



Cisco Unified Communications Solution Reference Network Design (SRND)

Based on Cisco Unified Communications Manager Release 5.x
October 30, 2008

Americas 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: 408 527-0883

ALL DESIGNS, SPECIFICATIONS, STATEMENTS, INFORMATION, AND RECOMMENDATIONS (COLLECTIVELY, "DESIGNS") IN THIS MANUAL ARE PRESENTED "AS IS," WITH ALL FAULTS. CISCO AND ITS SUPPLIERS DISCLAIM ALL WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE WARRANTY 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 THE DESIGNS, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

THE DESIGNS ARE SUBJECT TO CHANGE WITHOUT NOTICE. USERS ARE SOLELY RESPONSIBLE FOR THEIR APPLICATION OF THE DESIGNS. THE DESIGNS DO NOT CONSTITUTE THE TECHNICAL OR OTHER PROFESSIONAL ADVICE OF CISCO, ITS SUPPLIERS OR PARTNERS. USERS SHOULD CONSULT THEIR OWN TECHNICAL ADVISORS BEFORE IMPLEMENTING THE DESIGNS. RESULTS MAY VARY DEPENDING ON FACTORS NOT TESTED BY CISCO.

CCDE, CCENT, Cisco Eos, Cisco Lumin, Cisco Nexus, Cisco StadiumVision, Cisco TelePresence, Cisco WebEx, the Cisco logo, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn and Cisco Store are service marks; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, iQuick Study, IronPort, the IronPort logo, LightStream, Linksys, MediaTone, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo 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. (0809R)

Cisco Unified Communications SRND (Based on Cisco Unified Communications Manager 5.x)
Copyright © 2006 Cisco Systems, Inc. All rights reserved.



CONTENTS

Preface **xxv**

CHAPTER 1

Introduction **1-1**

- Overview of Cisco Unified Communications **1-2**
 - Cisco IP Network Infrastructure **1-3**
 - Quality of Service **1-4**
 - Call Processing Agent **1-4**
 - Communication Endpoints **1-5**
 - Conferencing, Messaging, and Collaboration Capabilities **1-6**
 - Applications **1-6**
 - Security **1-8**

CHAPTER 2

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**
- Multisite WAN with Centralized Call Processing **2-4**
 - Best Practices for the Centralized Call Processing Model **2-7**
 - Remote Site Survivability **2-8**
 - Voice Over the PSTN as a Variant of Centralized Call Processing **2-10**
 - VoPSTN Using AAR **2-12**
 - VoPSTN Using Dial Plan **2-13**
- Multisite WAN with Distributed Call Processing **2-14**
 - Benefits of the Distributed Call Processing Model **2-16**
 - Best Practices for the Distributed Call Processing Model **2-16**
 - Call Processing Agents for the Distributed Call Processing Model **2-17**
- Clustering Over the IP WAN **2-17**
 - WAN Considerations **2-18**
 - Intra-Cluster Communications **2-19**
 - Failover Between Subscriber Servers **2-20**
 - Unified CM Publisher **2-20**
 - Call Detail Records (CDR) **2-20**
 - Delay Testing **2-20**
 - Error Rate **2-21**

Troubleshooting	2-21
Local Failover Deployment Model	2-22
Unified CM Provisioning for Local Failover	2-24
Gateways for Local Failover	2-25
Voicemail for Local Failover	2-25
Music on Hold and Media Resources for Local Failover	2-25
Remote Failover Deployment Model	2-26
Design Considerations for Section 508 Conformance	2-27

CHAPTER 3

Network Infrastructure 3-1

What's New in This Chapter	3-3
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)	3-23
Category 3 Cabling	3-24
IBM Type 1A and 2A Cabling	3-25
LAN Quality of Service (QoS)	3-25
Traffic Classification	3-27
Interface Queuing	3-28
Bandwidth Provisioning	3-29
Impairments to IP Communications if QoS is Not Employed	3-29
WAN Infrastructure	3-30
WAN Design and Configuration	3-30
Deployment Considerations	3-30
Guaranteed Bandwidth	3-31
Best-Effort Bandwidth	3-32
WAN Quality of Service (QoS)	3-33
Traffic Prioritization	3-34
Link Efficiency Techniques	3-35
Traffic Shaping	3-37
Resource Reservation Protocol (RSVP)	3-40
RSVP Principles	3-40
RSVP and QoS in WAN Routers	3-43
RSVP Application ID	3-47
RSVP Design Best Practices	3-49

Bandwidth Provisioning	3-49
Provisioning for Bearer Traffic	3-51
Additional Considerations for Bearer Traffic with RSVP	3-54
Provisioning for Call Control Traffic with Centralized Call Processing	3-58
Provisioning for Call Control Traffic with Distributed Call Processing	3-63
Wireless LAN Infrastructure	3-63
WLAN Design and Configuration	3-64
Wireless Infrastructure Considerations	3-64
Wireless AP Configuration and Design	3-67
Wireless Security	3-68
WLAN Quality of Service (QoS)	3-70
Traffic Classification	3-70
Interface Queuing	3-71
Bandwidth Provisioning	3-71

CHAPTER 4
Gateways 4-1

What's New in This Chapter	4-1
Traffic Patterns and Gateway Sizing	4-2
Definitions and Terminology	4-2
PSTN Traffic Patterns	4-3
Normal Business Traffic Profile	4-3
Contact Center Traffic Profile	4-3
Gateway Sizing for Contact Center Traffic	4-4
Voice Activity Detection (VAD)	4-4
Codec	4-5
Performance Overload	4-5
Performance Tuning	4-5
Additional Information	4-6
Understanding Cisco Gateways	4-6
Cisco Access Analog Gateways	4-7
Cisco Access Digital Trunk Gateways	4-7
Gateway Selection	4-7
Core Feature Requirements	4-7
Gateway Protocols	4-8
Gateway Protocols and Core Feature Requirements	4-11
DTMF Relay	4-11
Supplementary Services	4-12
Unified CM Redundancy	4-15
Site-Specific Gateway Requirements	4-16

QSIG Support	4-24
Fax and Modem Support	4-25
Gateway Support for Fax Pass-Through and Fax Relay	4-25
Gateway Support for Modem Pass-Through and Modem Relay	4-26
Supported Platforms and Features	4-27
Platform Protocol Support	4-28
Gateway Combinations and Interoperability of Features	4-29
Feature Support Between Similar Gateways	4-30
Gateway Configuration Examples	4-30
Cisco IOS Gateway Configuration	4-30
Cisco VG248 Configuration	4-31
Unified CM Configuration for Cisco IOS Gateways	4-32
Clock Sourcing for Fax and Modem Pass-Through	4-32
T.38 Fax Relay	4-33
Loose Gateway Controlled with Named Service Event (NSE)	4-33
Gateway Controlled with Capability Exchange Through H.245 or Session Description Protocol (SDP)	4-33
Call-Agent-Controlled T.38 with H.323 Annex D	4-35
Gateways for Video Telephony	4-36
Routing Inbound Calls from the PSTN	4-37
Routing Outbound Calls to the PSTN	4-39
Automated Alternate Routing (AAR)	4-40
Least-Cost Routing	4-41
ISDN B-Channel Binding, Rollover, and Busy Out	4-42
Inbound Calls	4-43
Outbound Calls	4-43
Configuring the Gateways in Unified CM	4-44
Call Signaling Port Numbers	4-44
Call Signaling Timers	4-45
Voice Gateways Bearer Capabilities	4-45

CHAPTER 5

Cisco Unified CM Trunks	5-1
What's New in This Chapter	5-1
A Comparison of H.323 and SIP Trunks	5-1
Overview of H.323 Trunks	5-3
Overview of SIP Trunks	5-3
IP PSTN and IP Trunks to Service Provider Networks	5-4
H.323 Trunks	5-4
Intercluster Trunk (Non-Gatekeeper Controlled)	5-5

Intercluster Trunk (Gatekeeper Controlled)	5-5
H.225 Trunk (Gatekeeper Controlled)	5-6
Gatekeeper Trunk Redundancy, Resilience, and Load Balancing	5-6
H.323 Trunks with Media Termination Points	5-11
H323 Outbound FastStart Call Connections	5-11
Other MTP Uses	5-11
H.323 Operation in Unified CM	5-12
SIP Trunks	5-14
General Deployment Considerations:	5-14
DTMF Transport	5-15
SIP Delayed Offer and Early Offer	5-16
Media Termination Points	5-17
SIP Trunk Transport Protocols	5-17
SIP Intercluster Trunks	5-18

CHAPTER 6

Media Resources 6-1

What's New in This Chapter	6-2
Voice Termination	6-2
Medium and High Complexity Mode	6-2
Flex Mode	6-3
DSP Resources for Voice Termination	6-4
Audio Conferencing	6-8
Audio Conferencing Resources	6-8
Transcoding	6-11
Transcoding Resources	6-12
Media Termination Point (MTP)	6-13
Re-Packetization of a Stream	6-13
H.323 Supplementary Services	6-13
H.323 Outbound Fast Start	6-13
Named Telephony Events (RFC 2833)	6-14
When are Media Termination Points Required for Named Telephony Events?	6-14
Configuration of DTMF on SIP and H.323 Gateways	6-16
CTI Route Points	6-17
MTP Resources	6-17
Annunciator	6-18
Cisco RSVP Agent	6-19
Cisco IP Voice Media Streaming Application	6-19
Hardware and Software Capacities	6-20

PVDMs	6-20
Cisco 2800 and 3800 Series Platforms	6-21
Network Modules	6-21
Calculating DSP Requirements for the NM-HDV	6-22
General Design Guidelines	6-22
Media Resource Groups and Lists	6-23
Deployment Models	6-24
IP PSTN Access	6-26

CHAPTER 7

Music on Hold 7-1

Deployment Basics of MoH	7-2
Unicast and Multicast MoH	7-2
Co-resident and Standalone MoH	7-3
Fixed and Audio File MoH Sources	7-3
MoH Server as Part of the Unified CM 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 (Live) Audio Sources	7-9
Unicast and Multicast in the Same Unified CM Cluster	7-10
Redundancy	7-11
Quality of Service (QoS)	7-11
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 Multisite Deployments	7-14
Call Admission Control and MoH	7-15
Multicast MoH from Branch Routers	7-16
Distributed Multisite Deployments	7-19
Clustering Over the WAN	7-20
Detailed Unicast and Multicast MoH Call Flows	7-20
SCCP Call Flows	7-20

SIP Call Flows 7-23

CHAPTER 8**Call Processing 8-1**

- Unified CM Cluster Guidelines 8-2
 - Hardware Platforms 8-2
 - Unified CM Cluster Services 8-3
 - Intracuster Communications 8-4
 - Intracuster Security 8-5
 - Publisher 8-6
 - Call Processing Subscriber 8-7
 - TFTP Server 8-12
 - CTI Manager 8-12
 - IP Voice Media Streaming Application 8-13
 - Voice Activity Detection 8-13
 - Unified CM Applications 8-13
- Unified CM Platform Capacity Planning 8-15
 - Capacity Calculations 8-16
 - Cisco Unified CM Capacity Tool 8-17
- Gatekeeper Design Considerations 8-20
 - Hardware Platform Selection 8-21
 - Gatekeeper Redundancy 8-21
 - Hot Standby Router Protocol (HSRP) 8-21
 - Gatekeeper Clustering (Alternate Gatekeeper) 8-24
 - Directory Gatekeeper Redundancy 8-27
- Interoperability of Unified CM and Unified CM Express 8-30
 - Multisite IP Telephony Deployments Using Unified CM and Unified CME with SIP Trunks 8-31
 - Multisite IP Telephony Deployments Using Unified CM and Unified CME, with H.323 Trunks and an IP-to-IP Gateway 8-33

CHAPTER 9**Call Admission Control 9-1**

- Best Practices Summary 9-2
- Call Admission Control Principles 9-3
 - Topology-Unaware Call Admission Control 9-3
 - Topology-Aware Call Admission Control 9-7
 - Special Considerations for MPLS Networks 9-11
- Call Admission Control Elements 9-11
 - Unified CM Static Locations 9-12
 - Cisco IOS Gatekeeper Zones 9-14
 - Unified CM RSVP-Enabled Locations 9-16

Cisco RSVP Agent Provisioning	9-18
Cisco RSVP Agent Registration	9-21
RSVP Policy	9-23
Migrating from Static Locations to RSVP Call Admission Control	9-25
RSVP Application ID	9-27
Cisco IOS Gatekeeper and IP-to-IP Gateway with RSVP	9-28
Via-Zone Gatekeeper	9-29
Design Best Practices	9-30
Redundancy	9-31
Configuration Guidelines	9-32
Call Admission Control Design	9-36
Simple Hub-and-Spoke Topologies	9-36
Centralized Unified CM Deployments	9-37
Distributed Unified CM Deployments	9-38
Two-Tier Hub-and-Spoke Topologies	9-40
Centralized Unified CM Deployments	9-41
Distributed Unified CM Deployments	9-43
Simple MPLS Topologies	9-44
Centralized Unified CM Deployments	9-46
Distributed Unified CM Deployments	9-48
Generic Topologies	9-50
Centralized Unified CM Deployments	9-51
Distributed Unified CM Deployments	9-54

CHAPTER 10

Dial Plan 10-1

Planning Considerations	10-2
Dialed Pattern Recognition	10-3
On-Net versus Off-Net Dialing	10-3
Abbreviated Dialing	10-4
Avoiding Overlap of Extension Dialing	10-4
Dialing String Length	10-4
Uniform On-Net Dial Plan	10-5
Variable Length On-Net Dial Plan	10-6
On-Net and Off-Net Access Codes	10-7
Plan Ahead	10-8
Dial Plan Elements	10-8
User Input on SCCP Phones	10-9
User Input on Type-A SIP Phones	10-9
User Input on Type-B SIP Phones	10-11

SIP Dial Rules	10-13
Call Routing in Unified CM	10-15
Route Patterns	10-17
Route Lists	10-20
Route Groups	10-21
Route Group Devices	10-21
Calling Privileges in Unified CM	10-21
Partitions	10-22
Calling Search Spaces	10-23
Digit Manipulation in Unified CM	10-26
Automated Alternate Routing	10-27
Establish the PSTN Number of the Destination	10-27
Prefix the Required Access Codes	10-28
Select the Proper Dial Plan and Route	10-28
Special Considerations for Sites Located Within the Same Local Dialing Area	10-29
Extension Mobility	10-30
Immediate Divert (iDivert)	10-32
Hunt Lists and Line Groups	10-33
Hunt Pilot	10-35
Hunt List	10-36
Line Group	10-36
Line Group Devices	10-37
Time-of-Day Routing	10-37
Call Routing in Cisco IOS with H.323 Dial Peers	10-38
Call Routing in Cisco IOS with a Gatekeeper	10-41
Centralized Gatekeeper Configuration	10-45
Distributed Gatekeeper Configuration	10-47
Distributed Gatekeeper Configuration with Directory Gatekeeper	10-48
Calling Privileges in Cisco IOS with H.323 Dial Peers	10-50
Digit Manipulation in Cisco IOS with H.323 Dial Peers	10-52
Design Considerations	10-54
Design Guidelines for Multisite Deployments	10-55
Choosing a Dial Plan Approach	10-57
Deploying Uniform On-Net Dial Plans	10-58
Inter-Site Calls Within a Cluster	10-60
Outgoing PSTN and IP WAN Calls	10-60
Emergency Calls	10-60
Incoming Calls	10-60
Voicemail Calls	10-60
Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing	10-61

Inter-Site Calls Within a Cluster	10-63
Outgoing PSTN and IP WAN Calls	10-64
Incoming Calls	10-65
Voicemail Calls	10-65
Deploying Variable-Length On-Net Dial Plans with Flat Addressing	10-66
Inter-Site Calls Within a Cluster	10-68
Outgoing PSTN and IP WAN Calls	10-69
Incoming Calls	10-72
Voicemail Calls	10-72
Special Considerations for Deployments Without Site Codes	10-72
Deploying Dialed Pattern Recognition in the Phones with Unified CM 5.x	10-74
Building Classes of Service for Unified CM with the Traditional Approach	10-76
Building Classes of Service for Unified CM with the Line/Device Approach	10-80
Guidelines for the Line/Device Approach	10-84
Extension Mobility Considerations for the Line/Device Approach	10-85
Building Classes of Service in Cisco IOS with H.323	10-88
Deploying Call Coverage	10-91
Deploying Call Coverage in a Multisite Centralized Call Processing Model	10-92
Deploying Call Coverage in a Multisite Distributed Call Processing Model	10-93
Hunt Pilot Scalability	10-94

CHAPTER 11

Emergency Services 11-1

Planning for 911 Functionality	11-1
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)	11-4
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	11-10
Gateway Considerations	11-11
Gateway Placement	11-11
Gateway Blocking	11-11
Answer Supervision	11-12
Cisco Emergency Responder Considerations	11-13

Emergency Responder Version Compatibility with Unified CM	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	11-14
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
Third-Party Voicemail Design 12-1

SMDI	12-1
Cisco Messaging Interface	12-2
Cisco VG248	12-3
Considerations When Using FXS Ports	12-3
Digital Set Emulation	12-4
Dual PBX Integration	12-5
Centralized Voicemail	12-6
Positive Disconnect Supervision	12-11
Summary of Third-Party Voicemail Integration	12-11

CHAPTER 13
Cisco Unity 13-1

Messaging Deployment Models	13-2
Single-Site Messaging	13-2
Centralized Messaging	13-3
Distributed Messaging	13-3
Messaging Failover	13-4
Messaging System Infrastructure Components	13-4
Port Groups (Separate Integrations)	13-5
Managing Bandwidth	13-6
Cisco Unity and Unity Connection Native Transcoding Operation	13-8
Configuring a Unified CM SIP Trunk for Voicemail Integration	13-9
Voice Port Integration with a Unified CM Cluster	13-10
Voice Port Integration with Dedicated Unified CM Backup Servers	13-12
Centralized Messaging and Centralized Call Processing	13-13

