectivity on the other. The DPA enables you to maintain your existing voicemail system, as well as its interfaces, while connecting to Cisco CallManager.

The DPA is available in two variants:

- DPA 7630 for emulating the Avaya Definity G3 7400 series digital telephone set
- DPA 7610 for emulating the Nortel Meridian 1 2616 digital telephone set

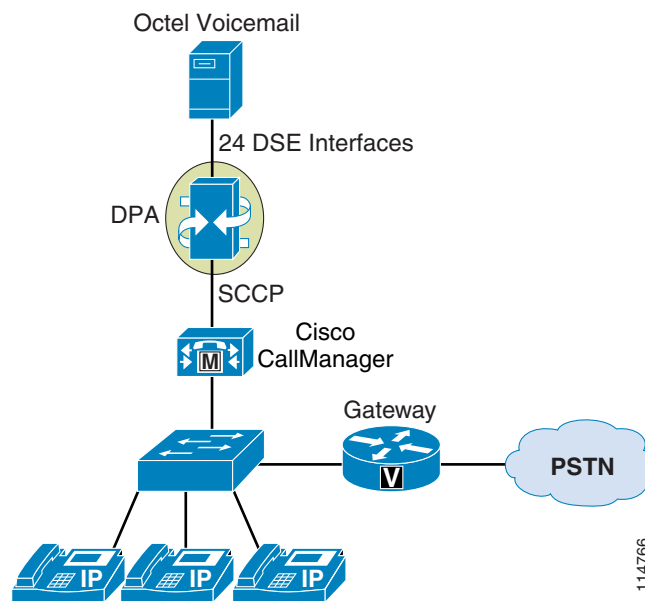


Note

The DPA will work with Octel Aria 250/350 or Serenade 200/300 voicemail systems only when using either Lucent/Avaya or Nortel digital set emulation.

Figure 12-7 illustrates the DPA integrating an Octel voicemail system with Cisco CallManager.

Figure 12-7 DPA Integrating Octel Voicemail with Cisco CallManager



Dual PBX Integration

Dual PBX integration is a useful option for enterprises that want to maintain existing voicemail services while migrating from their current PBX to IP Telephony.

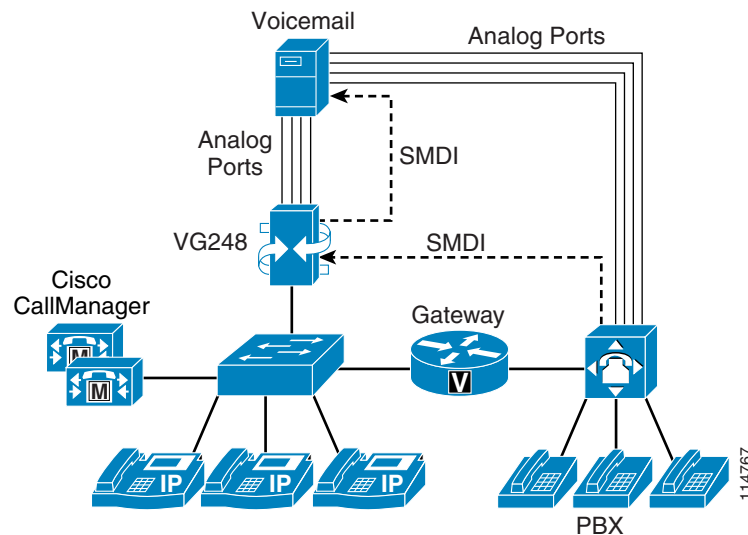


Note

Most voicemail vendors do not support this scenario due to its complex nature, but some will provide support on a "site-specific" basis if requested. Consult with your voicemail vendor before attempting to implement this solution.

The Cisco VG248 has inherent multiplexing capabilities that enable it to provide dual integration. The VG248 can combine information from an existing serial link with its own link, and then present a single serial stream to the voicemail system. (See [Figure 12-8](#).)

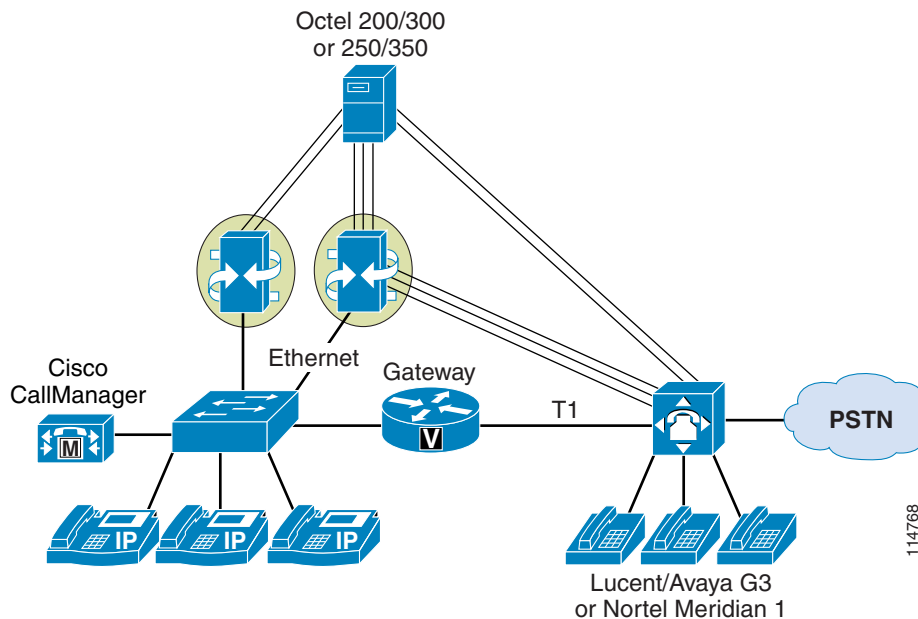
Figure 12-8 Dual Integration via the VG248 and SMDI



The VG248 works with any voicemail system that has SMDI capability and analog FXS ports. The following prerequisites are required prior to implementation, assuming a dual integration is required:

- Uniform dial plan
- Transfer and reconnect sequences
- Connectivity between the PBX and Cisco CallManager

The Cisco DPA also has the capability to provide dual integration through digital set emulation, as illustrated in [Figure 12-9](#).

Figure 12-9 Dual Integration via DPA

The DPA works with Octel digital set emulation. The following prerequisites are required prior to implementation, assuming a dual integration is required:

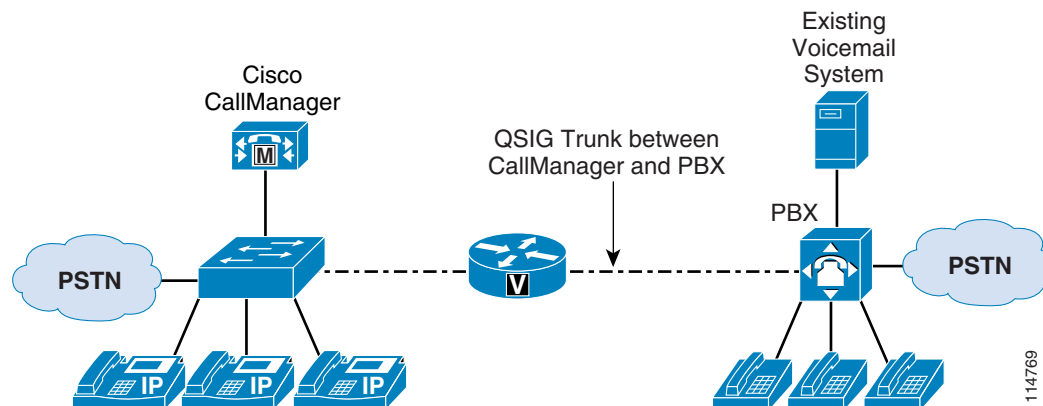
- Uniform dial plan
- Transfer and reconnect sequences
- Connectivity between the PBX and Cisco CallManager

Centralized Voicemail

In a centralized voicemail deployment, two or more PBXs share a single voicemail system. The sharing is achieved by integrating the voicemail system to only one PBX and then utilizing an inter-PBX private networking protocol to extend voicemail services to remote subscribers. The networked PBXs look and act like one large PBX to the voicemail system. Various PBX manufacturers have developed proprietary protocols that enable the delivery of such services as well as providing feature transparency to subscribers across a large network (for example, Avaya DCS, Nortel MCDN, Siemens CorNet, Alcatel ABC, NEC CCIS, and Fujitsu FIPN).

The primary motivation for using a centralized voicemail system stems from the desire to offer voicemail services to IP Telephony subscribers from the existing voicemail system so that the subscribers do not have to learn a new telephony user interface (TUI).

Some voicemail systems are capable of supporting multiple PBXs (dual PBX integration) via protocols such as Simple Messaging Desktop Interface (SMDI). Other solutions have been introduced, such as the Cisco Digital PBX Adapter (DPA), which also allow for dual integration. In some circumstances, these solutions are either not possible because the voicemail vendor may choose not to support this configuration, or a dual integration is simply not technically possible because the voicemail system cannot support dissimilar PBX integrations simultaneously. In such circumstances, a centralized voicemail deployment provides an alternative solution to dual integration. (See [Figure 12-10](#).)

Figure 12-10 Centralized Voicemail with Cisco CallManager and QSIG

Note that the term *centralized voicemail* does not refer to the voicemail system itself. Centralized voicemail is a function of the PBX's inter-PBX networking protocol (either a proprietary protocol such as Avaya DCS, Nortel MCDN, or Siemens CorNet or a standards-based protocol such as QSIG or DPNSS), which is needed to deliver the voicemail features.

The following important terms and concepts apply to centralized voicemail:

- Message center Private Integrated Services Network Exchange (PINX) — This is the PBX that is "hosting" the voicemail system (the PBX directly connected to the voicemail system).
- Subscriber PINX — This is the PBX that is "remote" from the voicemail system (not directly connected to the voicemail system).

A centralized voicemail configuration requires a suitable inter-PBX networking protocol such as QSIG. This protocol must also deliver the following minimum level of feature support:

- Message Waiting Indication (MWI)
- Transfer — Needed to ensure that the correct calling and called party IDs are delivered to the voicemail system.
- Divert — Needed to ensure that the correct calling and called party IDs are delivered to the voicemail system.

Other features might also be required, depending on how the voicemail system will be used. For example, if the voicemail system is also serving as an automated attendant, then the Path Replacement feature is needed to prevent calls from hair-pinning.

Not all PBXs are capable of serving as the message center PINX. In this case, consider relocating the voicemail system to Cisco CallManager and have Cisco CallManager act as the message center PINX, with the PBX acting as the subscriber PINX. (See [Figure 12-10](#).)

Support

Cisco cannot guarantee that another vendor's product will act in a particular manner, nor can Cisco specify what is required in terms of configuration changes or upgrades to another vendor's product. It is the responsibility of the customer to ask these questions and seek confirmation directly from the supplier and/or vendor of each product.

Cisco can assist you in determining which particular questions to ask your supplier and/or vendor, such as: "What do I have to do to my PBX to enable remote PBX users, connected via QSIG, to have a mailbox as well as full access to all voicemail features such as MWI?"

To help with PBX interoperability, Cisco has tested a number of different PBXs with Cisco CallManager and has documented these tests in the form of Application Notes. These documents, while not a guarantee of success, do provide some level of guidance in terms of features supported as well as configuration details for both Cisco CallManager and the PBX. Application Notes for Cisco CallManager have already been written for the leading PBXs, and they cover the scenario of centralized voicemail with Cisco CallManager acting as the message center PINX. The Application Notes are available at

<http://www.cisco.com/go/interoperability>

**Note**

It is not feasible for Cisco to test other vendors' PBXs acting as the message center PINX. Cisco has neither the facilities nor the expertise to configure these systems, therefore customers must request this information directly from their supplier and/or vendor.

Summary

- Centralized voicemail is a function of the inter-PBX networking protocol, not the voicemail system itself.
- Not all PBXs can act as the message center PINX. Customers *must* confirm this feature with their PBX supplier and/or vendor; Cisco cannot provide or support this feature on the PBX.
- Cisco CallManager can act as the message center PINX, thus providing customers with an alternative if their PBX cannot perform this function.
- Confirm if Path Replacement is needed. Cisco CallManager Release 4.1 and later supports this feature.

Positive Disconnect Supervision

Positive disconnect supervision is a signal sent from a PBX port to the voicemail system to indicate that the far-end device has gone on-hook. This signal typically takes the form of a drop in loop current for approximately 600 ms, causing the voicemail system to terminate the session.

Without this signal, the voicemail system would be unaware that the far-end device has gone on-hook and would continue to record whatever supervisory tones the PBX provides under this condition. (For example, some PBXs play dial tone while others play busy tone.) The voicemail system would continue to record these tones until the maximum message time has expired. (For example, if the mailbox has a limit is 3 minutes per message and a caller hangs up after 30 seconds, then the voicemail system would continue to record such tones for another 2.5 minutes in the absence of positive disconnect supervision.) This unnecessary recording can be annoying to subscribers and can impact system performance by increasing disk usage as well as causing higher port usage times. Some voicemail systems are able to deal with this scenario by monitoring for known tones and then deleting them, but system performance is still impacted in this case.

A similar issue exists when subscribers call into their mailboxes to check for messages. If a user simply hangs up without disconnect supervision, the voicemail system will stay on the line waiting for a valid response before any activity timers eventually expire. In this scenario, the main impact is from the additional port usage time incurred.

For these reasons, positive disconnect supervision must be provided by the analog ports connecting to the voicemail system.

Summary of Third-Party Voicemail Integration

There are other methods for connecting voicemail systems to Cisco CallManager (such as Microsoft TAPI and PRI ISDN trunks in conjunction with SMDI), but these methods are uncommon. The vast majority of third-party voicemail integrations use the Cisco VG248, and this is the recommended solution.

The method of integrating Cisco CallManager to the voicemail system depends on how the voicemail system is currently integrated to the PBX. If analog ports are currently in use, then the Cisco VG248 provides the best integration method. However, if Avaya or Nortel digital set emulation is currently in use, then the Cisco DPA can provide a solution at lower cost by integrating with your current voicemail system without the need to re-engineer it for analog FXS ports.

**Note**

Cisco does not test or certify any third-party voicemail systems. Within the industry, it is generally considered to be the responsibility of the voicemail vendor to test and/or certify their products with various PBX systems. Cisco does, of course, test its interfaces to such equipment and will support these interfaces regardless of which third-party voicemail system is connected.



Cisco Unity

This chapter focuses on the design aspects of integrating Cisco Unity with Cisco CallManager. The design topics covered in this chapter apply to both voicemail and unified messaging deployments.

Additionally, this chapter discusses design issues for deploying Cisco Unity with Microsoft Exchange 2000 or 2003 or Lotus Notes Domino message stores and Microsoft Windows 2000 or 2003. This chapter does *not* cover deployments or upgrades from Microsoft NT 4.0 and/or Exchange 5.5.

For additional design information about Cisco Unity, including integrations with other non-Cisco messaging systems, refer to the *Cisco Unity Design Guide*, available at

<http://www.cisco.com>

This chapter covers the following Cisco Unity design topics:

- [Messaging Deployment Models, page 13-2](#)
- [Messaging System Infrastructure Components, page 13-4](#)
- [Managing Bandwidth, page 13-4](#)
- [Cisco Unity Native Transcoding Operation, page 13-6](#)
- [Voice Port Integration with a Cisco CallManager Cluster, page 13-6](#)
- [Voice Port Integration with Dedicated Cisco CallManager Backup Servers, page 13-12](#)
- [Centralized Messaging and Centralized Call Processing, page 13-13](#)
- [Distributed Messaging with Centralized Call Processing, page 13-15](#)
- [Combined Messaging Deployment Models, page 13-17](#)
- [Centralized Messaging with Clustering Over the WAN, page 13-18](#)
- [Distributed Messaging with Clustering Over the WAN, page 13-20](#)
- [Cisco Unity Messaging Failover, page 13-21](#)
- [Cisco Unity Failover and Clustering Over the WAN, page 13-22](#)
- [Centralized Messaging with Multiple Cisco CallManager Servers, page 13-23](#)

Cisco CallManager Release 4.0 introduced *line groups*, *hunt lists*, and *hunt pilots*, which can impact the operation of existing voice ports. Before upgrading from older versions of Cisco CallManager to Cisco CallManager Release 4.0 or later, refer to the appropriate *Cisco CallManager Administration Guide* for your 4.x release, available at

<http://www.cisco.com>

All of the deployment models and design considerations discussed in this chapter are fully supported by Cisco CallManager Release 3.1 and later.

Messaging Deployment Models

Cisco Unity supports three primary messaging deployment models:

- [Single-Site Messaging, page 13-2](#)
- Multi-site WAN deployment with [Centralized Messaging, page 13-2](#)
- Multi-site WAN deployment with [Distributed Messaging, page 13-2](#)

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

In addition to the three messaging deployment models, Cisco Unity also supports messaging failover. (See [Messaging Failover, page 13-3](#).)

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

Single-Site Messaging

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

Centralized Messaging

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

Because messaging clients may be either local or remote from the messaging system, special design considerations apply to the following graphical user interface (GUI) clients: ViewMail for Outlook (VMO) with Microsoft Exchange and Domino Unified Communications Services (DUC) with Lotus Domino, and the use of the telephone record and playback (TRaP) and message streaming features. Remote clients should not use TRaP and should be configured to download messages before playback. Because different features and operations for local and remote clients can cause user confusion, TRaP should be disabled on the voice ports and GUI clients should be configured to download messages and not use TRaP, regardless of whether the client is local or remote. Cisco Unity inbox accessed via Cisco Unity Personal Assistant (CPCA) should be allowed only for local clients. The Cisco Unity telephone user interface (TUI) operates the same way for both local and remote clients.

Distributed Messaging

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

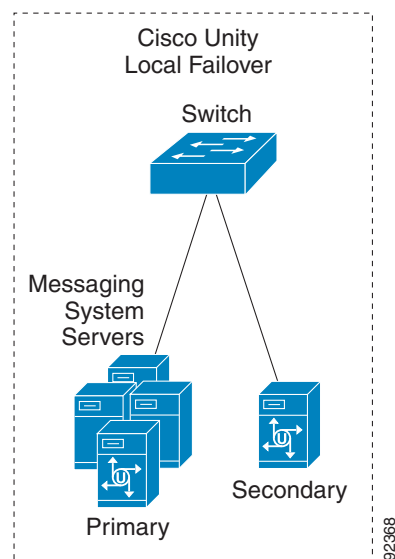
message routing infrastructure. No messaging infrastructure components should be separated by a WAN from the messaging system they service. Distributed messaging is essentially multiple, single-site messaging models with a common messaging backbone.

The distributed messaging model has the same design criteria as centralized messaging with regard to local and remote GUI clients, TRaP, and message downloads.

Messaging Failover

All three messaging deployment models support messaging failover. You can implement local messaging failover as illustrated in Figure 13-1. With local failover, both the primary and secondary Cisco Unity servers are located at the same site on the same highly available LAN.

Figure 13-1 Local Failover of Cisco Unity Messaging



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

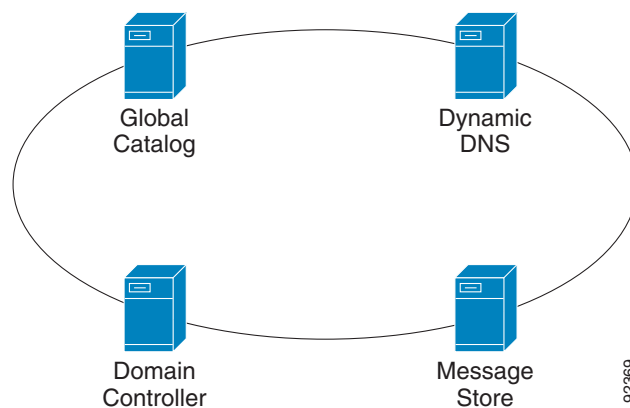
- Single-site messaging and single-site call processing
- Centralized messaging and centralized call processing
- Distributed messaging and centralized call processing
- Centralized messaging and distributed call processing
- Distributed messaging and distributed call processing

Refer to the *Cisco Unity Design Guide* at www.cisco.com for further details on site classification and a detailed analysis of supported combinations of messaging and call processing deployment models.

Messaging System Infrastructure Components

Cisco Unity interacts with various network resources, include Dynamic Domain Name Server (DDNS), directory servers, and message stores. (See [Figure 13-2](#).) Instead of viewing Cisco Unity as a single monolithic device, it is better to think of it as a system of interdependent components. Each messaging system infrastructure component in a Cisco Unity messaging system is required for normal operation, and it is critical that all of these components be on the same highly available LAN. (In most cases they will be physically co-located.) Any WAN links between these components can introduce delays that will impact the operation of Cisco Unity. These delays will manifest themselves as long delays and periods of silence during TUI operation. For more information, refer to the *Network and Infrastructure Considerations* chapter in the *Cisco Unity Design Guide*, available from www.cisco.com.

Figure 13-2 Cisco Unity Messaging System Infrastructure Components



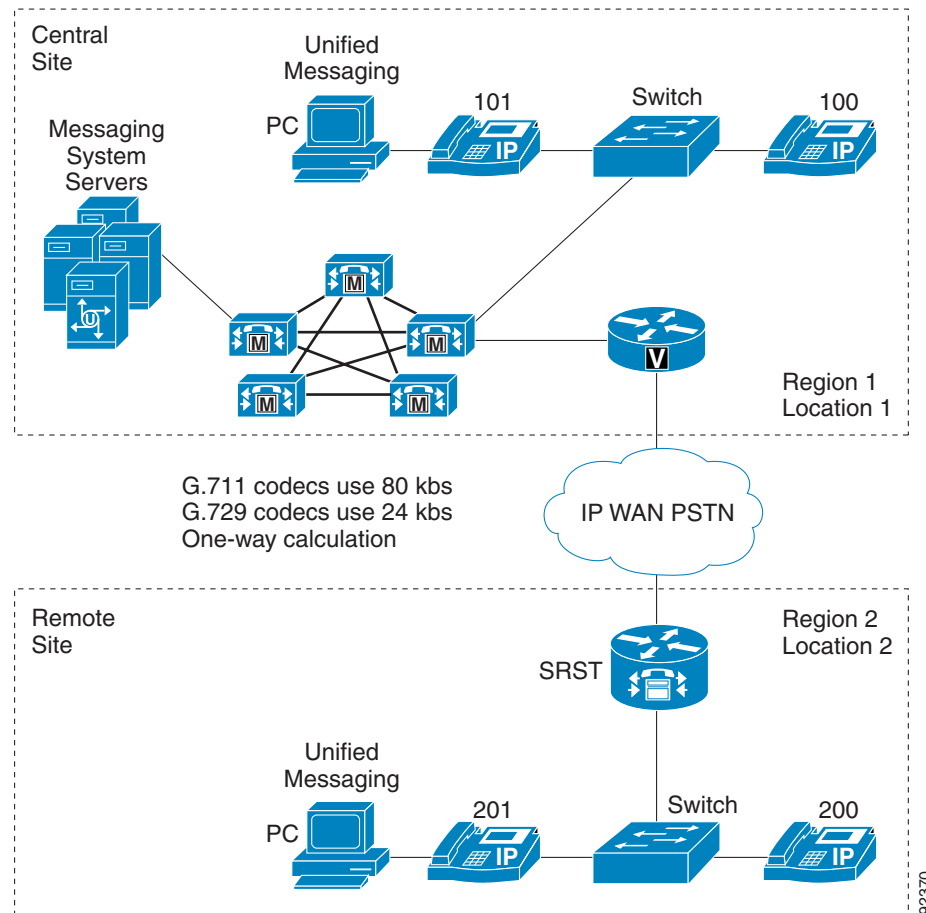
[Figure 13-2](#) is a logical representation of the message system infrastructure components. Some of these components can be located on the same server. In the case of Domino Lotus Notes, the message store and directory (Names.nsf) are located on the same server. Microsoft Windows, the global catalog server, and domain controller may also be on the same server. You can also use multiple instances of each component, as in message store clustering, which Cisco Unity supports for Microsoft Exchange 2000 or 2003 and Lotus Domino. All messaging system infrastructure components must be located on the same highly available LAN as the Cisco Unity server(s) that they service. When Cisco Unity is deployed with distributed messaging, each site (location) must have its own complete set of message system components.

Managing Bandwidth

Cisco CallManager provides a variety of features for managing bandwidth. Through the use of regions, locations, and even gatekeepers, Cisco CallManager can ensure that the number of voice calls going over a WAN link does not oversubscribe the existing bandwidth and cause poor voice quality. Cisco Unity relies on Cisco CallManager to manage bandwidth and to route calls. If you deploy Cisco Unity in an environment where calls or voice ports might cross WAN links, these calls will be transparent to gatekeeper-based call admission control. This situation occurs any time the Cisco Unity server is servicing either distributed clients (distributed messaging or distributed call processing) or when Cisco CallManager is remotely located (distributed messaging or centralized call processing). The only alternative to gatekeeper-based call admission control is to use Cisco CallManager regions and locations.

Figure 13-3 uses a small centralized messaging and centralized call processing site to illustrate how regions and locations work together to manage available bandwidth. For a more detailed discussion of regions and locations, refer to the chapter on [Call Admission Control](#), page 9-1.

Figure 13-3 *Locations and Regions*



In Figure 13-3, regions 1 and 2 are configured to use G.711 for intra-region calls and G.729 for inter-region calls. Locations 1 and 2 are both set to 24 kbps. Location bandwidth is budgeted only in the case of inter-location calls.

An intra-region (G.711) call would not be budgeted against the available bandwidth for the location. For example, when extension 100 calls extension 101, this call is not budgeted against the 24 kbps of available bandwidth for Location 1. However, an inter-region call using G.729 is budgeted against both bandwidth allocations of 24 kbps for Location 1 and Location 2. For example, when extension 100 calls extension 200, this call would be connected but any additional (simultaneous) inter-region calls would receive reorder (busy) tone.

Impact of Non-Delivery of RDNIS on Voicemail Calls Routed via AAR

Automated alternate routing (AAR), a Cisco CallManager feature, can route calls over the PSTN when the WAN is oversubscribed. However, when calls are rerouted over the PSTN, Redirected Dialed Number Information Service (RDNIS) can be affected. Incorrect RDNIS information can impact voicemail calls that are rerouted over the PSTN by AAR when Cisco Unity is remote from its messaging clients. If the RDNIS information is not correct, the call will not reach the voicemail box of the dialed

user but will instead receive the auto-attendant prompt, and the caller might be asked to reenter the extension number of the party they wish to reach. This behavior is primarily an issue when the telephone carrier is unable to ensure RDNIS across the network. There are numerous reasons why the carrier might not be able to ensure that RDNIS is properly sent. Check with your carrier to determine if they provide guaranteed RDNIS deliver end-to-end for your circuits. The alternative to using AAR for oversubscribed WANs is simply to let callers hear reorder tone in an oversubscribed condition.

Cisco Unity Native Transcoding Operation

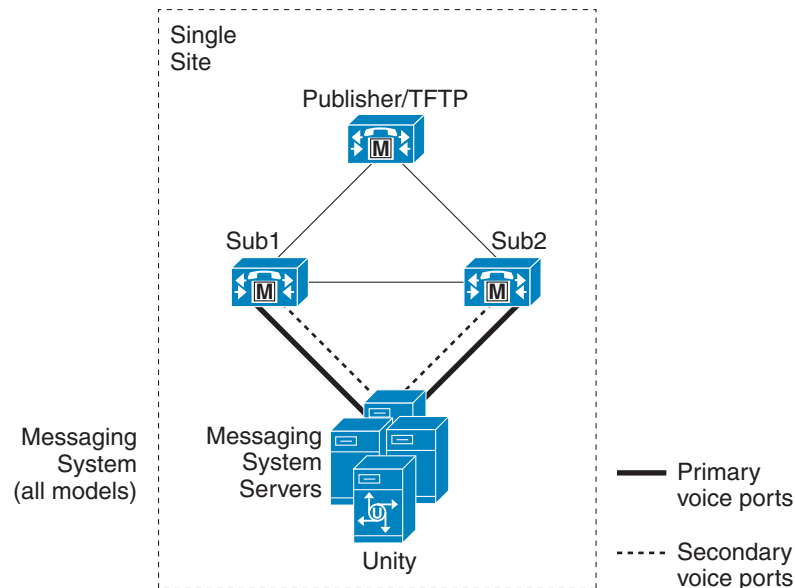
By default, the Cisco Unity server performs transcoding automatically. Currently, Cisco CallManager and Cisco Unity support only G.729 and G.711 for the SCCP TAPI Service Provider (TSP) voice ports, and other codecs are supported with legacy integration using Intel or Dialogic voice boards. Cisco Unity native transcoding does not use external hardware transcoders but instead uses the server's main CPU. In order for Cisco CallManager to assign hardware transcoders to voice port calls, the native transcoding must be disabled (turned off) on the Cisco Unity server via a registry setting. The registry setting is called "Audio - Enable G.729a codec support," and the tool for setting it is the *Advanced Settings Tool* available at www.CiscoUnityTools.com. By default, there is no codec registry key, and therefore native transcoding is enabled (on). The Advanced Settings Tool adds a new registry key that enables you to select either one of the two available codecs. Cisco Unity will then "advertise" to Cisco CallManager that it supports only one codec. If a call terminating or originating on the voice ports is using a different codec than the type configured for the Cisco Unity server, Cisco CallManager will assign an external transcoding resource for the call. For detailed information on how to configure transcoding resources on Cisco CallManager, refer to the chapter on [Media Resources](#), page 6-1.

When the Advanced Settings Tool is used to enable only one codec, the Cisco Unity server system prompts must be the same as the codec used. By default, the system prompts are G.711. If the codec is set to G.711, the system prompts will work correctly. However, if G.729 is selected, you will have to change the system prompts. Refer to the *Cisco Unity Administration Guide* on www.cisco.com for details on how to change the system prompts. If the system prompts are not the same as the codec selected by the registry, then the end users will hear unintelligible system prompts.

Voice Port Integration with a Cisco CallManager Cluster

When deploying Cisco Unity in a single-site messaging environment, integration with the Cisco CallManager cluster occurs through the SCCP voice ports. Design considerations must include proper deployment of the voice ports among the Cisco CallManager subscribers so that, in the event of a subscriber failure (Cisco CallManager failover), users and outside calls can continue to access voice messaging. (See [Figure 13-4](#).)

Figure 13-4 Cisco Unity Server(s) Integrated with a Cisco CallManager Cluster (No Dedicated Backup Servers)



92371

The Cisco CallManager cluster in [Figure 13-4](#) employs 1:1 server redundancy and 50/50 load balancing. During normal operations, each subscriber server is active and handles 50% of the total server call processing load. In the event of a subscriber server failure, the remaining subscriber server takes up the load of the failed server.

[Figure 13-5](#) and [Figure 13-6](#) show the voicemail port configurations in the Cisco Unity Telephone Integration Manager (UTIM). The configuration uses two groups of voicemail ports, with each group containing one-half of the total number of licensed voice ports. One group is configured so that its primary server is SUB1 and its secondary (backup) server is SUB2 (see [Figure 13-5](#)). The second group is configured so that SUB2 is the primary server and SUB1 is the backup (see [Figure 13-6](#)).

Make sure that MWI-only ports or any other special ports are equally distributed between the two groups. During the configuration of the voice ports, pay special attention to the naming convention. When configuring the two groups of ports in the Cisco UTIM utility, make sure that the device name prefix is unique for each group and that you use the same device name when configuring the voicemail ports in Cisco CallManager Administration. As shown in [Figure 13-5](#) and [Figure 13-6](#), the device name prefix is unique for each group of ports in this example, with group SUB1 using CiscoUM1 as the device name prefix and SUB2 using CiscoUM2.

For additional design information on the ratio of inbound to outbound voicemail ports (for MWI, message notification, and TRaP), refer to the *Cisco Unity Design Guide* available at www.cisco.com.

Figure 13-5 Cisco Unity Configuration for Voice Ports for Subscriber 1

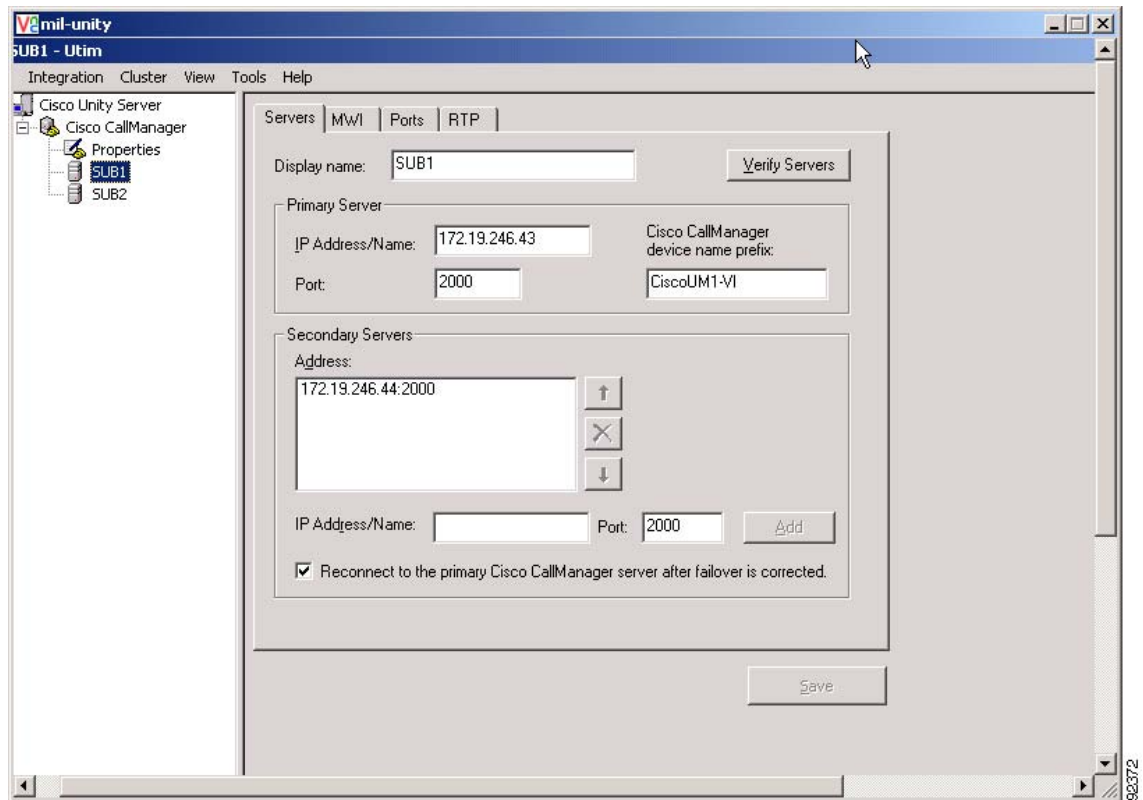
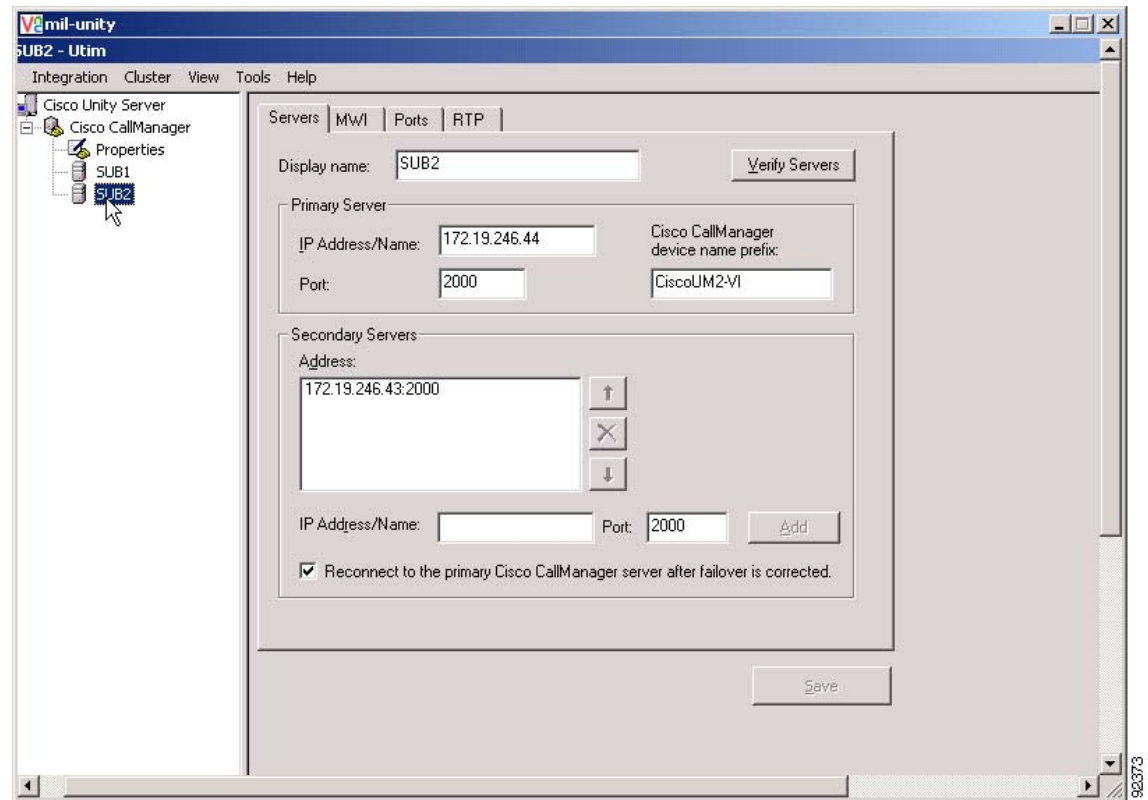


Figure 13-6 Cisco Unity Configuration for Voice Ports for Subscriber 2**Note**

The device name prefix is unique for each group of ports and must match the same naming convention for the voicemail ports configured in Cisco CallManager Administration.

In Cisco CallManager Administration, half of the ports in this example are configured to register using the unique device name prefix of CiscoUM1, and the other half are configured to register using CiscoUM2. (See [Figure 13-7](#).) When the ports register with Cisco CallManager, half will be registered with subscriber server Sub1 (see [Figure 13-4](#)), and the other half will be registered with Sub2.

Figure 13-7 Voicemail Port Configuration in Cisco CallManager Administration with Cisco CallManager Release 3.3 and Earlier

The screenshot shows the Cisco CallManager Administration interface. The main heading is "Find and List Voice Mail Ports". Below it, a search bar indicates "10 matching record(s) for Device Name begins with ''". The search criteria are set to "Device Name" and "begins with". The results table is as follows:

Device Name	Voicemail System	Voicemail Type	Subscript	IP Address
CiscoUM1-VI3	Unity	VM	Sub1	172.19.246.45
CiscoUM1-VI4	Unity	VM	Sub1	172.19.246.45
CiscoUM1-VI5	Unity-MWI Only	VM	Sub1	172.19.246.45
CiscoUM2-VI1	Unity	VM	Sub2	172.19.246.45
CiscoUM2-VI2	Unity	VM	Sub2	172.19.246.45
CiscoUM2-VI3	Unity	VM	Sub2	172.19.246.45
CiscoUM2-VI4	Unity	VM	Sub2	172.19.246.45



Note

The naming convention used for the voicemail ports in Cisco CallManager Administration must match the device name prefix used in Cisco UTIM, otherwise the ports will fail to register.

During normal operation of Cisco CallManager, the voicemail ports might be busy or might not answer (for example, during voicemail maintenance). With Cisco CallManager Release 3.3 and earlier, you must configure how the ports will operate in instances such as Forward Busy and Forward No Answer. Typically, each port would be forwarded to the next port in sequence (see Table 13-1). The exception to this general rule is the port just prior to the MWI-only port, which would skip the next port (the MWI-only port) and would be forwarded to the port after that. For example, Table 13-1 shows that CiscoUM1-VI4 (DN 1003) is forwarded to CiscoUM2-VI1 (DN 1005). The last port that is not an MWI-only port would be forwarded back to the first port, thus completing the circle. For optimum performance, limit the number of dedicated MWI ports to no more than 25% of the total number of voice ports, and disable MWI functionality on all other ports.

Table 13-1 Typical Voicemail Port Configuration for Forward Busy and Forward No Answer Conditions with Cisco CallManager Release 3.3 and Earlier

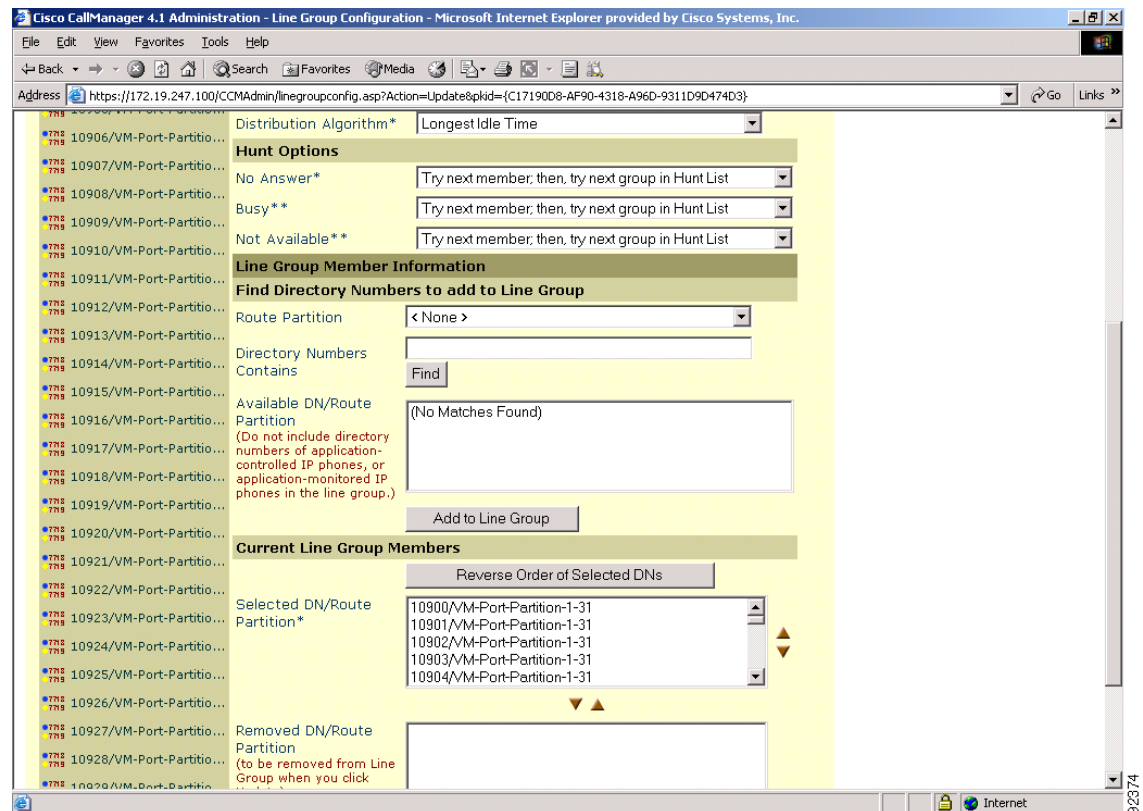
Voice Port Name	Directory Number	Forward Busy	Forward No Answer
CiscoUM1-VI1	1000	1001	1001
CiscoUM1-VI2	1001	1002	1002
CiscoUM1-VI3	1002	1003	1003
CiscoUM1-VI4	1003	1005	1005
CiscoUM1-VI5	1004	MWI-only port	

Table 13-1 Typical Voicemail Port Configuration for Forward Busy and Forward No Answer Conditions with Cisco CallManager Release 3.3 and Earlier (continued)

Voice Port Name	Directory Number	Forward Busy	Forward No Answer
CiscoUM2-VI1	1005	1006	1006
CiscoUM2-VI2	1006	1007	1007
CiscoUM2-VI3	1007	1008	1008
CiscoUM2-VI4	1008	1001	1001
CiscoUM2-VI5	1009	MWI-only port	

Cisco CallManager Release 4.0 introduced new features that substantially change how the voicemail ports are configured and operate. Release 4.0 and later use line groups, hunt lists, and hunt pilots to specify how the voicemail ports forward to each other based on availability, usage, and state. When configuring how the voicemail ports hunt, remove any MWI ports from the Current Line Group Members list (see Figure 13-8). This will keep the ports from being in the hunt group for answering calls and will ensure that the ports are free for MWI activity. Additionally, because the MWI ports should be configured via the Cisco UTIM utility not to answer calls, removing the ports from the line group will ensure that users accessing voicemail do not receive a ring-no-answer condition. For additional information about line groups, hunt lists and hunt pilots, refer to the *Cisco CallManager Administration Guide* for Release 4.0 and later, available at www.cisco.com.

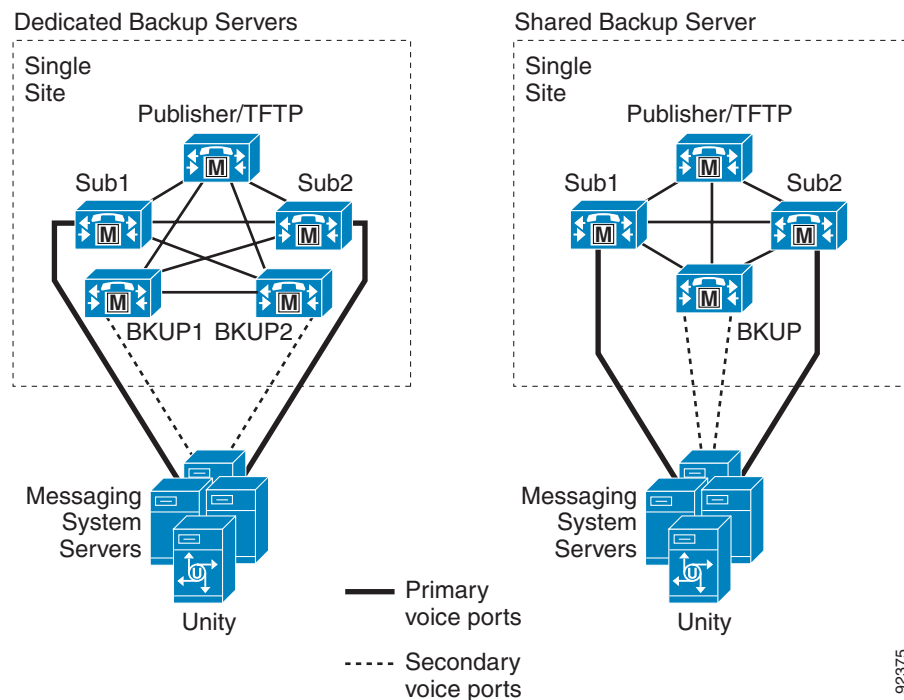
Figure 13-8 Line Group Configuration in Cisco CallManager Administration with Cisco CallManager Release 4.0 and Later



Voice Port Integration with Dedicated Cisco CallManager Backup Servers

This Cisco CallManager cluster configuration allows each subscriber server to operate at a call processing load higher than 50%. Each primary subscriber server has either a dedicated or shared backup server. (See [Figure 13-9](#).) During normal operation the backup server processes no calls, in the event of failure or maintenance of a Subscriber server, the backup server will then take the full load of that server.

Figure 13-9 Cisco Unity Server(s) Integrated with a Single Cisco CallManager Cluster with Backup Subscriber Server(s)

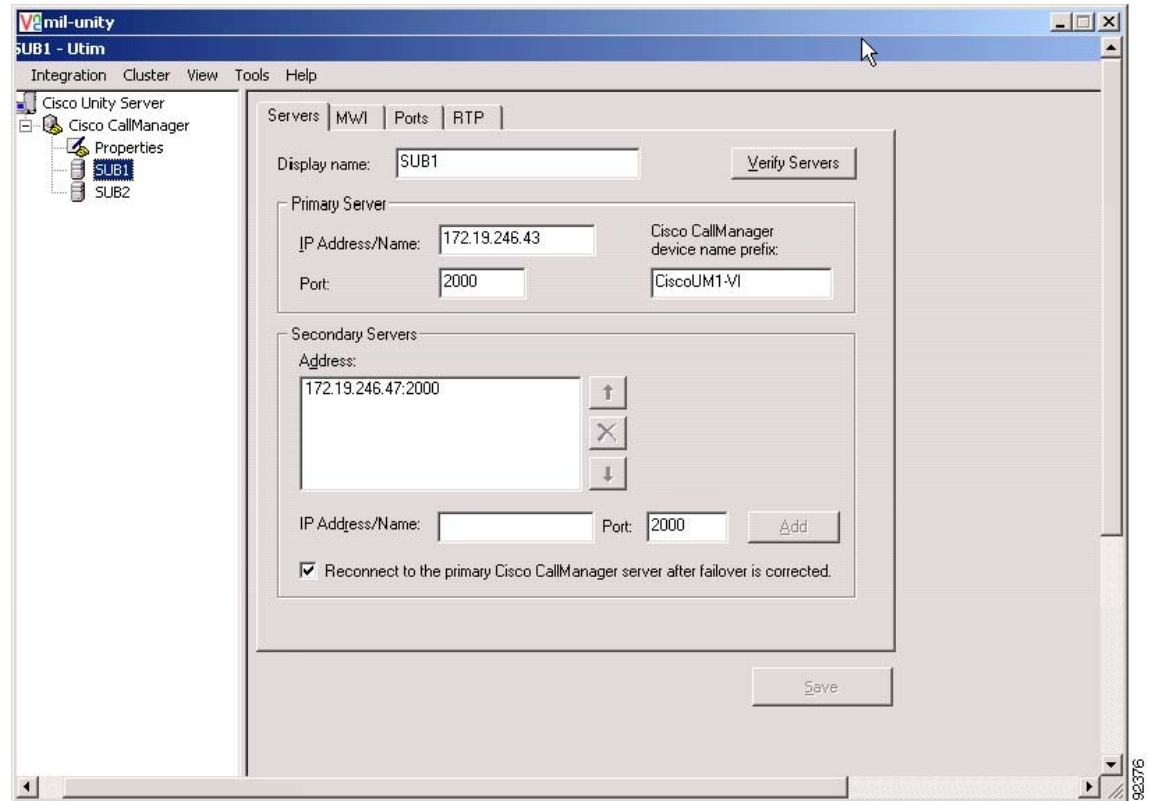


Configuration of the voicemail ports in this case is similar to the 50/50 load-balanced cluster. However instead of configuring the voice ports to use the opposite subscriber server as the secondary server, the individual shared or dedicated backup server is used. In the Cisco CallManager clustered with a shared backup server, both of the secondary ports for the subscribers servers are configured to use the single backup server.

The voice port names (device name prefix) must be unique for each Cisco UTIM group and must be the same as the device names used on the Cisco CallManager server.

The configuration for this example includes two groups of voicemail ports in the Cisco UTIM utility. The Secondary Servers box contains the IP address or DNS name of the backup server. (See [Figure 13-10](#).) If the cluster has a shared backup server, then the same server is used for both groups. When configuring the voice ports, make sure that there are equal numbers of MWI-only and fully functional voice ports in both groups of ports. Also make sure that Audio Messaging Interchange Specification (AMIS) and TRaP voice ports are equally distributed between the two groups. Configure the ports to call-forward-busy and call-forward-no-answer as indicated in [Table 13-1](#).

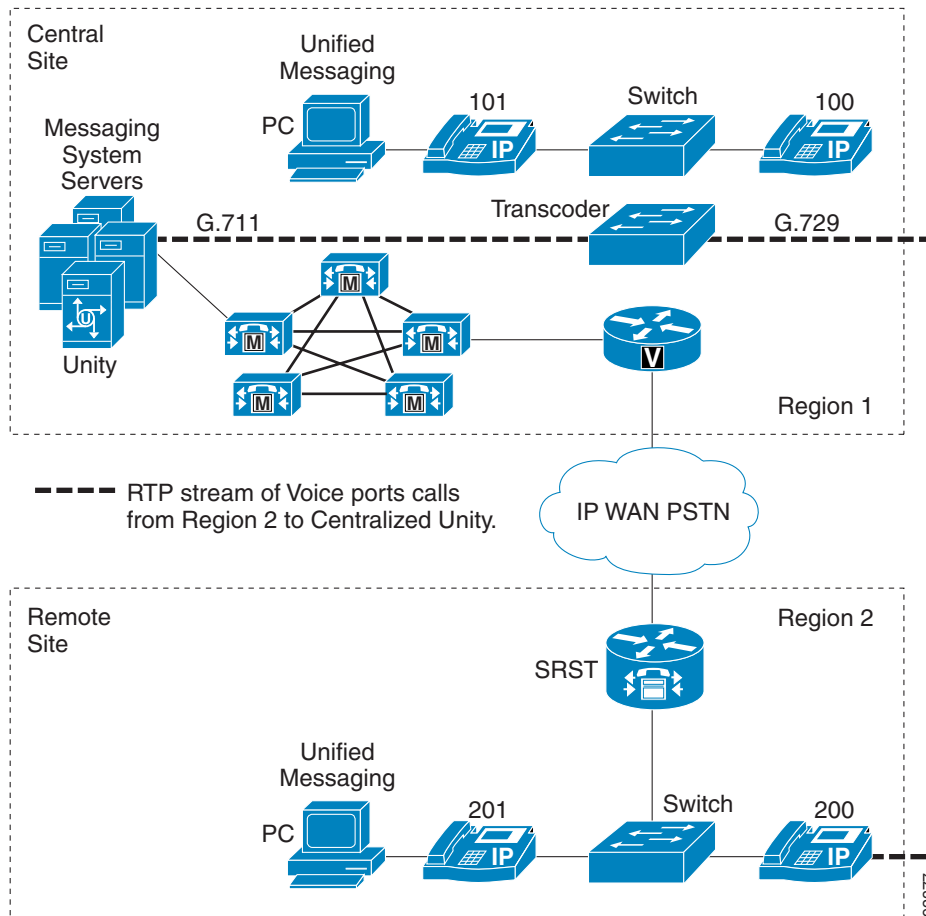
Figure 13-10 Cisco UTIM Configuration Screen for Subscriber Server (Sub1) Voice Ports with Backup Server (Bkup1) as the Secondary



To have the voice ports reconnect to the subscriber server after service has been restored, check the box labeled “Reconnect to the primary Cisco CallManager server after failover is corrected.”

Centralized Messaging and Centralized Call Processing

In centralized messaging, the Cisco Unity server is co-located with the Cisco CallManager cluster, but messaging clients are both local and remote in relation to the messaging system. (See [Figure 13-11](#).) When remote users access resources at the central site (such as voice ports, IP phones, or PSTN gateways, as in Tail-End Hop-Off (TEHO)), these calls are transparent to gatekeeper call admission control. Therefore, regions and locations must be configured in Cisco CallManager for call admission control. (See [Managing Bandwidth, page 13-4](#).) When making inter-region calls to IP phones or MGCP gateways, IP phones automatically select the inter-region codec that has been configured. If native transcoding is disabled, the Cisco Unity voice ports will require Cisco CallManager transcoding resources for calls that transverse the WAN (inter-region).

Figure 13-11 Centralized Messaging with Centralized Call Processing

In **Figure 13-11**, regions 1 and 2 are configured to use G.711 for intra-region calls and G.729 for inter-region calls. Native transcoding has been disabled on the Cisco Unity server.

As **Figure 13-11** shows, when a call is made from extension 200 to the voicemail ports in Region 1, the inter-region G.729 codec is used at the endpoint but the RTP stream is transcoded to use G.711 on the voice ports. Native transcoding on the Cisco Unity server has been disabled in this example. Cisco CallManager transcoding resources must be located at the same site as the voicemail system.

Hairpinning

Another issue to consider is hairpinning (or tromboning) of voice calls over multiple Cisco Unity voice ports. Hairpinning is not a concern in environments that use only SCCP TSP voice ports, but it is a concern with dual-integration environments, where hairpinning can occur between the voice ports of the legacy system and the SCCP TSP voice ports.

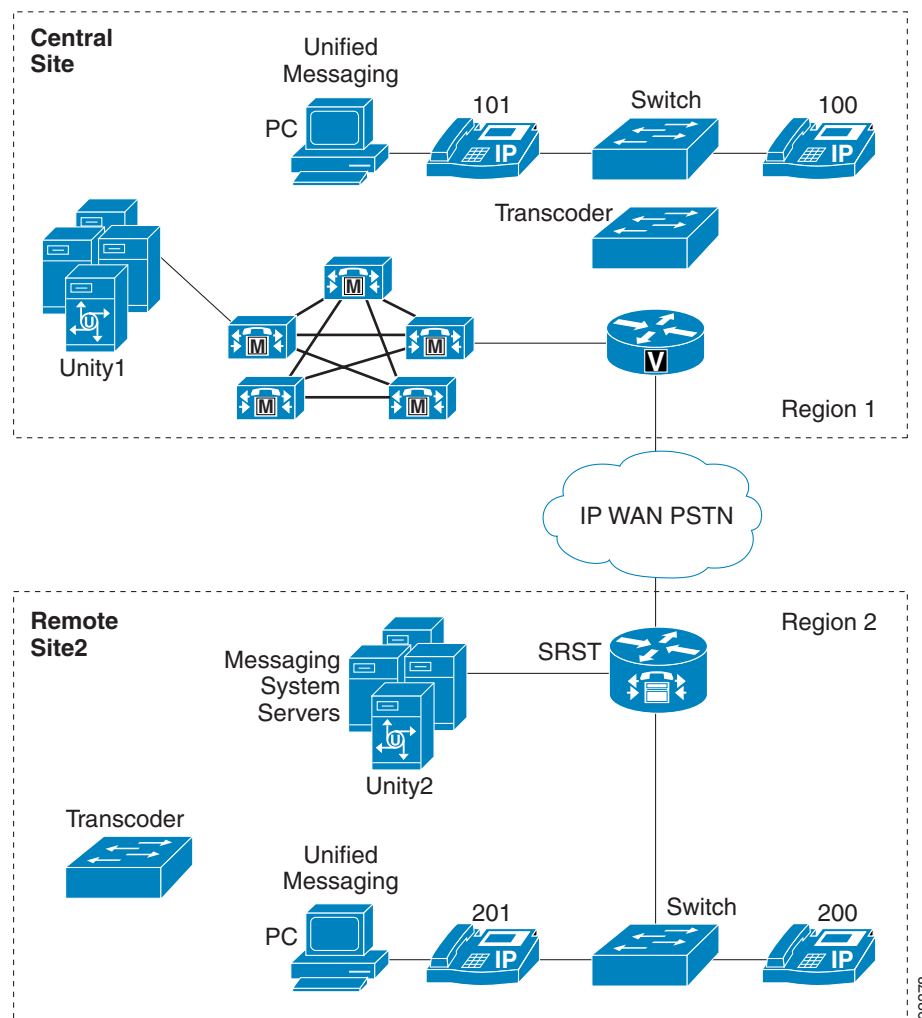
For more information about dual integration, refer to the *Cisco Unity Administration Guide*, available at

<http://www.cisco.com>

Distributed Messaging with Centralized Call Processing

Distributed messaging means that there are multiple messaging systems distributed within the telephony environment, and each messaging system services only local messaging clients. This model differs from centralized messaging, where clients are both local and remote from the messaging system. [Figure 13-12](#) illustrates the distributed messaging model with centralized call processing. As with other multi-site call processing models, the use of regions and locations is required to manage WAN bandwidth. For this model, you must also disable Cisco Unity native transcoding.

Figure 13-12 Distributed Messaging with Centralized Call Processing



For the configuration in [Figure 13-12](#), transcoder resources must be local to each Cisco Unity message system site. Regions 1 and 2 are configured to use G.711 for intra-region calls and G.729 for inter-region calls. Native transcoding has been disabled on the Cisco Unity servers.

Voice messaging ports for both Cisco Unity servers must be assigned the appropriate region and location by means of calling search spaces and device pools configured on the Cisco CallManager server. In addition, to associate telephony users with a specific group of voicemail ports, you must configure Cisco

CallManager voicemail profiles. For details on configuring calling search spaces, device pools, and voicemail profiles, refer to the applicable version of the *Cisco CallManager Administration Guide*, available at

<http://www.cisco.com>

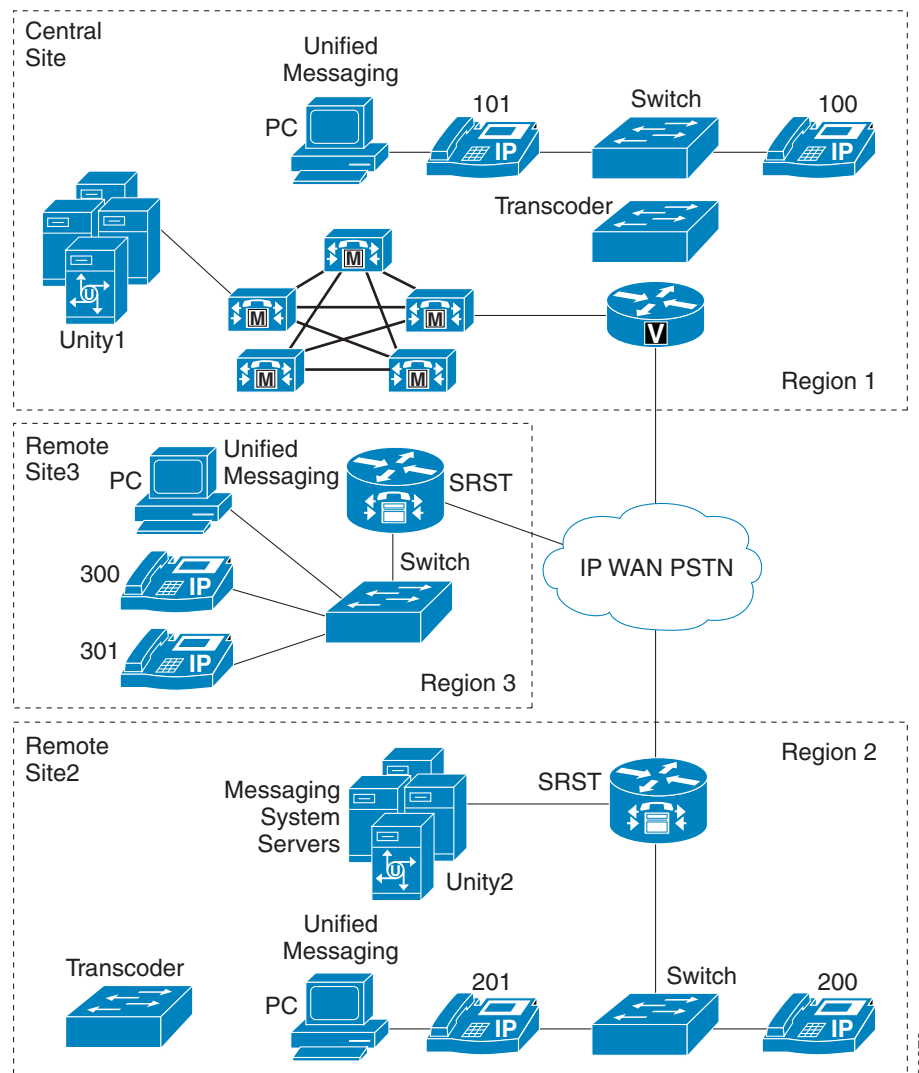
The messaging systems are "networked" together to present a single messaging system to both inside and outside users. For information about Cisco Unity Networking for the distributed Unity servers, Refer to the Networking in Cisco Unity Guide available from:

http://www.cisco.com/en/US/products/sw/voicesw/ps2237/products_installation_and_configuration_guides_list.html

Combined Messaging Deployment Models

It is possible to combine messaging models in the same deployment, provided that the deployment adheres to all the guidelines listed in the preceding sections. [Figure 13-13](#) shows a user environment in which both centralized and distributed messaging are employed simultaneously.

Figure 13-13 Combined Deployment Models



[Figure 13-13](#) shows the combination of two messaging models. Regions 1 and 3 use centralized messaging with centralized call processing, while Region 2 uses distributed messaging with centralized call processing. All regions are configured to use G.711 for intra-regions calls and G.729 for inter-region calls.

In [Figure 13-13](#), centralized messaging and centralized call control are used between the Central Site and Site3. The messaging system at the Central Site provides messaging services for clients at both the Central Site and Site3. Site2 uses the distributed messaging model with centralized call control. The messaging system (Unity2) located at Site2 provides messaging services for only those users located within Site2. In this deployment, both models adhere to their respective design guidelines as presented

in this chapter. Transcoding resources are located locally to each messaging system site, and they support clients who access messaging services from a remote site (relative to the messaging system), as in the case of a Site2 user leaving a message for a Central Site user.

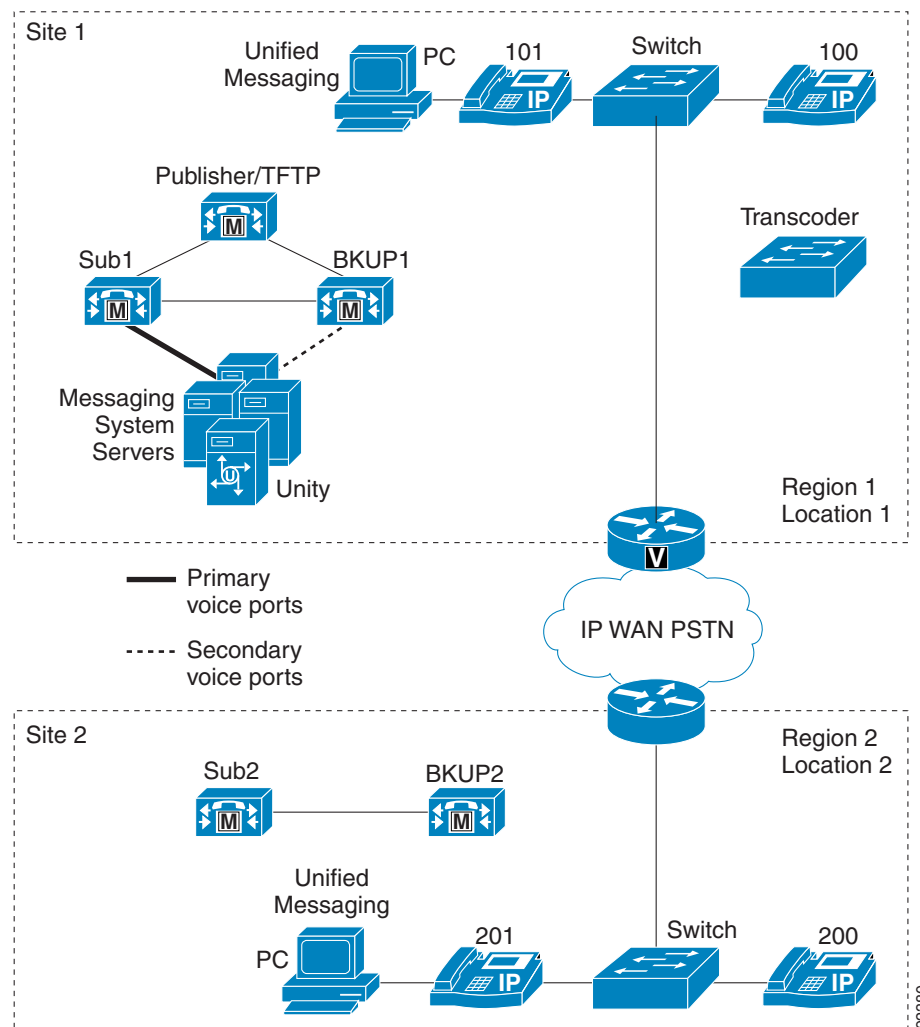
Centralized Messaging with Clustering Over the WAN

This section addresses Cisco Unity design issues for deploying centralized messaging with Cisco CallManager clustering over the WAN with local failover. In the case of a WAN failure with this model, all remote messaging sites will lose voicemail capability until the WAN is restored. (See [Figure 13-14](#).)

Clustering over the WAN supports local failover. With local failover, each site has a backup subscriber server physically located at the site. This section focuses on deploying Cisco Unity centralized messaging with local failover for clustering over the WAN.

For additional information, refer to the section on [Clustering Over the IP WAN](#), page 2-16.

Figure 13-14 Cisco Unity Centralized Messaging and Clustering Over the WAN with Local Failover



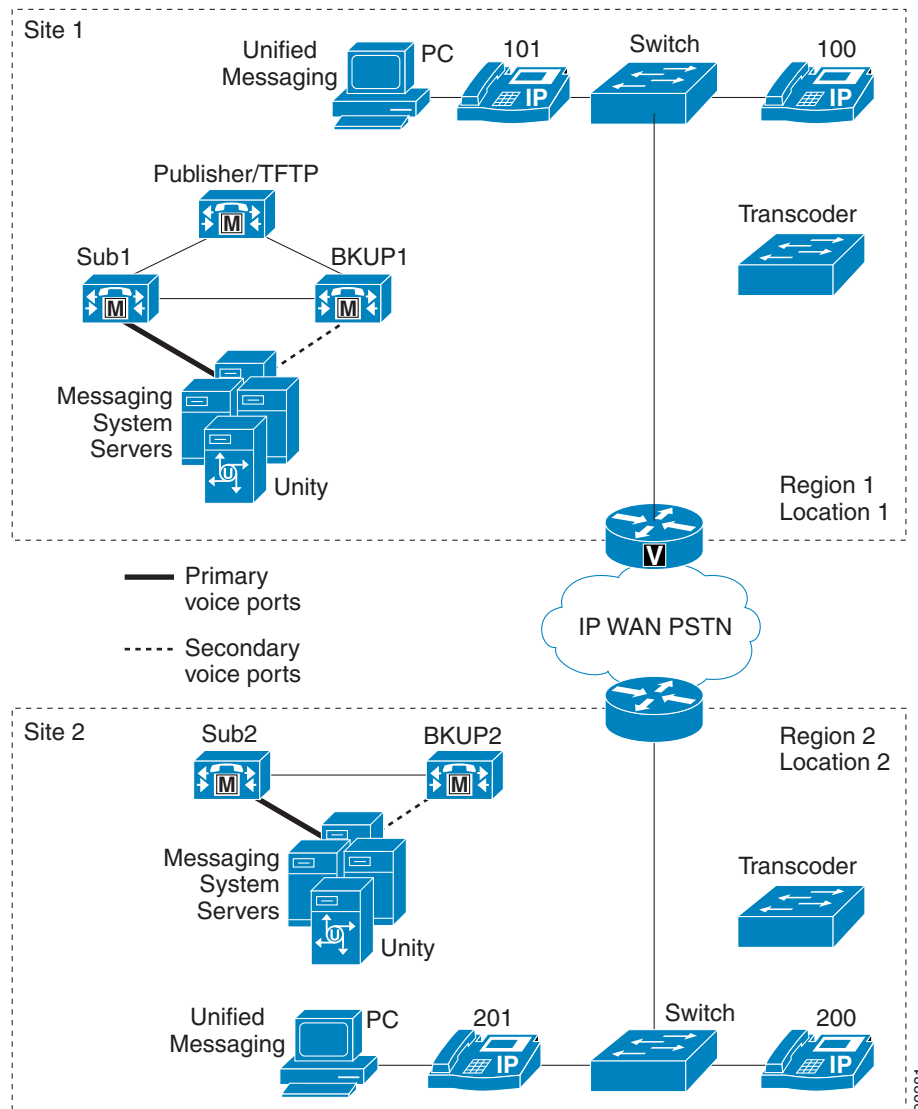
The minimum bandwidth required between cluster sites is a T1 line (1,544 MHz). This amount of bandwidth can support signaling and database traffic for up to 10,000 busy hour call attempts (BHCA), but it does not include the required media bandwidth.

Clustering over the WAN with Cisco CallManager Release 3.3(3) and earlier supports up to four sites per cluster, but Cisco CallManager Release 4.1 and higher supports up to eight sites. Cisco Unity supports up to the maximum number of sites in either case. The voicemail ports are configured only at the site where the Cisco Unity messaging system is located (see [Figure 13-14](#)). Voicemail ports do not register over the WAN to the remote site(s). Messaging clients at the other site(s) access all voicemail resources from the primary site. There is no benefit to configuring voice ports over the WAN to any of the remote sites because, in the event of a WAN failure, remote sites would lose access to the centralized messaging system. Because of bandwidth consideration, the voicemail ports should have TRaP disabled and all messaging clients should download voicemail messages to their local PCs (unified messaging only).

Distributed Messaging with Clustering Over the WAN

Local failover sites that also have Cisco Unity messaging server(s) deployed would have voice ports registered to the local Cisco CallManager subscriber server(s), similar to the centralized messaging model. For information about configuring the voice ports, see [Voice Port Integration with a Cisco CallManager Cluster](#), page 13-6, and [Voice Port Integration with Dedicated Cisco CallManager Backup Servers](#), page 13-12.

Figure 13-15 Cisco Unity Distributed Messaging and Clustering over the WAN with Local Failover



In a purely distributed messaging implementation with clustering over the WAN, each site in the cluster would have its own Cisco Unity messaging server with messaging infrastructure components. If not all of the sites have local Cisco Unity messaging systems but some sites have local messaging clients using a remote messaging server(s), this deployment would be a combination model with both distributed

messaging and centralized messaging. (See [Combined Messaging Deployment Models, page 13-17](#).) In the event of a WAN failure in this model, all remote sites that use centralized messaging will lose voicemail capability until the WAN is restored.

Each site that does not have a local messaging server must use a single messaging server for all of its messaging clients, but all such sites do not have to use the same messaging server. For example, suppose Site1 and Site2 each have a local messaging server. Site3 can then have all of its clients use (register with) the messaging server at Site2, while Site4 can have all of its clients use the messaging server at Site1. Transcoder resources are required at sites that have local Cisco Unity messaging server(s).

As with other distributed call processing deployments, calls going between these sites are transparent to gatekeeper call admission control, therefore you must configure regions and locations in Cisco CallManager to provide call admission control. (See [Managing Bandwidth, page 13-4](#).)

The distributed Cisco Unity servers may also be networked digitally. For more information on this topic, refer to the *Cisco Unity Networking Guide*, available at www.cisco.com. The networking guides are specific to the particular messaging store deployed.

Cisco Unity Messaging Failover

Cisco Unity failover provides voicemail survivability in the case of a Cisco Unity server failure. (See [Figure 13-1](#).) With Cisco Unity local failover, both the primary and secondary Unity messaging servers reside at the same physical location, and the messaging infrastructure components are co-located with the primary server. Optionally, messaging infrastructure components (such as messaging store servers, domain controller and global catalog (DC/GC) servers, and DNS servers) can also have redundant components that may be physically co-located with the Cisco Unity secondary server.

Cisco Unity failover is supported by all of the messaging deployment models.