Distributed Messaging with Centralized Call Processing	13-14
Combined Messaging Deployment Models	13-16
Centralized Messaging with Clustering Over the WAN	13-17
Distributed Messaging with Clustering Over the WAN	13-19
Cisco Unity Messaging Failover	13-20
Cisco Unity Failover and Clustering Over the WAN	13-20
Centralized Messaging with Multiple Unified CM Servers	13-21
Considerations for Distributed PIMG Deployments	13-22

CHAPTER 14

Cisco Unified MeetingPlace Integration	14-1
Cisco Unified MeetingPlace Components	14-2
Deployment Models	14-3
Single Site	14-3
Multisite WAN with Centralized Call Processing	14-4
Multisite WAN with Distributed Call Processing	14-5
Clustering Over the WAN	14-6
DMZ Deployment	14-7
Call Admission Control, QoS, and Bandwidth	14-8
Call Admission Control	14-8
QoS Markings	14-8
Bandwidth	14-8
Directory Integration	14-9
H.323 and SIP Direct Integration	14-10
H.323	14-10
SIP	14-10
Gatekeeper Integration	14-11
Dial Plan	14-12
Unified CM Dial Plan	14-12
Reservationless Single Number Access (RSNA)	14-12
Redundancy and Load Balancing	14-12
Cisco Unified MeetingPlace Audio Server	14-13
Cisco Unified MeetingPlace H.323/SIP IP Gateway	14-14
Cisco Unified MeetingPlace Web	14-15
Unified CM	14-16
Cisco Unified MeetingPlace Video	14-16
Video Administration	14-17
Video Deployment Models	14-18
Video Call Admission Control, QoS, and Bandwidth	14-19

Video Call Admission Control	14-19
Video QoS	14-20
Video Bandwidth	14-20
Video Gatekeeper Integration	14-20
H.323 Video Endpoints and Video Gatekeeper	14-21
Via-Zone Video Gatekeeper	14-21
Video Administration and Video Gatekeeper Overview	14-22
Video Administration and Video Gatekeeper Integration with Unified CM	14-23
Additional Video Gatekeeper Configuration Requirements	14-23
Zones and RASAggregator Redundancy Groups with H.323 Video Endpoints on Unified CM	14-24
Video Call Flows	14-27
Inbound Call from an H.323 Video Endpoint	14-27
Outbound Call to an H.323 Video Endpoint	14-28
Video Dial Plan	14-29
Fixed-Length Meeting IDs	14-29
Variable-Length Meeting IDs	14-29
Video Redundancy	14-30
Video Integration Component	14-30
Video Administration	14-30
Video Gatekeeper	14-31
Unified CM	14-31

CHAPTER 15

Cisco Unified MeetingPlace Express 15-1

Deployment Models	15-1
Single Site	15-2
Multisite WAN with Centralized Call Processing	15-3
Multisite WAN with Distributed Call Processing	15-4
Clustering Over the WAN	15-5
Security and DMZ Deployment	15-5
Call Admission Control, Bandwidth, and QoS	15-7
Call Admission Control	15-7
Bandwidth Considerations for Web Applications and Screen Sharing	15-7
QoS	15-8
Directory Integration	15-9
H.323 and SIP Direct Integration	15-9
H.323 Gateway	15-9
SIP Trunk	15-10
Gatekeeper Integration	15-11

Design Considerations	15-11
Redundancy	15-12
Cisco Unified MeetingPlace Express Server Redundancy	15-12
Physical Network Redundancy	15-12
Redundancy Using a Gatekeeper	15-12
Redundancy Using H.323 Gateway Integration	15-13
Redundancy Using SIP Trunk Integration	15-13
Other Important Design Considerations	15-14

CHAPTER 16

IP Video Telephony 16-1

IP Video Telephony Solution Components	16-1
Video Enhancements in Cisco Unified CM 5.x	16-2
Protocols	16-2
MTP and Transcoder Support for Video Calls	16-4
Topology-Aware Locations	16-4
Administration Considerations	16-5
Regions	16-5
Locations	16-8
Retry Video Call as Audio	16-8
Wait for Far-End to Send TCS	16-10
Multipoint Conferencing	16-13
SCCP MCU Resources	16-16
Media Resource Groups and Lists	16-17
H.323 and SIP MCU Resources	16-18
Sizing the MCU	16-20
IVR for Dial-In Conference	16-21
Gatekeepers	16-22
Supported Gatekeeper Platforms	16-25
Endpoint Gatekeepers	16-26
Provisioning H.323 Clients	16-26
Provisioning H.323 MCUs	16-31
Provisioning H.320 Gateways	16-33
Gatekeeper Zone Configuration	16-34
Summary of Endpoint Gatekeepers	16-41
Migrating from Cisco Unified CM 4.0	16-44
Applications	16-45
CTI Applications	16-45
Cisco Emergency Responder	16-45
Cisco Unified Communications Manager Assistant	16-46

Cisco Unified IP Interactive Voice Response and Cisco Unified Contact Center	16-46
Cisco Attendant Console	16-46
Cisco Personal Assistant	16-47
Cisco IP SoftPhone and Cisco IP Communicator	16-47
Collaboration Solutions	16-47
T.120 Application Sharing	16-47
Cisco Unified MeetingPlace	16-47
Wireless Networking Solutions	16-48
Cisco Aironet 802.11b Networking Solutions	16-48
Cisco Unified Wireless IP Phone 7920	16-48
XML Services	16-48

CHAPTER 17

LDAP Directory Integration 17-1

What's New in This Chapter	17-2
What is Directory Integration?	17-2
Directory Access for IP Telephony Endpoints	17-4
Directory Integration with Unified CM 5.x	17-6
Comparison with the Unified CM 4.x Approach	17-7
Unified CM 5.x Directory Architecture	17-8
LDAP Synchronization	17-11
Synchronization Mechanism	17-14
Security Considerations	17-16
Best Practices for LDAP Synchronization	17-17
Additional Considerations for Microsoft Active Directory	17-17
LDAP Authentication	17-19
Additional Considerations for Microsoft Active Directory	17-22
User Filtering for Directory Synchronization and Authentication	17-24
Optimizing Unified CM Database Synchronization	17-24
Using the LDAP Structure to Control Synchronization	17-25

CHAPTER 18

IP Telephony Migration Options 18-1

Phased Migration	18-1
Parallel Cutover	18-2
The Need for QSIG in Multisite Enterprises	18-3
Summary	18-5

CHAPTER 19

Voice Security 19-1

What's New in This Chapter	19-1

General Security	19-2
Security Policy	19-2
Security in Layers	19-3
Secure Infrastructure	19-4
Physical Security	19-4
IP Addressing	19-5
Phone Security	19-5
PC Port on the Phone	19-6
Gratuitous ARP	19-7
PC Voice VLAN Access	19-8
Web Access	19-9
Video Capabilities	19-9
Settings Access	19-10
Phone Authentication and Encryption	19-11
Access Security	19-12
Voice and Video VLANs	19-12
Switch Port	19-13
Port Security: MAC CAM Flooding	19-13
Port Security: Prevent Port Access	19-14
Port Security: Prevent Rogue Network Extensions	19-15
DHCP Snooping: Prevent Rogue DHCP Server Attacks	19-16
DHCP Snooping: Prevent DHCP Starvation Attacks	19-18
DHCP Snooping: Binding Information	19-19
Requirement for Dynamic ARP Inspection	19-20
Quality of Service	19-24
Access Control Lists	19-24
VLAN Access Control Lists	19-24
Router Access Control Lists	19-26
Gateways and Media Resources	19-28
Putting Firewalls Around Gateways	19-29
Firewalls and H.323	19-31
Firewalls	19-31
Routed ASA and PIX	19-34
Transparent ASA and PIX	19-34
ASA and PIX Configuration Example	19-35
FWSM Routed Mode	19-37
FWSM Transparent Mode	19-37
FWSM Configuration Example	19-37
Data Center	19-39

Applications Servers	19-39
Cisco Security Agent on the Unified CM and Application Servers	19-39
Unmanaged Cisco Security Agent	19-40
Managed Cisco Security Agent	19-40
Antivirus	19-40
General Server Guidelines	19-40
Deployment Examples	19-41
Lobby Phone Example	19-41
Firewall Deployment Example (Centralized Deployment)	19-44
Conclusion	19-45

CHAPTER 20

IP Telephony Endpoints 20-1

Analog Gateways	20-2
Analog Interface Module	20-2
Low-Density Analog Interface Module	20-2
High-Density Analog Interface Module	20-3
Supported Platforms and Cisco IOS Requirements for Analog Interface Modules	20-4
Cisco Communication Media Module (CMM)	20-5
WS-X6624-FXS Analog Interface Module	20-5
Cisco VG224 Gateway	20-5
Cisco VG248 Gateway	20-5
Cisco ATA 186 and 188	20-6
Cisco Unified IP Phones	20-6
Cisco Basic IP Phones	20-6
Cisco Unified IP Phone 7902G	20-6
Cisco Unified IP Phone 7905G	20-6
Cisco Unified IP Phone 7906G	20-6
Cisco Unified IP Phone 7911G	20-7
Cisco Unified IP Phone 7912G	20-7
Cisco Business IP Phones	20-7
Cisco Unified IP Phone 7940G	20-7
Cisco Unified IP Phone 7941G	20-8
Cisco Unified IP Phone 7941G-GE	20-8
Cisco Manager IP Phones	20-8
Cisco Unified IP Phone 7960G	20-8
Cisco Unified IP Phone 7961G	20-8
Cisco Unified IP Phone 7961G-GE	20-9
Cisco Executive IP Phones	20-9
Cisco Unified IP Phone 7970G	20-9

Cisco 7971G-GE	20-10
Cisco Unified IP Phone Expansion Module 7914	20-10
Software-Based Endpoints	20-10
Cisco Unified Personal Communicator	20-10
Maximum Cisco Unified Personal Communicator Configuration Limits	20-11
Codec Selection	20-11
Call Admission Control	20-11
Cisco IP Communicator	20-11
Maximum IP Communicator Configuration Limits	20-12
Codec Selection	20-12
Call Admission Control	20-13
Wireless Endpoints	20-13
Site Survey	20-14
Authentication	20-14
Capacity	20-15
Phone Configuration	20-16
Roaming	20-16
AP Call Admission Control	20-18
Device Mobility and Unified CM	20-18
Cisco IP Conference Station	20-19
Video Endpoints	20-20
SCCP Video Endpoints	20-20
Cisco Unified Video Advantage	20-20
Cisco IP Video Phone 7985G	20-23
Codecs Supported by Cisco Unified Video Advantage and Cisco IP Video Phone 7985G	20-23
Third-Party SCCP Video Endpoints	20-24
Third-Party SIP IP Phones	20-25
QoS Recommendations	20-25
Cisco VG224 and VG248	20-25
Cisco ATA 186 and IP Conference Station	20-26
Cisco ATA 188 and IP Phones	20-26
Software-Based Endpoints	20-31
Cisco Unified Wireless IP Phones	20-33
Video Telephony Endpoints	20-34
Cisco Unified Video Advantage with a Cisco Unified IP Phone	20-35
Cisco IP Video Phone 7985G	20-35
Sony and Tandberg SCCP Endpoints	20-35
H.323 and SIP Video Endpoints	20-36
Endpoint Features Summary	20-37

CHAPTER 21**Cisco Unified Presence 21-1**

Presence 21-1

Cisco Unified Presence Components 21-2

Cisco Unified Presence User 21-3

Unified CM Presence 21-4

Unified CM Presence with SIP 21-4

Unified CM Presence with SCCP 21-5

Unified CM Speed Dial Presence 21-5

Unified CM Call History Presence 21-6

Unified CM Presence Policy 21-7

Unified CM Subscribe Calling Search Space 21-7

Unified CM Presence Groups 21-7

Unified CM Presence Guidelines 21-8

Cisco Unified Presence Server 21-8

Cisco Unified Presence Server Cluster 21-9

Cisco Unified Presence Server Redundancy 21-10

Cisco Unified Presence Server Deployment Models 21-10

Cisco Unified Presence Server Performance 21-11

Cisco Unified Presence Server Licensing 21-11

Cisco Unified Presence Server Deployment 21-12

Cisco Unified Presence Server Guidelines 21-14

Cisco IP Phone Messenger Application 21-15

Cisco IP Phone Messenger Bandwidth Considerations 21-18

Cisco Unified Personal Communicator 21-19

Cisco Unified Personal Communicator Deployment 21-19

Design Considerations for Cisco Unified Personal Communicator 21-22

Third-Party Presence Server Integration 21-23

Microsoft Live Communications Server 2005 21-23

IBM Sametime 7.5 21-23

CHAPTER 22**Cisco Unified CM Applications 22-1**

IP Phone Services 22-2

IP Phone Services Phone Support 22-2

Unified CM Services and IP Phone Service Enterprise Service Parameters 22-2

Unified CM Services for IP Phone Services 22-2

IP Phone Service Enterprise Service Parameters 22-3

IP Phone Services Architecture 22-3

IP Phone Services Redundancy 22-6

IP Phone Services Scalability 22-6

Guidelines and Restrictions for IP Phone Services	22-7
Extension Mobility (EM)	22-7
EM Phone Support	22-7
Unified CM Services and EM Service Parameters	22-8
Unified CM Services for EM	22-8
EM Service Parameters	22-8
EM Architecture	22-9
EM Redundancy	22-11
Guidelines and Restrictions for EM	22-12
EM Performance and Capacity	22-13
EM Interactions: Unified CM Assistant, AC, and WebDialer	22-13
Unified CM Assistant	22-13
Unified CM Assistant Phone Support	22-13
Unified CM Services and Unified CM Assistant Service Parameters	22-14
Unified CM Services for Unified CM Assistant	22-14
Unified CM Assistant Service Parameters	22-14
Unified CM Assistant Functionality and Architecture	22-16
Unified CM Assistant Proxy Line Mode	22-16
Unified CM Assistant Share Lined Mode	22-17
Unified CM Assistant Architecture	22-17
Unified CM Assistant Dial Plan Considerations	22-19
Unified CM Assistant Console	22-23
Unified CM Assistant Console Installation	22-23
Unified CM Assistant Desktop Console QoS	22-23
Unified CM Assistant Console Directory Window	22-23
Unified CM Assistant Phone Console QoS	22-24
Unified CM Assistant Redundancy	22-24
Service and Component Redundancy	22-25
Device and Reachability Redundancy	22-26
Guidelines and Restrictions for Unified CM Assistant	22-27
Unified CM Assistant Performance and Capacity	22-28
Unified CM Assistant Interactions with EM	22-28
Attendant Console	22-28
AC Phone Support	22-29
Unified CM Services and AC Service Parameters	22-29
Unified CM Services for AC	22-29
AC Service Parameters	22-29
AC and AC Device Authentication Application Users	22-30
AC Functionality and Architecture	22-31

AC Architecture	22-32
Attendant Console Desktop Application	22-33
Attendant Console Installation	22-33
Attendant Console QoS	22-34
Attendant Console Directory Window	22-34
AC Redundancy	22-36
Service and Component Redundancy	22-36
Device and Reachability Redundancy	22-37
Guidelines and Restrictions for AC	22-38
AC Performance and Capacity	22-38
AC Interactions with EM	22-39
WebDialer	22-39
WebDialer Phone Support	22-39
Unified CM Services and WebDialer Service Parameters	22-40
Unified CM Services for WebDialer	22-40
WebDialer Service Parameters	22-40
WebDialer Functionality and Architecture	22-41
WebDialer Servlet	22-41
Redirector Servlet	22-42
WebDialer Architecture	22-44
WebDialer URLs	22-45
WebDialer Redundancy	22-47
Service and Component Redundancy	22-47
Device and Reachability Redundancy	22-47
Guidelines and Restrictions for WebDialer	22-48
WebDialer Performance and Scalability	22-48
WebDialer Interactions with EM	22-49

CHAPTER 23

Cisco Mobility Applications 23-1

Mobile Connect	23-1
Mobile Connect Phone Support	23-2
Unified CM Services and Mobility System Parameters	23-3
Unified CM Services for Mobile Connect	23-3
Unified Mobility System Parameters	23-3
Mobile Connect Functionality	23-4
Desk Phone Pickup	23-5
Remote Destination Phone Pickup	23-6
Single Enterprise Voicemail Box	23-7
Mobile Connect Architecture	23-8

Mobile Connect Dial Plan Considerations	23-9
Centralized Gateway Call Routing	23-11
Distributed Gateway Call Routing	23-12
Mobile Connect Redundancy	23-13
Unified Mobility Application Server Redundancy	23-14
Unified CM Server Redundancy	23-14
PSTN Gateway Redundancy	23-15
Mobile Voice Access	23-15
Mobile Voice Access Phone Support	23-15
Cisco Unified CM Services and Mobility System Parameters	23-15
Unified CM Services for Mobile Voice Access	23-15
Unified Mobility Mobile Voice Access System Parameters	23-16
Mobile Voice Access IVR VoiceXML Gateway URL	23-16
Mobile Voice Access Functionality	23-16
Mobile Voice Access Using Hairpinning	23-17
Desk and Remote Destination Phone Pickup	23-19
Enabling and Disabling Mobile Connect	23-19
Mobile Voice Access Architecture	23-20
Mobile Voice Access Dial Plan Considerations	23-20
Mobile Voice Access Redundancy	23-21
Unified Mobility	23-21
Guidelines and Restrictions for Unified Mobility	23-21
Unified Mobility Performance and Capacity	23-22

APPENDIX A

Recommended Hardware and Software Combinations A-1

GLOSSARY

INDEX



Preface

Last revised on: October 30, 2008

This document provides design considerations and guidelines for deploying a Cisco Unified Communications System based on Cisco Unified Communications Manager 5.x (release 5.0 and subsequent 5.x releases).

This document focuses on the following components of the Cisco Unified Communications System:

- Cisco Unified Communications Manager (Unified CM), formerly Cisco Unified CallManager
- Cisco Unity and Cisco Unity Connection
- Cisco Unified MeetingPlace and Cisco Unified MeetingPlace Express
- Cisco Unified Video Advantage
- Cisco Unified Presence
- Cisco Unified Mobility

This document should be used in conjunction with other documentation available at the following locations:

- For more information about the Cisco Unified Communications System:
<http://www.cisco.com/go/unified-techinfo>
<http://www.cisco.com>
- For more information about Cisco Unified Communications Manager:
http://www.cisco.com/en/US/products/sw/voicesw/ps556/tsd_products_support_series_home.html
<http://www.cisco.com>
- For other Solution Reference Network Design (SRND) documents:
<http://www.cisco.com/go/designzone>

New or Changed Information for This Release



Note

Unless stated otherwise, the information in this document applies to Cisco Unified Communications Manager 5.x (release 5.0 and subsequent 5.x releases). Any differences between the various releases of Cisco Unified Communications Manager are specifically noted in the text.