For information on configuring Cisco Unity failover, refer to the *Cisco Unity Failover Configuration and Administration Guide*, available at

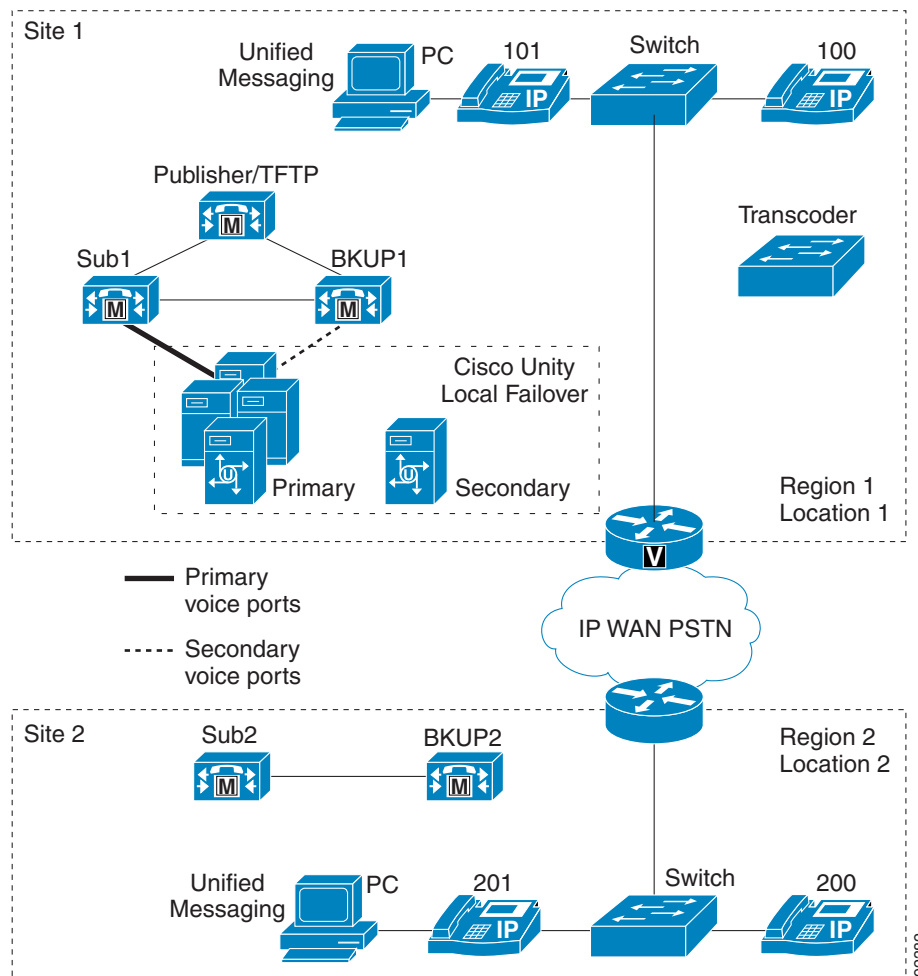
<http://www.cisco.com>

Cisco Unity Failover and Clustering Over the WAN

When deploying Cisco Unity local failover with clustering over the WAN, apply the same design practices described in [Centralized Messaging with Clustering Over the WAN, page 13-18](#), and [Distributed Messaging with Clustering Over the WAN, page 13-20](#). The voice ports from the primary Cisco Unity server should not cross the WAN during normal operation.

Figure 13-16 depicts Cisco Unity local failover. Note that the primary and secondary Cisco Unity servers are both physically located at the same site. Cisco Unity failover supports up to the maximum number of remote sites available with clustering over the WAN for Cisco CallManager.

Figure 13-16 Cisco Unity Local Failover and Clustering Over the WAN



For information on configuring Cisco Unity failover, refer to the *Cisco Unity Failover Configuration and Administration Guide*, available at

<http://www.cisco.com>

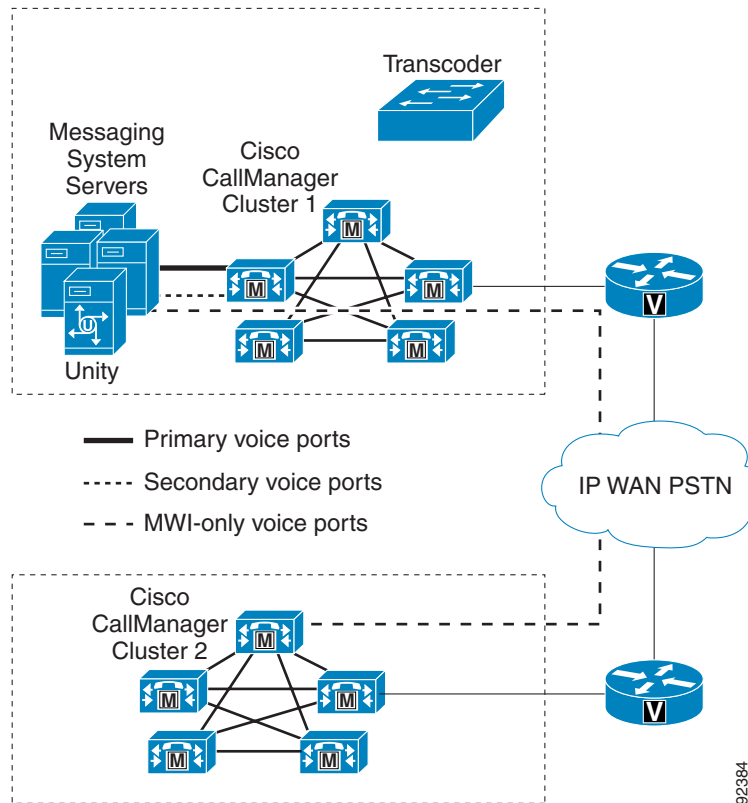
Centralized Messaging with Multiple Cisco CallManager Servers

For this model, configure fully functional voice ports and MWI ports on the Cisco CallManager cluster co-located with the Cisco Unity messaging server (Cluster 1 in [Figure 13-17](#)), and assign MWI-only ports on the remote Cisco CallManager cluster (Cluster 2 in [Figure 13-17](#)). Whenever a voicemail call crosses the WAN between Cluster 1 and Cluster 2, transcoding resources will be required. These resources must be co-located with the Cisco Unity messaging server. The Cisco Unity server must have native transcoding disabled in order to use Cisco CallManager transcoding resources.

MWI-only voicemail ports are configured on Cisco CallManager Cluster 2 in [Figure 13-17](#). In order for telephony subscribers to access the voicemail ports on Cisco CallManager Cluster 1, the pilot number must be something other than a DN so that the number can be routed via a route pattern. When an intercluster trunk is configured between the two clusters, the voicemail pilot can be routed over that intercluster trunk. Because a route pattern is used, gatekeeper call admission control can be used and the WAN bandwidth can be managed properly.

The following configuration information applies to [Figure 13-17](#):

Cluster 1	Cluster 2
Extensions: 2000 to 2999	Extensions: 12000 to 12999
Voicemail ports: 4	No voicemail ports
MWI-only ports: 1	MWI-only ports: 1
Port DNs: 2200 to 2205	Port DN: 14000
MWI On/Off DNs: 2249 and 2250, respectively	MWI On/Off DNs: 2249 and 2250, respectively
Cisco ports: 1 to 5	Cisco port: 6
Voicemail pilot: 2200	Voicemail pilot: 82000 (Route pattern strips the 8 and sends the pilot number over a gatekeeper-controlled trunk via G.729.)

Figure 13-17 Single Cisco Unity Server Servicing Multiple Clusters

92384

Messaging clients at both Cluster 1 and Cluster 2 sites use the Cisco Unity messaging infrastructure physically located at Cluster 1.

For more information on integrating a single Cisco Unity with multiple Cisco CallManager clusters, refer to the White Paper on *Cisco Unity Integration with Multiple Clusters of Cisco CallManager 3.1 and Later*, available at:

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_unity/whitpapr/clust_31.htm



Directory Access and Integration

Directories are specialized databases that are optimized for a high number of reads and searches, and occasional writes and updates. Directories typically store data that does not change often, such as employee information, user privileges on the corporate network, and so on.

Another aspect of directories is that they are extensible, which means that the type of information stored in them can be modified and extended. The term *directory schema* refers to the type of stored information and the rules it obeys. Many directories provide methods for extending the directory schema to accommodate information types defined by different applications. This capability enables enterprises to use the directory as a central repository for user information.

The Lightweight Directory Access Protocol (LDAP) provides applications with a standard method for accessing and potentially modifying the information stored in the directory. This capability enables companies to centralize all user information in a single repository, available to several applications, with a remarkable reduction in maintenance costs through the ease of adds, moves, and changes.

This chapter covers the main design principles for integrating Cisco CallManager with a corporate LDAP directory. It also summarizes design considerations for providing Cisco IP Telephony endpoints, such as Cisco IP Phones and Cisco IP SoftPhone, with access to a corporate LDAP directory. The main topics include:

- [Directory Access Versus Directory Integration, page 14-2](#)
- [Directory Access for Cisco IP Telephony Endpoints, page 14-3](#)
- [Directory Integration with Cisco CallManager, page 14-6](#)
- [Best Practices for Directory Integration, page 14-9](#)

The considerations presented in this chapter apply to Cisco CallManager as well as the following applications bundled with it: Extension Mobility, IP Manager Assistant, Web Dialer, Bulk Administration Tool, Real-Time Monitoring Tool, and Multi-Level Administration.

For all other Cisco voice applications, please refer to the respective product documentation available at <http://www.cisco.com>

In particular, for Cisco Unity, refer to the *Cisco Unity Design Guide* and to the following white papers: *Cisco Unity Data and the Directory*, *Active Directory Capacity Planning*, and *Cisco Unity Data Architecture and How Cisco Unity Works*.

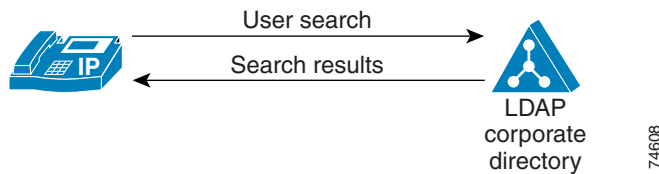
Directory Access Versus Directory Integration

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 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 14-1 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. One entry can then be selected and used to dial the corresponding person from the Cisco IP Phone.

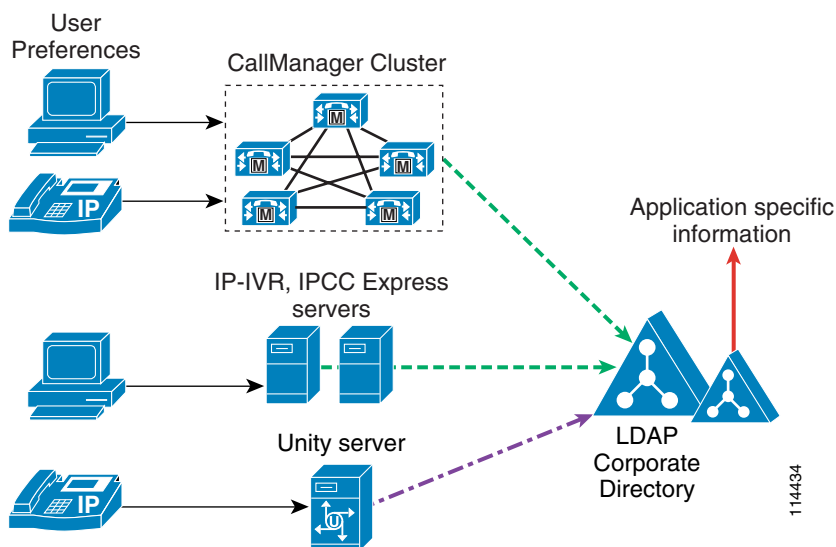
Figure 14-1 Directory Access for a Cisco IP Telephony Endpoints



Note that directory access, as defined here, involves only read operations on the directory and does not require you to make any directory schema extensions or other configuration changes.

In 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 separate, embedded databases. Figure 14-2 shows an example of directory integration as it is defined in this chapter.

Figure 14-2 Directory Integration for Cisco IP Telephony Applications



Note that directory integration involves read and write operations on the directory, so it requires you to make schema extensions and other configuration changes to your corporate LDAP directory.

By default, Cisco CallManager stores user information (such as devices controlled by the user, personal address book entries, 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 application-specific user information in the corporate directory.

**Note**

As of Cisco CallManager Release 3.1, directory integration is supported for Microsoft Active Directory (AD) 2000 and Netscape Directory Server Release 4.x. Support for iPlanet/Sun Directory Server 5.1 was added in Cisco CallManager Release 3.3(2), and support for Microsoft Active Directory (AD) 2003 was added in Cisco CallManager Releases 3.3(3) and 4.0(1)sr2.

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 good knowledge of the directory structure.
- The applications must be able to contact the directory at all times, and the directory must provide adequate response times. Availability of the directory service can affect application functionality.
- Additional load is introduced on the directory in terms of both data storage and read/write queries. Careful planning and sizing are recommended to avoid oversubscription of the servers when any new service or application is introduced.

Although directory integration across applications has numerous advantages, it is important to understand all its implications and to verify the business needs of each specific deployment.

Directory Access for Cisco IP Telephony Endpoints

The guidelines contained in this section apply regardless of whether Cisco CallManager or other IP Telephony applications have been integrated to a corporate directory. The end-user perception in both cases is the same because the differences affect only how 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 LDAPv3-compliant directory server for XML-capable phones such as Cisco IP Phone models 7940, 7960, and so on.

**Note**

Cisco IP SoftPhone, Release 1.2 and later, has a built-in mechanism to access and search LDAP directories, as does the Cisco IP Communicator. Refer to the product documentation for details on how to configure this feature.

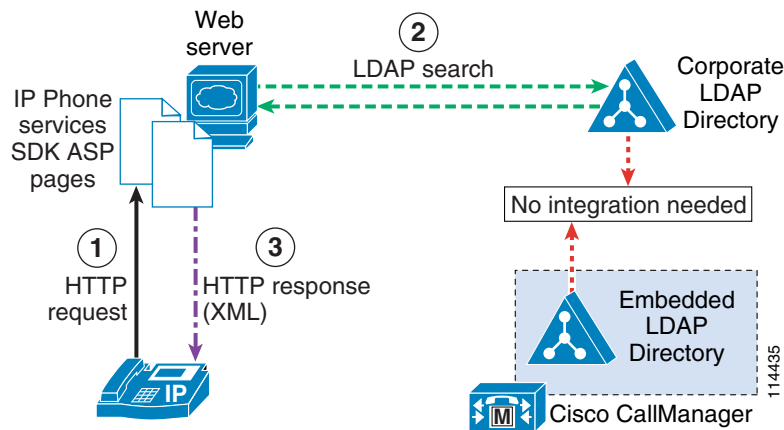
Directory Access for Cisco IP Phones

XML-capable Cisco IP Phones (such as models 7940, 7960, and so on) can search a corporate LDAP directory when a user presses the Directories button on the phone. The IP Phones use Hyper-Text Transfer Protocol (HTTP) to send requests to a web server. The responses from the web server must contain some specific Extensible Markup Language (XML) objects that the phone can interpret and

display. 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. After being encapsulated in the appropriate XML objects, the response is interpreted and sent back to the phone.

Figure 14-3 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 14-3 Message Exchange for Cisco IP Phone Corporate Directory Access without Directory Integration



The proxy function provided by the web server can be configured using the Cisco IP Phone Services Software Development Kit (SDK) version 2.0 or later, which includes the Cisco LDAP Search Component Object Model (COM) server. You can download the latest Cisco IP Phone Services SDK from Developer Support Central at

http://www.cisco.com/pcgi-bin/dev_support/access_level/product_support

Click the **Log in now** link, log in, and then select **CallManager - IP phone Services SDK** from the **Voice Technology/AVVID** list box.

Use the following steps to configure directory access for Cisco IP Phones.

- Step 1** Install the Cisco IP Phone Services SDK on a web server running Microsoft Internet Information Server (IIS). This server must be distinct from any Cisco CallManager server, but it can be an existing web server on the corporate network. (Refer to the SDK product documentation for more details on the installation procedure.)
- Step 2** Using the documentation provided with the SDK, create active server pages (ASPs) to interface with the LDAP Search COM object. Sample ASPs are provided with the SDK, but you can write your own if you need a higher level of customization. If you are using Release 3.3 of the IP Phone Services SDK, place the sample `ldapsearch.asp` page in an IIS virtual directory, and then edit the file to point to your corporate LDAP directory server by setting the following parameters:
 - `s.server`
Set this parameter to the LDAP server name or IP address (for example, `ldap.vse.lab`).
 - `s.port`
Set this parameter to the port used for LDAP requests on your LDAP server (the standard port is 389).

- s.base

Set this parameter to the search base for LDAP lookups. This search base should include all users who must be returned from a lookup (for example: cn=Users, dc=vse, dc=lab).

- s.AuthName

If your LDAP server requires authentication for lookups, set this parameter to the distinguished name for a user who has permissions to search the subtree specified in the search base (for example: cn=CCMDirMgr, ou=System Accounts, cn=Users, dc=vse, dc=lab).

- s.AuthPasswd

If your LDAP server requires authentication for lookups, set this parameter to the password for a user who has permissions to search the subtree specified in the search base.

- Step 3** In the Enterprise Parameter Configuration page in Cisco CallManager Administration (**System > Enterprise Parameters**), edit the **URL Directories** field, as shown in Figure 14-4. Set this field to the URL leading to the ldapsearch.asp file on the web server you previously set up.
- Step 4** Reset the IP phones to make the changes take effect.

Figure 14-4 Configuring the Enterprise Parameters on Cisco CallManager to Enable Directory Access

Phone URL Parameters		
Parameter Name	Parameter Value	Suggested Value
URL Authentication	http://SJCOCM1/CCMCIP/authenticate.asp	
URL Directories	http://web.vse.lac/ldapsearch.asp	
URL Idle		
URL Idle Time (sec)	0	0
URL Information	http://SJCOCM1/CCMCIP/GetTelecasterH	
URL Messages		
IP Phone Proxy Address		
URL Services	http://SJCOCM1/CCMCIP/getservicesmer	
Enable All User Search*	True	True
User Search Limit*	64	64
* indicates required item		
 Click for more information.		

114436

In addition, directory access for Cisco IP Phones has the following characteristics:

- All LDAPv3-compliant directories are supported.
- Cisco CallManager user preferences (speed dials, call forward all, personal address book) are not integrated with the corporate LDAP directory. Therefore, users have a separate login and password to access the Cisco CallManager User Options web page.

Directory Integration with Cisco CallManager

CallManager uses an embedded Microsoft SQL database to store system and device configuration data, such as dial plan information, phone and gateway configurations, and media resource utilization. It also uses an embedded LDAP directory to store user and application profiles, such as devices controlled by a user, computer telephony integration (CTI) user parameters, and personal address book entries.

Both the SQL database and the LDAP directory run on every 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.

**Note**

The examples and recommendations in this chapter are based on Cisco CallManager Releases 4.0 and 4.1, which introduced some new features and enhancements. Some behaviors might differ and some features might not be available if you are running an older version of Cisco CallManager.

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.

Because different directory vendors typically use different User object models with several additional, nonstandard attributes, Cisco CallManager uses only the standard LDAPv3 core attributes from the User object. The User object is then augmented with an auxiliary class, `ciscoocUser`, which contains the following attributes:

- `ciscoatGUID`

This attribute uniquely identifies a user within the directory.

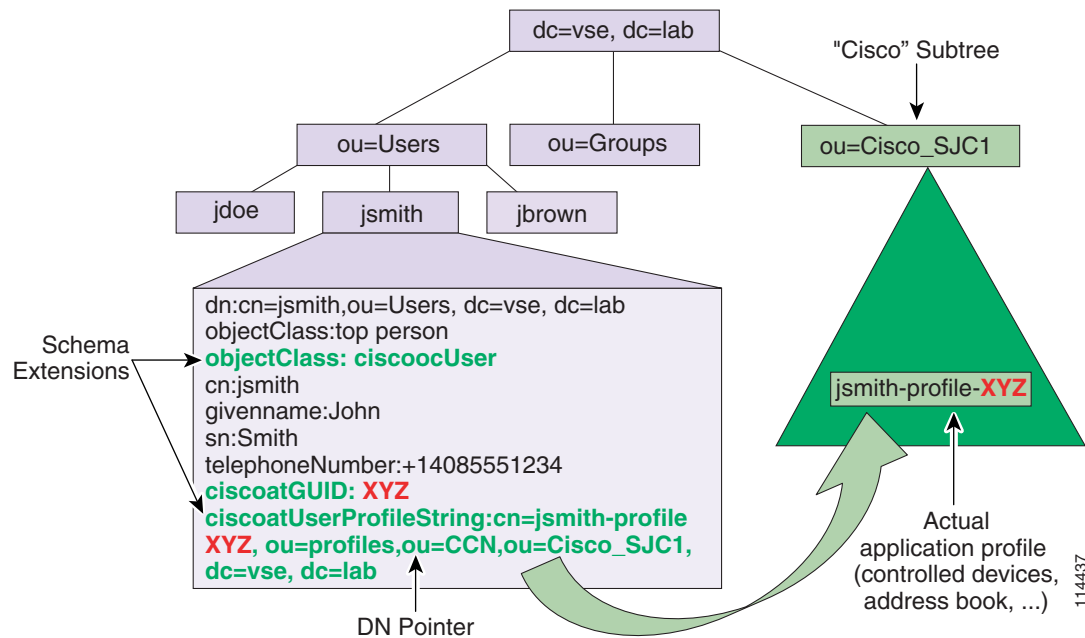
- `ciscoatUserProfile`

This attribute is used by earlier versions of Cisco CallManager and other applications. It is still present for backward compatibility.

- `ciscoatUserProfileString`

This attribute is a distinguished name pointer to another object in the directory, which contains the user's application-specific profile. This approach minimizes the impact on the core User object, and all the application-specific information can be stored in a separate organizational unit (OU) within the directory, usually called the Cisco subtree, `CISCOBASE`, or Cisco Directory Information Tree (DIT). [Figure 14-5](#) illustrates this process.

Figure 14-5 Cisco CallManager's Approach to Storing Application-Specific User Information in the Directory



The object pointed to by the `ciscoatUserProfileString` attribute belongs to a structural object class called `ciscoocUserProfile`. The main purpose of this object is to store some specific details for the user, including the user's locale, any Cisco IP Manager Assistant (IPMA) assistants for the user, and pointers to various specific profile objects for all Cisco applications that integrate with the directory. The application profile used by Cisco CallManager belongs to the auxiliary class called `ciscoCCNocAppProfile`, and it is where Cisco CallManager stores the user's extension mobility PIN, the list of devices controlled by this user, information on whether this user is permitted to use CTI applications, and so on. Both of these profile objects are created by Cisco CallManager under the "Cisco" subtree.



Note

The list of devices associated with a user is stored in the directory as a multi-valued attribute. If you are using the embedded Cisco CallManager directory, the maximum number of devices that can be associated with a user is 2,000 (2,500 if all subscribers in the cluster are dual-CPU servers). However, Microsoft Active Directory limits the number of values in a multi-valued attribute to 850. Therefore, if you are integrating Cisco CallManager with Microsoft Active Directory, the maximum number of devices you can associate with a user is 850.

The Cisco Customer Directory Configuration Plugin

To integrate Cisco CallManager with an external LDAP directory, run the Cisco Customer Directory Configuration Plugin, which is bundled with Cisco CallManager (**Applications > Install Plugins**). This plugin serves three main purposes:

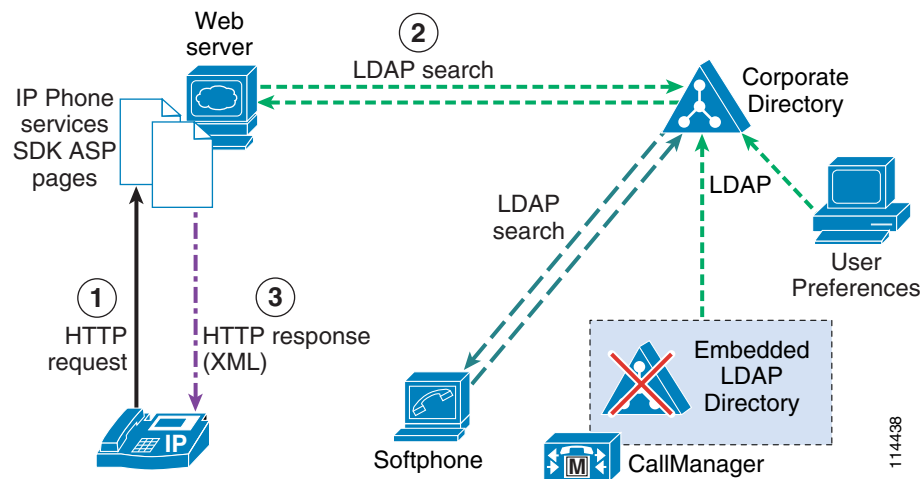
- It extends the corporate directory schema to accommodate the application-specific objects and attributes.
- It populates the "Cisco" subtree with the configuration objects needed by Cisco CallManager.

- It configures Cisco CallManager to use the corporate directory and disables its embedded directory.

Usually, running the plugin locally on Cisco CallManager performs the schema update. However, starting with Cisco CallManager Release 4.0, a new option exists to create the LDAP Data Interchange Format (LDIF) files separately so that the schema update can be performed directly on the corporate directory's Schema Master server using the LDIF files. This option allows different groups of people to perform the relevant parts of the work and reduces the need to update over the network when Cisco CallManager is not local to the Schema Master server.

After the plugin has been run, Cisco CallManager effectively uses the corporate directory to store user preferences. If the Cisco IP Telephony endpoints have also been enabled to access the corporate directory, as described in the preceding section, the resulting scenario is that shown in [Figure 14-6](#).

Figure 14-6 Message Exchange for Cisco IP Phone Corporate Directory Access When Cisco CallManager Is Integrated with the Corporate Directory



Security Considerations

Starting with Cisco CallManager Release 4.1, the communication between Cisco CallManager and the embedded directory is set by default to use LDAP over Secure Socket Layer (SSL), also known as LDAPS.

When integrating with a corporate directory, it is also possible to enable LDAP over SSL, thus ensuring that all sensitive LDAP data goes over a secure connection. The LDAP over SSL option can be configured as part of the Cisco Customer Directory Plugin, and it requires a certificate shared with the corporate directory and issued by the same Certificate Authority.

When LDAP over SSL is enabled for Cisco CallManager, the following additional Cisco IP Communications applications will also communicate with the directory over this secure channel:

- User pages within Cisco CallManager Administration
- Cisco Multi-Level Administration (MLA)
- Cisco IP Phone Options pages
- Extension Mobility application — SSL is used for communications between the Extension Mobility application, running on a Cisco CallManager server, and the corporate directory. However, SSL is not used for communications between the IP phones and the Extension Mobility application, which use HTTP instead.

- Cisco CTI Manager
- Serviceability and Cisco Real-Time Monitoring Tool (RTMT)
- Cisco CDR Analysis and Reporting (CAR)
- Cisco IP Manager Assistant (IPMA) service
- Cisco Bulk Administration Tool (BAT)

Adding Cisco CallManager Servers to a Domain

Adding the Cisco CallManager servers to a Microsoft Windows domain differs substantially from integrating Cisco CallManager with an external directory. Although these operations do not exclude each other, they are completely independent and have different implications:

- Adding Cisco CallManager servers to a Microsoft Windows Active Directory (AD) domain could cause domain policies to be applied to the Windows 2000 Server operating system, and the addition affects only the management of the Cisco CallManager server itself.
- Integrating Cisco CallManager with an external directory (such as Microsoft Active Directory or Netscape Directory Server) causes Cisco CallManager to store all its user information and preferences in that directory, but it does not affect management of the Cisco CallManager server itself.

Cisco recommends that you keep Cisco CallManager servers as workgroup servers. However, if you want to add the servers to a domain, avoid applying domain policies to the server that could interfere with its normal operation.

For additional recommendations about adding servers to a domain, refer to the latest Cisco CallManager product documentation available online at

<http://www.cisco.com>

Best Practices for Directory Integration

The directory integration process involves several components and services in your network and therefore should be carefully planned and implemented. This section addresses following topics:

- [Planning the Directory Integration, page 14-10](#)
- [Preparing the Directory for Integration, page 14-11](#)
- [Integrating Cisco CallManager with the Directory, page 14-15](#)
- [Maintaining the Directory Integration, page 14-18](#)



Note

Because the majority of known Cisco CallManager integrations are performed with Microsoft Active Directory (AD), this chapter focuses on best practices for AD. However, most recommendations and best practices mentioned in this section also apply to the other directory product supported by Cisco CallManager, the Sun/iPlanet Netscape Directory Server.

Planning the Directory Integration

Because the directory is an enterprise-wide resource that is used by a potentially large number of applications and end users, it is essential to plan the integration carefully to minimize the impact on all other applications.



Tip

Before starting the integration, ensure that your organization's directory team is involved in the planning, design, and implementation phases.

As mentioned previously in this chapter, integrating Cisco CallManager and other applications with an external directory involves extending the directory schema. Schema extension is a delicate operation. For example, in the case of Microsoft Windows 2000 Active Directory, schema changes cannot be undone. You should take the following precautions to avoid damaging the directory:

- Review the planned schema changes with your organization's directory team. This practice should be part of your organization's change control procedures.
- Create a replica of the production directory in a lab setup, and test the integration against it.
- Back up the production directory, both data as well as schema, before integration, and make sure a workable back-out plan exists to enable successful restoration of the data and schema if it becomes necessary.
- Plan and perform the schema extension during off-peak hours to minimize the impact on other applications and end users.

Although the preceding list of precautions might cause concern at first, in practice the schema extension rarely causes problems that would require it to be backed out. However, no matter how safe an operation is known to be, you should not add risk by skipping any possible precautions to ensure that you can recover from unplanned issues.

Another important consideration is that, as soon as the voice applications have been integrated with the directory, they rely on it for their correct operation, and inability to reach the directory server can negatively affect the voice system.

For example, if the directory suddenly becomes unavailable, end users cannot log into the Cisco CallManager User Options web page to configure their preferences; Extension Mobility users, Attendant Console operators, and IPCC Express agents cannot log in or out; and the dial-by-name function is unavailable.

To avoid these problems, you should design your directory infrastructure so that it is highly available to all Cisco voice applications. You can use any of the following methods to achieve this high availability:

- Leverage the directory replication mechanism to place a directory server or servers in the same location as the Cisco voice applications.
- Use a server load-balancing mechanism, such as Cisco IOS software server load balancing (SLB), to provide server redundancy in a specific campus or data center and to ensure that local servers are accessed by preference.
- Use Domain Name System (DNS) domain names instead of specific domain controller host names when configuring the directory plugin.

With redundant servers, the first name returned by DNS might be the name of a server that is not as local to Cisco CallManager as others returned later in the response. Also, if your DNS server has the round-robin feature enabled, by design it rotates the order in which addresses are returned in the response. Depending on mechanisms such as client-side DNS cache timeout, along with other possible clients querying for the same domain in the interim, Cisco CallManager could run two consecutive

operations against two different domain controllers (DCs). In addition to the locality problem already mentioned, using DNS redundancy could keep objects created in the first operation from being found by a search on a different DC by a later query if the directory has not replicated in the meantime. Therefore, before choosing to use DNS to make the implementation redundant, be sure that these issues do not affect your deployment.

Also note that DNS is needed for proper LDAP referral; Cisco CallManager must be able to resolve the host names of any of the DCs returned in an LDAP referral.

**Note**

Although Microsoft Windows 2000 DNS has a feature that returns local resources first (called LocalNetPriority), it is based on inspecting the requesting client's classful IP address. Therefore, it is of limited use in subnetted networks. Microsoft Knowledge Base article 177883 documents this feature (see <http://support.microsoft.com/>). If you are not using Windows 2000 DNS, you should check to see which features in your chosen implementation might alleviate some of these issues. These recommendations are based on the assumptions that Cisco CallManager is set to use DNS and that it is the same DNS infrastructure used by your Active Directory.

Preparing the Directory for Integration

With Microsoft Windows 2000 Active Directory, the domain controllers have to be set to allow schema changes. This requirement applies only to the domain controller that acts as the Schema Master (the location where change takes place), and it is fully documented in Microsoft Knowledge Base article 285172, which is available at

<http://support.microsoft.com>

If you are integrating Cisco CallManager with the Microsoft Windows 2000 Active Directory, and if Microsoft Exchange 2000 needs to coexist in the same forest, you must perform an additional preparation step. Cisco CallManager uses the labeledURI attribute specified by the iNetOrgPerson class, as defined in RFC 2798. Microsoft currently defines this attribute differently for Exchange 2000, which causes a naming clash with the Cisco CallManager schema. The problem is documented in Microsoft Knowledge Base article 314649 (available from <http://support.microsoft.com>). You can obtain the iNetOrgPerson kit from

<http://msdn.microsoft.com/library/en-us/dnactdir/html/inetopkit.asp>

**Caution**

Remember to back up your directory before extending the schema. Make sure you have tested the restore mechanism you intend to use *before* you need it.

To extend the directory schema for Cisco CallManager, run the Cisco Customer Directory Configuration plugin from the publisher server for the cluster, and follow these best practices:

- Always use the **Custom** setup type when running the plugin. The Express setup type is appropriate only for integration with a standalone domain used exclusively for Cisco CallManager and other Cisco voice applications. You should never use the Express setup when integrating with an existing domain.
- Select only the **Install Schema on the Schema Master** option in the plugin configuration screen.
- Try to ensure that the Schema Master server is relatively local to Cisco CallManager or that a high-speed connection exists between them. If neither is possible, then consider creating just the LDIF files with the plugin and using them directly to update the schema on the Schema Master.

- In this phase, provide the plugin with credentials (distinguished name and password) for a user who is a member of the Schema Admins group in Active Directory. These credentials are used only for the schema extension, not for normal Cisco CallManager operation.

**Note**

If you have a large AD forest or a complex topology, schema changes might take some time to propagate to all domains and all domain controllers in the forest. Allow enough time for this propagation to happen before continuing with the preparation process, or else force replication if required.

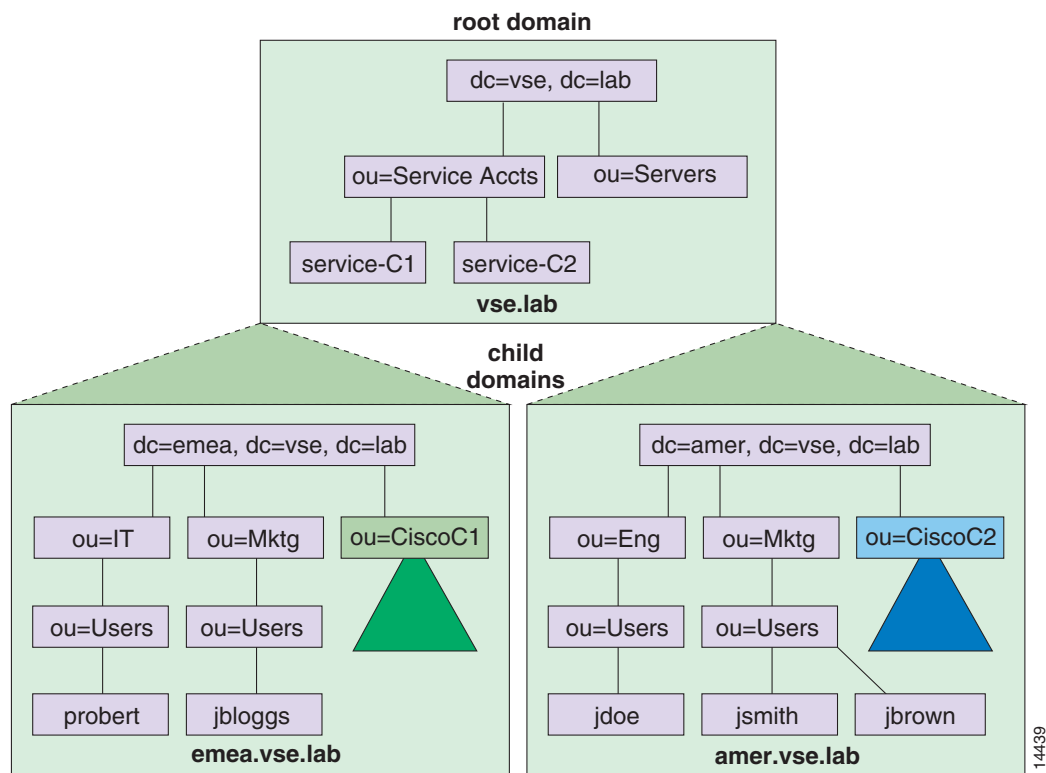
As soon as the schema has been extended, you can decide where to create the "Cisco" OUs (that is, the subtrees) that will be used by the Cisco CallManager clusters and the other Cisco voice applications. There must be one organizational unit (OU) for each Cisco CallManager cluster to be integrated.

For a deployment with a single domain AD or Netscape directory server, placement is not critical. The OU can effectively be placed anywhere in the tree.

In a multiple-domain AD forest, the root domain is often kept free of users and resources and is used as a placeholder domain, so the "Cisco" subtrees would typically reside in the child domains. In this type of multiple-domain topology, domains can be created based on geographic boundaries. Therefore, it is unlikely that every location has a local domain controller for each domain. To reduce replication traffic across the network, domain controllers are usually placed only where needed. With this in mind, it is best to try to place the Cisco OU for a given cluster in the domain containing the majority of users serviced by that cluster.

Figure 14-7 shows a multiple-domain, single-tree AD forest in which the "Cisco" OUs for two CallManager clusters have been created in the two respective child domains, emea.vse.lab and amer.vse.lab.

Figure 14-7 Multiple-Domain, Single-Tree AD Forest



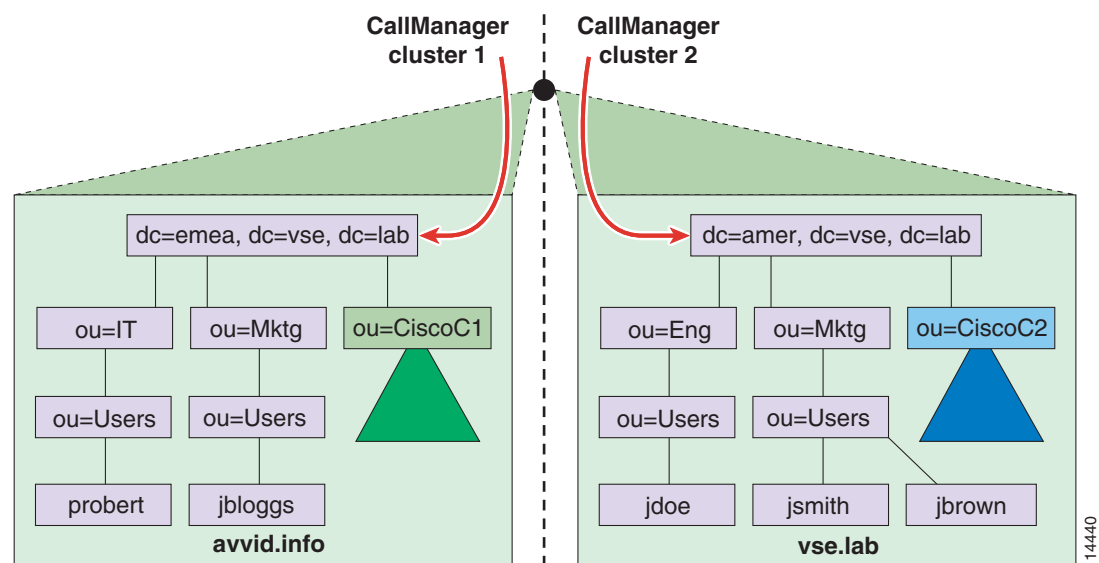
114439

In [Figure 14-7](#), each cluster services the corresponding geography in a centralized call processing model, thus ensuring that its user data stored in AD is also local. This design saves having to retrieve the information from a DC that is probably non-local, and it reduces the number of LDAP referrals required to find the relevant information in a search.