The following chapters are either new in the current release of this document, or they contain information that has changed significantly from previous releases of this document.

- [Network Infrastructure, page 3-1](#)
- [Gateways, page 4-1](#)
- [Cisco Unified CM Trunks, page 5-1](#)
- [Media Resources, page 6-1](#)
- [LDAP Directory Integration, page 17-1](#)
- [Voice Security, page 19-1](#)

Within each chapter, new and revised information is listed in a section titled *What's New in This Chapter*.

Revision History

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

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

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
October 30, 2008	Updated content as indicated in New or Changed Information for This Release, page xxvi .
September 27, 2007	Updated content for several topics.
August, 2007	Updated the document to change the product name from Cisco Unified CallManager to Cisco Unified Communications Manager.
February, 2007	Added a new chapter on Mobility and made other updates to existing chapters.
November, 2006	Modified and updated this document to apply to Cisco Unified Communications Manager releases 5.0 and 5.1.
August, 2006	Added sections on Cisco Unified Presence, Cisco Unified Personal Communicator, and Cisco Unified MeetingPlace Express.

Revision Date	Comments
April, 2006	Updated the text to reflect new product names and also fixed a few typographical errors.
March, 2006	Initial version of this document for Cisco Unified Communications Manager Release 5.0.

Obtaining Documentation, Obtaining Support, and Security Guidelines

For information on obtaining documentation, obtaining support, providing documentation feedback, security guidelines, and also recommended aliases and general Cisco documents, see the monthly *What's New in Cisco Product Documentation*, which also lists all new and revised Cisco technical documentation, at:

<http://www.cisco.com/en/US/docs/general/whatsnew/whatsnew.html>



CHAPTER 1

Introduction

Last revised on: September 27, 2007

The Cisco Unified Communications System delivers 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, the Cisco Unified Communications System delivers unparalleled performance and capabilities to address current and emerging communications needs in the enterprise environment. The Cisco Unified Communications family of products is designed to optimize feature functionality, reduce configuration and maintenance requirements, and provide interoperability with a wide variety of other applications. The Cisco Unified Communications System provides this capability while maintaining a high level of availability, quality of service (QoS), and security for your network.

The Cisco Unified Communications System incorporates and integrates the following major communications technologies:

- **IP telephony**
IP telephony refers to technology that transmits voice communications over a network using IP standards. Cisco Unified Communications includes a wide array of hardware and software products such as call processing agents, IP phones (both wired and wireless), voice messaging systems, video devices, and many special applications.
- **Customer contact center**
Cisco IP Contact Center products 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. They include access to a large pool of agents and multiple channels of communication as well as customer self-help tools.
- **Video telephony**
The Cisco Unified Video Advantage products enable real-time video communications and collaboration using the same IP network and call processing agent as Cisco Unified Communications. With Cisco Unified Video Advantage, making a video call is now as easy as dialing a phone number.
- **Rich-media conferencing**
Cisco Conference Connection and Cisco Unified MeetingPlace enhance the virtual meeting environment with a integrated set of IP-based tools for voice, video, and web conferencing.

- Third-party applications

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

The remainder of this document focuses on system design considerations for the following components of the Cisco Unified Communications System:

- Cisco Unified Communications Manager (Unified CM), formerly Cisco Unified CallManager
- Cisco Unified Video Advantage
- Cisco Unified MeetingPlace

For information about other aspects of the Cisco Unified Communications System, such as Cisco IP Contact Center, refer to the documentation available at the following locations:

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

<http://www.cisco.com/go/unified-techinfo>

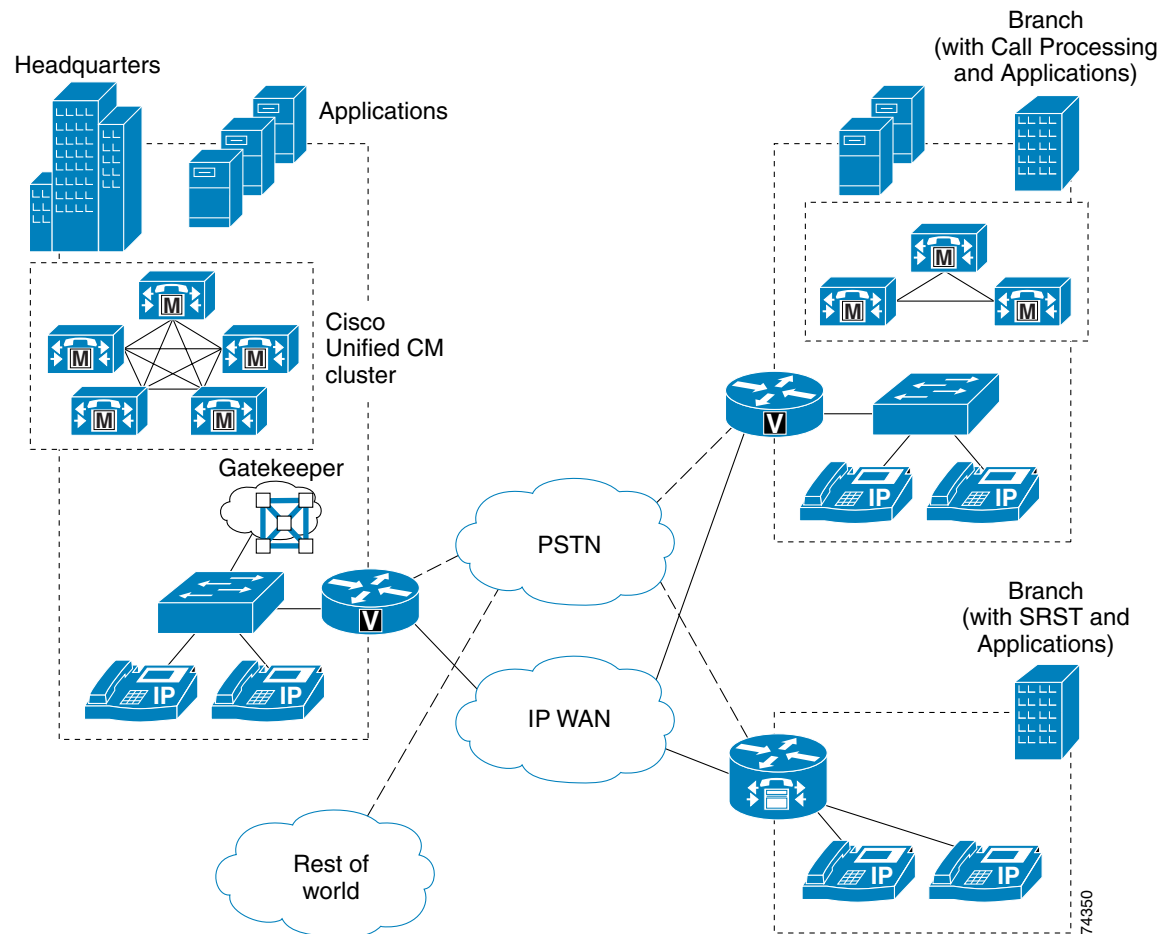
You can also find additional documentation for the Cisco Unified Communications family of products at the following location:

<http://www.cisco.com>

Overview of Cisco Unified Communications

Cisco IP Telephony is the leading converged network IP 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 deployment employing the Cisco IP network infrastructure, with Cisco Unified Communications Manager (Unified CM) as the call processing agent.

Figure 1-1 Typical IP Telephony Deployment

The foundation architecture for Cisco IP Telephony includes the following major components (see Figure 1-1):

- [Cisco IP Network Infrastructure, page 1-3](#)
- [Quality of Service, page 1-4](#)
- [Call Processing Agent, page 1-4](#)
- [Communication Endpoints, page 1-5](#)
- [Conferencing, Messaging, and Collaboration Capabilities, page 1-6](#)
- [Applications, page 1-6](#)
- [Security, 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.

For more information about the IP network infrastructure, see the chapter on [Network Infrastructure](#), page 3-1.

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, Cisco Unified Communications includes Quality of Service (QoS) features such as traffic classification, queuing, traffic shaping, compressed Real-Time Transport Protocol (cRTP), and Transmission Control Protocol (TCP) header compression.

The QoS components for Cisco Unified Communications are provided through the rich IP traffic management, queuing, 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 (bandwidth allocation)

For more information about QoS, see the various sections on QoS in the chapter on [Network Infrastructure](#), page 3-1.

Call Processing Agent

Cisco Unified Communications Manager (Unified CM) is the core call processing software for Cisco IP Telephony. It builds call processing capabilities on top of the Cisco IP network infrastructure. Unified CM software extends enterprise telephony features and capabilities to packet telephony network devices such as IP phones, media processing devices, voice gateways, and multimedia applications.

You can deploy the call processing capabilities of Unified CM 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 Unified CM or cluster of Unified CMs 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).

- Multisite WAN model with centralized call processing

In the multisite WAN model with centralized call processing, the Unified CM 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). 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.

- Multisite WAN model with distributed call processing

In the multisite WAN model with distributed call processing, each site has its own Unified CM 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 Unified CM 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.

For information on how to apply these deployment models in designing your Cisco IP Telephony network, see [Deployment Models, page 2-1](#).

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 Unified IP Phones, IP Telephony endpoints include the following devices:

- Software-based endpoints

Cisco IP Communicator and Cisco Unified Personal 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 IP Communicator is a software-based application that delivers enhanced telephony support through the PC. It is designed to meet diverse customer needs by serving as a supplemental telephone when traveling, as a telecommuting device, or as a primary desktop telephone. Cisco Unified Personal Communicator integrates a wide variety of communications applications and services into a single desktop computer application to provide quick and easy access to powerful communications tools for voice, video, web conferencing, call management, directories, and presence information.

- Video telephony endpoints

Video telephony capability is now fully integrated with Cisco Unified CM. In addition, Cisco Unified Video Advantage brings video telephony functionality to Cisco Unified IP Phones and the Cisco IP Communicator softphone application. The video telephony solution consists of a Windows-based application and USB camera. Users make calls from their Cisco Unified IP Phones using the familiar phone interface, and calls are displayed with video on their PCs without requiring any extra button pushes or mouse clicks.

- Wireless endpoints

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

For more information about the various types of endpoints, see [IP Telephony Endpoints, page 20-1](#).

Conferencing, Messaging, and Collaboration Capabilities

Cisco Unified Communications supports the following additional features and applications to provide conferencing, voice messaging, and multi-media collaboration functionality:

- Conferencing

Cisco Unified CM can interface with a number of other Cisco software and hardware devices to provide complete conferencing capability, including annunciator functions and music on hold. For more information about designing and provisioning conference capabilities in Unified CM, see [Media Resources, page 6-1](#).

- Voice messaging

Cisco Unified CM is capable of integrating with third-party voicemail systems and with Cisco Unity and Unity Connection to provide a full range of voicemail and voice messaging functionality. For more information about integrating third-party voicemail systems with Unified CM, see [Third-Party Voicemail Design, page 12-1](#). For more information about integrating Cisco Unity and Unity Connection with Unified CM, see [Cisco Unity, page 13-1](#).

- Video telephony

Video telephony capability is now fully integrated into Cisco Unified CM, and there are also new video endpoints available from Cisco and its strategic partners. Video calls and conferences are now just as easy to make as a voice call on an IP phone. For more information about video capability with Unified CM, see [IP Video Telephony, page 16-1](#).

- Multi-media collaboration

Cisco Unified MeetingPlace is a complete rich-media conferencing application that integrates voice, video, and web conferencing capabilities to make remote meetings as natural and effective as face-to-face meetings. For more information about integrating MeetingPlace with Cisco Unified CM, see [Cisco Unified MeetingPlace Integration, page 14-1](#).

Applications

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

- IP Phone Services

Cisco Unified IP Phone Services are applications that utilize the web client and/or server and XML capabilities of the Cisco Unified IP Phone. The Cisco Unified IP Phone firmware contains a micro-browser that enables limited web browsing capability. These phone service applications provide the potential for value-added services and productivity enhancement by running directly on the user's desktop phone.

- Extension Mobility

The Cisco Extension Mobility (EM) feature enables users to configure a Cisco Unified IP Phone as their own, on a temporary basis, by logging in to that phone. After a user logs in, the phone adopts the user's individual device profile information, including line numbers, speed dials, services links, and other user-specific properties of a phone. The EM feature dynamically configures a phone according to the authenticated user's device profile. The benefit of this application is that it allows users to be reached at their own extension from any phone within the Cisco Unified CM cluster, regardless of physical location, provided the phone supports EM.

- Cisco Unified Communications Manager Assistant

Cisco Unified CM Assistant is a Unified CM integrated application that enables assistants to handle incoming calls on behalf of one or more managers. With the use of the Unified CM Assistant Console desktop application, assistants can quickly determine a manager's status and decide what to do with a call. Assistants can manipulate calls using their phone's softkeys or via the PC interface with either keyboard shortcuts, drop-down menus, or by dragging and dropping calls to the managers' proxy lines.

- Attendant Console

The Cisco Unified Communications Manager Attendant Console (AC) application enables a receptionist to answer and transfer or dispatch calls within an organization. The attendant can install the attendant console, a client/server Java application, on a PC that runs Windows 2000 or Windows XP. The attendant console connects to the Cisco Unified CM Attendant Console Server for login services, line state, and directory services. Multiple attendant consoles can connect to a single AC server.

- WebDialer

Cisco WebDialer is a click-to-dial application for Cisco Unified CM that enables users to place calls easily from their PCs using any supported phone device. There is no requirement for administrators to manage CTI links or build JTAPI or TAPI applications because Cisco WebDialer provides a simplified web application and HTTP or Simple Objects Access Protocol (SOAP) interface for those who want to provide their own user interface and authentication mechanisms. Either way, the solution scales to support the entire Unified CM cluster with full redundancy.

For more information about these applications, see [Cisco Unified CM Applications, page 22-1](#).

In addition to the preceding applications for enterprise telephony users, Cisco mobility applications provide the ability to deliver features and functionality of the enterprise IP Communications environment to mobile workers wherever they might be. Mobility functionality delivered within the Cisco Unified Communications solution is provided via the Cisco Unified Mobility application server. The Unified Mobility server communicates with Cisco Unified CM via the Java Telephony Application Programming Interface (JTAPI) and AVVID XML Layer (AXL) and provides the following mobility application functionality:

- Mobile Connect

Mobile Connect or Single Number Reach provides Cisco Unified Communications users with the ability to be reached via a single enterprise phone number that will ring on both their IP desk phone and their cellular phone simultaneously. Mobile Connect users can pick up an incoming call on either their desk or cellular phones and at any point can move the in-progress call from one of these phones to the other without interruption.

- Mobile Voice Access

Mobile Voice Access or Enterprise Dial tone provides mobile users with the ability to make calls from their mobile phone as if they were calling from their enterprise IP desk phone. The feature provides a cost savings in terms of toll charges for long distance or international calls, and it provides the enterprise with an easy way to track phone calls made by employees via a uniform and centrally located set of call detail records.

For more information on Cisco Unified Mobility, Mobile Connect, and Mobile Voice Access, see [Cisco Mobility Applications, page 23-1](#).

Security

Security for a Cisco Unified Communications deployment involves the following main considerations, 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 Unified CM, 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

For more information about security for IP Telephony networks, see [Voice Security, page 19-1](#).



CHAPTER 2

Deployment Models

Last revised on: September 27, 2007

This chapter describes the deployment models for Cisco Unified Communications Manager 5.x. For design guidance with earlier releases of Cisco Unified Communications Manager (Unified CM), refer to the Unified Communications Solution Reference Network Design (SRND) documentation available at

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

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

- [Single Site, page 2-2](#)

The single-site model 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.

- [Multisite WAN with Centralized Call Processing, page 2-4](#)

The multisite 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.