Cisco does not recommend having a Cisco CallManager cluster service users in different domains because response times while user data is being retrieved might be less than optimal if domain controllers for all included domains are not local. However, this scenario is not a common one because the reasons for creating a multiple-domain AD (namely, geography, bandwidth, or organizational structure) are usually the same reasons for needing multiple clusters.

Currently clusters cannot span trees within an AD forest because they would not have a contiguous namespace, which is a requirement for LDAP referrals. A cluster can exist within a domain or a single tree, even in a multiple-tree forest (as described previously), but all the users for a specific cluster must be contained in the same namespace, as shown in [Figure 14-8](#).

Figure 14-8 Cisco CallManager Clusters Must Be Contained Within a Single Tree (Contiguous Namespace)



The User Search Base is another important element used by Cisco CallManager when integrating with a directory. The User Search Base refers to the root of the subtree used by Cisco CallManager to search for users who can potentially be associated with devices within the cluster. (The section on [Integrating Cisco CallManager with the Directory](#), page 14-15, discusses how to set this parameter.)

You should create a special user account that Cisco CallManager and the other Cisco voice applications can use to access and manage the directory. There should be one account per Cisco CallManager cluster because this arrangement allows each account to be granted specific permissions only when needed and allows for easier administration on a per-cluster basis without the risk of affecting other parts of the enterprise. In the examples in this chapter, this account is called the CCM Directory Manager, but the name you choose for this user account may be different.

Each CCM Directory Manager account should be granted at least the following permissions within your directory:

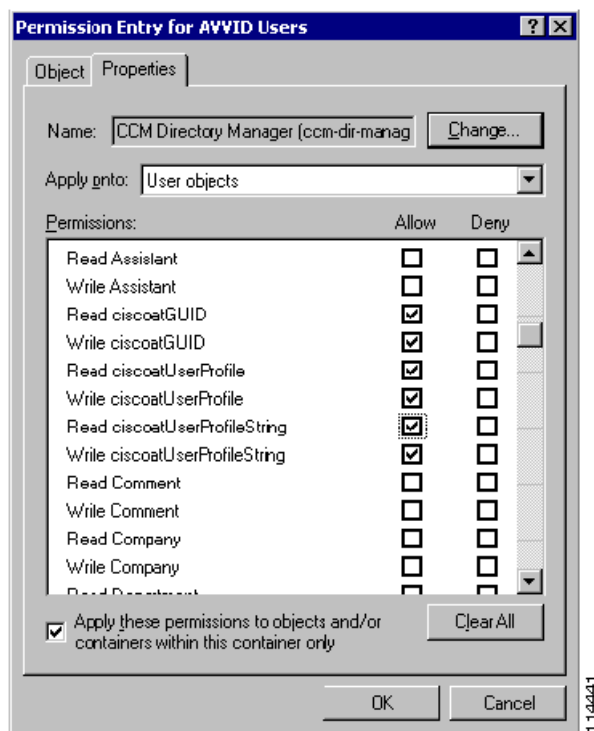
- **Read/Write/Create all child objects/Delete all child objects** privileges on the respective "Cisco" OU subtrees. The rights must be set to apply to **This object and all child objects** for both the object and the properties. In AD, you can set this privilege by using the advanced options for security within Active Directory Users and Computers (ADUC). The default is to apply to the object only, so you will have to change the rights.
- **Read** privileges on all the OUs contained at and below the User Search Base. You can set this privilege just at the User Search Base level, as long as inheritance is not blocked lower in the tree.
- **Read/Write** privileges on the ciscoatGUID, ciscoatUserProfile, and ciscoatUserProfileString attributes for all User objects contained below the User Search Base. In AD, you can set this privilege by using the advanced options for security within ADUC.



Tip

In AD, to set the permissions for the ciscoatGUID, ciscoatUserProfile, and ciscoatUserProfileString attributes for all User objects within the User Search Base, select the CCM Directory Manager user from the Advanced security window for the organizational unit (OU) at the root of the User Search Base. (In this chapter, the user is called CCM Directory Manager, but your user name may be different.) Then click **View/Edit** and go to the **Properties** tab of the new window, as shown in [Figure 14-9](#). From the **Apply onto** drop-down menu, select **User objects**, and then scroll down to the ciscoatGUID, ciscoatUserProfile, and ciscoatProfileString attributes. Allow **Write** permissions for all of them.

Figure 14-9 Setting Permissions for the User Account in Active Directory



**Tip**

While creating the CCM Directory Manager account, set the **Password never expires** option. When you want to change the password, run the CCMPwdChanger utility from Cisco CallManager. This method updates the password in AD, updates the registry in Cisco CallManager, and updates directory initialization files.

Integrating Cisco CallManager with the Directory

After you have prepared the directory as described in the preceding section, you can perform the integration by running the Cisco Customer Directory Configuration plugin again. The two key concepts to consider at this point are the User Search Base and the User Creation Base.

**Note**

The User Creation Base was introduced in Cisco CallManager Release 4.0. In previous versions of Cisco CallManager, the User Search Base is also used for user creation.

As mentioned in the preceding section, the User Search Base parameter represents the root of the subtree used by Cisco CallManager for all user searches.

The User Creation Base parameter tells Cisco CallManager where to create the following system accounts, which are needed by some applications and the features bundled with them:

- CCM Administrator, used by Cisco CallManager Multilevel Administration Access (MLA)
- CCM SysUser, used by CallBack and Extension Mobility
- IPMA SysUser, used by Cisco IP Manager Assistant

The User Creation Base must be contained within the User Search Base because Cisco CallManager has to be able to search for the system account users before authenticating them.

To set the User Search Base, look at where the users serviced by the cluster are located, and set the User Search Base to the first level that includes all of them. The lower you set the User Search Base, the better response times and performance you achieve because searches will not have to follow many referrals and cross slow WAN links to get to remote domain controllers. Also, user data not needed by the cluster does not have to be parsed in the search.

In a single-domain AD forest (or standalone Netscape Directory), set the User Search Base to the lowest organizational unit (OU) that contains all potential users for the Cisco CallManager cluster (for example: ou=AVVID Users, dc=vse, dc=lab). This OU might even be the root of the domain (for example: dc=vse, dc=lab, or o=avvid.lab) if users are spread across a set of OUs directly under this one.

In a multiple-domain AD forest, try to keep the users for a specific Cisco CallManager cluster within a single domain, and follow the guidelines described previously. If a single domain is not possible because users are spread across multiple domains, set the User Search Base to the lowest point in the tree containing all domains with users serviced by the Cisco CallManager cluster. In structures in which serviced child domains are under the top-level domain, the User Search Base must be set at the root of the entire AD forest. In all cases, though, try to ensure that a domain controller for each serviced domain is collocated with Cisco CallManager, or that the network is sufficiently resilient and fast to allow remote searches with no greater performance degradation than occurs with local searches.

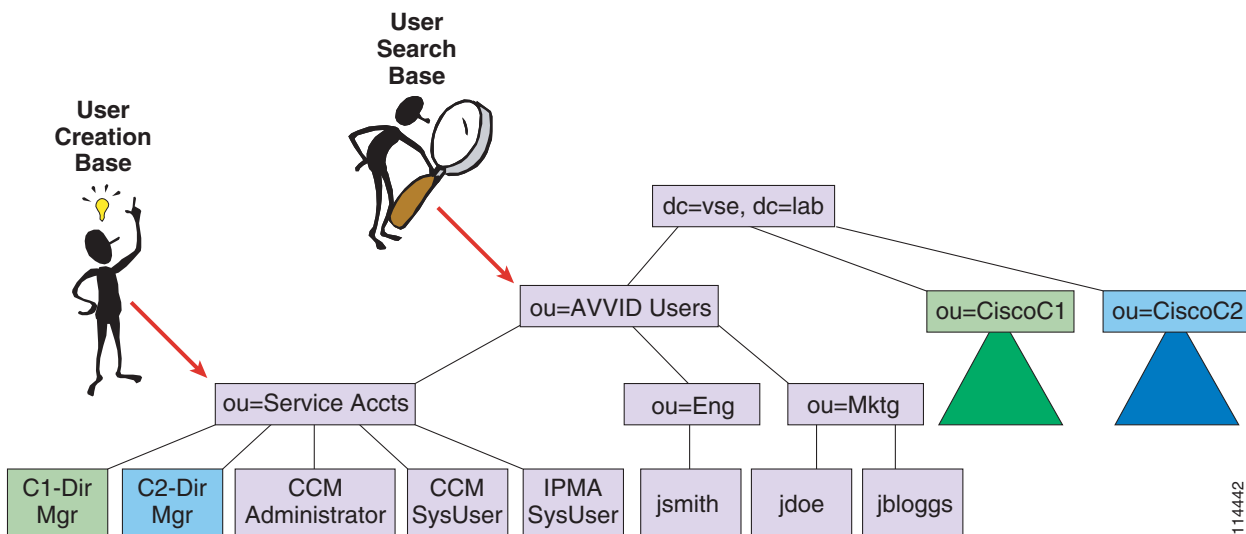
**Note**

Although the User Creation Base must exist within the User Search Base, take care to ensure that no existing Active Directory user policies are applied to the system accounts created there. A simple way to protect the system accounts from user policies is to put the accounts in a sub-OU and block inheritance of the Group Policy Object (GPO) to that OU.

It is possible to integrate multiple Cisco CallManager clusters with the same AD forest. However, due to the potentially large number of combinations of customer AD configurations and Cisco voice applications that can be deployed together with Cisco CallManager and that also use the directory, you must request specific support from the Cisco engineering team before proceeding with a multi-cluster integration. Contact your local Cisco account representative to initiate the support request. When deploying Cisco voice applications other than Cisco CallManager (such as the CDR Analysis and Reporting tool, Multilevel Administration Access, IP Contact Center, IPCC Express, and so forth), additional limitations may apply. Refer to the respective online documentation and release notes for details on these other products.

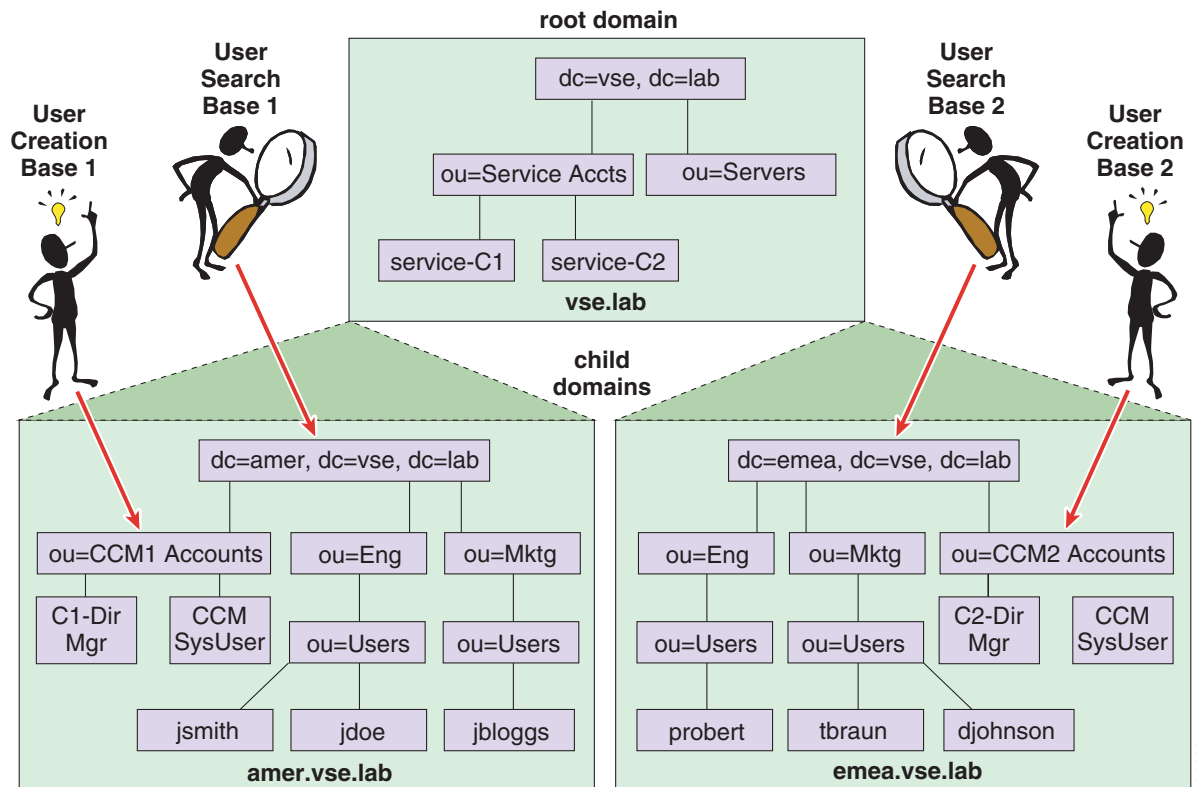
When integrating multiple Cisco CallManager clusters with the same AD domain (or standalone Netscape Directory), you can share the system accounts across clusters by specifying the same User Creation Base, as shown in [Figure 14-10](#).

Figure 14-10 Setting the User Search Base and User Creation Base in a Single AD Domain



When integrating multiple Cisco CallManager clusters with different AD domains within a forest, Cisco recommends that you define a different User Creation Base for each cluster, setting it to an OU within the relevant domain, as shown in [Figure 14-11](#).

Figure 14-11 Setting the User Search Base and User Creation Base in a Multiple-Cluster, Multiple-Domain Forest

**Tip**

You can prevent system and service accounts from appearing in Cisco CallManager Administration by adding the string **CiscoPrivateUser** to the user's Description field. The CCM Administrator, CCM SysUser, and IPMA SysUser accounts have this field set by default, but you can safely add the description to the CCM Directory Manager account as well. Use Microsoft ADSIEdit (Active Directory Service Interfaces, available as a part of the Windows 2000 Support Tools) or any other LDAP tool to update the Description field.

After you have decided how to set the User Search Base and User Creation Base, you can run the Cisco Customer Directory Configuration plugin again on the Cisco CallManager publisher server within your cluster. Follow these best practices when performing this step:

- Select the **Custom** plugin setup type, and select the **Configure Active Directory** and the **Enable CallManager Integration with Active Directory** options only.
- Specify the host name of a domain controller located within the same LAN as Cisco CallManager, or use the domain name if you are using DNS and all domain controllers for this domain are located in the same LAN as Cisco CallManager. Refer to [Planning the Directory Integration, page 14-10](#), for more details on how to provide high availability for the directory integration.

After completing these steps on the publisher server, set the passwords for the three system accounts by using either the CCMPwdChanger tool bundled with Cisco CallManager (open a DOS window by selecting **Start > Run**, type `cmd`, enter `CCMPwdChanger`, and press **Enter**) or your corporate directory's interface (for example, ADUC in the case of Active Directory). Cisco recommends that you set the password policy for these users so that their passwords never expire and are not set to change on first logon. This password policy is also one reason to avoid having any GPOs applied to these accounts.

If you enforce an expiration policy, Cisco CallManager will stop working when the password expires, and there will be no warning to alert you that the problem is an expired password. If you have policies that require password changes every three months, for example, you should run the CCMPwdChanger tool every three months.

After completing all the preceding steps, you can run the Cisco Customer Directory Configuration plugin on every subscriber server within the Cisco CallManager cluster.

**Note**

When you perform the integration, no data migration occurs between the Cisco CallManager embedded directory and the corporate directory. If you want to migrate users and profiles that you had configured in the embedded directory, Cisco has developed some migration scripts that can help you in this task. To obtain these scripts, contact your Cisco account team or channel partner. Note that the scripts are provided as-is and without support.

Maintaining the Directory Integration

After you have integrated Cisco CallManager with an external directory, you should set the user and password management procedures and policies accordingly.

Use the corporate directory's interface (or supported API) for the following administrative operations:

- Adding a user
- Removing a user
- Setting and changing the core user attributes (such as display name, department, address, and password)

In addition, use Cisco CallManager Administration for the following administrative operations:

- Configuring CallManager-specific user attributes (such as PIN and user locale)
- Associating a user with a device (such as an IP phone or a CTI port)

By default, you cannot use Cisco CallManager Administration to add or remove users. You also cannot modify any of their core user attributes, such as name and phone number. If you want to enable adding and removing users in Cisco CallManager Administration, you can modify the `UMDirectoryConfiguration.ini` file on the Cisco CallManager servers, as described in *Installing the Cisco Customer Directory Configuration Plugin for Cisco CallManager Release 4.0(1)*, available at

<http://www.cisco.com>

This functionality, provided for your convenience, does not replace your existing user/directory management tools. Be aware that this functionality is limited; Cisco expects that you typically will add or delete users by using other available tools.

Even if Cisco CallManager is integrated with Microsoft Active Directory, you still cannot set or modify user passwords through Cisco CallManager Administration because Active Directory does not allow passwords to be set through clear-text LDAP. To change directory passwords in a secure fashion, you can use either the CCMPwdChanger tool bundled with Cisco CallManager or the management interface provided by the directory vendor.

Even after modifying the `UMDirectoryConfiguration.ini` file on Cisco CallManager, you still have to give the CCM Directory Manager account sufficient permissions to create and delete users in AD.

When integrating multiple Cisco CallManager clusters with the same directory, keep in mind that you cannot associate the same user with devices in different clusters. Each user must be associated with a single, specific Cisco CallManager cluster at any point in time. (You can, of course, move users from one cluster to another by simply disassociating them from devices in the first cluster and associating them to devices in the second cluster.)

For user-initiated password changes and the setting of preferences, instruct your users to do the following:

- Use the directory application's interface to change their passwords. In the case of Microsoft Active Directory, password changes are done through the users' Windows workstations or by the administrator using the management tools.
- Use the Cisco CallManager User Options web page to change PINs and Cisco CallManager preferences (such as speed dials and call-forward-all numbers).

**Note**

Although the Cisco CallManager User Options web page allows users to change their passwords even after integration with AD, Cisco does not recommend this practice because users probably will not realize that they are changing their Windows passwords at the same time. Also, the communication between the client workstation and the Cisco CallManager server uses HTTP, so the password would travel across the network in clear text. You can remove the **Change your Password** option from the Cisco CallManager User Options web page simply by removing the relevant code from the active server page (ASP).

As far as adds, moves, and changes are concerned, be aware that the following operations are not supported when Cisco CallManager is integrated with the directory:

- Changing a username (in the case of AD, this is the sAMAccountName)
- Moving a user from one OU to another
- Moving or renaming the "Cisco" OU

However, to work around these restrictions, you can manually delete the user's CallManager-specific attributes (such as the profile in the "Cisco" OU and the data in the ciscoatGUID, ciscoatUserProfile, and ciscoatUserProfileString attributes for the user in question). Then the user can be renamed or moved using the directory management tools before being re-added as a Cisco CallManager subscriber. Although it is more cumbersome, this procedure preserves the user's file ownership and security principles in the directory application.

Cisco CallManager Upgrades

Because the schema might change with every major Cisco CallManager release, you should run the Cisco Customer Directory Integration plugin after every upgrade.

If you have integrated multiple Cisco CallManager clusters with the same directory, you need to extend the schema only once. If the clusters are running different Cisco CallManager releases, you should extend the schema from within the cluster that is running the most recent release.



IP Telephony Migration Options

This chapter describes the following main methods for migrating to an IP Telephony system (or any other phone system, for that matter):

- [Phased Migration, page 15-1](#)
- [Flash Cutover, or Big-Bang Migration, page 15-3](#)

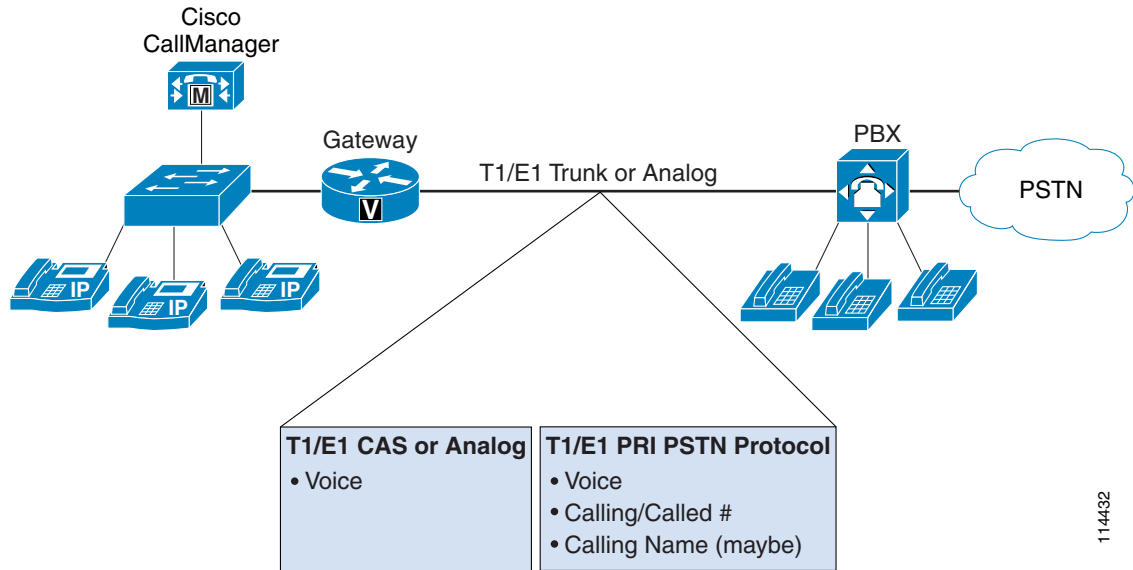
Neither method is right or wrong, and both depend upon individual customer circumstances and preferences to determine which option is most suitable.

This chapter also discusses [The Need for QSIG in Multiple-Site Enterprises, page 15-3](#).

Phased Migration

This approach typically entails a small initial IP Telephony deployment that is connected to the main corporate PBX. The choice of which signaling protocol to use is determined by the required features and functionality as well as by the cost of implementation. Cisco CallManager can provide either regular PSTN-type PRI or QSIG PRI.

PSTN-type PRI provides for basic call connectivity as well as Automatic Number Identification (ANI). In some instances, the protocol also supports calling name information, as illustrated in [Figure 15-1](#).

Figure 15-1 Features Supported by Signaling Protocols

114432

This level of connectivity is available to *all* PBXs; that is, if the PBX can connect to the public network via PRI, then it can connect to Cisco CallManager because Cisco CallManager can be configured as the "network" side of the connection.

Cisco CallManager, Release 3.3 and later, incorporates the International Standards Organization (ISO) variant of QSIG. The QSIG protocol allows for additional feature transparency between PBXs from different vendors, over and above the features that can be obtained from PSTN-type PRI, and it is therefore more appropriate for large enterprises that are already operating complex networks. (Refer to [The Need for QSIG in Multiple-Site Enterprises](#), page 15-3.)

With either PSTN-type PRI or QSIG, the process of phased migration is similar: move subscribers from the PBX to Cisco CallManager in groups, one group at a time, until the migration is complete.

The Cisco San Jose campus, consisting of some 23,000 subscribers housed in approximately 60 buildings, was migrated to IP Telephony in this manner and took just over one year from start to finish. We converted one building per weekend. All subscribers in the selected building were identified, and their extensions were deleted from the PBX on a Friday evening. At the same time, additions were made to the PBX routing tables so that anyone dialing those extension numbers would then be routed over the correct PRI trunk for delivery to Cisco CallManager. During the weekend, new extensions were created in Cisco CallManager for the subscribers, and new IP phones were delivered to their appropriate locations, ready for use by Monday morning. This process was simply repeated for each building until all subscribers had been migrated.

Flash Cutover, or Big-Bang Migration

This approach begins with implementation of a complete IP Telephony infrastructure that is redundant, highly available, QoS-enabled, and equipped with Ethernet ports that are powered inline. Once the infrastructure is complete, the IP Telephony application can then be deployed. All IP phones and gateways can be fully configured and deployed so that subscribers have two phones on their desk simultaneously, an IP phone as well as a PBX phone. This approach provides the opportunity to test the system as well as allowing subscribers time to familiarize themselves with their new IP phones. Outbound-only trunks can also be connected to the IP Telephony system, giving subscribers the opportunity to use their new IP phones to place external calls as well as internal calls.

Once the IP Telephony system is fully deployed, you can select a time slot for bringing the new system into full service by transferring the inbound PSTN trunks from the PBX to the IP Telephony gateways. You can also leave the PBX in place until such a time as you are confident with the operation of the IP Telephony system, at which point the PBX can be decommissioned.

A flash cutover has the following advantages over a phased migration:

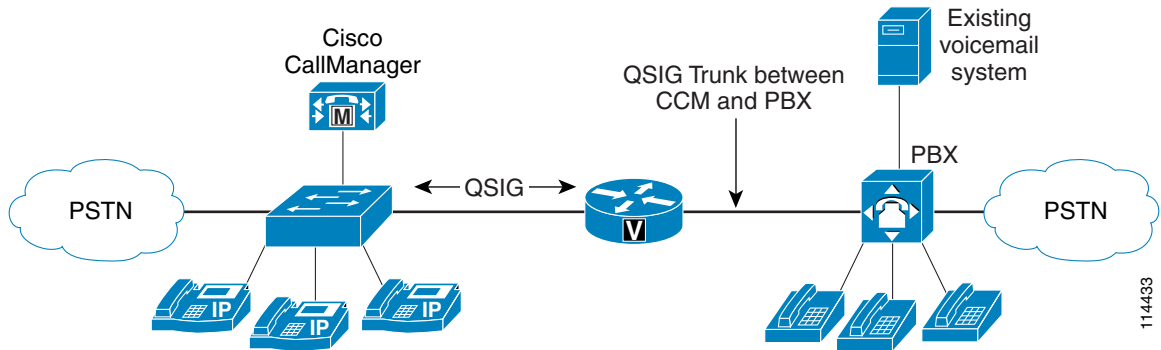
- If something unexpected occurs, the flash cutover provides a back-out plan that allows you to revert to the PBX system by simply transferring the inbound PSTN trunks from the IP Telephony gateways back to the PBX. You can also consider attempting something similar with a phased migration when you are 50% of the way through the migration.
- The flash cutover allows for verification of the IP Telephony database configuration before the system carries live PSTN traffic. This scenario can be run for any length of time prior to the cutover of the inbound PSTN trunks from the PBX to the IP Telephony gateways, thereby ensuring correct configuration of all subscriber information, phones, gateways, the dial plan, and so forth.
- Training can be carried out at a more relaxed pace by allowing subscribers to explore and use the IP Telephony system at their own leisure before the cutover of the inbound PSTN trunks.
- The system administrator does not have to make special provisions for "communities of interest." With a phased approach, you have to consider maintaining the integrity of call pick-up groups, hunt groups, shared lines, and so forth. These associations can be easily accounted for when moving the complete site in a flash cutover.

One disadvantage of the flash cutover is that it requires the IP Telephony solution to be fully funded from the beginning because the entire system must be ready prior to bringing it into service. With a phased migration, on the other hand, you can purchase individual components of the system when they are needed, and this approach does not prevent you from starting with a small trial system prior to moving to full deployment.

The Need for QSIG in Multiple-Site Enterprises

While some enterprises consist of only one location, others consist of many sites, some of which may potentially be spread over large distances. PBX networks for multi-site enterprises are usually connected using T1 or E1 PRI trunks (depending on location) running a proprietary protocol such as Avaya DCS, Nortel MCDN, Siemens CorNet, NEC CCIS, Fujitsu FIPN, or Alcatel ABC, among others. These proprietary networking protocols enable the PBXs to deliver a high level of feature transparency between subscribers.

QSIG was developed to enable the interconnection of PBXs from different vendors, thereby allowing similar levels of feature transparency. Cisco first added QSIG to Cisco CallManager Release 3.3 to enable Cisco CallManager to be part of a large enterprise network. (See [Figure 15-2](#).)

Figure 15-2 QSIG Used Between Cisco CallManager and a PBX

The following important notes apply to QSIG as implemented in Cisco CallManager Release 4.0(1):

- Cisco CallManager is currently supported only as an endpoint Private Integrated Network eXchange (PINX); that is, it has no transit/tandem functionality.
- Cisco CallManager Release 4.0(1) builds upon the previous release by adding support for Divert, Transfer, and Message Waiting Indication (MWI), thereby making centralized voice mail deployments possible.
- QSIG for Cisco CallManager Release 4.0(1) supports the following features:
 - Basic call
 - Direct Inward Dialing (DID)
 - Calling number
 - Called number
 - Connected name
 - Transfer (by join)
 - Message Waiting Indication (MWI)
 - Divert (by forward-switching)
 - Calling name restriction
 - Calling number restriction

By supporting QSIG, Cisco CallManager can be introduced into a large enterprise network while also maintaining feature transparency between subscribers. PBX locations can then be converted to IP Telephony whenever convenient.

However, unless you already have QSIG enabled on your PBX or have a specific need for its additional features and functionality, the cost of upgrading the PBX might be hard to justify if it will be retired within a short period of time. For example, why spend \$30,000 on enabling the PBX for QSIG if you plan to retire the PBX in two or three months?

Summary

Although both methods of migration work well and neither method is right or wrong, the flash cutover method usually works best in most cases. In addition, large enterprises can improve upon either migration method by using QSIG to enable Cisco CallManager to become part of the enterprise network.



IP Telephony Endpoints

This chapter summarizes various types of Cisco IP Telephony endpoints along with their features, related design considerations, and QoS recommendations. The Cisco IP Telephony endpoints can be categorized into the following major types:

- [Analog Gateway, page 16-1](#)
- [Cisco Desktop IP Phone, page 16-5](#)
- [Software-Based Endpoint, page 16-10](#)
- Specialized Cisco IP Phones, including:
 - [Wireless Endpoint, page 16-14](#)
 - [Cisco IP Conference Station, page 16-19](#)

In addition, the section on [QoS Recommendations, page 16-19](#), lists generic QoS configurations, and the [Endpoint Features Summary, page 16-30](#), lists all the endpoints features.

For information about video-enabled endpoints, refer to the *Cisco IP Video Telephony Solution Reference Network Design (SRND)*, available at

<http://www.cisco.com/go/srnd>

Analog Gateway

An analog gateway includes analog network modules, Cisco Communication Media Module (CMM) with 24-FXS port adapter, Catalyst 6500 24-FXS analog interface module, Cisco VG224, Cisco VG248, and Cisco Analog Telephone Adaptor (ATA) 186 and 188. The analog gateway is usually used to connect analog devices, such as fax machines, modems, TDD/TTYs, and analog phones, to the VoIP network so that the analog signal can be packetized and transmitted over the IP network.

Analog Network Module

Cisco analog network modules include low density voice/fax network modules (NM-1V and NM-2V), high-density voice/fax network modules (NM-HDA-4FXS), and Cisco IP Communication voice/fax modules (NM-HD-1V, NM-HD-2V, NM-HD-2VE, NM-HDV2, NM-HDV2-1T1/E1, and NM-HDV2-2T1/E1). Cisco analog network modules connect the PSTN and other legacy telephony equipment, including PBXs, analog telephones, fax machines, and key systems, to Cisco multiservice access routers. Cisco analog network modules are best suited for connecting low- to high-density analog devices to the IP network with limited call features.

Low-Density Voice/Fax Network Module

The low-density voice/fax network modules, which include NM-1V and NM-2V, contain one or two voice/fax interface cards (VIC). The voice/fax interface cards include two-port FXS VIC (VIC-2FXS); two-port FXO VIC (VIC-2FXO, VIC-2FXO-M1, VIC-2FO-M2, VIC-2FXO-M3, and VIC-2FXO-EU); two-port Direct Inward Dial VIC (VIC-2DID); two-port E&M VIC (VIC-2E/M); two-port Centralized Automated Message Accounting VIC (VIC-2CAMA); and two-port BRI VIC (VIC-2BRI-S/T-TE and VIC-2BRI-NT/TE). NM-1V and NM-2V can serve up to two and four FXS connections, respectively.

High-Density Voice/Fax Network Module

The high-density voice/fax network module, NM-HDA-4FXS, has four on-board FXS ports and room for two expansion modules (EM-HDA-8FXS or EM-HDA-4FXO) to provide up to 12 analog ports (4 FXS + 8 FXO) or 16 analog ports (12 FXS + 4 FXO). A configuration using two eight-port FXS EMs is *not* currently supported. The NM-HDA also has a connector for a daughter module (DSP-HDA-16) that provides additional DSP resources.

Cisco IP Communications Voice/Fax Network Module

Cisco IP Communications voice/fax network modules include NM-HD-1V, NM-HD-2V, NM-HD-2VE, NM-HDV2, NM-HDV2-1T1/E1, and NM-HDV2-2T1/E1. The NM-HD-1V and NM-HD-2V contain one and two VICs, respectively. The NM-HD-2VE contains two VICs or two voice/WAN interface cards (VWIC), or a combination of one VIC and one VWIC. The NM-HD-1V, NM-HD-2V, and NM-HD-2VE can serve up to four, eight, and eight FXS connections, respectively. NM-HDV2, NM-HDV2-1T1/E1, and NM-HDV2-2T1/E1 can be fitted with either digital or analog/BRI voice or WAN interface cards, with up to 4 FXS connections. The difference among these three network module is that NM-HDV2-1T1/E1 has one built-in T1/E1 port while NM-HDV2-2T1/E1 has two built-in T1/E1 ports.

The voice/fax interface cards include two-port and four-port FXS VICs (VIC2-2FXS and VIC-4FXS/DID); two-port and four-port FXO VICs (VIC2-2FXO and VIC2-4FXO); two-port Direct Inward Dial VIC (VIC-2DID); two-port E&M VIC (VIC2-2E/M); and two-port BRI VIC (VIC2-2BRI-NT/TE). The voice/wan interface cards include one-port and two-port RJ-48 multiflex trunk (MFT) T1 VWICs for both voice and WAN connections (VWIC-1MFT-T1, VWIC-2MFT-T1, and VWIC-2MFT-T1-DI); one-port and two-port RJ-4 and VWIC-2MFT-E1-D1); and one-port and two-port RJ-48 G703 VWICs for WAN connections only (VWIC-1MFT-G703 and VWIC-2MFT-G703).

Supported Platforms and Cisco IOS Requirements for Analog Network Modules

The supported platforms for Cisco analog network modules are the Cisco 2600XM, Cisco 2691, Cisco 3640 and 3660, and Cisco 3725 and 3745. [Table 16-1](#) lists the maximum number of network modules supported per platform and the minimum Cisco IOS software version required.

Table 16-1 Supported Platforms and Cisco IOS Requirements for Analog Network Modules

Platform	Maximum Number of Network Modules Supported				Minimum Cisco IOS Software Release Required			
	NM-1V, -2V	NM-HDA-4FXS	NM-HD-1V, -2V, -2VE	NM-HDV2, -1T1/E1, -2T1/E1	NM-1V, -2V	NM-HDA-4FXS	NM-HD-1V, -2V, -2VE	NM-HDV2, -1T1/E1, -2T1/E1
Cisco 2600XM	1	1	1	1	12.2(8)T or later	12.2(8)T or later	12.2(15)ZJ "Plus" images	12.3(7)T 12.3(11)T
Cisco 2691	1	1	1	1	12.2(8)T or later	12.2(8)T or later	12.2(15)ZJ "Plus" images	12.3(7)T 12.3(11)T
Cisco 3640	3	3	3	Not supported	11.3(1)T or later 12.0(1)T or later	12.2(2)XT or later 12.2(8)T or later	12.2(15)ZJ "Plus" images	N/A
Cisco 3660	6	6	6	Not supported	11.3(1)T or later 12.0(1)T or later	12.2(2)XT or later 12.2(8)T or later	12.2(15)ZJ "Plus" images	N/A
Cisco 3725	2	2	2	2	12.2(8)T or later	12.2(8)T or later	12.2(15)ZJ "Plus" images	12.3(7)T 12.3(11)T
Cisco 3745	4	4	4	4	12.2(8)T or later	12.2(8)T or later	12.2(15)ZJ "Plus" images	12.3(7)T 12.3(11)T

Cisco Communication Media Module (CMM)

The Cisco CMM is a line card that provides high-density analog, T1, and E1 gateway connections for Catalyst 6000 and Cisco 7600 series switches. The Cisco CMM can serve up to 72 FXS connections. The CMM operates as either an MGCP or H.323 gateway, and it provides Survivable Remote Site Telephony (SRST) service for up to 480 IP phones.

Cisco CMM can contain the following interface port adapters: 24-port FXS analog port adapter (WS-SVC-CMM-24FXS), six-port T1 interface port adapter (WS-SVC-CMM-6T1), six-port E1 interface port adapter (WS-SVC-CMM-6E1), and conference/transcoding port adapter (WS-SVC-CMM-ACT). [Table 16-2](#) lists the minimum software requirements for the compatible port adapters.

Table 16-2 Software Requirements for CMM Port Adapters

	WS-SVC-CMM-24FXS	WS-SVC-CMM-6T1	WS-SVC-CMM-6E1	WS-SVC-CMM-ACT
Cisco IOS Release	12.2(13)ZP	12.2(2)YK	12.2(2)YK	12.2(13)ZP
CatOS Release	8.1(1)	7.3(1)	7.3(1)	8.1(1)
Native IOS Release	12.1(15)E	12.1(14)E	12.1(14)E	12.1(13)E
Maximum number of port adapters per CMM	3	3	3	4

WS-X6624-FXS Analog Interface Module

The Cisco WS-X6624-FXS analog interface module is an MGCP-based device for connecting high-density analog devices to the IP telephony network, and it provides 24 analog ports.