- [Multisite WAN with Distributed Call Processing, page 2-14](#)

The multisite 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 2-17](#)

This model deploys a single Unified CM 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 2-27](#), 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/en/US/products/sw/voicesw/ps556/products_device_support_tables_list.html

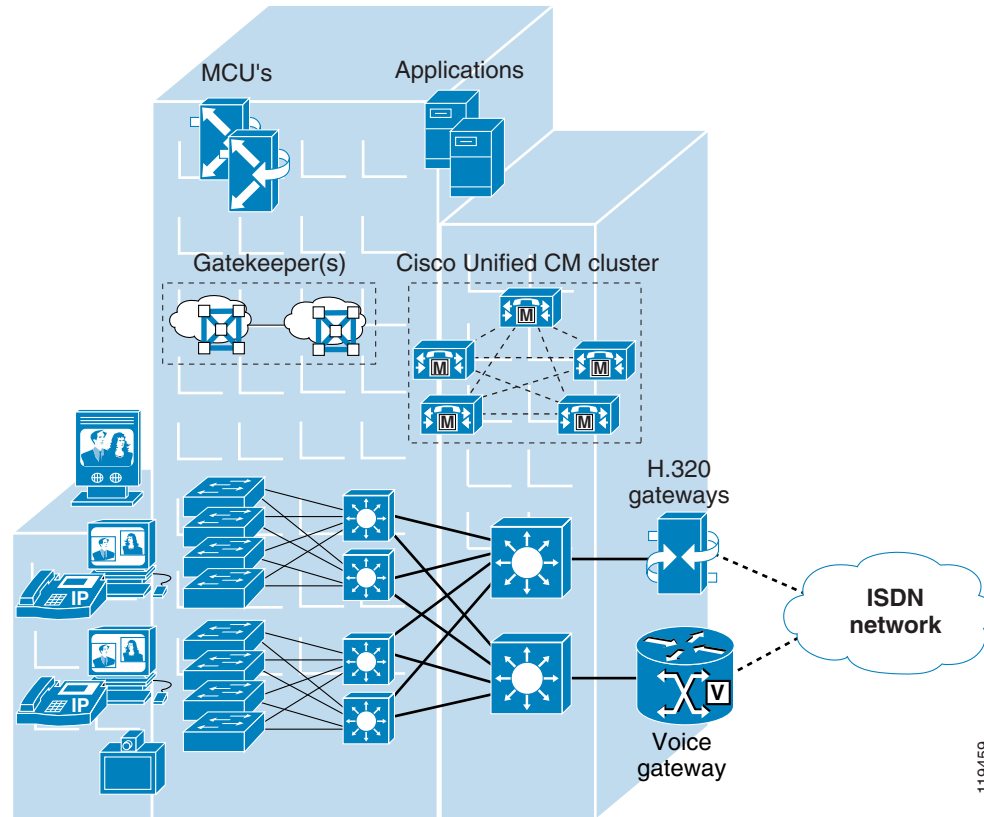
Single Site

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 Unified CM or Unified CM cluster.
- Maximum of 30,000 Skinny Client Control Protocol (SCCP) or Session Initiation Protocol (SIP) IP phones or SCCP video endpoints per cluster.
- Maximum of 500 H.323 devices (gateways, MCUs, trunks, and clients) per Unified CM cluster.
- PSTN for all external calls.
- Digital signal processor (DSP) resources for conferencing, transcoding, and media termination point (MTP).
- Voicemail or unified messaging components.
- Capability to integrate with legacy private branch exchange (PBX) and voicemail systems.
- H.323 clients, MCUs, and H.323/H.320 gateways that require a gatekeeper to place calls must register with a Cisco IOS Gatekeeper (Cisco IOS Release 12.3(8)T or greater). Unified CM then uses an H.323 trunk to integrate with the gatekeeper and provide call routing and bandwidth management services for the H.323 devices registered to it. Multiple Cisco IOS Gatekeepers may be used to provide redundancy.
- MCU resources are required for multipoint video conferencing. Depending on conferencing requirements, these resources may be either SCCP or H.323, or both.
- H.323/H.320 video gateways are needed to communicate with H.320 videoconferencing devices on the public ISDN network.
- High-bandwidth audio (for example, G.711, G.722, or Cisco Wideband Audio) between devices within the site.
- High-bandwidth video (for example, 384 kbps or greater) between devices within the site. The Cisco Unified Video Advantage Wideband Codec, operating at 7 Mbps, is also supported.

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

Figure 2-1 Single-Site Deployment Model

119459

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:

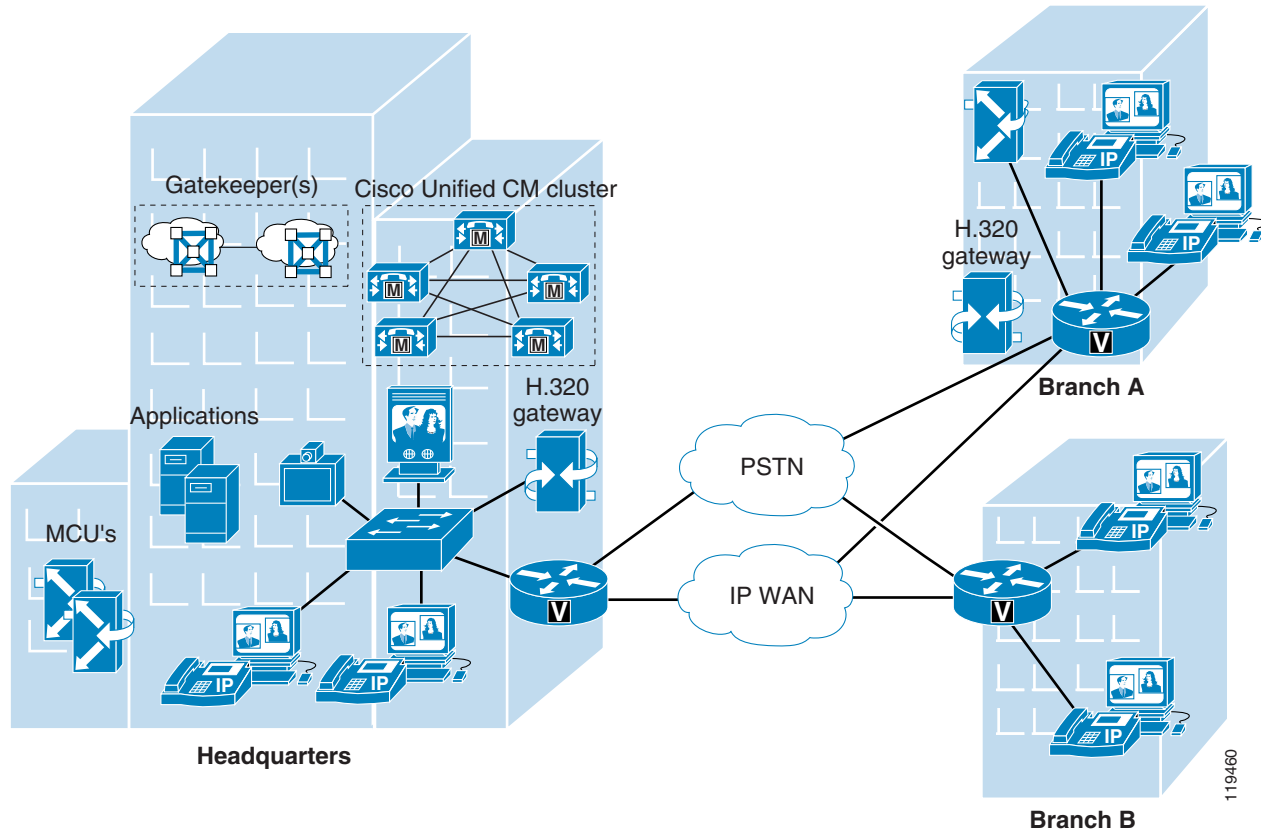
- Provide a highly available, fault-tolerant infrastructure based on a common infrastructure philosophy. A sound infrastructure is essential for easier migration to IP telephony, integration with applications such as video streaming and video conferencing, and expansion of your IP telephony deployment across the WAN or to multiple Unified CM 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 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](#).

Multisite WAN with Centralized Call Processing

The multisite 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 Unified CM 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 Unified CM 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 Multisite Deployment Model with Centralized Call Processing

The multisite model with centralized call processing has the following design characteristics:

- Single Unified CM or Unified CM cluster.
- Maximum of 30,000 Skinny Client Control Protocol (SCCP) or Session Initiation Protocol (SIP) IP phones or SCCP video endpoints per cluster.
- Maximum of 500 H.323 devices (gateways, MCUs, trunks, and clients) per Unified CM cluster.
- PSTN for all external calls.
- Digital signal processor (DSP) resources for conferencing, transcoding, and media termination point (MTP).
- Voicemail or unified messaging components.
- Capability to integrate with legacy private branch exchange (PBX) and voicemail systems.
- H.323 clients, MCUs, and H.323/H.320 gateways that require a gatekeeper to place calls must register with a Cisco IOS Gatekeeper (Cisco IOS Release 12.3(8)T or greater). Unified CM then uses an H.323 trunk to integrate with the gatekeeper and provide call routing and bandwidth management services for the H.323 devices registered to it. Multiple Cisco IOS Gatekeepers may be used to provide redundancy.
- MCU resources are required for multipoint video conferencing. Depending on conferencing requirements, these resources may be either SCCP or H.323, or both, and may all be located at the central site or may be distributed to the remote sites if local conferencing resources are required.

- H.323/H.320 video gateways are needed to communicate with H.320 videoconferencing devices on the public ISDN network. These gateways may all be located at the central site or may be distributed to the remote sites if local ISDN access is required.
- High-bandwidth audio (for example, G.711, G.722, or Cisco Wideband Audio) between devices in the same site, and low-bandwidth audio (for example, G.729 or G.728) between devices in different sites.
- High-bandwidth video (for example, 384 kbps or greater) between devices in the same site, and low-bandwidth video (for example, 128 kbps) between devices at different sites. The Cisco Unified Video Advantage Wideband Codec, operating at 7 Mbps, is recommended only for calls between devices at the same site.
- Minimum of 768 kbps or greater WAN link speeds. Video is *not* recommended on WAN connections that operate at speeds lower than 768 kbps.
- Unified CM locations provide call admission control, and automated alternate routing (AAR) is also supported for video calls.
- Maximum of 500 locations per Unified CM cluster.
- Survivable Remote Site Telephony (SRST) version 4.0 and higher support video. However, versions of SRST prior to 4.0 do *not* support video, and SCCP video endpoints located at remote sites become audio-only devices if the WAN connection fails.

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 Unified CM provides call admission control. (Refer to the section on [Unified CM Static Locations, page 9-12](#), 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 for SCCP and SIP phones 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 multisite WAN model with centralized call processing:

- Minimize delay between Unified CM and remote locations to reduce voice cut-through delays (also known as clipping).
- Use the locations mechanism in Unified CM 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 Unified CM Release 3.1 and later. This configuration can support a maximum of 30,000 IP phones when Unified CM 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 Unified CM server or running Cisco Unified CM Express on the branch router.

Remote Site Survivability

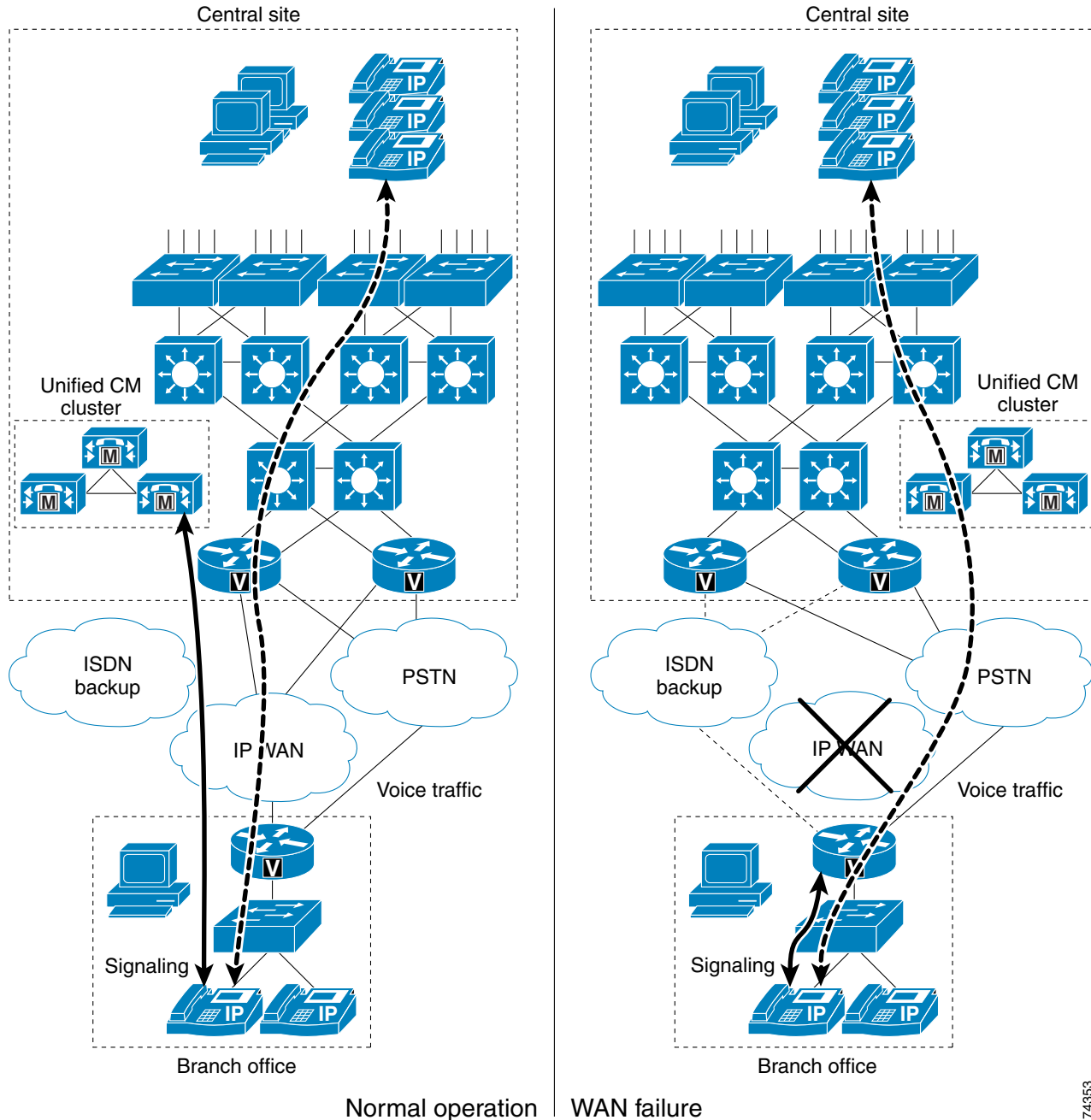
When deploying IP Telephony across a WAN with the centralized call processing model, you should take additional steps to ensure that data and voice services at the remote sites are highly available. [Table 2-1](#) summarizes the different strategies for providing high availability at the remote sites. 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
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 Unified CM 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

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 Unified CM 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 Unified CM 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

74353

automatically. The branch IP phones can then make and receive calls either internally or through the PSTN. The phone displays the message “Unified CM fallback mode,” and some advanced Unified CM 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 Unified CM cluster and resume normal operation. The branch router deletes its information about the IP phones and reverts to its standard routing or gateway configuration.

**Note**

When WAN connectivity to the central site is reestablished, or when Unified CM is reachable again, SRST phones with active calls will not immediately re-register to Unified CM until those active calls are terminated.

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 Unified CM 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 Unified CM 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 Unified CM 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 10-27](#).)
- Using a combination of dial plan constructs in both Unified CM 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 Unified CM 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 coming into the branch's gateway to a destination outside the branch site will result in hairpinning of the call through the gateway, thus using two trunk ports. This behavior applies to:
 - Calls forwarded to a voicemail system located outside the branch
 - Calls forwarded to an on-net abbreviated dialing destination located in a different branch

The gateway port utilization resulting from these call forwarding flows should be taken into account when sizing the trunks connecting the branch to the PSTN.

- 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.

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 2-12](#)
- [VoPSTN Using Dial Plan, page 2-13](#)

VoPSTN Using AAR

This method consists of configuring the Unified CM 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 Unified CM 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.

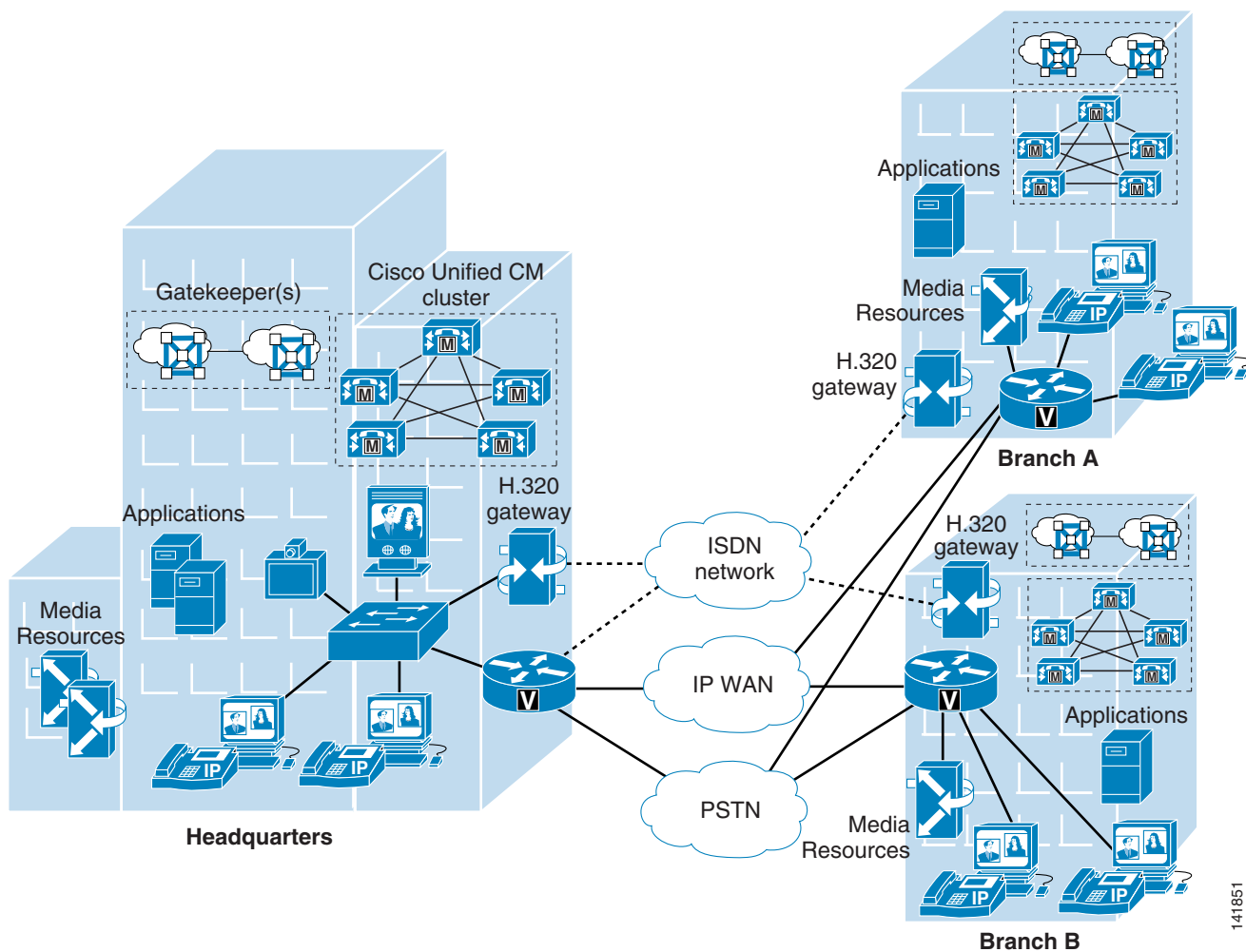
In addition to the general considerations listed for VoPSTN, the following design guidelines apply to the dial plan implementation method:

- 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.