Cisco VG224 Gateway

The Cisco VG224 analog gateway is a Cisco IOS-based, high-density 24-port gateway for connecting analog devices to the IP telephony network. It supports Cisco IOS Release 12.3.4-XD. The Cisco VG224 connects with Cisco CallManager using MGCP, and it has built-in MGCP failover to an H.323 connection to an SRST router. The Cisco VG224 integrates with Cisco CallManager Express using either H.323 or SIP. The Cisco VG224 supports Cisco CallManager Release 3.3(3)SR2 and later. The Cisco VG224 is best suited for high-density analog devices connecting to the IP network with limited call features, and it is more cost-effective than the Cisco VG248 for deployments of up to 24 analog ports.

Cisco VG248 Gateway

The Cisco VG248 analog gateway is a high-density Skinny Client Control Protocol (SCCP) gateway for connecting analog devices to an enterprise voice network, and it has 48 analog ports.

The Cisco VG248 has SRST capability in Release 1.2. By default, the VG248 uses its gateway as the SRST router. It can use a specific SRST router if IP address of the SRST router has been configured on the Cisco VG248. The Cisco VG248 is best suited for high-density analog devices with rich call features.

Cisco ATA 186 and 188

The Cisco Analog Telephone Adaptor (ATA) 186 or 188 is an analog telephony adaptor that can connect two analog devices to the IP telephony network. The difference between the Cisco ATA 186 and 188 is that the former has only one 10 Base-T Ethernet connection while the later has an integrated Ethernet switch providing two 10/100 Base-T Ethernet connections for itself and a co-located PC or other Ethernet-based device. The Cisco ATA 186 or 188 supports only unicast Music on Hold (MoH).

The Cisco ATA 186 and 188 can be configured in any of the following ways:

- Cisco ATA web configuration page
- Cisco ATA voice configuration menu
- Configuration file downloaded from the TFTP server

The SCCP-based ATA behaves like an SCCP IP phone. The Cisco ATA 186 or 188 can be configured as an H.323 client to Cisco CallManager or as an H.323 terminal to an H.323 gatekeeper. When the Cisco ATA registers with the gatekeeper as an H.323 terminal, the H.323 proxy must be disabled for the zone in which the ATA is located.

The Cisco ATA 186 or 188 can also be configured as a SIP client that registers with the SIP server to make phone calls with another endpoint. The Cisco ATA 186 or 188 can act as either a user agent client (UAC) when it initiates SIP requests or as a user agent server (UAS) when it responds to requests. The Cisco ATA 186 and 188 are best suited for low-density analog devices connecting to the IP network. In H.323 mode, the Cisco ATA 186 and 188 do not mark the voice bearer packets correctly by default. You must manually configure the ATA to mark voice bearer traffic with a Differentiated Services Code Point (DSCP) value of EF by changing the Type of Service (ToS) field from its default value to a ToS value of 0x000000B8. This configuration change is no longer necessary with Cisco ATA Release 3.1 and later.

Cisco Desktop IP Phone

The Cisco desktop IP phones include low-end, mid-range, and high-end IP phones.

Low-End Cisco Desktop IP Phone

The low-end Cisco desktop IP phone is best suited for low-traffic users with limited call features and budget requirements. The low-end Cisco desktop IP phones include the Cisco IP Phones 7902G, 7905G, 7910G, 7910G+SW, and 7912G.

Cisco IP Phone 7902G

The Cisco IP Phone 7902G supports a single line, and it has a single 10 Base-T Ethernet port on the back of the phone. The Cisco IP Phone 7902G does not have a liquid crystal display (LCD) screen.

Cisco IP Phone 7905G

The Cisco IP Phone 7905G supports a single line, and it has a single 10 Base-T Ethernet port on the back of the phone. The speaker operates in one-way listen mode only. The Cisco IP Phone 7905G also supports the cBarge function (see [cBarge](#), [page 16-7](#)).

Cisco IP Phone 7910G and 7910G+SW

The Cisco IP Phone 7910G supports a single line, and the speaker operates in one-way listen mode only. The Cisco IP Phone 7910G also has six fixed-feature access keys that can be configured in the customized phone button template by the administrator to provide various end-user call features. Because there are only six fixed-feature keys, the model 7910G cannot provide end users with all of the call features through one phone button template.

The only difference between the Cisco IP Phone 7910G and 7910G+SW is that the former has a single 10 Base-T Ethernet port and the latter has two 10/100 Base-T Ethernet ports.

Cisco IP Phone 7912G

The Cisco IP Phone 7912G supports a single line, and it has two 10/100 Base-T Ethernet connections. The speaker operates in one-way listen mode only. The Cisco IP Phone 7912G also supports the cBarge function (see [cBarge](#), page 16-7).

Mid-Range Cisco Desktop IP Phone

The mid-range Cisco desktop IP phone is best suited for high-traffic users with extensive call features, such as speakers, headset, and so forth. The mid-range Cisco desktop IP phones include the Cisco IP Phones 7940G and 7960G. The Cisco IP Phone 7940G can have up to two directory numbers, and the Cisco IP Phone 7960G can have a total of six directory numbers. Both phone models support the Barge and cBarge functions, and both are compatible with Cisco VT Advantage video-enabled endpoints for making video calls.

Barge

Barge uses the built-in conference bridge feature available on the Cisco IP Phone 7940G or 7960G, and it works only when a line is shared with another phone that also supports the Barge capability. A shared-line IP phone can join (barge in on) an established call between a IP phone with which it shares the line and another phone. The first shared-line phone in the call becomes a conference bridge to terminate the other two voice streams.

The Built-In-Bridge field on the phone configuration page can be set to On, Default, or Off mode. If set to Default, the phone derives its setting from the Enterprise Service Parameter "Built In Bridge - Enable" configuration. In order for Barge to work, the Built-In-Bridge field must be set to On mode on the phone that is going to be the virtual conference bridge, and the privacy setting must be set to Off mode so that the shared-line IP phone can initiate Barge.

[Figure 16-1](#) illustrates how Barge works in a situation where there are three IP phones, A, B, and C. In this example, phones B and C share a line with each other, and both phones must have the Built-In-Bridge capability. Because phone B is the virtual conference bridge, its Built-In-Bridge field must be set to On mode for Barge to work.

Figure 16-1 Barge Example

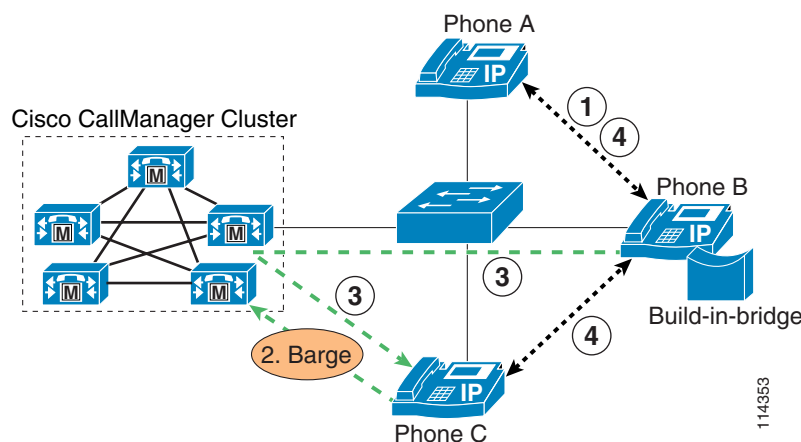


Figure 16-1 shows the following sequence of events:

1. A call is established between phones A and B.
2. Phone C tries to Barge into the existing call by sending a request to Cisco CallManager.
3. Cisco CallManager instructs both phones B and C to receive each other's media streams and to begin sending their media streams to each other.
4. Phone B becomes the conference bridge, and it terminates both phone A and C's media streams. When phone C leaves the conference, Cisco CallManager instructs phone B to close the media receive channel for phone C.

If phone A leaves the conference, the call will be terminated. If phone B leaves the conference, Cisco CallManager will instruct all three phones to stop media transmission and close the media receive channels, and it will instruct phones A and C to set up a point-to-point call. It takes about 400 ms from the time phone B activates the EndCall softkey to the time when both phones A and C set up two-way RTP streams.



Note

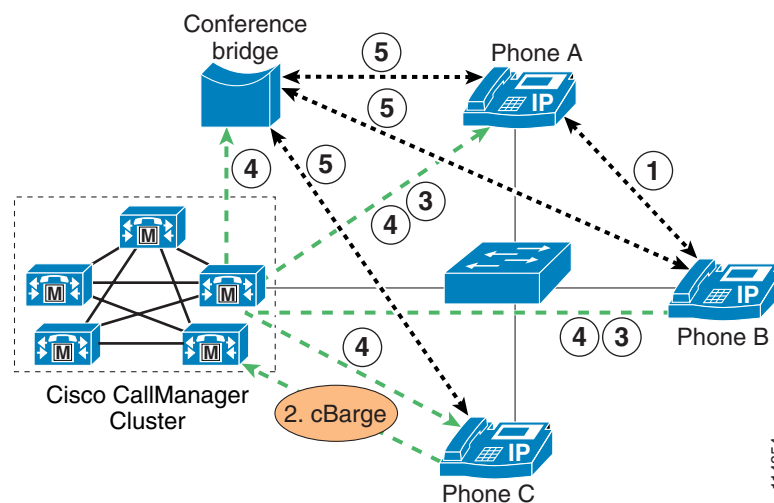
Currently, Barge is supported on G.711 calls only. For conference calls over the WAN, use the cBarge feature.

cBarge

cBarge also works on a shared-line phone, but it does not require the other shared-line phone to support the cBarge function at the same time. cBarge is similar to regular conferencing, which uses either software or hardware conference bridge resources. cBarge uses the shared conference bridge resources specified in the media resource group list (MRGL) of first shared-line phone in the call. For cBarge to work, the privacy setting must be set to Off mode so that the other shared-line IP phone can initiate the cBarge.

Figure 16-2 illustrates how cBarge works in a situation where there are three IP phones, A, B and C, with phones B and C sharing a line with each other.

Figure 16-2 cBarge Example



114354

Figure 16-2 shows the following sequence of events:

1. A call is established between phones A and B.
2. Phone C tries to cBarge into the existing call by sending a request to Cisco CallManager.
3. Cisco CallManager instructs both phones A and B to close their media receive channels and stop sending their media streams.
4. Cisco CallManager instructs all three IP phones to get ready for the incoming media streams and to send their own media streams to the shared conference bridge, which could be either the software bridge on Cisco CallManager or the hardware conference bridge configured on a conferencing device. Cisco CallManager also instructs the shared conference bridge to get ready for the incoming media streams from the IP phones and to relay these media streams to the IP phones as well.
5. The media streams from all three IP phones are terminated at the shared conference bridge, and the phones receive the incoming media streams from the shared conference bridge.

If any phone leaves the conference, Cisco CallManager will instruct all three phones to close their media receive channels and stop media transmission. Cisco CallManager will then instruct the remaining two phones to set up a point-to-point call. It takes about 400 ms from the time any phone activates the EndCall softkey to the time when the remaining two phones set up two-way RTP streams.

**Note**

For both Barge and cBarge, if phone C is disconnected or rebooting when the call is made between phones A and B, or if phone C disconnects and reboots after the call is set up between phones A and B, then phone C will not be able to Barge or cBarge into the existing call between phones A and B.

Security

The Cisco 7940G and 7960G IP Phones perform authentication transport layer security (TLS) via a locally significant certificate. The IP phone gets the locally significant certificate from the Certificate Authority Proxy Function (CAPF), which is an application running on the Cisco CallManager publisher. The certificate is a digital identity document that is hash encrypted by the Certificate Authority (CA), and it authenticates the identity of the device.

The Cisco 7940G and 7960G IP Phones have two secure modes: Authenticated mode and Encrypted mode. In the Authenticated mode, the IP phone does authentication TLS via the Label Switch Controller (LSC) and registers with Cisco CallManager on TCP port 2443. In the Encrypted mode, the IP phone not only does authentication TLS but also encrypts both its signaling and media packets. When Cisco CallManager becomes unavailable, the Cisco 7940G or 7960G IP Phone can make secure (TLS) signaling connections to SRST devices on TCP port 2443.

If a Cisco 7940G or 7960G IP Phone is in Encrypted mode and associated with a wideband codec region, Cisco CallManager will ignore the wideband codec during an encrypted call and instead will choose another supported codec from the codec list for the region. However, for a non-secure or authenticated call, Cisco CallManager will use the wideband codec to set up the call. A Cisco 7940G or 7960G IP Phone in Encrypted mode can barge into an existing encrypted call as well as non-secure and authenticated calls, but these phones cannot be the virtual bridge of an encrypted Barge call.

A Cisco 7940G or 7960G IP Phone displays two icons indicating its security modes for both the calls and connections with Cisco CallManager. A shield icon displays on the phone's screen when the call or connection is authenticated but not encrypted. For an encrypted call, a padlock icon displays. When an IP phone sets up an encrypted call and then is put on hold or transferred to another IP phone, the padlock icon is replaced by the off-hook phone icon, which means that the media streams associated with these tasks are not encrypted.

Encrypted conference bridging and cBarge are not supported currently. If there are three Cisco 7940G or 7960G IP Phones in encrypted mode setting up a conference, the media streams from these three participants are not encrypted. If any party leaves the conference, the remaining two participants will set up an encrypted point-to-point call. Encrypted video calls are not supported. If an IP phone in Encrypted mode makes a video call, the video RTP media traffic is not encrypted but the audio media packets are encrypted.

High-End Cisco Desktop IP Phone

The Cisco IP Phone 7970G, the high-end Cisco desktop IP phone, is best suited for high-traffic users with extensive call features. The Cisco IP Phone 7970G has a high-resolution, color touch-screen display as well as more function keys and more security features than the mid-range Cisco IP phones. The Cisco IP Phone 7970G can also have up to eight directory numbers.

With the current phone software load (TERM70.5-0-1-4DEV), the Cisco IP Phone 7970G supports the Barge and cBarge features. The Cisco IP Phone 7970G is also compatible with Cisco VT Advantage video-enabled endpoints for making video calls. Currently, the Cisco IP Phone 7970G is the only Cisco IP phone that supports both Cisco pre-standard Power-over-Ethernet (PoE) and the IEEE 802.3af PoE. For the Cisco IP Phone 7970G to have full display brightness, you must use the external power adaptor (CP-PWR-CUBE2) with both inline power and 802.3af PoE.

Security

The Cisco IP Phone 7970G has a manufacturing installed certificate (MIC) that is written into flash and cannot be erased by a hardware reset. The Cisco IP Phone 7970G has two secure modes, Authenticated mode and Encrypted mode. In the Authenticated mode, the phone does authentication TLS via the MIC and registers with Cisco CallManager on TCP port 2443. In the Encrypted mode, the phone not only does authentication TLS but also encrypts both its signaling and media packets. If Cisco CallManager becomes unavailable, the Cisco IP Phone 7970G can make secure (TLS) signaling connections to SRST devices on TCP port 2443. The Cisco IP Phone 7970G in Encrypted mode can Barge into an existing encrypted call as well as non-secure and authenticated calls. Unlike the Cisco 7940G or 7960G, the Cisco 7970G IP Phone can be the virtual bridge of an encrypted Barge call.

The Cisco IP Phone 7970G displays two icons indicating its security modes for both the calls and connections with Cisco CallManager. A shield icon appears on the phone's screen when the call or connection is authenticated but not encrypted. A padlock icon appears when the call or connection is both authenticated and encrypted. A non-secure call is indicated by an off-hook phone icon. When a Cisco IP Phone 7970G sets up an encrypted phone call and then is put on hold or transferred to another IP phone, the padlock icon is replaced by the off-hook phone icon, which means that the media streams associated with these tasks are not encrypted.

Encrypted conference bridging and cBarge are not supported currently. If there are three Cisco 7970G IP Phones in encrypted mode setting up a conference, the media streams from these three participants are not encrypted. If any party leaves the conference, the remaining two participants will set up an encrypted point-to-point call. Encrypted video calls are not supported. If an IP phone in Encrypted mode makes a video call, the video RTP media traffic is not encrypted but the audio media packets are encrypted.

Software-Based Endpoint

Software-based endpoints include Cisco IP SoftPhone and Cisco IP Communicator. A software-based endpoint is an application installed on a client PC, and it registers with, and is controlled by, Cisco CallManager.

Cisco IP SoftPhone

This section summarizes the following design considerations that apply when using Cisco IP SoftPhone with Cisco CallManager:

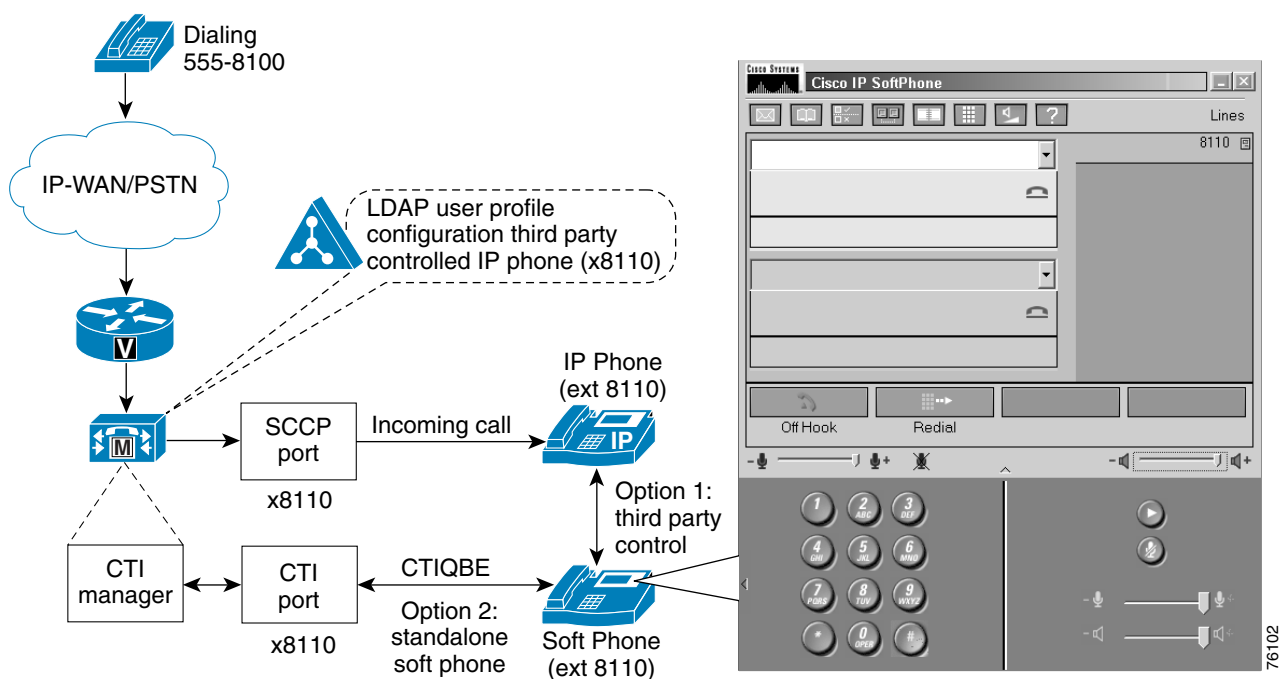
- [Maximum Cisco IP SoftPhone Configuration Limits, page 16-11](#)
- [Codec Selection, page 16-11](#)
- [Call Admission Control, page 16-12](#)

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/english/softphon/index.htm

Figure 16-3 illustrates that the Cisco IP SoftPhone application can either monitor or control the hardware IP phone associated with it. In the case of a third-party controlled phone, Cisco IP SoftPhone acts as a virtual extension of the desktop IP phone. The Cisco IP SoftPhone application is able to see and handle incoming and outgoing calls for the hardware phone. From the perspective of provisioning devices and CTI resources, each user with this configuration would be provisioned as a third-party controlled IP phone. As a CTI port, the Cisco IP SoftPhone is a dedicated line to process calls directly to the client machine without the additional control and monitoring of a desktop phone.

Figure 16-3 Cisco IP SoftPhone Device Association Options



The Cisco IP SoftPhone is not able to run a CTI port and a third-party control phone with the same directory number (DN) at the same time. As shown in [Figure 16-3](#), the user is able to have x8110 as either a CTI port or a control for the desktop phone.

For more information on device and resource provisioning, refer to the chapter on [Call Processing](#), page 8-1.

Maximum Cisco IP SoftPhone Configuration Limits

Regardless of the device limits allowed per server, there are maximum limits on the number CTI devices you can configure in Cisco CallManager. The CTI device limits as they apply to Cisco IP SoftPhone are as follows:

- Maximum of 800 Cisco IP SoftPhones per Cisco Media Convergence Server (MCS) 7825 or 7835; maximum of 3,200 Cisco IP SoftPhones per cluster of MCS 7825s or 7835s
- Maximum of 2,500 Cisco IP SoftPhones per Cisco Media Convergence Server (MCS) 7845; maximum of 10,000 Cisco IP SoftPhones per cluster of MCS 7845s

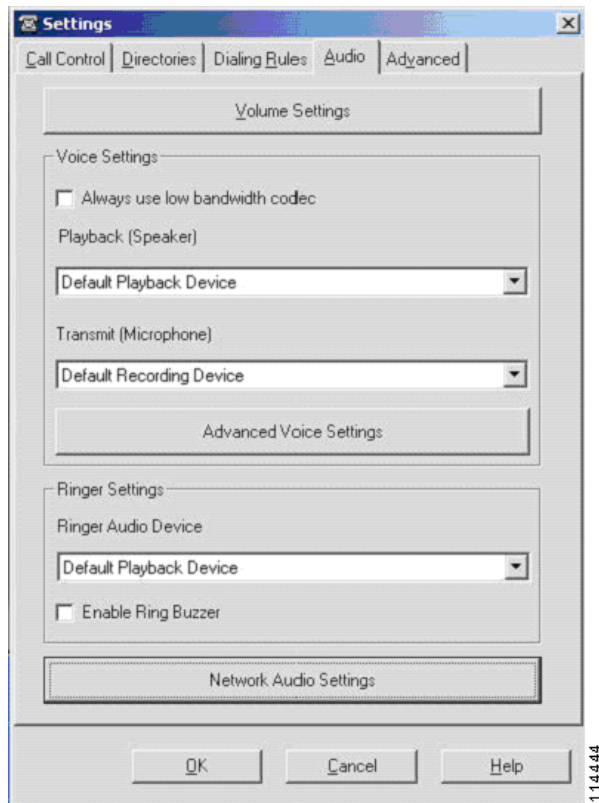
The following assumptions apply to the preceding maximum Cisco IP SoftPhone limits:

- Each Cisco IP SoftPhone is configured with one line appearance.
- Each Cisco IP SoftPhone is processing an estimated six or fewer busy hour call attempts (BHCA).
- No other CTI applications requiring CTI devices are configured in the Cisco CallManager cluster.

Codec Selection

Cisco IP SoftPhone supports G.711, G.723, and G.729a codecs. Cisco recommends G.729a low-bandwidth codec configurations in deployments with telecommuters connecting their Cisco IP SoftPhones across the WAN.

Because Cisco CallManager does not support the G.723 codec, Cisco IP Softphone has two bandwidth codec settings to use: G.711 is the default setting, with a user-configurable option to select the low-bandwidth codec setting of G.729 on the TAPI Service Provider (TSP) client (see [Figure 16-4](#)). Refer to the chapter on [Network Infrastructure](#), page 3-1, for details on provisioning network bandwidth.

Figure 16-4 Cisco IP Softphone Audio Setting

Cisco IP SoftPhone users with low-bandwidth connections across the WAN should consider selecting this low-bandwidth G.729 codec setting.

Call Admission Control

Call admission control ensures that there is enough bandwidth available to process IP phone calls over the network. While there are several mechanisms for implementing call admission control, Cisco IP SoftPhone uses only the *locations* mechanism configured in Cisco CallManager. Refer to the chapter on [IP Telephony Deployment Models, page 2-1](#), for more information on call admission control with Cisco CallManager locations.

Locations-based call admission control is effective in managing call bandwidth as long as the Cisco IP SoftPhone is mobile only within a single Cisco CallManager location. However, call admission control becomes problematic if the Cisco IP SoftPhone moves between Cisco CallManager locations. For more information on mobility, refer to [Device Mobility and Cisco CallManager, page 16-18](#).

Cisco IP Communicator

Cisco IP Communicator is similar to Cisco IP SoftPhone, and it is an ideal solution for remote users and telecommuters. Cisco IP Communicator is an SCCP-based standalone device. From the client PC screen, it looks like a Cisco IP Phone 7970G. This section summarizes the following design considerations that apply when using Cisco IP Communicator with Cisco CallManager:

- [Maximum IP Communicator Configuration Limits](#), page 16-13
- [Codec Selection](#), page 16-13
- [Call Admission Control](#), page 16-14

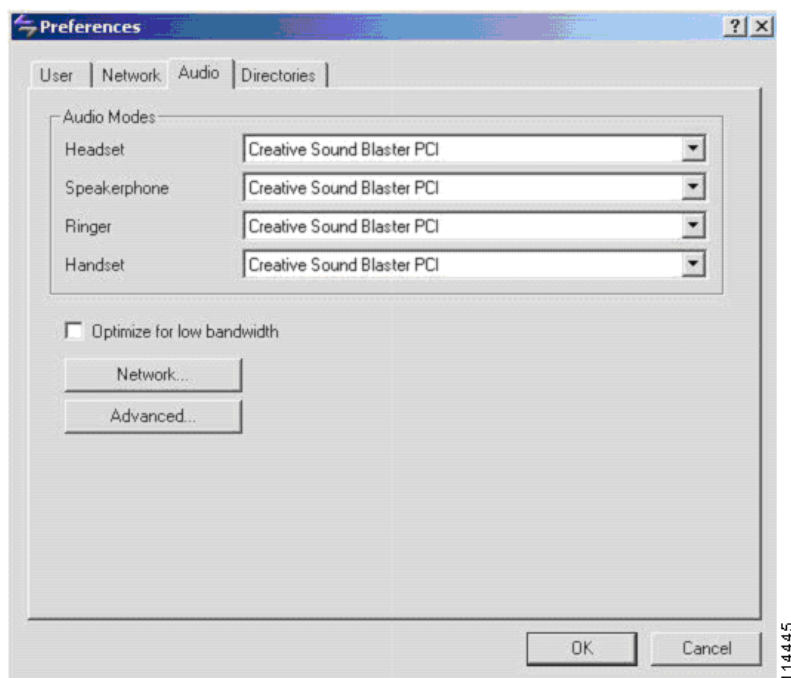
Maximum IP Communicator Configuration Limits

Because Cisco IP Communicator is an SCCP standalone device, the design guidelines for IP phones in the various IP Telephony deployment models still hold true for the Cisco IP Communicator. Refer to the chapter on [IP Telephony Deployment Models](#), page 2-1, for details.

Codec Selection

Cisco IP Communicator supports G.711 and G.729a codecs. The codec selection can be done by configuring the region in which Cisco IP Communicator is located. Cisco recommends G.729a low-bandwidth codec configurations in deployments with telecommuters connecting their Cisco IP Communicator across the WAN. Cisco IP Communicator also has a low-bandwidth codec overriding capability in a G.711 region, which can be enabled by checking the Optimize for Low Bandwidth option in the Audio setting window (see [Figure 16-5](#)). In this case, the Cisco IP Communicator will use a G.729 codec to set up a call with another phone in the same region.

Figure 16-5 Cisco IP Communicator Audio Setting



Call Admission Control

Call admission control ensures that there is enough bandwidth available to process IP phone calls over the network. While there are several mechanisms for implementing call admission control, Cisco IP Communicator uses only the *locations* mechanism configured in Cisco CallManager. Refer to the chapter on [IP Telephony Deployment Models, page 2-1](#), for more information on call admission control with Cisco CallManager locations.

Locations-based call admission control is effective in managing call bandwidth as long as the Cisco IP Communicator is mobile only within a single Cisco CallManager location. However, call admission control becomes problematic if the Cisco IP Communicator is moved between Cisco CallManager locations. For more information on mobility, refer to [Device Mobility and Cisco CallManager, page 16-18](#).

Wireless Endpoint

Cisco wireless endpoints use a wireless LAN (WLAN) infrastructure via wireless access points (APs) to provide telephony functionality and features. This type of endpoint is ideal for environments with the need for mobile users within a area where traditional wired phones are undesirable or problematic. (Refer to [Wireless LAN Infrastructure, page 3-40](#), for more information about wireless network design.)

The Cisco Wireless IP Phone 7920G is a hardware-based phone with a built-in radio antenna that enables 802.1b wireless LAN connectivity to the network. These phones register with Cisco CallManager using Skinny Client Control Protocol (SCCP), just like the hardware-based phones and Cisco IP Communicator. For more information, refer to the *Cisco Wireless IP Phone 7920 Design and Deployment Guide*, available at

<http://www.cisco.com/go/srnd>

Site Survey

Before deploying the Cisco Wireless IP Phone 7920G, you must perform a complete site survey to determine the appropriate number and location of APs required to provide radio frequency (RF) coverage. Your site survey should take into consideration which types of antennas will provide the best coverage, as well as where sources of RF interference might exist. A site survey requires the use of the Site Survey tool on the Cisco Wireless IP Phone 7920G (accessed via **Menu > Network Config > Site Survey**) and the Aironet Client Utility Site Survey Tool used with a Cisco Aironet NIC card on a laptop or PC. Additional third-party tools can also be used for site surveys; however, Cisco highly recommends that you conduct a final site survey using the Cisco Wireless IP Phone 7920G because each endpoint or client radio can behave differently depending on antenna sensitivity and survey application limitations.

Authentication

To connect to the wireless network, the Cisco Wireless IP Phone 7920G must first use one of the following authentication methods to associate and communicate with the AP:

- Cisco LEAP

This method allows the Cisco Wireless IP Phone 7920G and AP to be authenticated mutually based on a username and password. Upon authentication, the dynamic key is generated and used for encrypting traffic between the Cisco Wireless IP Phone 7920G and the AP. A LEAP-compliant Radius authentication server, such as the Cisco Secure Access Control Server (ACS), is required to provide access to the user database.

- Static Wired Equivalent Privacy (WEP)

This method involves the configuration of static 10 (40-bit) or 26 (128-bit) character keys on the Cisco Wireless IP Phone 7920G and the AP. This method is AP-based authentication in which access to the network is gained if the device has a matching key.

- Open Authentication

This method requires no exchange of identifying information between the Cisco Wireless IP Phone 7920G and the AP. Cisco does *not* recommend this method because it provides no secure exchange of voice or signaling, and it allows any rogue device to associate to the AP.

Capacity

Each AP can support a maximum of seven active G.711 voice calls or eight G.729 calls. If these numbers are exceeded, poor quality can result due to dropped or delayed voice packets or dropped calls. AP rates set lower than 11 Mbps will result in lower call capacity per AP.

Based on these active call capacity limits, and using Erlang ratios, you can calculate the number of Cisco Wireless 7920G IP Phones that each AP can support. For example, given a typical user-to-call capacity ratio of 3:1, a single AP can support 21 to 24 Cisco Wireless 7920G IP Phones, depending whether the codec used is G.711 or G.729. However, this number does not take into consideration the possibility that other Cisco Wireless 7920G IP Phones could roam to this AP, so a lower number of phones per AP is more realistic.

The number of APs per VLAN or Layer 2 subnet should also be considered. To optimize memory and performance on the APs, Cisco recommends deploying no more than 30 APs on a single VLAN or subnet. This recommendation, taken with typical user-to-call capacity ratios, limits the number of Cisco Wireless 7920G IP Phones per Layer 2 subnet to approximately 500 (or about 15 to 17 Cisco Wireless 7920G IP Phones per AP).

These capacities were calculated with voice activity detection (VAD) disabled and a packetization sample size of 20 milliseconds (ms). VAD is a mechanism for conserving bandwidth by not sending RTP packets while no speech is occurring during the call. However, enabling or disabling VAD is a global cluster-wide configuration parameter on Cisco CallManager. (It is referred to as Silence Suppression in Cisco CallManager.) Thus, if VAD is enabled for the Cisco Wireless IP Phone 7920G, then it will be enabled for all devices in the Cisco CallManager cluster. Cisco recommends leaving VAD (Silence Suppression) *disabled* to provide better overall voice quality.

At a sampling rate of 20 ms, a voice call will generate 50 packets per second (pps) in either direction. Cisco recommends setting the sample rate to 20 ms for almost all cases. By using a larger sample size (for example, 30 or 40 ms), you can increase the number of simultaneous calls per AP, but a larger end-to-end delay will result. In addition, the percentage of acceptable voice packet loss within a wireless

environment decreases dramatically with a larger sample size because more of the conversation is missing when a packet is lost. For more information about voice sampling size, see [Bandwidth Provisioning, page 3-28](#).

Phone Configuration

You can configure the Cisco Wireless IP Phone 7920G either through the phone's keypad or with the 7920 Configuration Utility running on a PC that is attached to the phone via a USB cable. In either case, you must configure the following parameters:

- **Network Configuration**
Configure either the DHCP server address or static settings such as IP address, subnet mask, default gateway, TFTP server, and DNS server, as appropriate for the network. These settings can be found on the Cisco Wireless IP Phone 7920G under **Menu > Network Config > Current Config**.
- **Wireless Configuration**
Configure the service set identifier (SSID) for the voice VLAN and the authentication type, including the WEP key and/or LEAP username and password when appropriate. These settings can be found on the Cisco Wireless IP Phone 7920G under **Menu > Network Config > 802.11b Configuration**.

Roaming

Currently the Cisco Wireless IP Phone 7920G is able to roam at Layer 2 (within the same VLAN or subnet) and still maintain an active call.

Layer 2 roaming occurs in the following situations:

- During the initial boot-up of the Cisco Wireless IP Phone 7920G, the phone roams to a new AP for the first time.
- If the Cisco Wireless IP Phone 7920G receives no beacons or responses from the AP to which it is currently associated, the phone assumes that the current AP is unavailable and it attempts to roam and associate with a new AP.
- The Cisco Wireless IP Phone 7920G maintains a list of eligible AP roam targets. If conditions change on the current AP, the phone consults the list of available AP roam targets. If one of the roam targets is determined to be a better choice, then the phone attempts to roam to the new AP.
- If the configured SSID or authentication type on the Cisco Wireless IP Phone 7920G is changed, the phone must roam to re-associate with an AP.

In trying to determine eligible AP roam targets for Layer 2 roaming, the wireless IP phone uses the following variables to determine the best AP to associate with:

- **Relative Signal Strength Indicator (RSSI)**
Used by the wireless IP phone to determine the signal strength and quality of available APs within an RF coverage area. The phone will attempt to associate with the AP that has the highest RSSI value and matching authentication/encryption type.
- **QoS Basis Service Set (QBSS)**
Enables the AP to communicate channel utilization information to the wireless phone. The phone will use the QBSS value to determine if it should attempt to roam to another AP, because APs with high channel utilization might not be able to handle VoIP traffic effectively.

- RSSI-Differential Threshold

The wireless IP phone will roam if the next AP RSSI is higher than the current AP RSSI by at least this threshold amount. The default threshold is 15.

- QBSS-Differential Threshold

The wireless IP phone will roam if the next AP QBSS is lower than the current AP QBSS by at least this threshold amount. The default threshold is 15.

The wireless IP phone uses the following steps to determine which AP it should roam to:

1. Find APs that are advertising QBSS in their beacons. If any of these APs meet the QBSS-differential threshold criteria, then begin the roaming process to one of them.
2. If no APs are advertising QBSS, or if the advertising APs do not meet the differential threshold criteria, then look for APs that are not advertising QBSS but that have acceptable RSSI levels, and begin the roaming process to one of them.

Layer 2 roaming times for the wireless IP phone depend on the type of authentication used. If static WEP keys are used for authentication between the phone and the AP, Layer 2 roaming occurs in less than 100 ms. If LEAP (with local Cisco Secure ACS authentication) is used, Layer 2 roaming occurs in 200 to 400 ms. Use of Fast Secure Roaming can decrease the LEAP authentication time to less than 150 ms for Layer 2 roams.

Layer 3 roaming occurs when the Cisco 7920 Wireless IP Phone moves from one AP to another AP and crosses a subnet boundary. With the release of the new Cisco Catalyst 6500 Series Wireless LAN Services Module (WLSM), the Cisco 7920 Wireless IP Phone now supports Layer 3 mobility with survivable calls while using Static WEP. Cisco Centralized Key Management (Cisco CKM) enables the Cisco 7920 Wireless IP Phone to achieve full Layer 3 mobility while using LEAP. For details about the Cisco WLSM, refer to the product documentation available at

<http://www.cisco.com>

AP Call Admission Control

Call admission control mechanisms in Cisco CallManager or in a gatekeeper can control WAN bandwidth utilization and provide QoS for existing calls, but both mechanisms are applied only at the beginning of a call. For calls between static devices, this type of call admission control is sufficient. However, for a call between two mobile wireless devices such the Cisco Wireless IP Phone 7920G, there must also be a call admission control mechanism at the AP level because the wireless devices may roam from one AP to another.

The AP mechanism for call admission control is QBSS, which is the beacon information element that enables the AP to communicate channel utilization information to the wireless IP phone. As previously mentioned, this QBSS value helps the phone determine whether it should roam to another AP. A lower QBSS value indicates that the AP is a good candidate to roam to, while a higher QBSS value indicates that the phone should not roam to this AP.

While this QBSS information is useful, it does not provide a 100% guarantee that calls will retain proper QoS during roaming. When a Cisco Wireless 7920 IP Phone is associated to an AP with a high QBSS, the AP will prevent a call from being initiated or received by rejecting the call setup and sending a Network Busy message to the initiating phone. However, once a call is set up between a wireless IP phone and another endpoint, the phone may roam and associate with an AP with a high QBSS, thus resulting in an oversubscription of the available bandwidth on that AP.

Device Mobility and Cisco CallManager

When a wireless IP phone becomes a mobile device and moves from one location to another, the following potential problems could arise:

- Inaccurate bandwidth accounting for Cisco CallManager locations-based call admission control

When a wireless IP phone roams from one location to another, there is currently no dynamic mechanism in Cisco CallManager to update the phone's location for purposes of call admission control. As a result, bandwidth can be subtracted from locations that are not actually using the bandwidth, and bandwidth can be used in other locations but not taken into account by the locations-based call admission control, thus causing oversubscription of the WAN bandwidth.

- Inappropriate codec selection

When a wireless IP phone roams from one location to another, there is currently no dynamic mechanism in Cisco CallManager to update its region and/or device pool for purposes of determining codec type. As a result, the wrong codec can be used throughout the telephony network.

- Inappropriate PSTN gateway selection

When a wireless IP phone roams from one location to another, there is currently no dynamic mechanism in Cisco CallManager to update the dial plan to specify the local PSTN gateway. As a result, the wireless IP phone might use a remote PSTN gateway for PSTN access. If the wireless IP phone places an emergency 911 call through this remote PSTN gateway, the emergency services will be directed to the location of the remote PSTN gateway and not to the location of the wireless IP phone that initiated the call.

**Note**

If Cisco Emergency Responder (ER) is deployed, then 911 calls will be routed to the local PSTN gateway and to the appropriate public safety answering point (PSAP). However, call admission control will still be unaware of the bandwidth used by this call, and the wrong codec might be selected.

To prevent these device mobility problems, you must manually reconfigure the following parameters of the wireless IP phone in Cisco CallManager each time the phone is moved from one physical location to another:

- Call admission control location
- Device pool and region
- Calling search space

These parameters must be adjusted appropriately for each location to which the wireless IP phone moves. Other parameters might have to be reconfigured manually if advanced or non-standard features are required. For example, the media resource group list (for conferencing, transcoding, and music-on-hold resources) and the automated alternate routing (AAR) calling search space and group (if AAR is configured) must be reconfigured to ensure that local media resources are used and that automatic alternate call routing is appropriate for each location.

Note that these device mobility issues come into play not only for the wireless IP phone but also for any device that moves between locations, including Cisco IP SoftPhone, Cisco IP Communicator, and any Cisco hardware IP phone that is physically moved from one place to another.

Finally, these device mobility issues affect both centralized and distributed call processing deployments.

Cisco IP Conference Station

The Cisco IP Conference Station combines conference room speaker-phone technology with Cisco IP Communications technology. The Cisco IP Conference Station is best suited for use in conferencing environments providing 360-degree room coverage.

Both the Cisco IP Conference Station 7935 and 7936 have an external speaker and three built-in microphones. The Cisco IP Conference Station 7936 requires Cisco CallManager Release 3.3(3) SR3 or later. The Cisco IP Conference Station 7936 also features a pixel-based LCD display with backlighting, and optional extension microphones can be connected to it for extended microphone coverage in larger rooms.

QoS Recommendations

This section provides the basic QoS guidelines and configurations for the Cisco Catalyst switches most commonly deployed with IP Telephony endpoints. For more details, refer to the *Quality of Service* design guide at

<http://www.cisco.com/go/srnd>

Cisco VG224 and VG248

Analog gateways are trusted endpoints. For Cisco VG224 and VG248 gateways, configure the switch to trust the DSCP value of the VG248 packets. The following sections list the commands to configure the most common Cisco Catalyst switches for the Cisco VG224 and VG248 analog gateways.



Note

In the following sections, *vvlan_id* is the voice VLAN ID and *dvlan_id* is the data VLAN ID.

Cisco 2950

```
CAT2950(config)#interface interface-id
CAT2950(config-if)#mls qos trust dscp
CAT2950(config-if)#switchport mode access
CAT2950(config-if)#switchport access vlan vvlan_id
```



Note

The **mls qos trust dscp** command is available only with Enhanced Image (EI).

Cisco 2970 or 3750

```
CAT2970(config)#mls qos
CAT2970(config)#interface interface-id
CAT2970(config-if)#mls qos trust dscp
CAT2970(config-if)#switchport mode access
CAT2970(config-if)#switchport access vlan vvlan_id
```

Cisco 3550

```
CAT3550(config)#mls qos
CAT3550(config)#interface interface-id
CAT3550(config-if)#mls qos trust dscp
CAT3550(config-if)#switchport mode access
```

```
Cat3550(config-if)#switchport access vlan vvlan_id
```

Cisco 4500 with SUP3II, IV, or V

```
CAT4500(config)#qos
CAT4500(config)#interface interface-id
CAT4500(config-if)#qos trust dscp
CAT4500(config-if)#switchport mode access
CAT4500(config-if)#switchport access vlan vvlan_id
```

Cisco 6500

```
CAT6500>(enable)set qos enable
CAT6500>(enable)set port qos 2/1 vlan-based
CAT6500>(enable)set vlan vvlan_id mod/port
CAT6500>(enable)set port qos mod/port trust trust-dscp
```

Cisco ATA 186 and Conference Station

Because the Cisco Analog Telephone Adaptor (ATA) 186 and IP Conference Station are trusted endpoints, their QoS configurations are identical to those described in the section on [Cisco VG224 and VG248](#), page 16-19.

Cisco ATA 188 and IP Phones

For the Cisco Analog Telephone Adaptor (ATA) 188 and IP Phones, Cisco recommends segregating the voice VLAN from the data VLAN. For the Cisco ATA 186, 7902, 7905, 7910, and IP Conference Station, Cisco still recommends configuring voice and data VLAN segregation and an auxiliary voice VLAN. In this way, the same access-layer configurations can be used with different IP phone models and ATAs, and end users can plug their IP phones or ATAs into different access ports on the switch and get the same treatment. For the Cisco ATA 186, 7902, 7905, 7910, and IP Conference Stations, the command to override the CoS value of the frames from the attached PC has no effects because these devices do not have a PC connected to them.

The following sections list the configuration commands for IP phones on the most commonly deployed Cisco Catalyst switches.

Cisco 2950

```
CAT2950(config)#
CAT2950(config)#class-map VVLAN
CAT2950(config-cmap)# match access-group name VVLAN
CAT2950(config-cmap)#class-map VLAN
CAT2950(config-cmap)# match access-group name DVLAN
CAT2950(config-cmap)#exit
CAT2950(config)#
CAT2950(config)#policy-map IPPHONE-PC
CAT2950(config-pmap)# class VVLAN
CAT2950(config-pmap-c0# set ip dscp 46
CAT2950(config-pmap-c)# police 1000000 8192 exceed-action-drop
CAT2950(config-pmap)# class DVLAN
CAT2950(config-pmap-c0# set ip dscp 0
CAT2950(config-pmap-c)# police 5000000 8192 exceed-action-drop
CAT2950(config-pmap-c)#exit
CAT2950(config-pmap)#exit
```

```

CAT2950(config)#
CAT2950(config)#interface interface-id
CAT2950(config-if)#mls qos trust device cisco-phone
CAT2950(config-if)#mls qos trust cos
CAT2950(config-if)#switchport mode access
CAT2950(config-if)#switchport voice vlan vvlan_id
CAT2950(config-if)#switchport access vlan dvlan_id
CAT2950(config-if)#service-policy input IPPHONE-PC
CAT2950(config-if)#exit
CAT2950(config)#
CAT2950(config)#ip access-list standard VVLAN
CAT2950(config-std-nacl)# permit voice_IP_subnet wild_card_mask
CAT2950(config-std-nacl)#exit
CAT2950(config)#ip access-list standard DVLAN
CAT2950(config-std-nacl)# permit data_IP_subnet wild_card_mask
CAT2950(config-std-nacl)#end

```

**Note**

The **mls qos map cos-dscp** command is available only with Enhanced Image (EI). With Standard Image (SI), this command is not available and the default CoS-to-DSCP mapping is as follows:

Cos Value	0	1	2	3	4	5	6	7
DSCP Value	0	8	16	24	32	40	48	56

Cisco 2970 or 3750

```

CAT2970(config)# mls qos map cos-dscp 0 8 16 24 34 46 48 56
CAT2970(config)# mls qos map policed-dscp 0 24 to 8
CAT2970(config)#
CAT2970(config)#class-map match-all VVLAN-VOICE
CAT2970(config-cmap)# match access-group name VVLAN-VOICE
CAT2970(config-cmap)#
CAT2970(config-cmap)#class-map match-all VVLAN-CALL-SIGNALING
CAT2970(config-cmap)# match access-group name VVLAN-CALL-SIGNALING
CAT2970(config-cmap)#
CAT2970(config-cmap)#class-map match-all VVLAN-ANY
CAT2970(config-cmap)# match access-group name VVLAN-ANY
CAT2970(config-cmap)#
CAT2970(config-cmap)# policy-map IPPHONE-PC
CAT2970(config-pmap)#class VVLAN-VOICE
CAT2970(config-pmap-c)# set ip dscp 46
CAT2970(config-pmap-c)# police 128000 8000 exceed-action drop
CAT2970(config-pmap-c)# class VVLAN-CALL-SIGNALING
CAT2970(config-pmap-c)# set ip dscp 24
CAT2970(config-pmap-c)# police 32000 8000 exceed-action policed-dscp-transmit
CAT2970(config-pmap-c)# class VVLAN-ANY
CAT2970(config-pmap-c)# set ip dscp 0
CAT2970(config-pmap-c)# police 32000 8000 exceed-action policed-dscp-transmit
CAT2970(config-pmap-c)# class class-default
CAT2970(config-pmap-c)# set ip dscp 0
CAT2970(config-pmap-c)# police 5000000 8000 exceed-action policed-dscp-transmit
CAT2970(config-pmap-c)# exit
CAT2970(config-pmap)# exit
CAT2970(config)#
CAT2970(config)#
CAT2970(config)#interface interface-id
CAT2970(config-if)# switchport voice vlan vvlan_id
CAT2970(config-if)# switchport access vlan dvlan_id
CAT2970(config-if)# mls qos trust device cisco-phone

```

```

CAT2970(config-if)# service-policy input IPPHONE-PC
CAT2970(config-if)# exit
CAT2970(config)#
CAT2970(config)#ip access list extended VVLAN-VOICE
CAT2970(config-ext-nacl)# permit udp Voice_IP_Subnet Subnet_Mask any range 16384 32767
dscp ef
CAT2970(config-ext-nacl)# exit
CAT2970(config)#ip access list extended VVLAN-CALL-SIGNALING
CAT2970(config-ext-nacl)# permit tcp Voice_IP_Subnet Subnet_Mask any range 2000 2002 dscp
cs3
CAT2970(config-ext-nacl)# exit
CAT2970(config)#ip access list extended VVLAN-ANY
CAT2970(config-ext-nacl)# permit ip Voice_IP_Subnet Subnet_Mask any
CAT2970(config-ext-nacl)# end
CAT2970#

```

Cisco 3550

```

CAT3550(config)# mls qos map cos-dscp 0 8 16 24 34 46 48 56
CAT3550(config)# mls qos map policed-dscp 0 24 to 8
CAT3550(config)#class-map match-all VOICE
CAT3550(config-cmap)# match ip dscp 46
CAT3550(config-cmap)#class-map match-any CALL SIGNALING
CAT3550(config-cmap)# match ip dscp 26
CAT3550(config-cmap)# match ip dscp 24
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all VVLAN-VOICE
CAT3550(config-cmap)# match vlan vvlan_id
CAT3550(config-cmap)# match class-map VOICE
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all VVLAN-CALL-SIGNALING
CAT3550(config-cmap)# match vlan vvlan_id
CAT3550(config-cmap)# match class-map CALL SIGNALING
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all ANY
CAT3550(config-cmap)# match access-group name ACL_Name
CAT3550(config-cmap)#
CAT3550(config-cmap)# class-map match-all VVLAN-ANY
CAT3550(config-cmap)# match vlan vvlan_id
CAT3550(config-cmap)# match class-map ANY
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all DVLAN-ANY
CAT3550(config-cmap)# match vlan dvlan_id
CAT3550(config-cmap)# match class-map ANY
CAT3550(config-cmap)#
CAT3550(config-cmap)#policy-map IPPHONE-PC
CAT3550(config-pmap)# class VVLAN-VOICE
CAT3550(config-pmap-c)# set ip dscp 46
CAT3550(config-pmap-c)# police 128000 8000 exceed-action drop
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class VVLAN-CALL-SIGNALING
CAT3550(config-pmap-c)# set ip dscp 24
CAT3550(config-pmap-c)# police 32000 8000 exceed-action drop
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class VVLAN-ANY
CAT3550(config-pmap-c)# set ip dscp 0
CAT3550(config-pmap-c)# police 32000 8000 exceed-action drop
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class DVLAN-VOICE
CAT3550(config-pmap-c)# set ip dscp 0
CAT3550(config-pmap-c)# police 5000000 8000 exceed-action drop
CAT3550(config-pmap-c)#exit
CAT3550(config-pmap)#exit

```

```

CAT3550(config)#interface interface-id
CAT3550(config-if)# switchport voice vlan vvlan_id
CAT3550(config-if)# switchport access vlan dvlan_id
CAT3550(config-if)# mls qos trust device cisco-phone
CAT3550(config-if)# service-policy input IPPHONE-PC
CAT3550(config-if)# exit
CAT3550(config)#
CAT3550(config)#ip access list standard ACL_ANY
CAT3550(config-std-nacl)# permit any
CAT3550(config-std-nacl)# end
CAT3550#

```

Cisco 4500 with SUPIII, IV, or V

```

CAT4500(config)# qos map cos 5 to dscp 46
CAT4500(config)# qos map cos 0 24 to dscp 8
CAT4500(config)#
CAT4500(config)#class-map match-all VVLAN-VOICE
CAT4500(config-cmap)# match access-group name VVLAN-VOICE
CAT4500(config-cmap)#
CAT4500(config-cmap)#class-map match-all VVLAN-CALL-SIGNALING
CAT4500(config-cmap)# match access-group name VVLAN-CALL-SIGNALING
CAT4500(config-cmap)#
CAT4500(config-cmap)#class-map match-all VVLAN-ANY
CAT4500(config-cmap)# match access-group name VVLAN-ANY
CAT4500(config-cmap)#
CAT4500(config-cmap)# policy-map IPPHONE-PC
CAT4500(config-pmap)#class VVLAN-VOICE
CAT4500(config-pmap-c)# set ip dscp 46
CAT4500(config-pmap-c)# police 128 kps 8000 byte exceed-action drop
CAT4500(config-pmap-c)# class VVLAN-CALL-SIGNALING
CAT4500(config-pmap-c)# set ip dscp 24
CAT4500(config-pmap-c)# police 32 kps 8000 byte exceed-action policed-dscp-transmit
CAT4500(config-pmap-c)# class VVLAN-ANY
CAT4500(config-pmap-c)# set ip dscp 0
CAT4500(config-pmap-c)# police 32 kps 8000 byte exceed-action policed-dscp-transmit
CAT4500(config-pmap-c)# class class-default
CAT4500(config-pmap-c)# set ip dscp 0
CAT4500(config-pmap-c)# police 5 mpbs 8000 byte exceed-action policed-dscp-transmit
CAT4500(config-pmap-c)# exit
CAT4500(config-pmap)# exit
CAT4500(config)#
CAT4500(config)#
CAT4500(config)#interface interface-id
CAT4500(config-if)# switchport voice vlan vvlan_id
CAT4500(config-if)# switchport access vlan dvlan_id
CAT4500(config-if)# qos trust device cisco-phone
CAT4500(config-if)# service-policy input IPPHONE-PC
CAT4500(config-if)# exit
CAT4500(config)#
CAT4500(config)#ip access list extended VVLAN-VOICE
CAT4500(config-ext-nacl)# permit udp Voice_IP_Subnet Subnet_Mask any range 16384 32767
dscp ef
CAT4500(config-ext-nacl)# exit
CAT4500(config)#ip access list extended VVLAN-CALL-SIGNALING
CAT4500(config-ext-nacl)# permit tcp Voice_IP_Subnet Subnet_Mask any range 2000 2002 dscp
cs3
CAT4500(config-ext-nacl)# exit
CAT4500(config)#ip access list extended VVLAN-ANY
CAT4500(config-ext-nacl)# permit ip Voice_IP_Subnet Subnet_Mask any
CAT4500(config-ext-nacl)# end
CAT4500#

```

Cisco 6500

```

CAT6500> (enable) set qos cos-dscp-map 0 8 16 24 32 46 48 56
CAT6500> (enable) set qos policed-dscp-map 0, 24, 46:8
CAT6500> (enable)
CAT6500> (enable) set qos policer aggregate VVLAN-VOICE rate 128 burst 8000 drop
CAT6500> (enable) set qos policer aggregate VVLAN-CALL-SIGNALING rate 32 burst 8000
policed-dscp
CAT6500> (enable) set qos policer aggregate VVLAN-ANY rate 5000 burst 8000 policed-dscp
CAT6500> (enable) set qos policer aggregate PC-DATA rate 5000 burst 8000 policed-dscp
CAT6500> (enable)
CAT6500> (enable) set qos acl ip IPPHONE-PC dscp 46 aggregate VVLAN-VOICE udp
Voice_IP_Subnet Subnet_Mask any range 16384 32767
CAT6500> (enable) set qos acl ip IPPHONE-PC dscp 24 aggregate VVLAN-CALL-SIGNALING tcp
Voice_IP_Subnet Subnet_Mask any range 2000 2002
CAT6500> (enable) set qos acl ip IPPHONE-PC dscp 0 aggregate VVLAN-ANY Voice_IP_Subnet
Subnet_Mask any
CAT6500> (enable) set qos acl ip IPPHONE-PC dscp 0 aggregate PC-DATA any
CAT6500> (enable) commit qos acl IPPHONE-PC
CAT6500> (enable) set vlan vvlan_id mod/port
CAT6500> (enable) set port qos mod/port trust-device ciscoipphone
CAT6500> (enable) set qos acl map IPPHONE-PC mod/port
CAT6500> (enable)

```

Software-Based Endpoints

Even though Cisco IP SoftPhone and IP Communicator mark their signaling packets and media packets respectively, Cisco recommends that you re-mark the DSCP value of the packets from PC that is running the Cisco IP SoftPhone or IP Communicator because the PC is not a trusted device in the network. Instead of using the full range of User Datagram Protocol (UDP) ports (16384 to 32767) for the media packets, Cisco IP Softphone and Communicator can be configured to use a specific UDP port.

For Cisco IP SoftPhone, you can specify the UDP port and port range under **Network Audio Settings > Audio Output Port**. For Cisco IP Communicator, you can specify the UDP port by using one of the following options:

- Specify the **RTP Port Range Start** and **RTP Port Range End** in the product-specific section of the IP Communicator configuration page.
- Select **Preferences > Audio Settings > Network > Port Range** and specify the port range.

If you use both options to set the UDP port and port range, settings from the second option will override the first option.

The following sections list the QoS configuration commands for Cisco IP SoftPhone and IP Communicator on the most commonly deployed Cisco Catalyst switches.

Cisco 2950 with Enhanced Image

The Cisco Catalyst 2950 Series switches are not recommended for software-based endpoint QoS implementations due to two limitations:

- The Cisco 2950 does not support the **range** keyword to specify the UDP port range within an ACL configuration. The workaround to this limitation is to configure a single static UDP port for the Cisco IP SoftPhone to use, as described in preceding section.
- The Cisco 2950 supports only 1-Mbps increments on FastEthernet ports, which can create a fairly large hole to admit onto the network unauthorized traffic that might be mimicking call signaling or media.

Cisco 2970 or 3750

```

CAT2970 (config) #mls qos
CAT2970 (config) #mls qos map policed-dscp 0 24 46 to 8
CAT2970 (config) #
CAT2970 (config) #class-map match-all SOFTPHONE-VOICE
CAT2970 (config-cmap) # match access-group name SOFTPHONE-VOICE
CAT2970 (config-cmap) #class-map match-all SOFTPHONE-SIGNALING
CAT2970 (config-cmap) # match access-group name SOFTPHONE-SIGNALING
CAT2970 (config-cmap) #exit
CAT2970 (config) #
CAT2970 (config) #policy-map SOFTPHONE-PC
CAT2970 (config-pmap-c) #class SOFTPHONE-VOICE
CAT2970 (config-pmap-c) # set ip dscp 46
CAT2970 (config-pmap-c) # police 128000 8000 exceed-action policed-dscp-transmit
CAT2970 (config-pmap-c) #class SOFTPHONE-SIGNALING
CAT2970 (config-pmap-c) # set ip dscp 24
CAT2970 (config-pmap-c) # police 32000 8000 exceed-action policed-dscp-transmit
CAT2970 (config-pmap-c) #class class-default
CAT2970 (config-pmap-c) # set ip dscp 0
CAT2970 (config-pmap-c) # police 5000000 8000 exceed-action policed-dscp transmit
CAT2970 (config-pmap-c) # exit
CAT2970 (config-pmap) #exit
CAT2970 (config) #
CAT2970 (config) #interface FastEthernet interface-id
CAT2970 (config-if) # switchport access vlan vvlan_id
CAT2970 (config-if) # switchport mode access
CAT2970 (config-if) # service-policy input SOFTPHONE-PC
CAT2970 (config-if) # exit
CAT2970 (config) #ip access list extended SOFTPHONE-VOICE
CAT2970 (config-ext-nacl) # permit udp host PC_IP_address eq fixed_port_number any
CAT2970 (config-ext-nacl) # exit
CAT2970 (config) #ip access-list extended SOFTPHONE-SIGNALING
CAT2970 (config-ext-nacl) # permit tcp host PC_IP_address host CallManager_IP_address eq
2748 or 2000
CAT2970 (config-ext-nacl) # exit

```

Cisco 3550

```

CAT3550 (config) #mls qos
CAT3550 (config) #mls qos map policed-dscp 0 24 46 to 8
CAT3550 (config) #
CAT3550 (config) #class-map match-all SOFTPHONE-VOICE
CAT3550 (config-cmap) # match access-group name SOFTPHONE-VOICE
CAT3550 (config-cmap) #class-map match-all SOFTPHONE-SIGNALING
CAT3550 (config-cmap) # match access-group name SOFTPHONE-SIGNALING
CAT3550 (config-cmap) #exit
CAT3550 (config) #
CAT3550 (config) #policy-map SOFTPHONE-PC
CAT3550 (config-pmap-c) #class SOFTPHONE-VOICE
CAT3550 (config-pmap-c) # set ip dscp 46
CAT3550 (config-pmap-c) # police 128000 8000 exceed-action policed-dscp-transmit
CAT3550 (config-pmap-c) #class SOFTPHONE-SIGNALING
CAT3550 (config-pmap-c) # set ip dscp 24
CAT3550 (config-pmap-c) # police 32000 8000 exceed-action policed-dscp-transmit
CAT3550 (config-pmap-c) #class class-default
CAT3550 (config-pmap-c) # set ip dscp 0
CAT3550 (config-pmap-c) # police 5000000 8000 exceed-action policed-dscp transmit
CAT3550 (config-pmap-c) # exit
CAT3550 (config-pmap) #exit
CAT3550 (config) #
CAT3550 (config) #interface FastEthernet interface-id
CAT3550 (config-if) # switchport access vlan vvlan_id

```

```

CAT3550 (config-if) # switchport mode access
CAT3550 (config-if) # service-policy input SOFTPHONE-PC
CAT3550 (config-if) # exit
CAT3550 (config) #ip access list extended SOFTPHONE-VOICE
CAT3550 (config-ext-nacl) # permit udp host PC_IP_address eq fixed_port_number any
CAT3550 (config-ext-nacl) # exit
CAT3550 (config) #ip access-list extended SOFTPHONE-SIGNALING
CAT3550 (config-ext-nacl) # permit tcp host PC_IP_address host CallManager_IP_address eq
2748 or 2000
CAT3550 (config-ext-nacl) # exit

```

Cisco 4500 with SUP3, IV, or V

```

CAT4500 (config) #qos
CAT4500 (config) #qos map policed-dscp 0 24 46 to 8
CAT4500 (config) #
CAT4500 (config) #class-map match-all SOFTPHONE-VOICE
CAT4500 (config-cmap) # match access-group name SOFTPHONE-VOICE
CAT4500 (config-cmap) #class-map match-all SOFTPHONE-SIGNALING
CAT4500 (config-cmap) # match access-group name SOFTPHONE-SIGNALING
CAT4500 (config-cmap) #exit
CAT4500 (config) #
CAT4500 (config) #policy-map SOFTPHONE-PC
CAT4500 (config-pmap-c) #class SOFTPHONE-VOICE
CAT4500 (config-pmap-c) # set ip dscp EF
CAT4500 (config-pmap-c) # police 128 kps 8000 byte exceed-action policed-dscp-transmit
CAT4500 (config-pmap-c) #class SOFTPHONE-SIGNALING
CAT4500 (config-pmap-c) # set ip dscp CS3
CAT4500 (config-pmap-c) # police 32000 kps 8000 byte exceed-action policed-dscp-transmit
CAT4500 (config-pmap-c) #class class-default
CAT4500 (config-pmap-c) # set ip dscp default
CAT4500 (config-pmap-c) # police 5 mpbs 8000 byte exceed-action policed-dscp transmit
CAT4500 (config-pmap-c) # exit
CAT4500 (config-pmap) #exit
CAT4500 (config) #
CAT4500 (config) #interface FastEthernet interface-id
CAT4500 (config-if) # switchport access vlan vvlan_id
CAT4500 (config-if) # switchport mode access
CAT4500 (config-if) # service-policy input SOFTPHONE-PC
CAT4500 (config-if) # exit
CAT4500 (config) #ip access list extended SOFTPHONE-VOICE
CAT4500 (config-ext-nacl) # permit udp host PC_IP_address eq fixed_port_number any
CAT4500 (config-ext-nacl) # exit
CAT4500 (config) #ip access-list extended SOFTPHONE-SIGNALING
CAT4500 (config-ext-nacl) # permit tcp host PC_IP_address host CallManager_IP_address eq
2748 or 2000
CAT4500 (config-ext-nacl) # exit

```

Cisco 6500

```

CAT6500> (enable) set qos enable
CAT6500> (enable) set qos policed-dscp-map 0, 24, 46:8
CAT6500> (enable)
CAT6500> (enable) set qos policer aggregate SOFTPHONE-VOICE rate 128 burst 8000
policed-dscp
CAT6500> (enable) set qos policer aggregate SOFTPHONE-SIGNALING rate 32 burst 8000
policed-dscp
CAT6500> (enable) set qos policer aggregate PC-DATA rate 5000 burst 8000 policed-dscp
CAT6500> (enable)
CAT6500> (enable) set qos acl ip SOFTPHONE-PC dscp 46 aggregate SOFTPHONE-VOICE udp host
PC_IP_address eq fixed_port_number any

```



```

CAT6500> (enable) set qos acl ip SOFTPHONE-PC dscp 24 aggregate SOFTPHONE-SIGNALING tcp
host PC_IP_address host CallManager_IP_address eq 2748 or 2000
CAT6500> (enable) set qos acl ip SOFTPHONE-PC dscp 0 aggregate PC-DATA any
CAT6500> (enable) commit qos acl SOFTPHONE-PC
CAT6500> (enable) set vlan vvlan_id mod/port
CAT6500> (enable) set port qos mod/port trust untrusted
CAT6500> (enable) set qos acl map SOFTPHONE-PC mod/port
CAT6500> (enable)

```

**Note**

The DSCP re-marking must be done by a Layer 3 capable switch. If the access layer switch (such as the Cisco Catalyst 2950 with Standard Image or the Cisco 3524XL) does not have this capability, then the DSCP re-marking must be done at the distribution layer switch.

Cisco Wireless IP Phone 7920G

By default, the Cisco Wireless IP Phone 7920 marks its SCCP signaling messages using a Per-Hop Behavior (PHB) value of AF31 or a Differentiated Services Code Point (DSCP) value of 26 (this corresponds to a ToS value of 0x68), and it marks RTP voice packets using a PHB value of EF or a DSCP value of 46 (ToS of 0xB8). With proper queueing on the AP and configuration on the upstream first-hop switch to trust the AP's port, the wireless IP phone traffic will receive the same treatment as wired IP phone traffic. This practice allows the QoS settings to be consistent from LAN to WLAN environments.

In addition, the Cisco Wireless IP Phone 7920 will automatically announce its presence to the AP using the Cisco Discovery Protocol (CDP). The CDP packets are sent from the wireless IP phone to the AP, and they identify the phone so that the AP can place all traffic to the phone in the high-priority queue.

While Ethernet switch ports can typically transmit and receive at 100 Mbps, 802.11b APs have a lower throughput rate that allows for a maximum data rate of 11 Mbps. Furthermore, wireless LANs are a shared medium and, due to contention for this medium, the actual throughput is substantially lower. This throughput mismatch means that, with a burst of traffic, the AP will drop packets, thus adding excessive processor burden to the unit and affecting performance.

By taking advantage of the policing and rate limiting capabilities of the Cisco Catalyst 3550 and 6500 Series switches, you can eliminate the need for the AP to drop excessive packets by configuring the upstream switch port to rate-limit or police traffic going to the AP. The switch port configurations in the following sections rate-limit the port(s) to a practical throughput of 7 Mbps for 802.11b and guarantee 1 Mbps for high-priority voice and control traffic. Furthermore, as indicated in the configuration examples, packets coming from the AP should be trusted and, based on the VLAN tag of each packet, the DSCP marking should either be maintained or marked down. Thus, packets sourced from the Cisco 7920 Wireless IP Phone on the voice VLAN should maintain the appropriate DSCP marking, and packets source from data devices on the data VLAN should be remarked to a DSCP value of 0.

Cisco 3550

```

CAT3550(config)#mls qos
CAT3550(config)#mls qos map cos-dscp 0 8 16 24 32 46 48 56
CAT3550(config)#mls qos map policed-dscp 24 26 46 to 8
CAT3550(config)#mls qos aggregate-policer AGG-POL-1M-VOICE-OUT 1000000 8000 exceed-action
policed-dscp-transmit
CAT3550(config)#mls qos aggregate-policer AGG-POL-6M-DEFAULT-OUT 6000000 8000
exceed-action drop
CAT3550(config)#
CAT3550(config)#class-map match-all EGRESS-DSCP-0
CAT3550(config-cmap)#match ip dscp 0
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all EGRESS-DSCP-8

```

```

CAT3550(config-cmap)#match ip dscp 8
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all EGRESS-DSCP-16
CAT3550(config-cmap)#match ip dscp 16
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all EGRESS-DSCP-32
CAT3550(config-cmap)#match ip dscp 32
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all EGRESS-DSCP-48
CAT3550(config-cmap)#match ip dscp 48
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all EGRESS-DSCP-56
CAT3550(config-cmap)#match ip dscp 56
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-any VOICE-SIGNALING
CAT3550(config-cmap)#match ip dscp 24
CAT3550(config-cmap)#match ip dscp 26
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all VOICE
CAT3550(config-cmap)#match ip dscp 46
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all INGRESS-DATA
CAT3550(config-cmap)#match any
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all INGRESS-VVLAN-VOICE
CAT3550(config-cmap)#match vlan vvlan-id
CAT3550(config-cmap)#match class-map VOICE
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all INGRESS-VVLAN-VOICE-SIGNALING
CAT3550(config-cmap)#match vlan vvlan-id
CAT3550(config-cmap)#match class-map VOICE-SIGNALING
CAT3550(config-cmap)#
CAT3550(config-cmap)#class-map match-all INGRESS-DVLAN
CAT3550(config-cmap)#match vlan dvlan-id
CAT3550(config-cmap)#match class-map INGRESS-DATA
CAT3550(config-cmap)#
CAT3550(config-cmap)#policy-map EGRESS-RATE-LIMITER
CAT3550(config-pmap)#class EGRESS-DSCP-0
CAT3550(config-pmap-c)#police aggregate AGG-POL-6M-DEFAULT-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class EGRESS-DSCP-8
CAT3550(config-pmap-c)#police aggregate AGG-POL-6M-DEFAULT-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class EGRESS-DSCP-16
CAT3550(config-pmap-c)#police aggregate AGG-POL-6M-DEFAULT-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class EGRESS-DSCP-32
CAT3550(config-pmap-c)#police aggregate AGG-POL-6M-DEFAULT-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class EGRESS-DSCP-48
CAT3550(config-pmap-c)#police aggregate AGG-POL-6M-DEFAULT-OUT
CAT3550(config-pmap-c)#class EGRESS-DSCP-56
CAT3550(config-pmap-c)#police aggregate AGG-POL-6M-DEFAULT-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class EGRESS-VOICE
CAT3550(config-pmap-c)#police aggregate AGG-POL-1M-VOICE-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class EGRESS-VOICE-SIGNALING
CAT3550(config-pmap-c)#police aggregate AGG-POL-1M-VOICE-OUT
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#policy-map INGRESS-QOS
CAT3550(config-pmap)#class INGRESS-VVLAN-VOICE
CAT3550(config-pmap-c)#set ip dscp 46
CAT3550(config-pmap-c)#

```

```

CAT3550(config-pmap-c)#class INGRESS-VVLAN-CALL-SIGNALING
CAT3550(config-pmap-c)#set ip dscp 24
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class INGRESS-DVLAN
CAT3550(config-pmap-c)#set ip dscp 0
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class class-default
CAT3550(config-pmap-c)#set ip dscp 0
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#interface interface id
CAT3550(config-if)#description 11Mb towards Wireless Access Point
CAT3550(config-if)#switchport access dvlan-id
CAT3550(config-if)#switchport voice vvlan-id
CAT3550(config-if)#mls qos trust dscp
CAT3550(config-if)#service-policy output EGRESS-RATE-LIMITER
CAT3550(config-if)#service-policy input INGRESS-QOS

```

Cisco 6500

```

CAT6500> (enable) set qos enable
CAT6500> (enable) set qos cos-dscp-map 0 8 16 24 32 46 48 56
CAT6500> (enable) set qos policed-dscp-map 24,26,46:0
CAT6500> (enable)
CAT6500> (enable) set qos policer microflow VOICE-OUT rate 1000 burst 32 policed-dscp
CAT6500> (enable) set qos policer microflow DATA-OUT rate 6000 burst 32 drop
CAT6500> (enable)
CAT6500> (enable) set qos acl ip AP-VOICE-EGRESS dscp 24 microflow VOICE-OUT ip any any
dscp-field 24
CAT6500> (enable) set qos acl ip AP-VOICE-EGRESS dscp 24 microflow VOICE-OUT ip any any
dscp-field 26
CAT6500> (enable) set qos acl ip AP-VOICE-EGRESS dscp 46 microflow VOICE-OUT ip any any
dscp-field 46
CAT6500> (enable) set qos acl ip AP-DATA-EGRESS dscp 0 microflow DATA-OUT ip any any
CAT6500> (enable)
CAT6500> (enable) set qos acl ip AP-VOICE-INGRESS dscp 24 ip any any dscp-field 26
CAT6500> (enable) set qos acl ip AP-VOICE-INGRESS trust-dscp ip any any
CAT6500> (enable) set qos acl ip AP-DATA-INGRESS dscp 0 ip any any
CAT6500> (enable)
CAT6500> (enable) set qos acl map AP-VOICE-EGRESS vvlan-id output
CAT6500> (enable) set qos acl map AP-DATA-EGRESS dvlan-id output
CAT6500> (enable) set qos acl map AP-VOICE-INGRESS vvlan-id input
CAT6500> (enable) set qos acl map AP-DATA-INGRESS dvlan-id input
CAT6500> (enable)
CAT6500> (enable) set port qos mod/port vlan-based
CAT6500> (enable)
CAT6500> (enable) set port qos mod/port trust trust-dscp
CAT6500> (enable)

```

Endpoint Features Summary

Table 16-3 summarizes the Cisco IP Telephony features for Cisco analog gateways, Table 16-4 summarizes the features for Cisco IP Phones, and Table 16-5 summarizes the features for Cisco IP software devices.

Table 16-3 Analog Gateway Features

Feature	Analog NMs	Ws-svc-cmm-24fxs	Ws-x6624-fxs	VG224	VG248	ATA 186 and 188
Ethernet Connection	N	N	N	Y ¹	Y ²	Y ³
Maximum number of Analog Ports	16	72	24	24	48	2
Caller ID	Y	N	N	Y	Y	N
Call Waiting	N	N	N	N	Y	Y
Caller ID on Call Waiting	N	N	N	N	Y	N
Call Hold	N	N	N	Y ⁴	N	Y
Call Transfer	N	N	N	Y ⁴	Y	Y
Call Forward	N	N	N	N	Y ⁵	Y
Auto-Answer	N	N	N	N	N	N
Ad Hoc Conference	N	N	N	N	Y	Y
Meet-Me Conference	N	N	N	N	N	Y
Call Pickup	N	N	N	N	N	Y
Group Pickup	N	N	N	N	N	Y
Redial	N	N	N	N	Y ⁶	Y ⁶
Speed Dial	N	N	N	N	Y	Y
On-hook Dialing	N	N	N	N	N	N
Voice Mail Access	Y	Y	Y	Y	Y	Y ⁷
Message Waiting Indicator (MWI)	N	N	N	N	Y	Y ⁷
Survivable Remote Site Telephony (SRST) Support	N	N	N	Y	Y	Y
Music on Hold (MoH)	Y	Y	Y	N	Y	Y ⁸
Mute	N	N	N	N	N	N
Multilevel Precedence and Preemption (MLPP)	N	N	N	N	N	N
Barge	N	N	N	N	N	N
cBarge	N	N	N	N	N	N
Call Preservation	N	N	N	N	Y ⁹	N
Call Admission Control	Y	N	N	N	N	N
Local Voice Busy-Out	Y	N	N	N	N	N
Private Line Automatic Ringdown (PLAR)	Y	N	N	N	N	N

Table 16-3 Analog Gateway Features (continued)

Feature	Analog NMs	Ws-svc-cmm-24fxs	Ws-x6624-fxs	VG224	VG248	ATA 186 and 188
Hunt Group	Y	N	N	N	N	N
Dial Plan Mapping	Y	N	N	N	N	N
Supervisory Disconnect	Y	N	N	N	N	N
Signaling Packet ToS Value Marking	0x68	0x68 ¹⁰	0x68	0x68	0x68	0x68
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0xB8	0xB8	0xA0
Fax Pass-Through	Y ¹¹	Y	Y ¹²	Y	Y ¹¹	Y
Fax Relay	Y	Y	N	Y	Y	N
Skinny Client Control Protocol (SCCP)	N	N	N	N	Y	Y
Session Initiation Protocol (SIP)	N	N	N	Y	N	Y
H.323	Y	Y	N	Y	N	Y
Media Gateway Control Protocol (MGCP)	Y	Y	Y	Y	N	Y ¹³
G.711	Y	Y	Y	Y	Y	Y
G.723	Y	N	N	N	N	Y
G.726	Y	N	N	N	N	N
G.729	Y	Y	Y	Y	Y	Y
Voice Activity Detection (VAD)	Y	Y	N	Y	N	Y
Comfort Noise Generation (CNG)	Y	Y	N	Y	N	Y

- Two 10/100 Base-T
- One 10/100 Base-T
- Two 10/100 Base-T for ATA 188; one 10 Base-T for ATA 186
- H.323 and SIP call control
- Call Forward All
- Last Number Redial
- Only on SCCP and SIP version
- Supports only unicast MoH
- Supported on VG248 version 1.2 or later
- It marks MGCP signaling on UDP port 2427, but it marks the MGCP keepalive packets as best-effort on TCP port 2428.
- Fax pass-through and fax relay
- Fax pass-through
- Cisco CallManager does not support MGCP with the ATA.