Multisite WAN with Distributed Call Processing

The multisite 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 Multisite Deployment Model with Distributed Call Processing



141851

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 Unified Communications Manager (Unified CM)
 - Cisco Unified Communications Manager Express (Unified CME)
 - Other IP PBX
- A centralized call processing site and all of its associated remote sites
- A legacy PBX with Voice over IP (VoIP) gateway

The multisite model with distributed call processing has the following design characteristics:

- Maximum of 30,000 Skinny Client Control Protocol (SCCP) or Session Initiation Protocol (SIP) IP phones or SCCP video endpoints per cluster.
- Maximum of 500 H.323 devices (gateways, MCUs, trunks, and clients) per Unified CM cluster.
- PSTN for all external calls.
- Digital signal processor (DSP) resources for conferencing, transcoding, and media termination point (MTP).
- Voicemail or unified messaging components.
- Capability to integrate with legacy private branch exchange (PBX) and voicemail systems.
- H.323 clients, MCUs, and H.323/H.320 gateways that require a gatekeeper to place calls must register with a Cisco IOS Gatekeeper (Cisco IOS Release 12.3(8)T or greater). Unified CM then uses an H.323 trunk to integrate with the gatekeeper and provide call routing and bandwidth management services for the H.323 devices registered to it. Multiple Cisco IOS Gatekeepers may be used to provide redundancy. Cisco IOS Gatekeepers may also be used to provide call routing and bandwidth management between the distributed Unified CM clusters. In most situations, Cisco recommends that each Unified CM cluster have its own set of endpoint gatekeepers and that a separate set of gatekeepers be used to manage the intercluster calls. It is possible in some circumstances to use the same set of gatekeepers for both functions, depending on the size of the network, complexity of the dial plan, and so forth. (For details, see [Gatekeepers](#), page 16-22.)
- MCU resources are required in each cluster for multipoint video conferencing. Depending on conferencing requirements, these resources may be either SCCP or H.323, or both, and may all be located at the regional sites or may be distributed to the remote sites of each cluster if local conferencing resources are required.
- H.323/H.320 video gateways are needed to communicate with H.320 videoconferencing devices on the public ISDN network. These gateways may all be located at the regional sites or may be distributed to the remote sites of each cluster if local ISDN access is required.
- High-bandwidth audio (for example, G.711, G.722, or Cisco Wideband Audio) between devices in the same site, but low-bandwidth audio (for example, G.729 or G.728) between devices in different sites.
- High-bandwidth video (for example, 384 kbps or greater) between devices in the same site, but low-bandwidth video (for example, 128 kbps) between devices at different sites. The Cisco Unified Video Advantage Wideband Codec, operating at 7 Mbps, is recommended only for calls between devices at the same site. Note that the Cisco VT Camera Wideband Video Codec is not supported over intercluster trunks.
- Minimum of 768 kbps or greater WAN link speeds. Video is *not* recommended on WAN connections that operate at speeds lower than 768 kbps.
- Call admission control is provided by Unified CM locations for calls between sites controlled by the same Unified CM cluster, and by the Cisco IOS Gatekeeper for calls between Unified CM clusters (that is, intercluster trunks). Automated alternate routing (AAR) is also supported for both intra-cluster and inter-cluster video calls.

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

Gatekeeper or Session Initiation Protocol (SIP) proxy servers are among the key elements in the multisite 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. (See [Gatekeeper Design Considerations, page 8-20](#).)
- 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. Unified CM 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 Unified CM cluster containing two servers can provide one-to-one redundancy, with the backup server being used as a publisher and Trivial File Transfer Protocol (TFTP) server.

The requirement for IP-based applications also greatly affects the choice of call processing agent because only Unified CM 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 Unified Communications Manager Express (Unified CME)	Up to 240 phones	• For small remote sites • Capacity depends on Cisco IOS platform
Cisco Unified CM	50 to 30,000 phones	• Small to large sites, depending on the size of the Unified CM 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 Unified CM 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 2-22](#)

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

- [Remote Failover Deployment Model, page 2-26](#)

Remote failover allows you to deploy the backup servers over the WAN. Using this deployment model, you may have up to eight sites with Unified CM subscribers being backed up by Unified CM 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 Unified CM 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 24 or PHB CS3). 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 2-19](#), 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 Unified CM servers 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 2-20](#). 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 to provide sufficient prioritized bandwidth for all ICCS traffic, especially the priority ICCS traffic. Standard QoS mechanisms must be implemented to avoid congestion and packet loss. If packets are lost due to line errors or other “real world” conditions, the ICCS packet will be retransmitted because it uses the TCP protocol for reliable transmission. The retransmission might result in a call being delayed during setup, disconnect (teardown), or other

supplementary services during the call. Some packet loss conditions could result in a lost call, but this scenario should be no more likely than errors occurring on a T1 or E1, which affect calls via a trunk to the PSTN/ISDN.

- **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 IBM Informix Dynamic Server (IDS) database that provides the main configuration information. The IDS database is replicated from the publisher server to all other servers in the cluster using best-effort. The IDS 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 IDS database configuration.
- Firewall management traffic, which is used to authenticate the subscribers to the publisher to access the publisher's database. The management traffic flows between all servers in a cluster. The management traffic may be 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).
- 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 Unified CM servers. This traffic is marked as priority ICCS traffic and exists between the Unified CM server with the CTI Manager and the Unified CM server with the CTI device.

Failover Between Subscriber Servers

With Unified CM Release 5.x, the device configuration records are cached during the initialization or boot-up time. The effect is that Unified CM might take a longer time to initialize, but any failover or failback for all devices is not affected by the delay in accessing the publisher database.

Unified CM 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 Unified CM 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
- Unified CM User page options requiring the database
- Unified CM 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 Unified CM 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 Unified CM servers, ideally the access switch to which the server is attached. Cisco IOS provides a

extended ping capable to set 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 Unified CM 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 Unified CM 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 Unified CM will tolerate random errors, they should be avoided to avoid impairing the performance of the cluster.

Troubleshooting

If the Unified CM 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 Unified CM 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 Unified CM 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::>
```

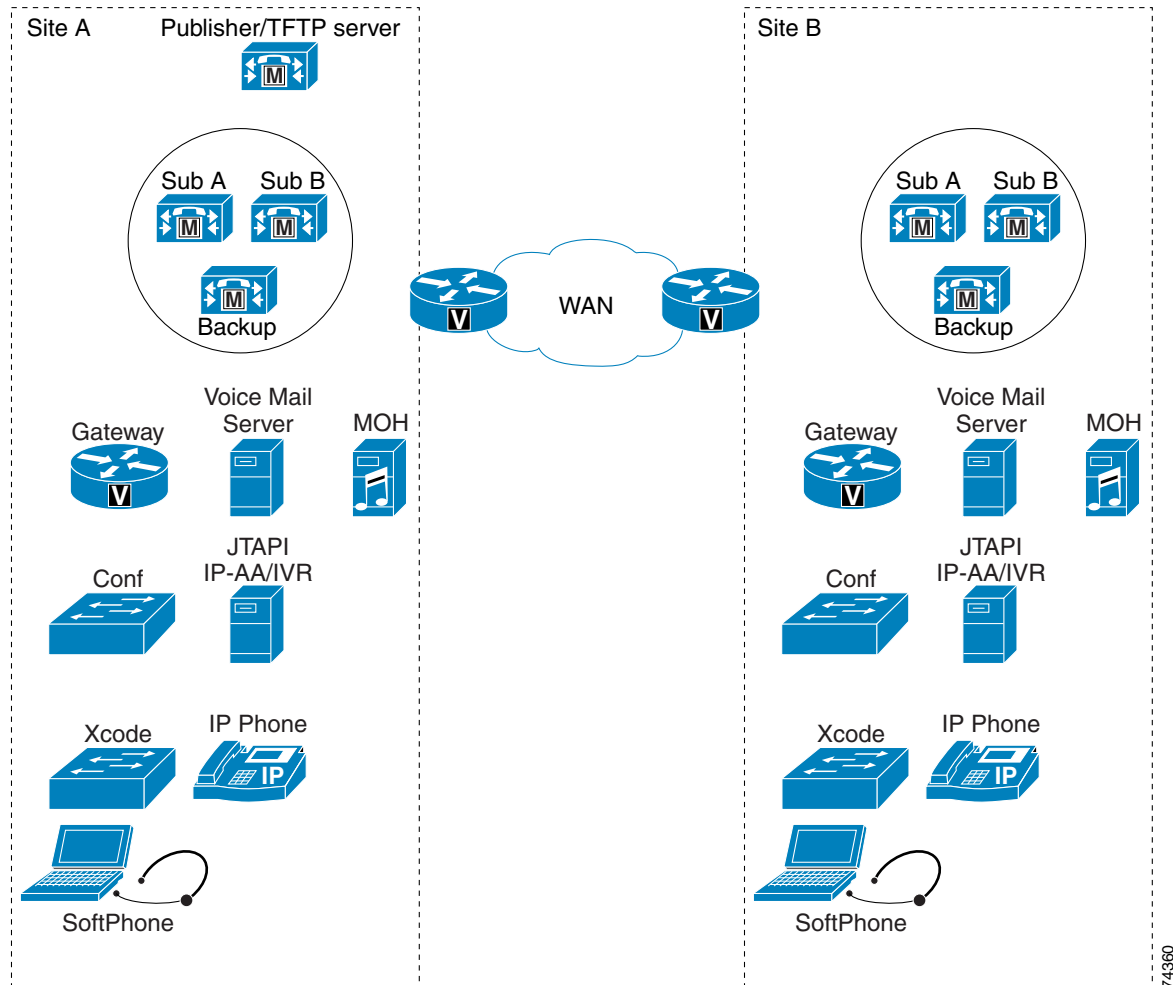
```
|Digit analysis: Remove remote pattern /5000020 , PID:
7:34:1|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>
|Digit analysis: Remove remote pattern /5000066 , PID:
7:34:2|<CLID::VO30-7835-1-Cluster><NID::VO30-7835-2><CT::0,0,0,0.0><IP::><DEV::>
.
.
|Digit analysis: Remove 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 Unified CM 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 Unified CM subscriber and one backup subscriber.
- Configure Unified CM *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 24 or PHB CS3). 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 database and other inter-server traffic.
- A maximum round-trip time (RTT) of 40 ms is allowed between any two servers in the Unified CM 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.
- 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 `Hub_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 Unified CM 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 `Hub_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.

Unified CM Provisioning for Local Failover

Provisioning of the Unified CM 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 Unified CM *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 two Unified CM servers. Cisco also recommends that you do not enable more than one server with the Cisco TFTP service at a location remote to publisher, over a low-speed WAN link. More than two TFTP servers can be deployed in a cluster, however this configuration can result in an extended period for rebuilding all the TFTP files on all TFTP servers.

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 in the DHCP servers at each site. If DHCP is not in use or if the TFTP server is manually configured, you should configure the correct address for the site. During an upgrade of the cluster or a restart of the Cisco TFTP service, the TFTP servers that are remote from the publisher will take longer to upgrade and rebuild all the configuration files. This time depends on the latency between the TFTP server and the publisher and on the quantity of devices configured in the database.

Other services, which may affect normal operation of Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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).

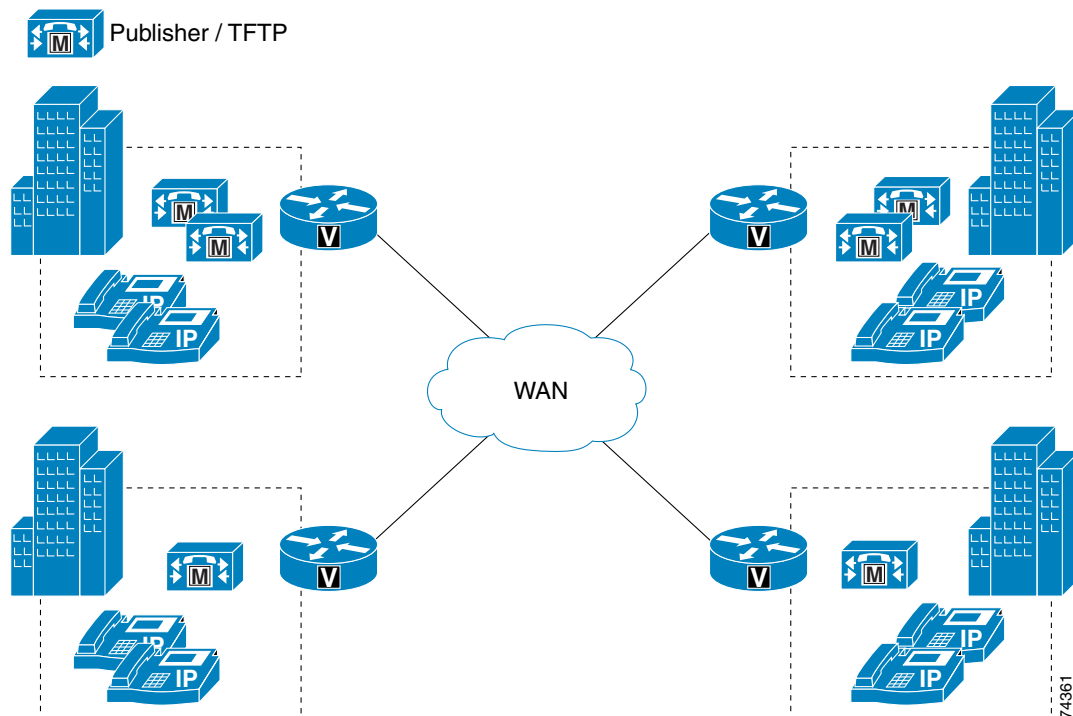
Music on Hold and Media Resources for Local Failover

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 Unified CM 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 2-22](#)), with the following modifications:

- Configure each site to contain at least one primary Unified CM subscriber and an optional backup subscriber as desired.
- You may configure Unified CM *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 Unified CM 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 3-49](#).

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 Unified CM to ensure that the TTY/TDD devices always use G.711 codecs.
- 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 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.



CHAPTER 3

Network Infrastructure

Last revised on: October 30, 2008

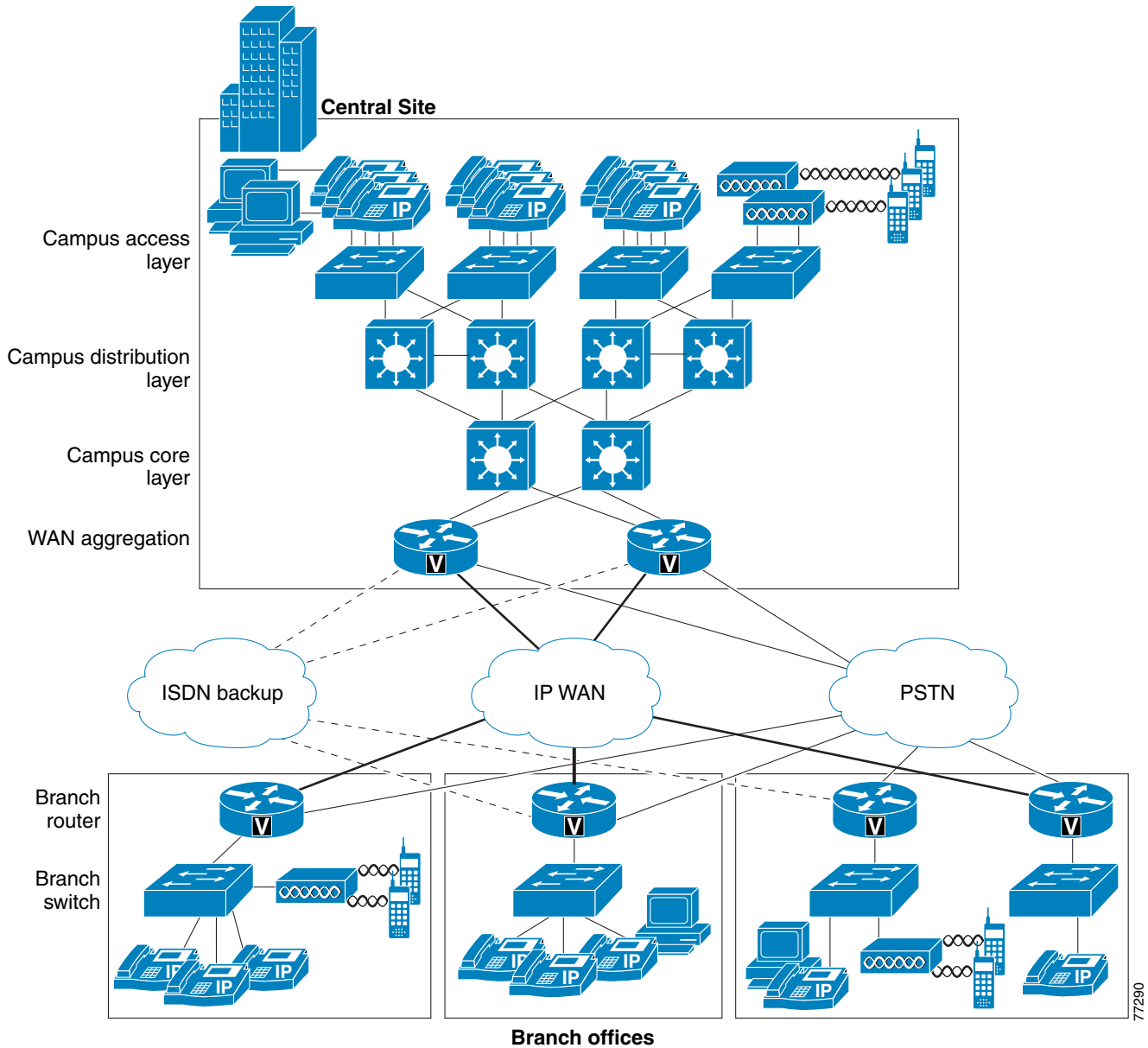
This chapter describes the requirements of the network infrastructure needed to build an IP telephony system in an enterprise environment. [Figure 3-1](#) illustrates the roles of the various devices that form the network infrastructure, and [Table 3-1](#) summarizes the features required to support 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 3-30](#)
- [Wireless LAN Infrastructure, page 3-63](#)

Figure 3-1 Typical Campus Network Infrastructure



77290

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

What's New in This Chapter

[Table 3-2](#) lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 3-2 New or Changed Information Since the Previous Release of This Document

New or Revised Topic	Described in:
Trivial File Transfer Protocol (TFTP)	Trivial File Transfer Protocol (TFTP) , page 3-15

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 3-25](#)

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:

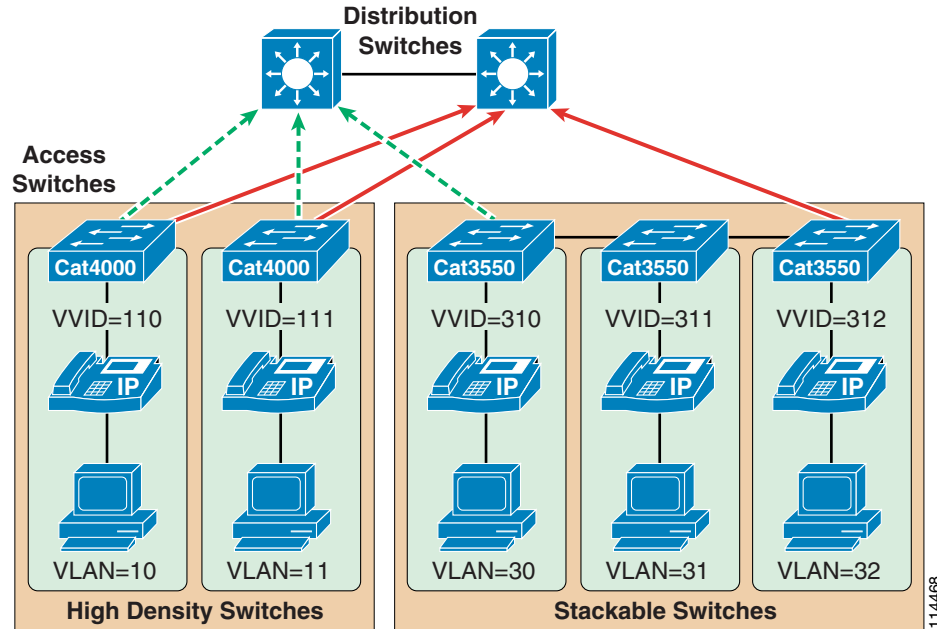
- [Campus Access Layer, page 3-4](#)
- [Campus Distribution Layer, page 3-7](#)
- [Campus Core Layer, page 3-9](#)
- [Network Services, page 3-10](#)

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/cpsa/gcnd_wp.pdf](http://www.cisco.com/warp/public/cc/so/neso/lnso/cpsa/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. 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 Cisco 3560 and 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.