Table 16-4 IP Phone Features

Feature	7902G	7905G	7910G	7910 +SW	7912G	7920G	7935G, 7936G	7940G	7960G	7970G
Ethernet Connection	Y ¹	Y ¹	Y ¹	Y ²	Y ²	N	Y ³	Y ²	Y ²	Y ²
Ethernet Switch	N	N	N	Y	Y	N	N	Y	Y	Y
Cisco Power-Over-Ethernet (PoE)	Y	Y	Y	Y	Y	N	N	Y	Y	Y
IEEE 802.3af Power-Over-Ethernet (PoE)	N	N	N	N	N	N	N	N	N	Y
Localization	N	Y	N	N	Y	N	N	Y	Y	Y
Directory Number	1	1	1	1	1	6	1	2	6	8
Liquid Crystal Display	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Caller ID	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Call Waiting	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Caller ID on Call Waiting	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Call Hold	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Call Transfer	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Call Forward	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Auto-Answer	Y	Y	N	N	Y	Y	N	Y	Y	Y
Ad Hoc Conference	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Meet-Me Conference	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Call Pickup	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Group Pickup	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Redial	⁴ Y	Y ⁴	Y ⁴	Y ⁴	Y ⁴	Y	Y	Y	Y	Y
Speed Dial	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
On-hook Dialing	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Voice Mail Access	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Message Waiting Indicator (MWI)	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Video call	N	N	N	N	N	N	N	Y	Y	Y
Survivable Remote Site Telephony (SRST) Support	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Music on Hold (MoH)	Y	Y	Y	Y	Y	Y ⁵	Y	Y	Y	Y
Speaker	N	Y ⁶	Y ⁶	Y ⁶	Y ⁶	N	Y	Y	Y	Y
Headset Jack	N	N	N	N	N	Y ⁷	N	Y	Y	Y
Mute	N	N	Y	Y	N	Y	Y	Y	Y	Y
Multilevel Precedence and Preemption (MLPP)	Y	Y	Y	Y	Y	N	N	Y	Y	Y
Barge	N	N	N	N	N	N	N	Y	Y	Y
cBarge	N	Y	N	N	Y	N	N	Y	Y	N

Table 16-4 IP Phone Features (continued)

Feature	7902G	7905G	7910G	7910 +SW	7912G	7920G	7935G, 7936G	7940G	7960G	7970G
Disable General Attribute Registration Protocol (GARP)	Y	Y	Y	Y	Y	N	N	Y	Y	Y
Signaling and Media Encryption	N	N	N	N	N	Y ⁸	N	N	N	Y
Signaling Integrity	N	N	N	N	N	N	N	Y	Y	Y
Manufacturing-Installed Certificate (X.509v3)	N	N	N	N	N	N	N	N	N	Y
Field-Installed Certificate	N	N	N	N	N	N	N	Y	Y	N
Third-Party XML Service	N	Y	N	N	Y	N	N	Y	Y	Y
External Microphone and Speaker	N	N	N	N	N	N	N	N	N	Y
Hunt Group	N	N	N	N	N	N	N	N	N	N
Dial Plan Mapping	N	N	N	N	N	N	N	N	N	N
Signaling Packet ToS Value Marking	0x68	0x68	0x68	0x68	0x68	0x68	0x68	0x60	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8
Skinny Client Control Protocol (SCCP)	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Session Initiation Protocol (SIP)	N	Y	N	N	Y	N	N	Y	Y	N
H.323	N	Y	N	N	N	N	Y	N	N	N
Media Gateway Control Protocol (MGCP)	N	N	N	N	N	N	N	Y	Y	N
G.711	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
G.723	N	Y	N	N	N	N	N	N	N	N
G.726	N	Y	N	N	N	N	N	N	N	N
G.729	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Voice Activity Detection (VAD)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

1. One 10 Base-T
2. Two 10/100 Base-T
3. One 10/100 Base-T
4. Last Number Redial
5. Supports only unicast MoH
6. One-way listen mode
7. The only supported headset for the Cisco IP Phone 7920 is one with a 2.5 mm jack, available at <http://www.cisco.getheadsets.com>.
8. Signaling and Media Encryption are available with Static WEP and LEAP security configurations.

Table 16-5 Software Device Features

Feature	IP Communicator	IP SoftPhone
Directory Number	8	6 ¹
Caller ID	Y	Y
Call Waiting	Y	Y
Caller ID on Call Waiting	Y	Y
Call Hold	Y	Y
Call Transfer	Y	Y
Call Forward	Y	Y
Auto-Answer	Y	Y
Ad Hoc Conference	Y	Y
Meet-Me Conference	Y	N
Call Pickup	Y	N
Group Pickup	Y	N
Redial	Y	Y
Speed Dial	Y	N
On-hook Dialing	Y	Y
Voice Mail Access	Y	Y
Message Waiting Indicator (MWI)	Y	Y
Video call	N	N
Survivable Remote Site Telephony (SRST) Support	Y	N
Music on Hold (MoH)	Y	Y
Speaker	Y	Y
Mute	Y	Y
Multilevel Precedence and Preemption (MLPP)	Y	Y
Barge	Y	N
cBarge	N	N
Disable General Attribute Registration Protocol (GARP)	Y	N
Signaling and Media Encryption	N	N
Signaling Integrity	N	N
Manufacturing-Installed Certificate (X.509v3)	N	N
Field-Installed Certificate	N	N
Third-Party XML Service	Y	N
Signaling Packet ToS Value Marking	0x60	0x68
Media Packet ToS Value Marking	0xB8	0xB8

Table 16-5 Software Device Features (continued)

Feature	IP Communicator	IP SoftPhone
Skinny Client Control Protocol (SCCP)	Y	N
Session Initiation Protocol (SIP)	N	N
H.323	N	Y
Media Gateway Control Protocol (MGCP)	N	N
Telephony Application Programming Interface (TAPI)	N	Y
G.711	Y	Y
G.723	N	Y
G.726	N	N
G.729	Y	Y
Voice Activity Detection (VAD)	Y	N
Comfort Noise Generation (CNG)	Y	N

1. As a standalone CTI device



Recommended Hardware and Software Combinations

For the most recent information on recommended hardware platforms, software releases, and firmware versions for Cisco IP Communications solutions based on Cisco CallManager Release 4.1, frequently refer to the latest Release Notes and other documentation at the following location:

http://www.cisco.com/univercd/cc/td/doc/product/voice/ip_tele/gblink/system/gbtst4x/index.htm



Note

The platforms and software versions recommended in the Cisco IP Communications Systems Release Notes are not the only supportable deployment options. They represent the combinations of hardware and software that Cisco has 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 Cisco IP Telephony solutions, contact your Cisco account representative.



GLOSSARY

A

AA	Automated attendant
AAD	Alerts and Activities Display
AAR	Automated Alternate Routing
ACD	Automatic call distribution
ACF	Admission Confirm
ACL	access control list
ACS	Access Control Server
AD	Microsoft Active Directory
ADUC	Active Directory Users and Computers
AFT	ALI Formatting Tool
AGM	Cisco Access Gateway Module
ALG	Application Layer Gateway
ALI	Automatic Location Identification
AMI	Alternate mark inversion
AMIS	Audio Messaging Interchange Specification
ANI	Automatic Number Identification
AP	Access point
API	Application Programming Interface
ARJ	Admission Reject
ARP	Address Resolution Protocol
ARQ	Admission Request
ASP	Active server page
ASR	Automatic speech recognition

ATA	Cisco Analog Telephone Adapter
ATM	Asynchronous Transfer Mode
AVVID	Cisco Architecture for Voice, Video, and Integrated Data

B

BAT	Cisco Bulk Administration Tool
BBWC	Battery-backed write cache
BHCA	Busy hour call attempts
BHCC	Busy hour call completions
BPDU	Bridge protocol data unit
bps	Bits per second
BRI	Basic Rate Interface
BTN	Bill-to number

C

CA	Certificate Authority
CAC	Call admission control
CAMA	Centralized Automatic Message Accounting
CAPF	Certificate Authority Proxy Function
CAR	Cisco CDR Analysis and Reporting
CAS	Channel Associated Signaling
CBWFQ	Class-Based Weighted Fair Queuing
CCS	Common channel signaling
CDP	Cisco Discovery Protocol
CDR	Call detail record
CIR	Committed information rate
CKM	Cisco Centralized Key Management
CLEC	Competitive local exchange carrier

CLID	Calling line identifier
CMC	Client Matter Code
CME	Cisco CallManager Express
CMI	Cisco Messaging Interface
CMM	Cisco Communication Media Module
CNG	Comfort noise generation
CO	Central office
COM	Component Object Model
COR	Class of restriction
CoS	Class of service
CPCA	Cisco Unity Personal Assistant
CPI	Cisco Product Identification tool
CPN	Calling party number
CRS	Cisco Customer Response Solution
cRTP	Compressed Real-Time Transport Protocol
CSUF	Cross-Stack UplinkFast
CTI	Computer telephony integration
CUE	Cisco Unity Express

D

DC	Domain controller
DDNS	Dynamic Domain Name Server
DHCP	Dynamic Host Configuration Protocol
DID	Direct inward dial
DIT	Directory Information Tree
DN	Directory number
DNIS	Dialed number identification service
DNS	Domain Name System

DoS	Denial of service
DPA	Digital PBX Adapter
DSCP	Differentiated Services Code Point
DSE	Digital set emulation
DSP	Digital signal processor
DTMF	Dual tone multifrequency
DTPC	Dynamic Transmit Power Control
DUC	Domino Unified Communications Services

E

E&M	Receive and transmit, or ear and mouth
EAP	Extensible Authentication Protocol
EC	Echo cancellation
ECM	Error correction mode
ECS	Empty Capabilities Set
EI	Enhanced Image
EIGRP	Enhanced Interior Gateway Routing Protocol
ELIN	Emergency location identification number
EM	Extension Mobility
ER	Cisco Emergency Responder
ERL	Emergency response location
ESF	Extended Super Frame

F

FAC	Forced Account Code
FCC	Federal Communications Commission
FIFO	First-in, first-out
FR	Frame Relay

FXO	Foreign Exchange Office
FXS	Foreign Exchange Station

G

GARP	General Attribute Registration Protocol
GC	Global catalog
GKTMP	Gatekeeper Transaction Message Protocol
GMS	Greeting management system
GPO	Group Policy Object
GUI	Graphical user interface
GUP	Gatekeeper Update Protocol

H

H.225D	H.225 daemon
HP	Hewlett-Packard
HSRP	Hot Standby Router Protocol
HTTP	Hyper-Text Transfer Protocol
Hz	Hertz

I

IANA	Internet Assigned Numbers Authority
IAPP	Inter-Access Point Protocol
ICCS	Intra-Cluster Communication Signaling
ICS	IBM Cabling System
ICT	Intercluster trunk
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IIS	Microsoft Internet Information Server

IntServ	Integrated Services
IntServ/DiffServ	Integrated Services/Differentiated Services
IP	Internet Protocol
IPCC	Cisco IP Contact Center
IPIPGW	IP-to-IP gateway
IPMA	Cisco IP Manager Assistant
IPSec	IP Security
ISO	International Standards Organization
ITEM	CiscoWorks IP Telephony Environment Monitor
ITU	International Telecommunication Union
IVR	Interactive voice response

J

JTAPI	Java Telephony Application Programming Interface

K

kbps	Kilobits per second

L

LAN	Local area network
LBR	Low bit-rate
LCD	Liquid crystal display
LCF	Location Confirm
LDAP	Lightweight Directory Access Protocol
LDIF	LDAP Data Interchange Format
LDN	Listed directory number
LEAP	Lightweight Extensible Authentication Protocol
LEC	Local Exchange Carrier

LFI	Link fragmentation and interleaving
LLQ	Low-latency queuing
LRJ	Location Reject
LRQ	Location Request
LSC	Label Switch Controller

M

MAC	Media Access Control
MAN	Metropolitan area network
Mbps	Megabits per second
MCS	Cisco Media Convergence Server
MCU	Multipoint Control Unit
MFT	Multiflex trunk
MGCP	Media Gateway Control Protocol
MIC	Manufacturing installed certificate
MIPS	Millions of instructions per second
MISTP	Multiple Instance Spanning Tree Protocol
MLA	Cisco Multi-Level Administration
MLP	Multilink Point-to-Point Protocol
MLPP	Multilevel Precedence and Preemption
MLTS	Multi-line telephone system
MoH	Music on hold
MPLS	Multiprotocol Label Switching
MRG	Media resource group
MRGL	Media resource group list
ms	Millisecond
MTP	Media termination point

mW	Milli-Watt
MWI	Message Waiting Indicator

N

NAT	Network Address Translation
NENA	National Emergency Number Association
NFAS	Non-Facility Associated Signaling
NIC	Network interface card
NPA	Numbering Plan Area
NSE	Network Services Engine
NSF	Network Specific Facilities
NTP	Network Time Protocol

O

OSPF	Open Shortest Path First
OU	Organizational unit

P

PBX	Private branch exchange
PC	Personal computer
PCI	Peripheral Component Interconnect
PCM	Pulse code modulation
PD	Powered device
PHB	Per-hop behavior
PINX	Private integrated services network exchange
PLAR	Private Line Automatic Ringdown
PoE	Power over Ethernet
POTS	Plain old telephone service

pps	Packets per second
PQ	Priority Queue
PRI	Primary Rate Interface
PSAP	Public safety answering point
PSE	Power source equipment
PSTN	Public switched telephone network
PVC	Permanent virtual circuit

Q

QBE	Quick Buffer Encoding
QBSS	QoS Basis Service Set
QoS	Quality of Service
QSIG	Q signaling

R

RADIUS	Remote Authentication Dial-In User Service
RAS	Registration Admission Status
RDNIS	Redirected Dialed Number Information Service
RF	Radio frequency
RFC	Request for Comments
RSSI	Relative Signal Strength Indicator
RSTP	Rapid Spanning Tree Protocol
RSVP	Resource Reservation Protocol
RTMT	Cisco Real-Time Monitoring Tool
RTP	Real-Time Transport Protocol
RTT	Round-trip time

S

S1, S2, S3, and S4	Severity levels for service requests
SCCP	Skinnny Client Control Protocol
SCSI	Small Computer System Interface
SDK	Software Development Kit
SDL	Signal Distribution Layer
SDP	Session Definition Protocol
SE	Cisco Systems Engineer
SF	Super Frame
SI	Standard Image
SIP	Session Initiation Protocol
SIW	Service Inter-Working
SLB	Server load balancing
SMDI	Simplified Message Desk Interface
SNMP	Simple Network Management Protocol
SQL	Structured Query Language
SRND	Solution Reference Network Design
SRST	Survivable Remote Site Telephony
S RTP	Secure Real-Time Transport Protocol
SS7	Signaling System 7
SSID	Service set identifier
SSL	Secure Socket Layer
STP	Spanning Tree Protocol
SUP1	Cisco Supervisor Engine 1
SUP2	Cisco Supervisor Engine 2
SUP2+	Cisco Supervisor Engine 2+
SUP3	Cisco Supervisor Engine 3

T

TAC	Cisco Technical Assistance Center
TAPI	Telephony Application Programming Interface
TCD	Telephony Call Dispatcher
TCER	Total Character Error Rate
TCL	Tool Command Language
TCP	Transmission Control Protocol
TCS	Terminal Capabilities Set
TDD	Telephone Device for the Deaf
TDM	Time-division multiplexing
TEHO	Tail-end hop-off
TFTP	Trivial File Transfer Protocol
TLS	Transport layer security
ToD	Time of day
ToS	Type of service
TRaP	Telephone record and playback
TSP	Telephony Service Provider
TTL	Time to live
TTS	Text-to-speech
TTY	Terminal teletype
TUI	Telephony user interface

U

UAC	User agent client
UAS	User agent server
UDC	Universal data connector
UDLD	UniDirectional Link Detection
UDP	User Datagram Protocol

UNC	Universal Naming Convention
UPS	Uninterrupted power supply
URI	Uniform resource identifier
UTIM	Cisco Unity Telephone Integration Manager
UTP	Unshielded twisted pair
UUIE	User-to-User Information Element

V

V3PN	Cisco Voice and Video Enabled Virtual Private Network
VAD	Voice activity detection
VAF	Voice-Adaptive Fragmentation
VATS	Voice-Adaptive Traffic Shaping
VIC	Voice interface card
VLAN	Virtual local area network
VMO	ViewMail for Outlook
VoIP	Voice over IP
VoPSTN	Voice over the PSTN
VPN	Virtual private network
VWIC	Voice/WAN interface card

W

WAN	Wide area network
WEP	Wired Equivalent Privacy
WFQ	Weighted fair queuing
WLAN	Wireless local area network
WLSM	Cisco Wireless LAN Services Module

X

XML Extensible Markup Language



Symbols

- ! in route patterns [10-11](#)
- @ in route patterns [10-10](#)

Numerics

- 1A and 2A cabling [3-20](#)
- 4ESS [4-13, 4-14](#)
- 508 conformance [2-25](#)
- 5ESS [4-13, 4-14](#)
- 7902G IP Phone [16-5](#)
- 7905G IP Phone [16-5](#)
- 7910G+SW IP Phone [16-5](#)
- 7910G IP Phone [16-5](#)
- 7912G IP Phone [16-6](#)
- 7920G Wireless IP Phone [16-14, 16-27](#)
- 7935 IP Conference Station [16-19](#)
- 7936 IP Conference Station [16-19](#)
- 7940G IP Phone [16-6](#)
- 7960G IP Phone [16-6](#)
- 7970G IP Phone [16-9](#)
- 802.1s [3-4](#)
- 802.1w [3-4, 3-6](#)
- 802.3af PoE [3-19](#)
- 9.@ route pattern [10-10](#)
- 911 calls [11-1](#)

A

- AA [12-1](#)
- AAD [1-8](#)
- AAR [2-9, 2-11, 10-19, 10-79, 13-5, 16-18](#)

- abbreviated dialing [10-3](#)
- abbreviations [GL-1](#)
- access codes [10-6, 10-20](#)
- Access Control Server (ACS) [3-45, 3-46](#)
- accessibility of IP Telephony features [2-25](#)
- Access Layer [3-4](#)
- access point (AP) [3-40, 3-44, 16-14](#)
- ACF [10-32](#)
- acronyms [GL-1](#)
- ACS [3-45, 3-46](#)
- Active Directory (AD) [3-46, 14-2, 14-9](#)
- Active Server Page (ASP) [14-3](#)
- AD [3-46, 14-2, 14-9](#)
- additional information [xv, xviii](#)
- address resolution [10-32, 10-33](#)
- Address Resolution Protocol (ARP) [3-44](#)
- Admission Confirm (ACF) [10-32](#)
- Admission Reject (ARJ) [10-32](#)
- Admission Request (ARQ) [10-32](#)
- advanced formulas for bandwidth calculations [3-33](#)
- agents for call processing [1-4, 2-15](#)
- Alerts and Activities Display (AAD) [1-8](#)
- ALI [11-4](#)
- allocating media resources [6-12](#)
- all trunks busy [11-12](#)
- alternate
 - endpoints [5-11](#)
 - gatekeeper [5-11, 8-19](#)
 - TFTP file locations [3-15](#)
- analog
 - gateways [4-1, 4-11, 4-20, 16-1](#)
 - network modules [16-1, 16-3](#)
- Analog Telephone Adaptor (ATA) [16-4, 16-20](#)

ANI [4-10, 11-4, 11-5, 11-6, 11-7](#)
 Annex M1 [5-10](#)
 annunciator [6-15](#)
 answer supervision [11-12](#)
 AP [3-40, 3-44, 16-14](#)
 applications [1-2, 1-6](#)
 architecture [1-2](#)
 area code [10-20](#)
 ARJ [10-32](#)
 ARP [3-44](#)
 ARQ [10-32](#)
 ASP [14-3](#)
 assistance, obtaining [xvii](#)
 Asynchronous Transfer Mode (ATM) [2-5, 2-14, 3-27](#)
 ATA [16-4, 16-20](#)
 ATM [2-5, 2-14, 3-27](#)
 audio sources [7-3, 7-9](#)
 authentication [3-45, 16-15](#)
 AuthName [14-4](#)
 AuthPasswd [14-4](#)
 auto-detection [8-25](#)
 automated alternate routing (AAR) [2-9, 2-11, 10-19, 10-79, 13-5, 16-18](#)
 automated attendant (AA) [12-1](#)
 Automatic Location Identification (ALI) [11-4](#)
 Automatic Number Identification (ANI) [4-10, 11-4, 11-5, 11-6, 11-7](#)
 AUTO negotiate [3-20](#)

B

BackboneFast [3-6](#)
 bandwidth
 advanced formulas [3-33](#)
 best-effort [3-26](#)
 call control traffic [3-31, 3-32, 3-33](#)
 consumption [3-28, 3-29, 3-30](#)
 for Cisco Unity [13-4](#)
 guaranteed [3-25](#)
 management of [9-6](#)
 provisioning [3-23, 3-25, 3-28, 3-48](#)
 request for [5-11](#)
 requirements for call admission control [9-6](#)
 requirements for gatekeepers [9-6](#)
 voice class requirements [3-36](#)
 Barge feature [16-6](#)
 BAT [14-8](#)
 benefits of
 distributed call processing [2-14](#)
 single-site deployment [2-3](#)
 best-effort bandwidth [3-26](#)
 best practices for
 centralized call processing [2-6](#)
 Cisco CallManager Express (CME) [8-27](#)
 Cisco Unity Express (CUE) [12-6](#)
 directory integration [14-9](#)
 distributed call processing [2-14](#)
 fax support [4-18](#)
 IP-to-IP gateway [9-16](#)
 line/device approach to building classes of service [10-71](#)
 modem support [4-20](#)
 music on hold [7-8](#)
 RSVP [9-13](#)
 single-site deployment [2-4](#)
 tandem gateway [8-28](#)
 WAN design [3-24](#)
 BHCA [2-22, 8-14, 8-15, 10-81](#)
 BHCC [10-81](#)
 big-bang migration [15-3](#)
 bill-to number (BTN) [11-5](#)
 BPDU [3-6](#)
 branch office router [7-16](#)
 bridge protocol data unit (BPDU) [3-6](#)
 BTN [11-5](#)
 bugs, reporting [xvii](#)
 Bulk Administration Tool (BAT) [14-8](#)
 bursting [3-39](#)

busy hour call attempts (BHCA) [2-22, 8-14, 8-15, 10-81](#)

busy hour call completions (BHCC) [10-81](#)

C

C5421 chipset [6-5](#)

C542 chipset [6-6](#)

C549 chipset [6-5](#)

C5510 chipset [6-4](#)

CA [16-8](#)

cabling

Category 3 [3-20](#)

IBM type 1A and 2A [3-20](#)

CAC (*see* call admission control)

calculations for server capacities [8-12](#)

call admission control

bandwidth management [9-6](#)

bandwidth requirements [9-6](#)

case study [9-41](#)

centralized call processing [9-22, 9-27, 9-36](#)

components [9-2](#)

described [9-1](#)

design considerations [9-20](#)

distributed call processing [9-22, 9-31, 9-39](#)

elements [9-2](#)

for Cisco IP Communicator [16-14](#)

for Cisco IP SoftPhone [16-12](#)

for music on hold [7-15](#)

for wireless access points [16-17](#)

gatekeeper [8-16, 9-6, 10-30](#)

hybrid centralized/distributed call processing [9-25, 9-33, 9-39](#)

locations [9-2](#)

moving devices to a new location [11-13](#)

topologies [9-20](#)

callback

for emergency services [11-8, 11-14](#)

from the PSAP [11-8, 11-14](#)

call control traffic [3-31, 3-33](#)

call detail record (CDR) [2-19](#)

call flows for music on hold [7-5](#)

calling line ID (CLID) [4-10, 10-11](#)

calling party number (CPN) [11-5](#)

calling privileges [10-14, 10-39](#)

calling restrictions [10-14, 10-39](#)

calling search spaces [10-14, 10-16, 10-50, 10-66, 10-70](#)

CallManager (*see* Cisco CallManager)

CallManager Capacity Tool [8-11, 8-12, 8-13](#)

CallManager Express (CME) [2-15, 8-25](#)

call processing

agents [1-4, 2-15](#)

centralized [2-4, 9-22, 9-27, 9-36, 13-13, 13-15](#)

distributed [2-13, 9-22, 9-31, 9-39](#)

guidelines [8-1](#)

hybrid centralized/distributed deployments [9-25, 9-33, 9-39](#)

redundancy [4-2, 8-5](#)

subscriber server [8-5](#)

with Cisco CallManager release 3.3 [8-12](#)

with Cisco CallManager releases 3.1 and 3.2 [8-12](#)

with gatekeeper [8-16](#)

calls

911 [11-1](#)

classification of [10-12](#)

coverage of [10-78](#)

forwarding [10-17, 10-74](#)

H.323 [5-10](#)

hairpinning [13-14](#)

hold [7-6](#)

incoming [10-49, 10-54, 10-61](#)

music on hold [7-1](#)

number of calls per DSP resource [6-4, 6-5, 6-6, 6-7](#)

outbound [10-49, 10-53, 10-58](#)

privileges [10-14](#)

restrictions [10-39](#)

routing [10-8, 10-27, 10-30](#)

tromboning [13-14](#)

within a cluster [10-49, 10-52, 10-57](#)

- CAMA [11-6](#)
- campus
 - access switch [3-3](#)
 - infrastructure requirements [3-1](#)
- cancellation of echo [4-19](#)
- CanMapAlias [5-11](#)
- capability exchange for T.38 fax relay [4-28](#)
- capacity planning for
 - Cisco CallManager servers [8-11, 8-12, 8-13](#)
 - gateways [8-14](#)
 - music on hold [7-13](#)
 - phones [8-14](#)
 - trunks [8-15](#)
 - wireless networks [16-15](#)
- Capacity Tool [8-11, 8-12, 8-13](#)
- CAPF [16-8](#)
- CAR [14-8](#)
- case study of call admission control [9-41](#)
- Category 3 cabling [3-20](#)
- cBarge feature [16-7](#)
- CDR [2-19](#)
- CDR Analysis and Reporting (CAR) [14-8](#)
- Centralized Automatic Message Accounting (CAMA) [11-6](#)
- centralized call processing [2-4, 2-9, 9-22, 9-27, 9-36, 10-79, 13-13, 13-15](#)
- centralized gatekeeper deployment [10-34](#)
- centralized messaging [13-2, 13-13, 13-18, 13-23](#)
- Certificate Authority (CA) [16-8](#)
- Certificate Authority Proxy Function (CAPF) [16-8](#)
- changes for this release [xv](#)
- channels for wireless devices [3-41](#)
- chipsets
 - C542 [6-6](#)
 - C5421 [6-5](#)
 - C549 [6-5, 6-13](#)
 - C5510 [6-4, 6-13](#)
- CIR [3-39](#)
- Cisco.com [xvi](#)
- ciscoatGUID [14-6](#)
- ciscoatUserProfile [14-6](#)
- ciscoatUserProfileString [14-6](#)
- Cisco CallManager
 - Capacity Tool [8-11, 8-12, 8-13](#)
 - configuration of Cisco IOS gateways for fax/modem support [4-25](#)
 - current release [xv](#)
 - described [1-4](#)
 - H.323 [5-8](#)
 - new for this release [xv](#)
 - Release 3.3 [8-12](#)
 - Releases 3.1 and 3.2 [8-12](#)
 - upgrades [14-19](#)
- Cisco CallManager Express (CME) [2-15, 8-25](#)
- Cisco Emergency Responder (ER) [11-9, 11-13, 16-18](#)
- Cisco IOS
 - calling privileges [10-39](#)
 - call routing [10-27, 10-30](#)
 - classes of service [10-75](#)
 - digit manipulation [10-41](#)
 - DSP resources supported [6-4, 6-5, 6-6, 6-13](#)
 - gateways [4-23, 4-25](#)
 - minimum release required [16-3](#)
- Cisco IP Communicator [16-24, 16-30](#)
- Cisco IP Conference Station [16-19, 16-20](#)
- Cisco IP SoftPhone [11-14, 16-10, 16-24, 16-30](#)
- Cisco IP Voice Media Streaming Application [6-15, 6-16](#)
- Cisco LEAP [3-45, 16-15](#)
- Cisco Messaging Interface (CMI) [12-8](#)
- Cisco Product Identification (CPI) tool [xvii](#)
- Cisco Technical Assistance Center (TAC) [xvii](#)
- Cisco Unity [1-6, 12-1, 13-1](#)
- Cisco Unity Express (CUE) [12-1](#)
- Cisco Unity Personal Assistant (CPCA) [13-2](#)
- Cisco Unity Telephone Integration Manager (UTIM) [13-6, 13-12](#)
- CiscoWorks2000 [1-8](#)
- classes of service for users [10-63, 10-67, 10-75](#)

- classification of
 - calls [10-12](#)
 - traffic [3-4, 3-22, 3-47, 16-19](#)
- class of restriction (COR) [10-39, 10-75](#)
- Class of Service (CoS) [3-4, 16-20](#)
- CLEC [11-4](#)
- CLID [4-10, 10-11](#)
- Client Matter Code (CMC) [10-12](#)
- clipping [2-6](#)
- clocking source for fax/modem support [4-27](#)
- clustering over the WAN
 - Cisco Unity [13-18, 13-20](#)
 - described [2-16](#)
 - failover with Cisco Unity [13-22](#)
 - local failover [2-21](#)
 - music on hold [7-19](#)
 - remote failover [2-24](#)
 - WAN considerations [2-17](#)
- clusters [8-1, 8-3, 8-6](#)
- CMC [10-12](#)
- CME [2-15, 8-25](#)
- CMI [12-8](#)
- CMM [7-3, 16-3](#)
- codecs
 - complexity modes [6-1, 6-2](#)
 - flex mode [6-2](#)
 - for music on hold [7-8](#)
 - G.711 [2-2](#)
 - low bit-rate (LBR) [6-18, 6-23](#)
 - selection of [16-11, 16-13](#)
 - types [7-3, 16-11, 16-13](#)
 - types supported by complexity mode [6-3](#)
- COM [14-3](#)
- combined deployment models for messaging [13-17](#)
- Committed Information Rate (CIR) [3-39](#)
- Communication Media Module (CMM) [7-3, 16-3](#)
- Communicator [16-13, 16-24, 16-30](#)
- competitive local exchange carrier (CLEC) [11-4](#)
- complexity modes for codecs [6-1, 6-2](#)
- complex topologies [9-40](#)
- Component Object Module (COM) [14-3](#)
- components of messaging system [13-4](#)
- compressed Real-Time Transport Protocol (cRTP) [3-27, 3-36](#)
- Computer Telephony Integration (CTI) [8-9, 8-12, 12-2](#)
- Conference Station [16-19, 16-20](#)
- conferencing
 - application scenarios [6-17](#)
 - Barge [16-6](#)
 - cBarge [16-7](#)
 - described [6-8](#)
 - guidelines [6-17](#)
 - hardware resources [6-10](#)
 - MeetingPlace [1-6](#)
 - minimum software releases [6-11](#)
 - number of sessions per DSP [6-12](#)
 - rich media [1-2](#)
 - software resources [6-14](#)
- configuration examples for
 - Cisco CallManager Express [8-25](#)
 - fax/modem support [4-23, 4-25](#)
 - gatekeeper [8-16](#)
 - QoS [16-19](#)
- conformance with Section 508 [2-25](#)
- connectivity options [2-5, 2-14](#)
- control signaling [3-31, 3-33](#)
- COR [10-39, 10-75](#)
- Core Layer [3-9](#)
- coresident servers [3-12, 7-3](#)
- core switch [3-3](#)
- CoS [3-4, 16-20](#)
- coverage of calls [10-78](#)
- CPCA [13-2](#)
- CPI [xvii](#)
- CPN [11-5](#)
- cRTP [3-27, 3-36](#)
- CTI [8-9, 8-12, 12-2](#)
- CTI Manager [8-1, 8-9](#)

CTI-QBE [12-2](#)
 CUE [12-1](#)
 customer contact [1-2](#)
 Customer Directory Configuration Plugin [14-7](#)
 customer support [xvii](#)
 cutover [15-1, 15-3](#)

D

database [8-4, 14-6](#)
 data center [3-10](#)
 DC Directory [14-9](#)
 delay
 of packets [2-19, 4-18, 4-20](#)
 variation (jitter) [4-18, 4-20](#)
 deployment models
 CallManager Express [8-26](#)
 Cisco Unity [13-2](#)
 Cisco Unity Express [12-2](#)
 clustering over the WAN [2-16, 7-19](#)
 combined for messaging [13-17](#)
 conferencing guidelines [6-17](#)
 described [2-1](#)
 DHCP [3-13](#)
 multi-site dial plan [10-44](#)
 multi-site WAN with centralized call processing [2-4, 6-18, 6-23, 7-14, 10-79](#)
 multi-site WAN with distributed call processing [2-13, 6-21, 6-25, 7-19, 10-45, 10-80](#)
 music on hold [7-14](#)
 single site [2-2, 6-17, 6-23, 7-14](#)
 transcoding guidelines [6-23](#)
 voice over the PSTN [2-9](#)
 desktop phones [16-5](#)
 destination of a call [10-19](#)
 devices
 hunt list [10-81](#)
 limits per server [8-13](#)
 line group [10-26](#)
 mobility [11-13, 16-18](#)
 pools [2-21, 2-24](#)
 route group [10-13](#)
 DHCP [3-11](#)
 dial peers [10-27, 10-39, 10-41](#)
 dial plan
 911 calls [11-1](#)
 abbreviated dialing [10-3](#)
 access codes [10-6](#)
 approaches to [10-46](#)
 calling privileges [10-14, 10-39](#)
 calling search space [10-66, 10-70](#)
 call routing [10-8](#)
 classes of service [10-63, 10-67, 10-75](#)
 dial peers [10-27, 10-39, 10-41](#)
 distribution of digits [10-4](#)
 elements [10-7](#)
 emergency call string [11-11](#)
 Extension Mobility [10-22, 10-66, 10-72](#)
 for distributed call processing [10-45](#)
 for multi-site deployments [10-44](#)
 for VoPSTN [2-12](#)
 functions [10-1](#)
 hunt lists [10-23, 10-25](#)
 international calls [10-11](#)
 line groups [10-23, 10-25](#)
 number of digits [10-3](#)
 on-net vs. off-net [10-2](#)
 overlapping extensions [10-3, 10-50](#)
 partitions [10-66, 10-70](#)
 planning considerations [10-2](#)
 shared line appearance [11-14](#)
 site codes [10-5](#)
 string length [10-3](#)
 uniform on-net dialing [10-4, 10-48](#)
 variable length on-net dialing [10-5, 10-50, 10-54](#)
 voicemail [10-49, 10-54, 10-61](#)
 DID [4-10, 11-5](#)
 differential threshold [16-16](#)

differentiated services code point (DSCP) [3-4, 3-34](#)
 digital gateways [4-2, 4-13, 4-20](#)
 Digital PBX Adapter (DPA) [12-10, 12-12](#)
 digital set emulation (DSE) [12-9](#)
 digital signal processor (*see* DSP resources)
 digit manipulation [10-11, 10-18, 10-41](#)
 Direct Inward Dial (DID) [4-10, 11-5](#)
 directories
 access [14-1, 14-3](#)
 Domain Controller (DC) [14-9](#)
 integration with IP telephony system [14-1, 14-6, 14-9](#)
 LDAP [14-1](#)
 maintaining [14-18](#)
 permissions [14-14](#)
 schema [14-1](#)
 security [14-8](#)
 directory gatekeeper [8-22, 10-37](#)
 Directory Information Tree (DIT) [14-6](#)
 directory number (DN) [10-81](#)
 distributed call processing [2-13, 9-22, 9-31, 9-39, 10-80](#)
 distributed gatekeeper deployment [10-36](#)
 distributed messaging [13-2, 13-15, 13-20](#)
 Distribution Layer [3-7](#)
 distribution of digits in a dial plan [10-4](#)
 DIT [14-6](#)
 DN [10-81](#)
 DNS [3-10, 14-9](#)
 documentation
 feedback [xvii](#)
 obtaining [xvi, xviii](#)
 ordering [xvi](#)
 related [xv, xviii](#)
 Domain Controller (DC) [14-9](#)
 Domain Name System (DNS) [3-10, 14-9](#)
 domains [14-9](#)
 Domino Unified Communications Services (DUC) [13-2](#)
 DPA [12-10, 12-12](#)
 DS0 [8-15](#)
 DSCP [3-4, 3-34](#)

DSE [12-9](#)
 DSP resources
 allocating a DSP farm [6-12](#)
 allocation in Cisco CallManager [6-20](#)
 C5421 chipset [6-5](#)
 C542 chipset [6-6](#)
 C549 chipset [6-5, 6-13](#)
 C5510 chipset [6-4, 6-13](#)
 described [6-1](#)
 for voice termination [6-3](#)
 hardware platforms [6-13](#)
 in single-site deployment model [2-2](#)
 NM-HDV module [6-13](#)
 number of calls [6-4, 6-5, 6-6, 6-7](#)
 number of sessions per DSP [6-12](#)
 voice termination [6-12](#)
 DTMF [4-2, 4-5, 5-8](#)
 dual PBX integration [12-11, 12-12](#)
 dual tone multifrequency (DTMF) [4-2, 4-5, 5-8](#)
 DUC [13-2](#)
 dynamic ANI interface [11-8](#)
 Dynamic Host Configuration Protocol (DHCP) [3-11](#)

E

E.164 address [10-58, 10-59, 11-4, 11-5, 11-7](#)
 E911 [11-1, 11-3](#)
 EAP [3-45, 3-46](#)
 echo cancellation [4-19](#)
 ECM [4-19](#)
 efficiency of links [3-36](#)
 elements of a dial plan [10-7](#)
 ELIN [11-6, 11-7](#)
 EM [1-6, 8-10, 8-15, 10-22, 10-66, 10-72](#)
 emergency call string [11-11](#)
 emergency location identification number (ELIN) [11-6, 11-7](#)
 Emergency Responder [11-9, 11-13](#)
 emergency response location (ERL) [11-6, 11-7, 11-13](#)

emergency services [11-1](#)

encryption [3-32, 3-33](#)

endpoints

alternate [5-11](#)

analog gateways [16-1](#)

defined [1-5](#)

directory access [14-3](#)

features [16-30](#)

gatekeeper output [8-22](#)

gatekeeper registration [8-22](#)

hiding IP addresses [6-26](#)

line group devices [10-26](#)

types of [16-1](#)

video [1-5](#)

wireless [16-14](#)

Enterprise MCM [8-16](#)

equations for calculating

bandwidth [3-32, 3-33, 3-34](#)

calling search spaces [10-66, 10-70](#)

partitions [10-66, 10-70](#)

ER [16-18](#)

ERL [11-6, 11-7, 11-13](#)

Error Correction Mode (ECM) [4-19](#)

error rate [2-20](#)

Extensible Authentication Protocol (EAP) [3-45, 3-46, 16-15](#)

eXtensible Markup Language (XML) [14-3](#)

Extension Mobility (EM) [1-6, 8-10, 8-15, 10-22, 10-66, 10-72](#)

extensions, overlapping [10-3](#)

F

FAC [10-12](#)

failover

between subscriber servers [2-18](#)

Cisco Unity [13-3, 13-21, 13-22](#)

clustering over the WAN [2-21, 2-24](#)

to PSTN [10-58, 10-59](#)

fax

clocking source [4-27](#)

Error Correction Mode [4-19](#)

features supported [4-23](#)

gateway support for [4-2, 4-18](#)

interoperability of features [4-22](#)

network modules [16-1, 16-2](#)

pass-through mode [4-18](#)

protocols supported [4-21](#)

relay mode [4-18](#)

supported platforms and features [4-20](#)

T.38 [4-27](#)

features of endpoints [16-30](#)

feedback on this document [xvii](#)

flash cutover [15-3](#)

flash used for music on hold [7-16](#)

flat addressing [10-46, 10-54](#)

flex mode for codecs [6-2](#)

Forced Account Codes (FAC) [10-12](#)

Foreign Exchange Office (FXO) [11-6](#)

Foreign Exchange Station (FXS) [12-9](#)

forwarding

calls [10-17, 10-74](#)

voicemail calls [13-10](#)

Frame Relay [2-5, 2-14, 3-27](#)

French national numbering plan [10-70](#)

full-duplex [3-20](#)

FXO [11-6](#)

FXS [12-9](#)

G

G.711 [2-2](#)

gatekeeper

alternate [5-11, 8-19](#)

call admission control [2-14, 9-6](#)

call routing [10-30](#)

centralized deployment [10-34](#)

clustering [8-19](#)

- configuration examples 8-16
- design considerations 8-16
- directory 8-22, 10-37
- distributed deployment 10-36
- endpoints 8-22
- H.225 trunks 5-3, 5-10
- intercluster trunks 5-2
- legacy 9-18
- output example 8-22
- redundancy 8-16, 8-22
- trunk redundancy 5-3
- via-zone 9-15, 9-19, 10-33
- zones 9-6
- gatekeeper controlled H.225 trunks 5-3, 5-10
- gatekeeper controlled intercluster trunks 5-2
- Gatekeeper Transaction Message Protocol (GKTMP) 5-11
- Gatekeeper Update Protocol (GUP) 5-3, 8-19
- gateways
 - 911 services 11-11
 - all trunks busy 11-12
 - analog 4-1, 4-11, 4-20, 16-1
 - blocking 11-12
 - capacity calculations 8-14
 - Cisco IOS 4-23, 4-25
 - configuration examples for fax/modem support 4-23
 - controlled with Network Services Engine (NSE) 4-27
 - core feature requirements 4-2
 - digital 4-2, 4-13, 4-20
 - DS0 8-15
 - fax support 4-18
 - features 16-30
 - for local failover 2-23
 - for music on hold 7-3
 - IP-to-IP 9-14, 9-19, 10-33
 - modem support 4-19
 - placement 11-11
 - protocols 4-3
 - QoS configuration examples 16-19

- QSIG support 4-17
- redundancy 4-8
- selection of 4-2
- site-specific requirements 4-10
- V.34 modem support 4-20
- V.90 modem support 4-20
- VG224 4-11, 16-4
- VG248 4-24, 16-4
- voice applications 4-1, 16-1, 16-4
- GKTMP 5-11
- glossary GL-1
- GPO 14-15
- Group Policy Object (GPO) 14-15
- guaranteed bandwidth 3-25
- GUP 5-3, 8-19

H

- H.225 trunks 5-3, 5-10
- H.245 4-28
- H.323
 - analog gateways 4-11
 - Annex M1 5-10
 - call hairpinning 8-25
 - calls 5-10
 - classes of service 10-75
 - dial peers for call routing 10-27
 - digital gateways 4-13, 4-14, 4-15
 - fax and modem support 4-21
 - gateways 4-3
 - in Cisco CallManager 5-8
 - in single-site deployment model 2-4
 - T.38 fax relay 4-29
 - trunks 5-2, 5-7
- hairpinning 8-25, 13-14
- half-duplex 3-20
- hardware platforms
 - DSP resources 6-4, 6-5, 6-6, 6-7, 6-10, 6-13
 - for analog network modules 16-3

- for gatekeepers [8-16](#)
- for music on hold [7-12](#)
- memory requirements [8-1, 8-11](#)
- recommendations [2-2](#)
- types [8-1](#)
- headers for voice packets [3-29](#)
- hiding IP addresses of endpoints [6-26](#)
- high availability of network services [3-4](#)
- high-availability server [8-2](#)
- high-density voice/fax network module [16-2](#)
- high-performance server [8-2](#)
- history of revisions [xvi](#)
- hold [7-1, 7-6](#)
- holdee [7-5](#)
- holder [7-5](#)
- Hot Standby Router Protocol (HSRP) [2-14, 3-7, 8-16, 8-17](#)
- HSRP [2-14, 3-7, 8-16, 8-17](#)
- HTTP [14-3](#)
- hub-and-spoke topology [2-6, 2-14, 3-3, 9-6, 9-21, 10-30](#)
- hunt
 - groups [10-23](#)
 - lists [10-23, 10-25, 10-81](#)
 - pilot [10-23, 10-24, 10-81](#)
- hybrid centralized/distributed call processing [9-25, 9-33, 9-39](#)
- Hyper Text Transfer Protocol (HTTP) [14-3](#)

- IBM Cabling System (ICS) [3-20](#)
- ICCS [2-17, 2-22, 8-4, 8-6](#)
- ICS [3-20](#)
- impairments without QoS [3-23](#)
- incoming calls [10-49, 10-54, 10-61](#)
- infrastructure (*see* network infrastructure)
- inline power [3-19](#)
- Integrated Services (IntServ) model [9-10, 9-11, 9-13](#)
- Integrated Services/Differentiated Services
(IntServ/DiffServ) model [9-10, 9-12, 9-13](#)

- Interactive Voice Response (IVR) [2-4](#)
- intercluster trunks
 - gatekeeper controlled [5-2](#)
 - non-gatekeeper controlled [5-2](#)
- interface types for 911 calls [11-4](#)
- interference to wireless communications [3-43](#)
- international calls [10-11](#)
- interoperability of fax and modem features [4-22](#)
- Intra-Cluster Communication Signaling (ICCS) [2-17, 2-22, 8-4, 8-6](#)
- IntServ/DiffServ model [9-10, 9-12, 9-13](#)
- IntServ model [9-10, 9-11, 9-13](#)
- invia [9-15, 10-33](#)
- IOS
 - calling privileges [10-39](#)
 - call routing [10-27, 10-30](#)
 - classes of service [10-75](#)
 - digit manipulation [10-41](#)
 - DSP resources supported [6-4, 6-5, 6-6, 6-13](#)
 - minimum release required [16-3](#)
- IP/H323 feature set [8-16](#)
- IP addresses, hiding [6-26](#)
- IPCC Enterprise Edition [1-7](#)
- IPCC Express [1-7](#)
- IP Communications [1-1](#)
- IP Communications voice/fax network modules [16-2](#)
- IP Communicator [16-13, 16-24, 16-30](#)
- IP Conference Station [16-19, 16-20](#)
- IP Contact Center (IPCC) [1-7](#)
- IP-IP gateway (IPIP GW) [9-14, 9-19, 10-33](#)
- IPIP GW [9-14, 9-19, 10-33](#)
- IPMA [14-8](#)
- IP Manager Assistant (IPMA) [14-8](#)
- IP phones [16-5](#)
- IP Precedence [3-4, 3-34](#)
- IP PSTN [6-26](#)
- IPSec [2-5, 2-14](#)
- IP Security Protocol (IPSec) [2-5, 2-14](#)
- IP Telephony [1-2](#)

IP Telephony Environment Monitor (ITEM) [1-8](#)
 IP-to-IP gateway (IPIP GW) [9-14, 9-19, 10-33](#)
 IP VOICE feature set [8-25](#)
 IP Voice Media Streaming Application [6-15, 6-16, 6-23](#)
 ISDN backup [2-6, 2-9](#)
 ITEM [1-8](#)
 IVR [2-4](#)

J

jitter [4-18, 4-20](#)
 JTAPI [8-9](#)

L

LAN infrastructure [3-4](#)
 Layer 2 [2-14, 3-4](#)
 Layer 3 [3-4](#)
 LBR [6-18, 6-23](#)
 LCF [8-22, 10-33](#)
 LDAP [8-4, 14-1](#)
 LDAP Data Interchange Format (LDIF) [14-7, 14-11](#)
 LDAP over Secure Socket Layer (LDAPS) [14-8](#)
 LDAPS [14-8](#)
 ldapsearch.asp [14-4](#)
 LDIF [14-7, 14-11](#)
 LDN [11-5](#)
 LEAP [3-45, 3-46, 16-15](#)
 leased lines [2-5, 2-14, 3-27](#)
 lease times for DHCP [3-12](#)
 LEC [11-2, 11-11](#)
 legacy gatekeeper [9-18](#)
 LFI [3-27, 3-36, 3-37](#)
 Lightweight Directory Access Protocol (LDAP) [8-4, 14-1](#)
 line/device approach to classes of service [10-67](#)
 line appearances [8-14](#)
 line group devices [10-26](#)
 line groups [10-23, 10-25, 10-81](#)

line speed mismatch [3-38](#)
 link efficiency [3-36](#)
 link fragmentation and interleaving (LFI) [3-27, 3-36, 3-37](#)
 listed directory number (LDN) [11-5](#)
 LLQ [3-27, 3-34](#)
 LMHOSTS file [3-10](#)
 load balancing [3-17, 5-3, 5-6, 8-8](#)
 local dialing area [10-21](#)
 local exchange carrier (LEC) [11-2, 11-11](#)
 local failover deployment model [2-21](#)
 Location Confirm (LCF) [8-22, 10-33](#)
 Location Reject (LRJ) [10-33](#)
 Location Request (LRQ) [8-22, 10-33](#)
 locations for call admission control [9-2](#)
 loose gateway [4-27](#)
 loss of packets [4-18, 4-20](#)
 low bit-rate (LBR) codecs [6-18, 6-23](#)
 low-density voice/fax network modules [16-2](#)
 low-latency queuing (LLQ) [3-27, 3-34](#)
 LRJ [10-33](#)
 LRQ [8-22, 10-33](#)

M

MACs [11-9, 14-18](#)
 management of the network [1-8](#)
 manipulation of digits [10-18, 10-41](#)
 manufacturing installed certificate (MIC) [16-9](#)
 masking IP addresses of endpoints [6-26](#)
 MCM [8-16](#)
 MCS-7815 Sound Card [7-4](#)
 Media Gateway Control Protocol (MGCP) [2-4, 4-3, 4-12, 4-16, 4-21, 4-29](#)
 media resource group (MRG) [6-17, 6-20](#)
 media resource group list (MRGL) [6-17, 6-20](#)
 media resources [2-23, 6-1, 8-10](#)
 Media Streaming Application [6-15, 6-16](#)

- media termination point (MTP)
 - described [6-9](#)
 - for PSTN calls [6-26](#)
 - hardware resources [6-10](#)
 - hiding IP addresses of endpoints [6-26](#)
 - in single-site deployment model [2-2](#)
 - minimum software releases [6-11](#)
 - number of supported sessions [6-12](#)
 - requirements for trunks [8-16](#)
 - software resources [6-23](#)
 - with H.323 trunk [5-7](#)
 - with SIP trunk [5-8](#)
- MeetingPlace [1-6](#)
- memory requirements for servers [8-1, 8-11](#)
- Message Waiting Indicator (MWI) [12-2, 12-12](#)
- messaging
 - bandwidth management [13-4](#)
 - centralized [13-2, 13-13, 13-18, 13-23](#)
 - Cisco Unity [13-1](#)
 - combined deployment models [13-17](#)
 - deployment models [13-2](#)
 - distributed [13-2, 13-15, 13-20](#)
 - failover [13-3, 13-21, 13-22](#)
 - system components [13-4](#)
- MGCP [2-4, 4-3, 4-12, 4-16, 4-21, 4-29](#)
- MIC [16-9](#)
- Microsoft Active Directory (AD) [14-2, 14-9](#)
- Microsoft ViewMail for Outlook (VMO) [13-2](#)
- migration to IP Telephony [15-1](#)
- MISTP [3-4](#)
- MLA [14-8](#)
- MLP [3-27](#)
- MLPP [6-15](#)
- MLTS [11-2](#)
- mobile devices [16-18](#)
- models for deployment (*see* deployment models)
- modem
 - clocking source [4-27](#)
 - features supported [4-23](#)
 - gateway support for [4-2, 4-19](#)
 - interoperability of features [4-22](#)
 - pass-through mode [4-19](#)
 - protocols supported [4-21](#)
 - relay mode [4-19](#)
 - supported platforms and features [4-20](#)
 - upspeed [4-19](#)
 - V.34 [4-20](#)
 - V.90 [4-20](#)
- MoH [2-23, 7-1](#)
- moves, adds, and changes (MACs) [11-9, 14-18](#)
- MPLS [2-5, 2-6, 2-14, 3-24, 3-27, 9-34](#)
- MPLS-based topology [9-34](#)
- MRG [6-17, 6-20](#)
- MRGL [6-17, 6-20](#)
- MTP
 - described [6-9](#)
 - for PSTN calls [6-26](#)
 - hardware resources [6-10](#)
 - hiding IP addresses of endpoints [6-26](#)
 - in single-site deployment model [2-2](#)
 - minimum software releases [6-11](#)
 - number of supported sessions [6-12](#)
 - requirements for trunks [8-16](#)
 - software resources [6-23](#)
 - with H.323 trunk [5-7](#)
 - with SIP trunk [5-8](#)
- multicast music on hold [7-2, 7-8, 7-10, 7-16, 7-19](#)
- multicast traffic on WLAN [3-44](#)
- Multi-Level Administration (MLA) [14-8](#)
- Multilevel Precedence Preemption (MLPP) [6-15](#)
- multi-line telephone system (MLTS) [11-2](#)
- Multilink Point-to-Point Protocol (MLP) [3-27](#)
- Multimedia Conference Manager (MCM) [8-16](#)
- multi-party voice conferencing [1-6](#)
- multiple Cisco CallManager servers [13-23](#)
- Multiple Instance Spanning Tree Protocol (MISTP) [3-4](#)
- Multiprotocol Label Switching (MPLS) [2-5, 2-6, 2-14, 3-24, 3-27, 9-34](#)

multi-site dial plan [10-44](#)
 multi-site WAN with centralized call processing [2-4, 6-18, 6-23, 7-14, 10-79](#)
 multi-site WAN with distributed call processing [2-13, 6-21, 6-25, 7-19, 10-80](#)
 music on hold (MoH) [2-23, 7-1](#)
 MWI [12-2, 12-12](#)

N

National Emergency Number Association (NENA) [11-6](#)
 native transcoding with Cisco Unity [13-6](#)
 NENA [11-6](#)
 Netscape Directory Server [14-9](#)
 network hold [7-6](#)
 network infrastructure
 access layer [3-4](#)
 core layer [3-9](#)
 distribution layer [3-7](#)
 high availability [3-4](#)
 LAN [3-4](#)
 overview [1-4](#)
 requirements [3-1](#)
 roles [3-3](#)
 WAN [3-24](#)
 WLAN [3-40](#)
 network management [1-8](#)
 network modules [16-1](#)
 network services [3-10](#)
 Network Services Engine (NSE) [4-21, 4-27](#)
 Network Specific Facilities (NSF) [4-14](#)
 Network Time Protocol (NTP) [3-18](#)
 new for this release [xv](#)
 NFAS [2-4, 4-14](#)
 NM-HDV module [6-13](#)
 nomadic phones [11-9](#)
 Non-Facility Associated Signaling (NFAS) [2-4, 4-14](#)
 non-gatekeeper controlled intercluster trunks [5-2](#)
 non-IOS hardware platforms [6-7](#)

NPA [10-20](#)
 NSE [4-21, 4-27](#)
 NSF [4-14](#)
 NTP [3-18](#)
 Numbering Plan Area (NPA) [10-20](#)
 number of digits dialed [10-3](#)

O

off-net dialing [10-2](#)
 on-net dialing [10-2, 10-4, 10-5, 10-48, 10-50, 10-54](#)
 open authentication [3-45, 16-15](#)
 Option 150 [3-11, 3-12](#)
 organizational unit (OU) [14-6, 14-11, 14-15](#)
 OU [14-6, 14-11, 14-15](#)
 outbound calls [10-49, 10-53, 10-58](#)
 outvia [9-15, 10-33](#)
 overlapping channels [3-41](#)
 overlapping dial plans [10-50](#)
 overlapping extensions [10-3](#)
 overlap receiving [10-11](#)
 overlap sending [10-11](#)
 oversubscription [3-39](#)

P

packets
 delay [4-20](#)
 headers [3-29](#)
 loss of [4-18](#)
 partitioned addressing [10-46, 10-50](#)
 partitions [10-14, 10-15, 10-50, 10-66, 10-70](#)
passive-interface command [3-9](#)
 performance
 call rate [8-1](#)
 of servers [8-11](#)
 permissions for directory access [14-14](#)
 phased migration [15-1](#)

- phones
 - capacity calculations [8-14](#)
 - configuration [16-16](#)
 - desktop IP models [16-5](#)
 - features [16-30](#)
 - line appearances [8-14](#)
 - location of for 911 purposes [11-9](#)
 - nomadic [11-9](#)
 - QoS [16-20](#)
 - roaming [3-41, 16-16, 16-18](#)
 - software based [1-5](#)
 - web services [1-6](#)
 - wireless [16-14, 16-27](#)
- pilot number for hunt lists [10-23, 10-24, 10-81](#)
- ping utility [2-19](#)
- PINX [12-12](#)
- plain old telephone service (POTS) [11-6](#)
- platforms [2-2, 8-1, 8-16](#)
- plugins [14-7](#)
- PoE [3-19](#)
- PortFast [3-6](#)
- ports
 - for integration of Cisco Unity with Cisco CallManager [13-6, 13-12](#)
 - voicemail configuration example [13-10](#)
- positive disconnect supervision [12-14](#)
- POTS [11-6](#)
- Power over Ethernet (PoE) [3-19](#)
- precedence settings for network traffic [3-4, 3-34](#)
- prefix
 - for access code [10-20](#)
 - for gatekeeper zones [9-18, 9-19, 9-24, 9-34](#)
- PRI [11-5](#)
- Primary Rate Interface (PRI) [11-5](#)
- prioritization of traffic [3-34](#)
- priority, urgent [10-12](#)
- Private Integrated Services Network Exchange (PINX) [12-12](#)
- privileges for making calls [10-14, 10-39](#)
- problems, reporting [xvii](#)
- profiles for Extension Mobility [8-15](#)
- progress_ind alert enable 8** command [11-13](#)
- propagation of database [8-4](#)
- Protocol Auto Detect [5-10](#)
- protocols
 - ARP [3-44](#)
 - cRTP [3-27, 3-36](#)
 - DHCP [3-11](#)
 - GKTMP [5-11](#)
 - GUP [5-3, 8-19](#)
 - H.225 [5-3, 5-10](#)
 - H.245 [4-28](#)
 - H.323 [2-4, 4-3, 4-11, 4-13, 4-14, 4-15, 4-21, 4-29, 5-2, 5-7, 8-25, 10-27, 10-75](#)
 - HSRP [2-14, 3-7, 8-16, 8-17](#)
 - HTTP [14-3](#)
 - IPSec [2-5, 2-14](#)
 - LDAP [8-4, 14-1](#)
 - MGCP [2-4, 4-3, 4-12, 4-16, 4-21, 4-29](#)
 - MISTP [3-4](#)
 - MLP [3-27](#)
 - NTP [3-18](#)
 - RAS [10-30](#)
- routing [3-9](#)
- RSTP [3-4, 3-6](#)
- RSVP [9-7, 9-14](#)
- RTP [2-14](#)
- SCCP [4-3, 4-21](#)
- SDP [4-28](#)
- SIP [2-14, 4-11, 4-13, 4-14, 4-15, 5-8, 6-10, 6-15](#)
- SMDI [12-7](#)
- SNMP [11-9](#)
- SRTP [3-29](#)
- STP [3-6](#)
- TFTP [3-14, 8-1, 8-8, 8-12](#)
- UDP [2-14, 5-3](#)
- provisioning servers [8-11, 8-12, 8-13](#)
- proxy [8-16](#)

PSAP [11-2, 11-8, 11-14, 16-18](#)
 PSTN [2-2, 2-5, 2-9, 2-14, 6-26, 10-19, 11-2](#)
 public safety answering point (PSAP) [11-2, 11-8, 11-14, 16-18](#)
 Public Switched Telephone Network (PSTN) [2-2, 2-5, 2-14, 10-19, 11-2](#)
 publisher server [2-18, 8-4, 14-6](#)

Q

Q.SIG [5-10](#)
 QBE [12-2](#)
 QBSS [16-16](#)
 QBSS-Differential Threshold [16-16](#)
 QoS
 configuration examples [16-19](#)
 for LAN [3-21](#)
 for music on hold [7-12](#)
 for WAN [3-24, 3-26](#)
 for wireless LAN [3-47](#)
 general [1-4](#)
 QoS Basis Service Set (QBSS) [16-16](#)
 QSIG [4-13, 4-17, 12-12, 15-3](#)
 Quality of Service (QoS)
 configuration examples [16-19](#)
 for LAN [3-21](#)
 for music on hold [7-12](#)
 for WAN [3-24, 3-26](#)
 for wireless LAN [3-47](#)
 general [1-4](#)
 quantity of
 calls [8-15](#)
 gateways [8-14](#)
 phones [8-14](#)
 trunks [8-15](#)
 queuing of voice traffic [3-23, 3-47](#)
 Quick Buffer Encoding (QBE) [12-2](#)

R

radio frequency (RF) [16-14](#)
 RADIUS [3-45, 3-46](#)
 Rapid Spanning Tree Protocol (RSTP) [3-4, 3-6](#)
 RAS [5-3, 9-6, 10-30](#)
 RBOC [11-2](#)
 RDNIS [13-5](#)
 Real-Time Monitoring Tool (RTMT) [14-8](#)
 Real-time Transport Protocol (RTP) [2-14](#)
 recommended hardware and software combinations [2-2](#)
 Redirected Dialed Number Information Service (RDNIS) [13-5](#)
 redundancy
 call processing [8-5](#)
 cluster configurations [8-6](#)
 for music on hold [7-11](#)
 for trunks [5-3](#)
 gatekeeper [8-16](#)
 gateway support for [4-2, 4-8](#)
 IP-to-IP gateways [9-17](#)
 load balancing [8-8](#)
 TFTP services [3-16](#)
 Regional Bell Operating Company (RBOC) [11-2](#)
 Registration Admission Status (RAS) [5-3, 9-6, 10-30](#)
 related documentation [xv](#)
 Relative Signal Strength Indicator (RSSI) [16-16](#)
 releases of software [xv](#)
 Remote Authentication Dial-In User Service (RADIUS) [3-45, 3-46](#)
 remote failover deployment model [2-24](#)
 remote site survivability [2-6](#)
 replication of database [8-4](#)
 request for bandwidth [5-11](#)
 request for technical service [xviii](#)
 resilience [5-3, 8-1](#)
 resolution of addresses [10-32, 10-33](#)
 Resource Reservation Protocol (RSVP) [9-7, 9-14](#)
 revision history [xvi](#)

RF [16-14](#)
 RFC 2833 [6-10](#)
 rich-media conferencing [1-2](#)
 RJ-45 [3-20](#)
 roaming [3-41, 11-9, 16-16, 16-18](#)
 roles in the network infrastructure [3-3](#)
 root guard [3-6](#)
 Round Trip Time (RTT) [2-19, 2-22](#)
 route
 filters [10-10](#)
 group devices [10-13](#)
 groups [10-11, 10-13](#)
 lists [10-13](#)
 patterns [10-8, 10-10](#)
 routers
 branch office [7-16](#)
 flash [7-16](#)
 roles and features [3-3](#)
 selective for E911 [11-3](#)
 routing
 calling line ID [10-11](#)
 calls [10-8, 10-27, 10-30](#)
 digit manipulation [10-11](#)
 protocols [3-9](#)
 time-of-day (ToD) [10-26](#)
 RSSI [16-16](#)
 RSSI-Differential Threshold [16-16](#)
 RSTP [3-4, 3-6](#)
 RSVP [9-7, 9-14](#)
 RTMT [14-8](#)
 RTP [2-14](#)
 RTT [2-19, 2-22](#)

S

s.AuthName parameter [14-4](#)
 s.AuthPasswdparameter [14-4](#)
 s.base parameter [14-4](#)
 s.port parameter [14-4](#)
 s.server parameter [14-4](#)
 scale [8-1](#)
 SCCP [4-3, 4-21](#)
 schema [14-1](#)
 Schema Master [14-7, 14-11](#)
 SDK [14-3](#)
 SDP [4-28](#)
 Search COM Server [14-3](#)
 Section 255 [2-25](#)
 Section 508 [2-25](#)
 Secure Socket Layer (SSL) [14-8](#)
 security
 directories [14-8](#)
 in general [1-7](#)
 phones [16-8, 16-9](#)
 wireless network [3-45](#)
 selective router [11-3](#)
 Sequenced Routing Update Protocol (SRTP) [3-29](#)
 Server Load Balancing (SLB) [14-10](#)
 servers
 capacity planning [8-11, 8-12, 8-13](#)
 coresident [3-12, 7-3](#)
 CTI Manager [8-9](#)
 data center [3-10](#)
 domain [14-9](#)
 farm [3-10](#)
 for Cisco CallManager [8-2](#)
 for media resources [6-1](#)
 for music on hold [7-3, 7-4, 7-12](#)
 high-performance [8-2](#)
 maximum number of devices [8-13](#)
 memory requirements [8-1, 8-11](#)
 multiple Cisco CallManager servers [13-23](#)
 performance [8-11](#)
 publisher [2-18, 8-4](#)
 standalone [3-12, 7-3](#)
 subscriber [2-18, 8-5](#)
 TFTP [8-8, 8-12](#)
 types [8-2](#)

- Service Inter-Working (SIW) [2-5, 2-14, 3-27](#)
 - service request [xviii](#)
 - services, supplementary [4-2](#)
 - service set identifier (SSID) [3-41, 3-44](#)
 - services within a cluster [8-3](#)
 - Session Definition Protocol (SDP) [4-28](#)
 - Session Initiation Protocol (SIP)
 - analog gateways [4-11](#)
 - annunciator [6-15](#)
 - digital gateways [4-13, 4-14, 4-15](#)
 - for distributed call processing [2-14](#)
 - media termination point for RFC 2833 [6-10](#)
 - trunks [5-8](#)
 - severity level of service request [xviii](#)
 - shaping traffic [3-38](#)
 - shared line appearance [11-14](#)
 - shielded twisted-pair (STP) [3-20](#)
 - signaling encryption [3-32, 3-33](#)
 - Signaling System 7 [2-4](#)
 - Simple Network Management Protocol (SNMP) [11-9](#)
 - Simplified Message Desk Interface (SMDI) [12-7](#)
 - single-site deployment model [2-2, 6-17, 6-23, 7-14](#)
 - single-site messaging model [13-2](#)
 - SIP
 - analog gateways [4-11](#)
 - annunciator [6-15](#)
 - digital gateways [4-13, 4-14, 4-15](#)
 - for distributed call processing [2-14](#)
 - media termination point for RFC 2833 [6-10](#)
 - trunks [5-8](#)
 - site codes [10-5, 10-61](#)
 - site survey [16-14](#)
 - SIW [2-5, 2-14, 3-27](#)
 - sizing servers [8-11, 8-12, 8-13](#)
 - Skinny Client Control Protocol (SCCP) [4-3, 4-21](#)
 - SLB [14-10](#)
 - SMDI [12-7](#)
 - SNMP [11-9](#)
 - soft clients [11-14](#)
 - SoftPhone [11-14, 16-10, 16-24, 16-30](#)
 - software-based phones [16-10, 16-30](#)
 - software conferencing resources [6-14](#)
 - Software Development Kit (SDK) [14-3](#)
 - software MTP resources [6-23](#)
 - software recommendations [2-2](#)
 - software versions [xv, 2-2, 6-11, 16-3](#)
 - Solution Reference Network Design (SRND) [xv](#)
 - Spanning Tree Protocol (STP) [3-6](#)
 - SQL [2-22, 14-6](#)
 - SRND [xv](#)
 - SRST [2-5, 2-6, 7-16, 10-79, 11-4](#)
 - SRTP [3-29](#)
 - SS7 [2-4](#)
 - SSID [3-41, 3-44](#)
 - SSL [14-8](#)
 - standalone server [3-12, 7-3](#)
 - standard server [8-2](#)
 - standby preempt** command [3-7](#)
 - standby track** command [3-7](#)
 - star topology [9-21](#)
 - static ANI interface [11-8](#)
 - Static Wire Equivalent Privacy (WEP) [3-45](#)
 - STP [3-6, 3-20](#)
 - string length [10-3](#)
 - subscriber server [2-18, 8-5, 14-6](#)
 - supplementary services on gateways [4-2, 4-6](#)
 - support [xvii](#)
 - survey of wireless network [16-14](#)
 - Survivable Remote Site Telephony (SRST) [2-5, 2-6, 7-16, 10-79, 11-4](#)
 - switches, roles and features [3-3](#)
-
- ## T
- T.38 fax relay [4-27](#)
 - TAC [xvii](#)
 - Tail-End Hop-Off (TEHO) [10-44](#)
 - TAPI [8-9](#)

- TCS 8-26
 - technical assistance xvii
 - Technical Assistance Center (TAC) xvii
 - TEHO 10-44
 - Telecommunications Act 2-25
 - telephone record and playback (TRaP) 13-2
 - telephone user interface (TUI) 13-2
 - Telex P-800 USB 7-4
 - Terminal Capabilities Set (TCS) 8-26
 - termination of calls 6-1
 - test calls for 911 11-14
 - TFTP 3-14, 8-1, 8-8, 8-12
 - third-party
 - software applications 1-2
 - voicemail systems 12-7, 12-15
 - threshold, differential 16-16
 - time-of-day (ToD) routing 10-26
 - TLS 16-8
 - ToD 10-26
 - Token Ring 3-20
 - topology
 - case study 9-41
 - complex 9-40
 - for call admission control 9-20
 - hub-and-spoke 9-6, 9-21, 10-30
 - MPLS-based 9-34
 - star 9-21
 - two-tier hub-and-spoke 9-26
 - traditional approach to classes of service 10-63
 - traffic
 - call control 3-31, 3-33
 - classification 3-4, 3-22, 3-47, 16-19
 - prioritization 3-34
 - provisioning for 3-29
 - queuing 3-23, 3-47
 - shaping 3-38
 - transcoding
 - application scenarios 6-23
 - Cisco Unity 13-6
 - described 6-8
 - guidelines 6-23
 - hardware resources 6-10
 - IP PSTN 6-26
 - minimum software releases 6-11
 - number of sessions per DSP 6-12
 - translation patterns 10-18
 - translation profiles 10-41
 - transport layer security (TLS) 16-8
 - TRaP 13-2
 - Trivial File Transfer Protocol (TFTP) 3-14, 8-1, 8-8, 8-12
 - tromboning 13-14
 - troubleshooting for clustering over the WAN 2-20
 - trunks
 - capacity calculations 8-15
 - described 5-1
 - H.225 5-3, 5-10
 - H.323 5-2, 5-7
 - intercluster, gatekeeper controlled 5-2
 - intercluster, non-gatekeeper controlled 5-2
 - load balancing 5-3, 5-6
 - MTP requirements 8-16
 - redundancy 5-3
 - SIP 5-8
 - trust 16-19
 - TUI 13-2
 - Tunneled Q.SIG 5-10
 - two-tier hub-and-spoke topology 9-26
-
- ## U
- UDC 3-20
 - UDLD 3-6
 - UDP 2-14, 3-36, 5-3
 - UMDirectoryConfiguration.ini file 14-18
 - unicast music on hold 7-2, 7-8, 7-10, 7-19
 - UniDirectional Link Detection (UDLD) 3-6
 - unified communications 1-1
 - unified messaging (*see also* messaging) 1-6, 12-1, 13-1

uniform on-net dial plan [10-4, 10-48](#)
 uninterrupted power supplies (UPS) [3-19](#)
 Unity [1-6, 12-1, 13-1](#)
 Unity Express [12-1](#)
 Unity Telephone Integration Manager (UTIM) [13-6, 13-12](#)
 universal data connector (UDC) [3-20](#)
 upgrades to Cisco CallManager [14-19](#)
 UplinkFast [3-6](#)
 UPS [3-19](#)
 upspeed [4-19](#)
 Urgent Priority [10-12](#)
 user classes of service [10-63, 10-67, 10-75](#)
 User Creation Base [14-15](#)
 User Datagram Protocol (UDP) [2-14, 3-36, 5-3](#)
 user hold [7-6](#)
 User Search Base [14-11, 14-15](#)
 User-to-User Information Element (UUIE) [5-10](#)
 utilization of
 DSOs [8-15](#)
 phones [8-14](#)
 trunks [8-15](#)
 UTIM [13-6, 13-12](#)
 UUIE [5-10](#)

V

V.34 modems [4-20](#)
 V.90 modems [4-20](#)
 V3PN [2-5, 2-14](#)
 VAD [4-19, 8-10](#)
 VAF [3-37](#)
 variable length on-net dial plan [10-5, 10-50, 10-54](#)
 VATS [3-39](#)
 versions of software [xv](#)
 VG224 Voice Gateway [4-11, 16-4, 16-19](#)
 VG248 Analog Phone Gateway [4-24, 12-8, 16-4, 16-19](#)
 via-zone gatekeeper [9-15, 9-19, 10-33](#)
 VIC [16-2](#)
 video telephony [1-2, 1-5](#)

ViewMail for Outlook (VMO) [13-2](#)
 virtual LAN (VLAN) [3-4, 3-41, 16-19](#)
 Virtual Private Network (VPN) [2-5, 2-14](#)
 VLAN [3-4, 3-41, 16-19](#)
 VMO [13-2](#)
 voice/WAN interface card (VWIC) [16-2](#)
 voice activity detection (VAD) [4-19, 8-10](#)
 Voice-Adaptive Fragmentation (VAF) [3-37](#)
 Voice-Adaptive Traffic Shaping (VATS) [3-39](#)
 Voice and Video Enabled IPsec VPN (V3PN) [2-5, 2-14](#)
 voice class bandwidth requirements [3-36](#)
 voice gateways [4-1, 16-1, 16-4](#)
 voice interface card (VIC) [16-2](#)
 voicemail
 centralized [12-12](#)
 Cisco Unity [12-1, 13-1](#)
 Cisco Unity Express [12-1](#)
 dial plan [10-49, 10-54, 10-61](#)
 dual PBX integration [12-11, 12-12](#)
 for local failover [2-23](#)
 integration with IP telephony system [12-1](#)
 port configurations for forwarding conditions [13-10](#)
 positive disconnect supervision [12-14](#)
 third-party systems [12-7, 12-13, 12-15](#)
 unified messaging [12-1, 13-1](#)
 Voice Media Streaming Application [6-23](#)
 voice over IP (VoIP) [3-29](#)
 voice over the PSTN (VoPSTN) [2-9](#)
 voice port integration [13-6, 13-12](#)
voice rtp send-recv command [11-13](#)
 voice termination [6-1, 6-12](#)
 voice translation profiles [10-41](#)
 VoIP [3-29](#)
 VoPSTN [2-9](#)
 VPN [2-5, 2-14](#)
 VWIC [16-2](#)

W

WAN aggregation router [3-3](#)

WAN infrastructure [3-24](#)

web services [1-6](#)

weighted fair queuing [3-34](#)

WEP [3-45, 16-15](#)

wildcards [10-10](#)

Wired Equivalent Privacy (WEP) [16-15](#)

wireless IP phones [16-14, 16-27](#)

wireless LAN (WLAN) [3-40](#)

WLAN infrastructure [3-40](#)

WS-X6624-FXS analog interface module [16-4](#)

X

XML [14-3](#)

Z

zone prefix [9-18, 9-19, 9-24, 9-34](#)

zones for gatekeepers [9-6](#)