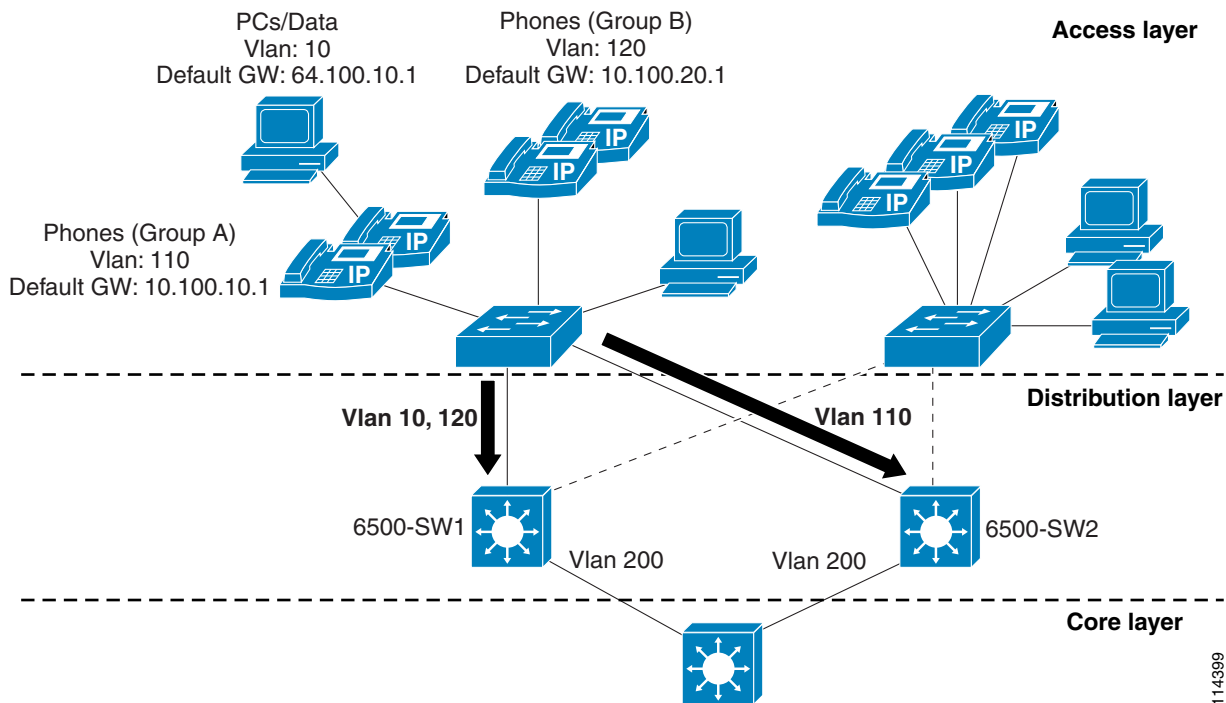
HSRP should also be configured in such a way as to load-balance traffic between both HSRP routers. To provide load balancing, configure each HSRP device as the active HSRP router for one VLAN or interface, and configure the standby router for another VLAN or interface. Evenly distributing active and 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
```

114399

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 Unified Communications Manager (Unified CM) 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 Unified CM 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

The deployment of an IP Communications system requires the coordinated design of a well structured, highly available, and resilient network infrastructure as well as an integrated set of network services including Domain Name System (DNS), Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP), and Network Time Protocol (NTP).

Domain Name System (DNS)

DNS enables the mapping of host names and network services to IP addresses within a network or networks. DNS server(s) deployed within a network provide a database that maps network services to hostnames and, in turn, 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.

Complete reliance on a single network service such as DNS can introduce an element of risk when a critical IP Communications system is deployed. 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, Cisco highly recommends that you do not rely on DNS name resolution for any communications between Unified CM and the IP Communications endpoints. The use of DNS to provide simplified system management and Server (SRV) records is supported, and Cisco recommends defining each Unified CM cluster as a member of a valid sub-domain within the larger organizational DNS domain whenever possible.

Configure Unified CM(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 in endpoint device configurations. During the initial installation of the publisher in a Unified CM cluster, the publisher will be referenced in the server table by the hostname you provided for the system. Before installation and configuration of any subsequent subscribers or the definition of any endpoints, you should change this server entry to the IP address of the publisher rather than the hostname. Each subscriber added to the cluster should be defined in this same server table via IP address initially and not by hostname. Each subscriber should be added to this server table one device at a time, and there should be no definitions for non-existent subscribers at any time other than the new subscriber being installed.

During installation of the publisher and subscriber, Cisco recommend that you do not select the option to enable DNS unless DNS is specifically required for system management purposes. If DNS is enabled, Cisco still highly recommend that you do not use DNS names in the configuration of the IP Communications endpoints, gateways, and Unified CM servers. Even if DNS is enabled on the servers in the cluster, it is never used for any intra-cluster server-to-server communications and is used only for communications to devices external to the cluster itself.

**Note**

Enabling the DNS service is not required for the use of SRV records by Unified CM.

Cisco Unified CM 5.0 does not permit the manual configuration of HOSTS or LHOSTS files. A local version of the HOSTS table is automatically built by the publisher in each cluster and distributed to all subscriber nodes via a secure communications channel. This local table is used for managing secure intra-cluster communications and does not contain addresses or names of any endpoints other than the Unified CM servers themselves. LMHOSTS files do not exist and are not used by Cisco Unified CM 5.0.

There are some situations in which configuring and using DNS might be unavoidable. For example, if Network Address Translation (NAT) is required for communications between the IP phones and Unified CM in the IP Communications 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 geographically redundant fashion so that a single DNS server failure will not prevent network communications 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**

Hostname resolution within the cluster via either the local HOSTS file or DNS queries is performed only at subsystem 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 in either the HOSTS file or 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 obtain initial configuration information, including IP address, subnet mask, default gateway, and TFTP server address. 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 Communications 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 Communications 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 Unified CM 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 would not affect 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 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 Unified 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 Unified CM 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.)

**Note**

The term *co-located* refers to two or more devices in the same physical location, with no WAN or MAN connection between them.

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 multisite Unified CM 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 multisite Unified CM 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 3-3](#) lists the basic Cisco IOS DHCP server configuration commands.

Example 3-3 Cisco IOS DHCP Server Configuration Commands

```
! Activate DHCP Service on the IOS Device

service dhcp

! Specify any IP Address or IP Address Range to be excluded from the DHCP pool

ip dhcp excluded-address <ip-address>|<ip-address-low> <ip-address-high>

! Specify the name of this specific DHCP pool, the subnet and mask for this
! pool, the default gateway and up to four TFTP

ip dhcp pool <dhcp-pool name>
  network <ip-subnet> <mask>
  default-router <default-gateway-ip>
  option 150 ip <tftp-server-ip-1> ...

! Note: IP phones use only the first two addresses supplied in the option 150
! field even if more than two are configured.
```

Unified CM DHCP Sever (Standalone versus Co-Resident DHCP)

Typically DHCP servers are dedicated machine(s) in most network infrastructures, and they run in conjunction with the DNS and Windows Internet Naming Service (WINS) services used by that network. In some instances, given a small Unified CM deployment with no more than 1000 devices registering to the cluster, you may run DHCP on a Unified CM server to support those devices. 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 *not* be run on a Unified CM server but instead must be run on a dedicated or standalone server(s).

**Note**

The term *co-resident* refers to two or more services or applications running on the same server.

Trivial File Transfer Protocol (TFTP)

Within a Cisco Unified CM system, endpoints such as IP phones rely on a TFTP-based process to acquire configuration files, software images, and other endpoint-specific information. The Cisco TFTP service is a file serving system that can run on one or more Unified CM servers. It builds configuration files and serves firmware files, ringer files, device configuration files, and so forth, to endpoints.

The TFTP file systems can hold several file types, such as the following:

- Phone configuration files
- Phone firmware files
- Certificate Trust List (CTL) files
- Tone localization files
- User interface (UI) localization and dictionary files
- Ringer files
- Softkey files
- Dial plan files for SIP phones

The TFTP server manages and serves two types of files, those that are not modifiable (for example, firmware files for phones) and those that can be modified (for example, configuration files).

A typical configuration file contains a prioritized list of Unified CMs for a device (for example, an SCCP or SIP phone), the TCP ports on which the device connects to those Unified CMs, and an executable load identifier. Configuration files for selected devices contain locale information and URLs for the messages, directories, services, and information buttons on the phone.

When a device's configuration changes, the TFTP server rebuilds the configuration files by pulling the relevant information from the Unified CM database. The new file(s) is then downloaded to the phone once the phone has been reset. As an example, if a single phone's configuration file is modified (such as during Extension Mobility login or logout), only that file is rebuilt and downloaded to the phone. However, if the configuration details of a device pool are changed (for example, if the primary Unified CM server is changed), then all devices in that device pool need to have their configuration files rebuilt and downloaded. For device pools that contain large numbers of devices, this file rebuilding process can impact server performance.

When a device requests a configuration file from the TFTP server, the TFTP server searches for the configuration file in its internal caches, the disk, and then alternate Cisco file servers (if specified). If the TFTP server finds the configuration file, it sends it to the device. If the configuration file provides Unified CM names, the device resolves the name by using DNS and opens a connection to the Unified CM. If the device does not receive an IP address or name, it uses the TFTP server name or IP address to attempt a registration connection. If the TFTP server cannot find the configuration file, it sends a "file not found" message to the device.

A device that requests a configuration file while the TFTP server is rebuilding configuration files or while it is processing the maximum number of requests, will receive a message from the TFTP server that causes the device to request the configuration file later. The Maximum Serving Count service parameter, which can be configured, specifies the maximum number of requests that can be handled concurrently by the TFTP server. (Default value = 500 requests.) Use the default value if the TFTP service is run along with other Cisco CallManager services on the same server. For a dedicated TFTP server, use the following suggested values for the Maximum Serving Count:

- 1500 requests for a single-processor system
- 3000 requests for a dual-processor system

An Example of TFTP in Operation

Every time an endpoint reboots, the endpoint will request a configuration file (via TFTP) whose name is based on the requesting endpoint's MAC address. (For a Cisco Unified IP Phone 7961 with MAC address ABCDEF123456, the file name would be SEPABCDEF123456.cnf.xml.) The received configuration file includes the software version that the phone must run and a list of Cisco Unified CM servers with which the phone should register. The endpoint might also download, via TFTP, ringer files, softkey templates, and other miscellaneous files to acquire the necessary configuration information before becoming operational.

If the configuration file includes software file(s) version numbers that are different than those the phone is currently using, the phone will also download the new software file(s) from the TFTP server to upgrade itself. The number of files an endpoint must download to upgrade its software varies based on the type of endpoint and the differences between the phone's current software and the new software.

TFTP File Transfer Times

Each time an endpoint requests a file, there is a new TFTP transfer session. For centralized call processing deployments, the time to complete each of these transfers will affect the time it takes for an endpoint to start and become operational as well as the time it takes for an endpoint to upgrade during a scheduled maintenance. While TFTP transfer times are not the only factor that can affect these end states, they are a significant component.

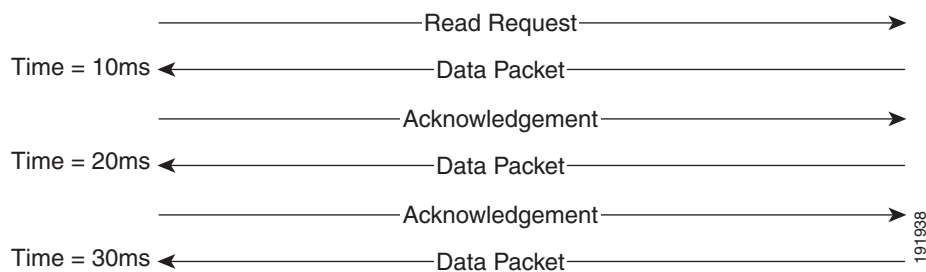
The time to complete each file transfer via TFTP is predictable as a function of the file size, the percentage of TFTP packets that must be retransmitted, and the network latency or round-trip time.

At first glance, network bandwidth might seem to be missing from the previous statement, but it is actually included via the percentage of TFTP packets that must be retransmitted. This is because, if there is not enough network bandwidth to support the file transfer(s), then packets will be dropped by the network interface queuing algorithms and will have to be retransmitted.

TFTP operates on top of the User Datagram Protocol (UDP). Unlike Transmission Control Protocol (TCP), UDP is not a reliable protocol, which means that UDP does not inherently have the ability to detect packet loss. Obviously, detecting packet loss in a file transfer is important, so RFC 1350 defines TFTP as a lock-step protocol. In other words, a TFTP sender will send one packet and wait for a response before sending the next packet (see [Figure 3-4](#)).

Figure 3-4 Example of TFTP Packet Transmission Sequence

Round Trip Time = 10ms

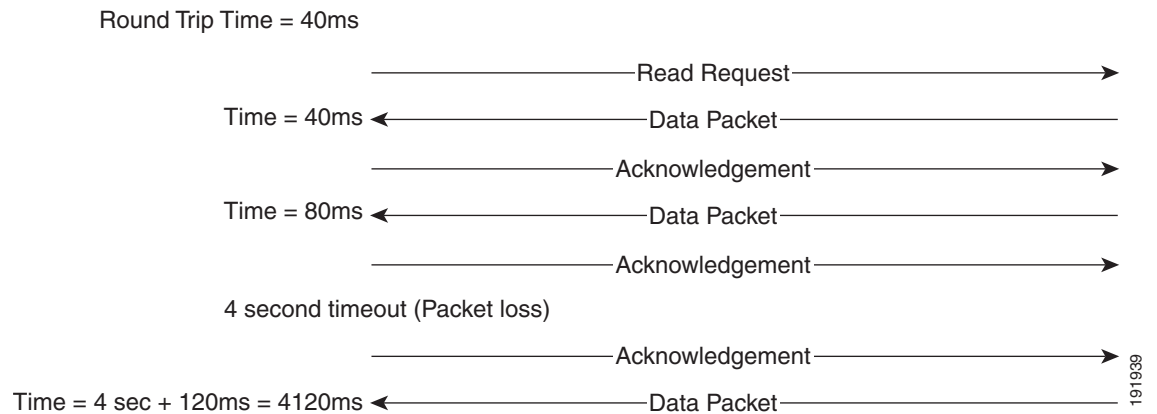


If a response is not received in the timeout period (4 seconds by default), the sender will resend the data packet or acknowledgment. When a packet has been sent five times without a response, the TFTP session fails. Because the timeout period is always the same and not adaptive like a TCP timeout, packet loss can significantly increase the amount of time a transfer session takes to complete.

Because the delay between each data packet is, at a minimum, equal to the network round-trip time, network latency also is a factor in the maximum throughput that a TFTP session can achieve.

In [Figure 3-5](#), the round-trip time has been increased to 40 ms and one packet has been lost in transit. While the error rate is high at 12%, it is easy to see the effect of latency and packet loss on TFTP because the time to complete the session increased from 30 ms (in [Figure 3-4](#)) to 4160 ms (in [Figure 3-5](#)).

Figure 3-5 Effect of Packet Loss on TFTP Session Completion Time



Use the following formula to calculate how long a TFTP file transfer will take to complete:

$$\text{FileTransferTime} = \text{FileSize} * [(\text{RTT} + \text{ERR} * \text{Timeout}) / 512000]$$

Where:

FileTransferTime is in seconds.

FileSize is in bytes.

RTT is the round-trip time in milliseconds.

ERR is the error rate, or percentage of packets that are lost.

Timeout is in milliseconds.

$$512000 = (\text{TFTP packet size}) * (1000 \text{ milliseconds per second}) = (512 \text{ bytes}) * (1000 \text{ milliseconds per second})$$

[Table 3-3](#) and [Table 3-4](#) illustrate the use of this equation to calculate transfer times for the software files for various endpoint device types, protocols, and network latencies.

Table 3-3 TFTP File Transfer Times for SCCP Devices

Device Type (Cisco Unified IP Phone)	Firmware Size (bytes, rounded up to next 100k)	Time to Complete Transfer (1% error rate)				
		40 ms RTT	80 ms RTT	120 ms RTT	160 ms RTT	200 ms RTT
7985	15,000,000	39 min 3 sec	58 min 35 sec	78 min 7 sec	97 min 39 sec	117 min 11 sec
7921	9,700,000	25 min 15 sec	37 min 53 sec	50 min 31 sec	63 min 9 sec	75 min 46 sec
7970 or 7971	6,300,000	16 min 24 sec	24 min 36 sec	32 min 48 sec	41 min 0 sec	49 min 13 sec
7941 or 7961	6,100,000	15 min 53 sec	23 min 49 sec	31 min 46 sec	39 min 42 sec	47 min 39 sec
7911 or 7906	6,100,000	15 min 53 sec	23 min 49 sec	31 min 46 sec	39 min 42 sec	47 min 39 sec

Table 3-3 TFTP File Transfer Times for SCCP Devices (continued)

Device Type (Cisco Unified IP Phone)	Firmware Size (bytes, rounded up to next 100k)	Time to Complete Transfer (1% error rate)				
		40 ms RTT	80 ms RTT	120 ms RTT	160 ms RTT	200 ms RTT
7920	1,200,000	3 min 7 sec	4 min 41 sec	6 min 15 sec	7 min 48 sec	9 min 22 sec
7936	1,800,000	4 min 41 sec	7 min 1 sec	9 min 22 sec	11 min 43 sec	14 min 3 sec
7940 or 7960	900,000	2 min 20 sec	3 min 30 sec	4 min 41 sec	5 min 51 sec	7 min 1 sec
7910	400,000	1 min 2 sec	1 min 33 sec	2 min 5 sec	2 min 36 sec	3 min 7 sec
7912	400,000	1 min 2 sec	1 min 33 sec	2 min 5 sec	2 min 36 sec	3 min 7 sec
7905	400,000	1 min 2 sec	1 min 33 sec	2 min 5 sec	2 min 36 sec	3 min 7 sec
7902	400,000	1 min 2 sec	1 min 33 sec	2 min 5 sec	2 min 36 sec	3 min 7 sec

Table 3-4 TFTP File Transfer Times for SIP Devices

Device Type (Cisco Unified IP Phone)	Firmware Size (bytes, rounded up to next 100k)	Time to Complete Transfer (1% error rate)				
		40 ms RTT	80 ms RTT	120 ms RTT	160 ms RTT	200 ms RTT
7970 or 7971	6,700,000	17 min 26 sec	26 min 10 sec	34 min 53 sec	43 min 37 sec	52 min 20 sec
7941 or 7961	6,500,000	16 min 55 sec	25 min 23 sec	33 min 51 sec	42 min 19 sec	50 min 46 sec
7911 or 7906	6,400,000	16 min 40 sec	25 min 0 sec	33 min 20 sec	41 min 40 sec	50 min 0 sec
7940 or 7960	900,000	2 min 20 sec	3 min 30 sec	4 min 41 sec	5 min 51 sec	7 min 1 sec
7912	400,000	1 min 2 sec	1 min 33 sec	2 min 5 sec	2 min 36 sec	3 min 7 sec
7905	400,000	1 min 2 sec	1 min 33 sec	2 min 5 sec	2 min 36 sec	3 min 7 sec

The values in [Table 3-3](#) and [Table 3-4](#) are the approximate times to download the necessary firmware files to the phone. This is *not* an estimate of the time that it will take for a phone to upgrade to the new firmware and become operational.

**Note**

The exact process each phone goes through on startup and the size of the files downloaded will depend on the phone model, the signaling type configured for the phone (SCCP, MGCP, or SIP), and the previous state of the phone. While there are differences in which files are requested, the general process each phone follows is the same, and in all cases TFTP is used to request and deliver the appropriate files. The general recommendations for TFTP server deployment do not change based on the protocol and/or phone models deployed.

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.

TFTP Load Sharing

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_Primary, TFTP1_Secondary
- In subnet 10.1.2.0/24: Option 150: TFTP1_Secondary, TFTP1_Primary

Under normal operations, a phone in subnet 10.1.1.0/24 will request TFTP services from TFTP1_Primary, while a phone in subnet 10.1.2.0/24 will request TFTP services from TFTP1_Secondary. If TFTP1_Primary fails, then phones from both subnets will request TFTP services from TFTP1_Secondary.

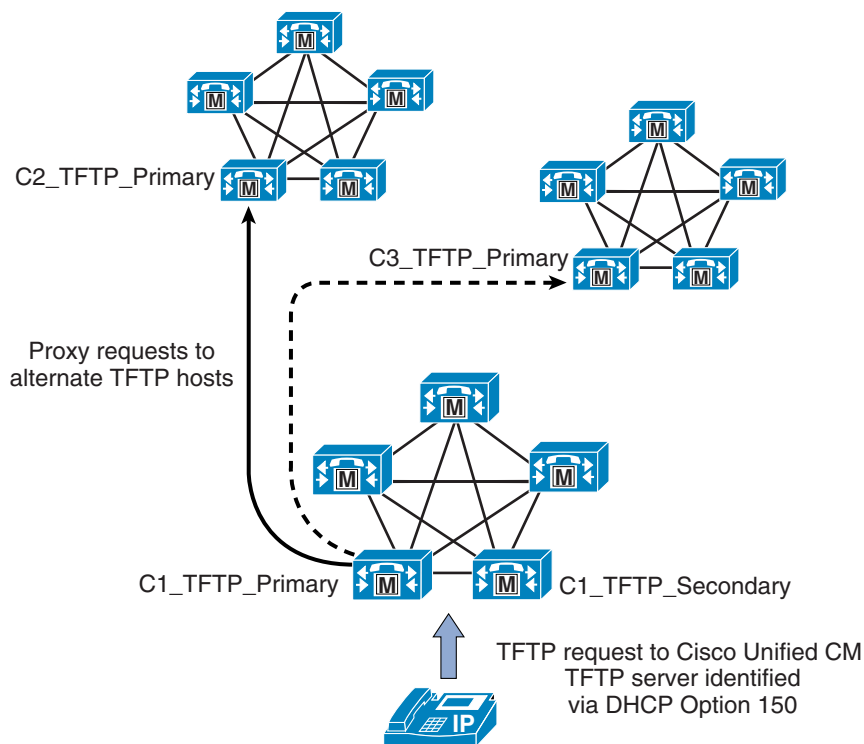
Load balancing avoids having a single TFTP server hot-spot, where all phones from multiple clusters rely on the same server for service. TFTP load balancing is especially important when phone software loads are transferred, such as during a Unified CM upgrade, because more files of larger size are being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

Centralized 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 servers whose addresses are 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. In a centralized TFTP deployment, a set of TFTP servers associated with one of the clusters must provide TFTP services to all the phones in the multi-cluster system.

In order to provide this single point of file access, each cluster's TFTP server must be able to serve files via the central proxy TFTP server. With Cisco Unified CM 5.0 and later releases, this proxy arrangement is accomplished by configuring a set of possible redirect locations in the central TFTP server, pointing to each of the other clusters' TFTP servers. This configuration uses a HOST redirect statement in the Alternate File Locations on the centralized TFTP server, one for each of the other clusters. Each of the redundant TFTP servers in the centralized cluster should point to one of the redundant servers in each of the child clusters. It is not necessary to point the centralized server to both redundant servers in the child clusters because the redistribution of files within each individual cluster and the failover mechanisms of the phones between the redundant servers in the central cluster provide for a very high degree of fault tolerance.

Figure 3-6 shows an example of the operation of this process. A request from a phone registered to Cluster 3 is directed to the centralized TFTP server configured in Cluster 1 (C1_TFTP_Primary). This server will in turn query each of the configured alternate TFTP servers until one responds with a copy of the file initially requested by the phone. Requests to the centralized secondary TFTP server (C1_TFTP_Secondary) will be sent by proxy to the other clusters' secondary TFTP servers until either the requested file is found or all servers report that the requested file does not exist.

Figure 3-6 Centralized TFTP Servers

153371

Centralized TFTP in a Mixed Environment, with Servers Running Different Releases of Cisco Unified CM

During migration from an older Unified CM release to Unified CM 5.0 and later releases, it is likely that a large centralized TFTP environment will have to operate in a mixed mode. Prior to Unified CM 5.0, the centralized TFTP servers did not request files from the child servers but rather had the TFTP directories of each of those clusters remotely mounted to the central server, which in turn was able to search all of the local and remote directories for the requested file. During a period of migration, it is necessary to provide a centralized TFTP server that can operate in both modes (mixed mode): *remote mount* for releases prior to Unified CM 5.0 and *proxy request* for Unified CM 5.0 and later releases. Because the servers for Unified CM 5.0 and later releases do not support remotely mounted file systems in a mixed environment, it is necessary to position Cisco Unified CM 4.1(3)SR3a or a later Windows OS-based release of Unified CM as the centralized mixed-mode TFTP cluster.

**Note**

Cisco Unified CM Release 4.1(3)SR3a, and subsequent Unified CM releases for the Windows OS platform, contain an upgrade to the cTFTP server daemon that allows it to support mixed-mode centralized TFTP designs. With these releases, the centralized TFTP server supports both remote mount and proxy request as methods for reaching Alternate TFTP Files servers in other clusters.

When configuring the mixed-mode TFTP server, it is necessary to specify the servers for Unified CM 5.0 and later releases via the HOST proxy request and to specify any servers prior to Unified CM 4.1(3)SR3a by using the remote mount configuration process, as shown in [Figure 3-7](#). (For details on the remote mount configuration, see [Centralized Configuration for Remote-Mount Servers Prior to Cisco Unified CM 4.1\(3\)SR3r](#), page 3-21.) Any child cluster supporting mixed mode can be configured as either a remote mount or a proxy cluster.

For centralized TFTP configurations, ensure that the main TFTP server exists in the cluster that runs the highest version of Cisco Unified Communications Manager. For example, if you are using a centralized TFTP server between a compatible Cisco Unified CM 4.x (mixed mode) cluster and a Unified CM 5.x cluster, ensure that the central TFTP server exists in the Cisco Unified CM 5.x cluster.

If the centralized TFTP server exists in the cluster that runs the lower version of Cisco Unified Communications Manager, all phones use the locale files from this centralized TFTP server. These older locale files can create display issues for phones registered to clusters running a higher version of Cisco Unified CM because the newer localized phrases are not included in the locale files that are served from the main cluster's TFTP server.

Figure 3-7 Dual-Mode Configuration

Parameter Name	Parameter Value	Suggested Value
Alternate File Location 1	HOST://10.104.28.10	
Alternate File Location 2	c:\Program Files\Cisco\TFTPpath\Skate3	
Alternate File Location 3	HOST://10.104.5.10	
Alternate File Location 4	HOST://10.104.8.10	
Alternate File Location 5		
Alternate File Location 6		
Alternate File Location 7		
Alternate File Location 8		
Alternate File Location 9		
Alternate File Location 10		
File Location*	C:\Program Files\Cisco\TFTPpath	C:\Program Files\Cisco\TFTPpath

Some parameters in this group are hidden, click on Advanced button to see hidden parameters

Centralized Configuration for Remote-Mount Servers Prior to Cisco Unified CM 4.1(3)SR3r

If a 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. To support an environment prior to Unified CM 4.1(3)SR3, the centralized TFTP server must be configured to search through remotely mounted subdirectories associated with the other clusters.

Example 3-4 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 form at \\<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.

With Cisco Unified CM 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)

**Note**

Beginning with Cisco Unified CM Release 5.1, the Enable Caching of Configuration Files service parameter is no longer available and configuration files are always cached in memory (rather than written to disk).

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.

Unified CM NTP Time Synchronization

Time synchronization is especially critical on Unified CM servers. In addition to ensuring that CDR records are accurate and that log files are synchronized, having an accurate time source is necessary for any future IPsec features to be enabled within the cluster and for communications with any external entity.

Unified CM 5.0 automatically synchronizes the NTP time of all subscribers in the cluster to the publisher. During installation, each subscriber is automatically configured to point to an NTP server running on the publisher. The publisher considers itself to be a master server and provides time for the cluster based on its internal hardware clock unless it is configured to synchronize from an external server. Cisco highly recommends configuring the publisher to point to a Stratum-1, Stratum-2, or Stratum-3 NTP server to ensure that the cluster time is synchronized with an external time source.

**Note**

Manual configuration of the NTP.conf file is no longer possible, and any changes made to the file will be automatically replaced by the system configuration.

For additional information about NTP time synchronization in a Cisco Unified Communications environment, refer to the *Cisco IP Telephony Clock Synchronization: Best Practices* white paper, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_white_paper0900aecd8037fdb5.shtml

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 3-5](#) illustrates the configuration of NTP time synchronization on Cisco IOS and CatOS devices.

Example 3-5 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 might be necessary to configure time zones using the **clock timezone** command (in Cisco IOS) and/or the **set timezone** command (in CatOS).

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 Unified IP Phones support the 802.3af PoE standard, see the [Endpoint Features Summary, page 20-37](#).

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 Unified IP Phones 7971, 7970, 7961, 7960, 7941, 7940, 7911, and 7910+SW) 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 Unified IP Phones 7902, 7905, and 7910) 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 Unified IP Phones 7971, 7970, 7961, 7960, 7941, 7940, 7912, 7911, and 7910+SW) 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 Unified IP Phone 7912 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.

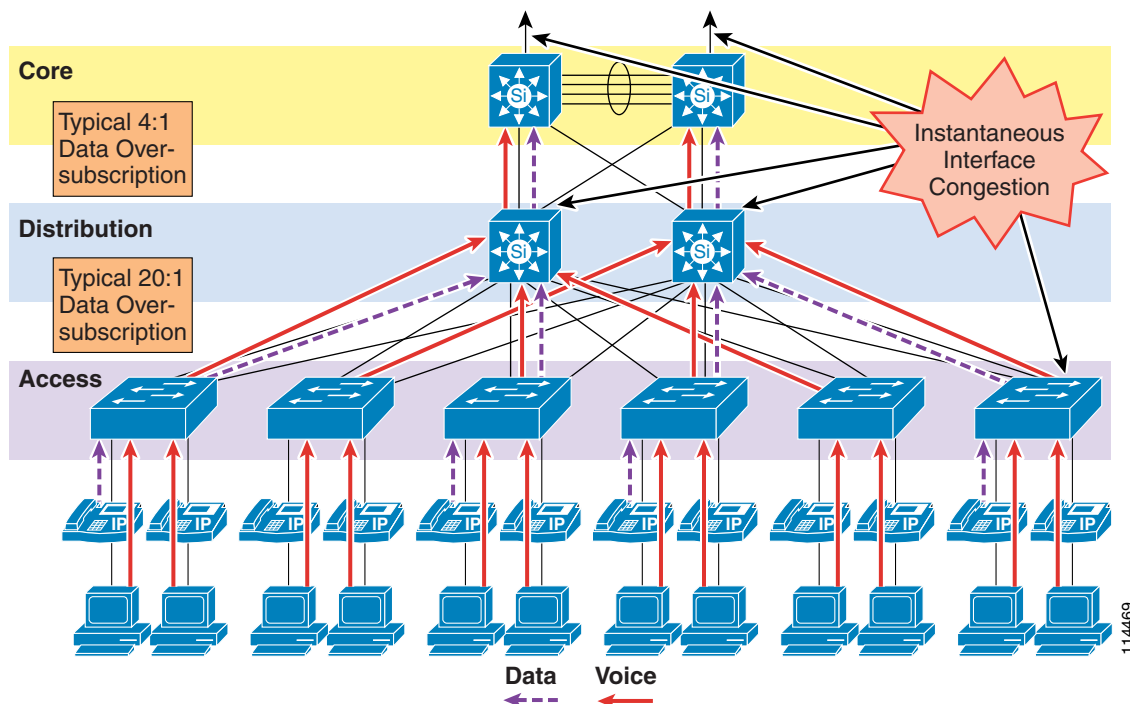
For more information about the use of IBM cabling, refer to the Product Bulletin *Shielded Twisted-Pair Cabling Support for Cisco Fast Ethernet Products*, available at

<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 3-8 illustrates the typical oversubscription that occurs in LAN infrastructures.

Figure 3-8 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 3-27](#)
- [Interface Queuing, page 3-28](#)
- [Bandwidth Provisioning, page 3-29](#)
- [Impairments to IP Communications if QoS is Not Employed, page 3-29](#)

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 3-5](#). As such, the IP phone can and should classify traffic flows.

[Table 3-5](#) lists the traffic classification requirements for the LAN infrastructure.

Table 3-5 Traffic Classification Guidelines for Various Types of Network Traffic

Application	Layer-3 Classification			Layer-2 Classification
	IP Precedence (IPP)	Per-Hop Behavior (PHB)	Differentiated Services Code Point (DSCP)	Class of Service (CoS)
Routing	6	CS6	48	6
Voice Real-Time Transport Protocol (RTP)	5	EF	46	5
Videoconferencing	4	AF41	34	4
Streaming video	4	CS4	32	4
Call signaling ¹	3	CS3 (currently) AF31 (previously)	24 (currently) 26 (previously)	3
Transactional data	2	AF21	18	2
Network management	2	CS2	16	2
Scavenger	1	CS1	8	1
Best effort	0	0	0	0

1. The recommended DSCP/PHB marking for call 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.

For more information about traffic classification, refer to the *Enterprise QoS Solution Reference Network Design (SRND)*, available at

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

Traffic Classification for Video Telephony

The main classes of interest for IP Video Telephony are:

- Voice
Voice is classified as CoS 5 (IP Precedence 5, PHB EF, or DSCP 46).
- Videoconferencing
Videoconferencing is classified as CoS 4 (IP Precedence 4, PHB AF41, or DSCP 34).
- Call signaling
Call signaling for voice and videoconferencing is now classified as CoS 3 (IP Precedence 3, PHB CS3, or DSCP 24) but was previously classified as PHB AF31 or DSCP 26.

Cisco highly recommends these classifications as *best practices* in a Cisco Unified Communications network.

The voice component of a call can be classified in one of two ways, depending on the type of call in progress. A voice-only (or normal) telephone call would have the media classified as CoS 5 (IP Precedence 5 or PHB EF), while the audio channel of a video conference would have the media classified as CoS 4 (IP Precedence 4 or PHB AF41). All the Cisco IP Video Telephony products adhere to the Cisco Corporate QoS Baseline standard, which requires that the audio and video channels of a video call both be marked as CoS 4 (IP Precedence 4 or PHB AF41). The reasons for this recommendation include, but are not limited to, the following:

- To preserve lip-sync between the audio and video channels
- To provide separate classes for audio-only calls and video calls

The signaling class is applicable to all voice signaling protocols (such as SCCP, MGCP, and so on) as well as video signaling protocols (such as SCCP, H.225, RAS, CAST, and so on). These protocols are discussed in more detail in the section on [Software-Based Endpoints](#), page 20-31.

Given the recommended classes, the first step is decide where the packets will be classified (that is, which device will be the first to mark the traffic with its QoS classification). There are essentially two places to mark or classify traffic:

- On the originating endpoint — the classification is then trusted by the upstream switches and routers
- On the switches and/or routers — because the endpoint is either not capable of classifying its own packets or is not trustworthy to classify them correctly

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, page 3-30](#)
- [WAN Quality of Service \(QoS\), page 3-33](#)
- [Resource Reservation Protocol \(RSVP\), page 3-40](#)
- [Bandwidth Provisioning, page 3-49](#)

WAN Design and Configuration

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 3-30](#)
- [Guaranteed Bandwidth, page 3-31](#)
- [Best-Effort Bandwidth, page 3-32](#)

Deployment Considerations

WAN deployments for voice networks may be hub-and-spoke or an arbitrary topology. A hub-and-spoke topology consists of 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. An arbitrary topology may contain multiple WAN links and any number of hops between the sites. In this scenario there may be many different paths to the same site or there may be different links used for communication with some sites compared to other sites. The simplest example is three sites, each with a WAN link to the other two sites, forming a triangle. In that case there are two potential paths between each site to each other site.

Topology-unaware call admission control requires the WAN to be hub-and-spoke, or a spoke-less hub in the case of MPLS VPN. This topology ensures that call admission control, provided by Unified CM's locations or a gatekeeper, 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.

Topology-aware call admission control may be used with either hub-and-spoke or an arbitrary WAN topology. This form of call admission control requires parts of the WAN infrastructure to support Resource Reservation Protocol (RSVP). For details, see [Resource Reservation Protocol \(RSVP\), page 3-40](#), and [Call Admission Control, page 9-1](#).

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

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. Topology-unaware call admission control normally requires redundant paths to be over-provisioned and under-subscribed to allow for failures that reduce the available bandwidth between sites without the call admission control mechanism being aware of those failures or the reduction in bandwidth. Topology-aware call admission control is able to adjust dynamically to many of the topology changes and allows for efficient use of the total available bandwidth.

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 Unified Communications Manager Express (Unified 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. For deployments that use clustering over the WAN, the one-way delay for signaling traffic between clusters should not exceed 20 milliseconds (see [Clustering Over the IP WAN, page 2-17](#)).

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 3-6](#) lists the QoS features and tools required for the WAN infrastructure to achieve this goal.

Table 3-6 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

Table 3-6 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
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

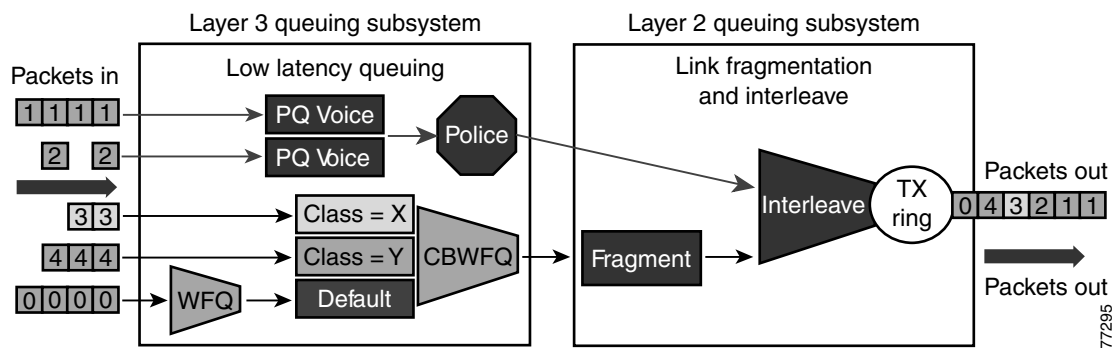
The following sections highlight some of the most important features and techniques to consider when designing a WAN to support both voice and data traffic:

- [Traffic Prioritization, page 3-34](#)
- [Link Efficiency Techniques, page 3-35](#)
- [Traffic Shaping, page 3-37](#)

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.

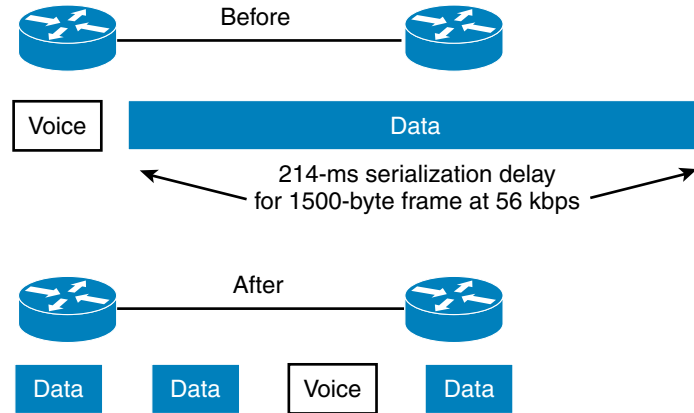
It should also be noted that, beginning in Cisco IOS Release 12.2(13)T, cRTP can be configured as part of the voice class with the Class-Based cRTP feature. This option allows cRTP to be specified within a class, attached to an interface via a service policy. This new feature provides compression statistics 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://www.cisco.com/go/designzone>

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.

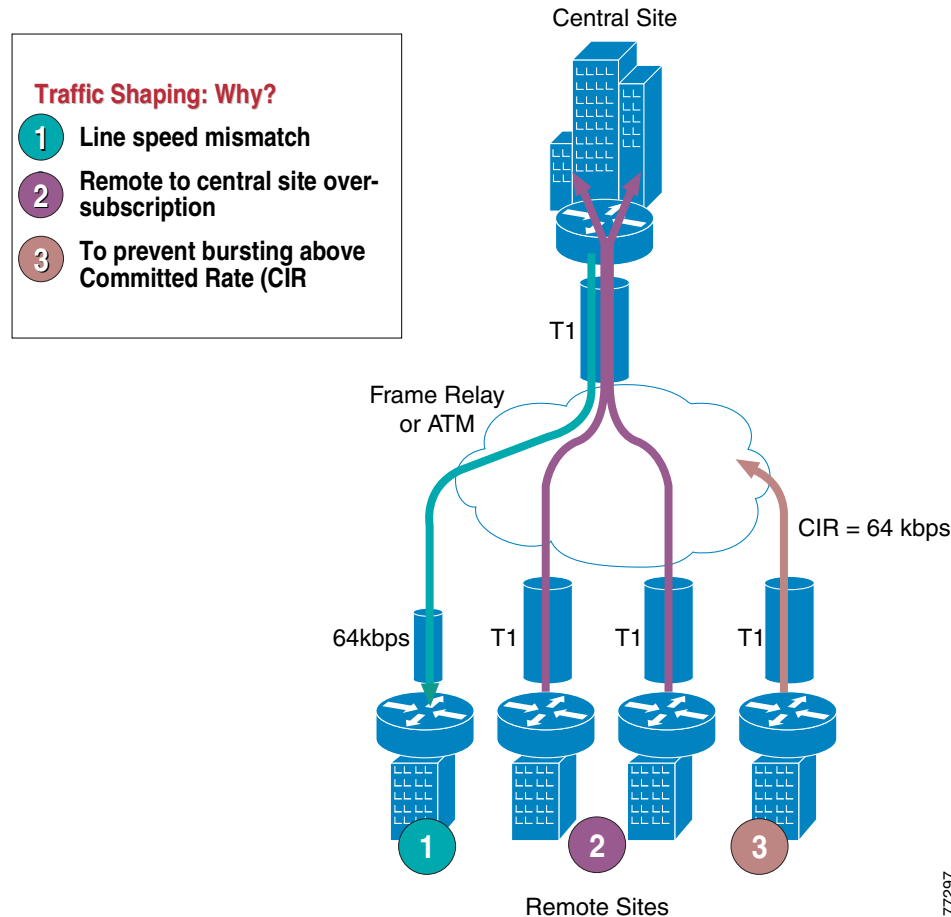
Figure 3-11 Traffic Shaping with Frame Relay and ATM

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.

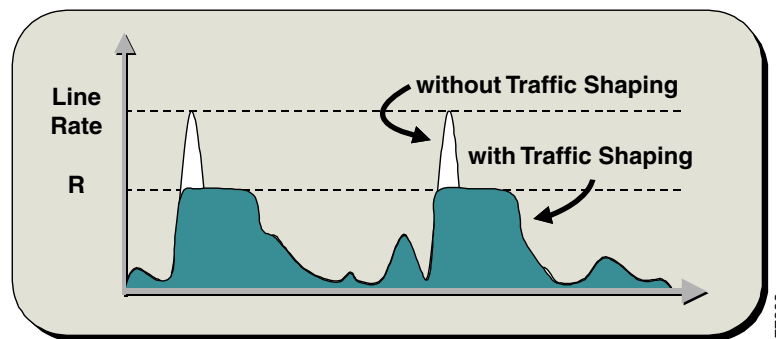
3. Bursting above Committed Information Rate (CIR)

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 3-36), 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/en/US/products/sw/iosswrel/ps1839/products_feature_guide09186a00801541c6.html

Resource Reservation Protocol (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, was first introduced by the IETF in RFC 2205, and it 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, therefore it can be used to provide topology-aware call admission control for voice and video calls.

This section focuses on the RSVP protocol principles and its interactions with the WAN infrastructure, specifically the QoS aspects, while the motivation and the mechanisms for call admission control based on RSVP are described in the chapter on [Call Admission Control, page 9-1](#).

This section covers the following specific topics:

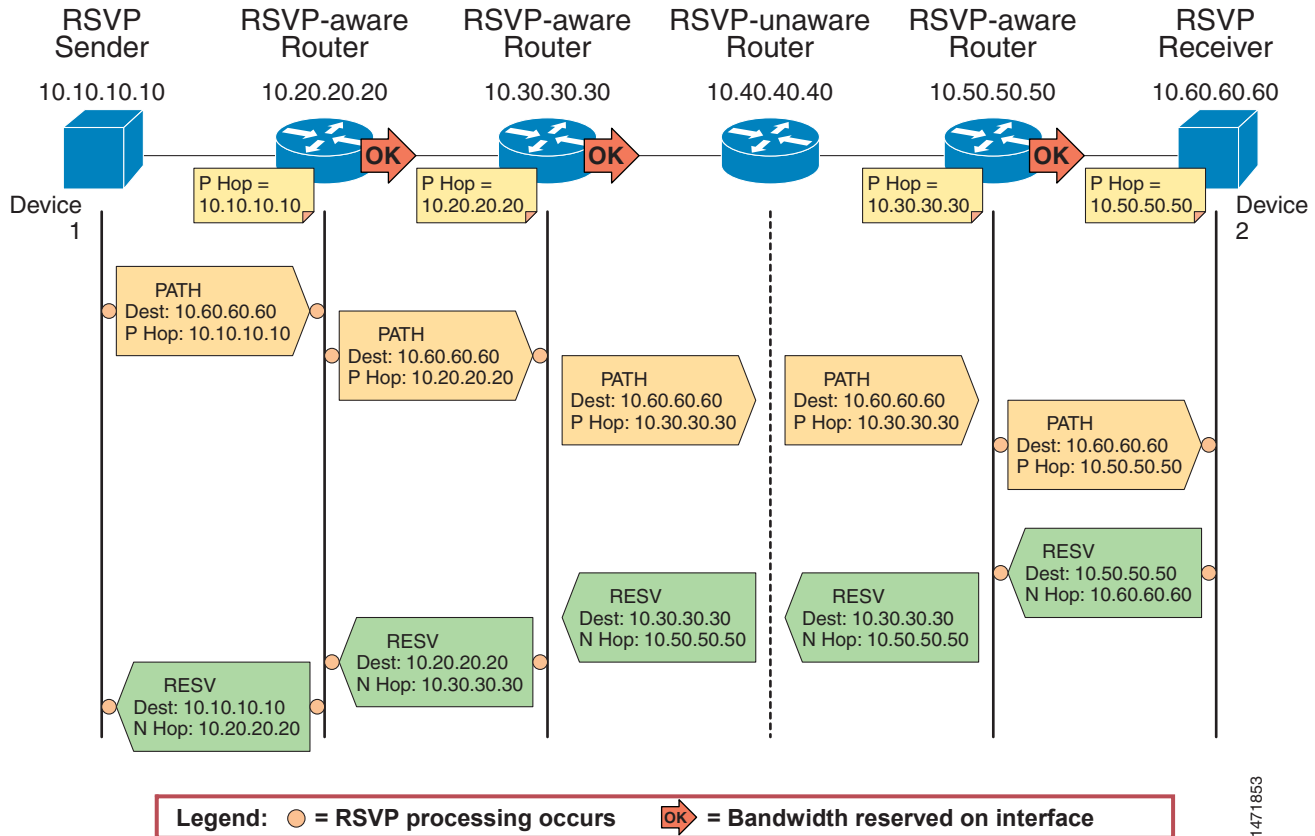
- [RSVP Principles, page 3-40](#)
- [RSVP and QoS in WAN Routers, page 3-43](#)
- [RSVP Application ID, page 3-47](#)
- [RSVP Design Best Practices, page 3-49](#)

RSVP Principles

RSVP performs resource reservation for a given data flow across a network by defining signaling messages that are exchanged between the source and destination devices for the data flow and that are processed by intermediate routers along the path. The RSVP signaling messages are IP packets with the protocol number in the IP header set to 46, and they are routed through the network according to the existing routing protocols.

Not all routers on the path are required to support RSVP because the protocol is designed to operate transparently across RSVP-unaware nodes. On each RSVP-enabled router, the RSVP process intercepts the signaling messages and interacts with the QoS manager for the router interfaces involved in the data flow in order to "reserve" bandwidth resources. When the available resources are not sufficient for the data flow anywhere along the path, the routers signal the failure back to the application that originated the reservation request.

The principles of RSVP signaling can be explained by using the example shown in [Figure 3-13](#). In this diagram, an application wishes to reserve network resources for a data stream flowing from Device 1, whose IP address is 10.10.10.10, to Device 2, whose IP address is 10.60.60.60.

Figure 3-13 Example of RSVP Path and Resv Message Flow

1471853

The following steps describe the RSVP signaling process for the example in Figure 3-13:

1. The application residing on Device 1 originates an RSVP message called Path, which is sent to the same destination IP address as the data flow for which a reservation is requested (that is, 10.60.60.60) and is sent with the "router alert" option turned on in the IP header. The Path message contains, among other things, the following objects:
 - The "session" object, consisting of destination IP address, protocol number, and UDP/TCP port, which is used to identify the data flow in RSVP-enabled routers.
 - The "sender Tspec" (traffic specification) object, which characterizes the data flow for which a reservation will be requested. This typically translates into a token bucket model that specifies a data rate and a burst size (or bucket depth).
 - The "P Hop" (or previous hop) object, which contains the IP address of the router interface that last processed the Path message. In this example, the P Hop is initially set to 10.10.10.10 by Device 1.
2. By means of the "router alert" option, the Path message is intercepted by the CPU of the RSVP-aware router identified as 10.20.20.20 in Figure 3-13, which sends it to the RSVP process. RSVP creates a path state for this data flow, storing the values of the session, sender Tspec, and P Hop objects contained in the Path message. Then it forwards the message downstream, after having replaced the P Hop value with the IP address of its outgoing interface (10.20.20.20 in this example).

3. Similarly, the Path message is intercepted by the CPU of the following RSVP-aware router, identified as 10.30.30.30 in [Figure 3-13](#). After creating the path state and changing the P Hop value to 10.30.30.30, this router also forwards the message downstream.
4. The Path message now arrives at the RSVP-unaware router identified as 10.40.40.40 in [Figure 3-13](#). Because RSVP is not enabled on this router, it just routes this message according to the existing routing protocols like any other IP packet, without any additional processing and without changing the content of any of the message objects.
5. Therefore, the Path message gets to the RSVP-aware router identified as 10.50.50.50, which processes the message, creates the corresponding path state, and forwards the message downstream. Notice that the P Hop recorded by this router still contains the IP address of the last RSVP-aware router along the network path, or 10.30.30.30 in this example.
6. The RSVP Receiver at Device 2 receives the Path message with a P Hop value of 10.50.50.50, and it can now initiate the actual reservation by originating a message called Resv. For this reason, RSVP is known as a receiver-initiated protocol. The Resv message carries the reservation request hop-by-hop from the receiver to the sender, along the reverse paths of the data flow for the session. At each hop, the IP destination address of the Resv message is the IP address of the previous-hop node, obtained from the path state. Hence, in this case Device 2 sends the Resv message with a destination IP address of 10.50.50.50. The Resv message contains, among other things, the following objects:
 - The "session" object, which is used to identify the data flow.
 - The "N Hop" (or next hop) object, which contains the IP address of the node that generated the message. In this example, the N Hop is initially set to 10.60.60.60 by Device 2.
7. When RSVP-aware router 10.50.50.50 receives the Resv message for this data flow, it matches it against the path state information using the received session object, and it verifies if the reservation request can be accepted based on the following criteria:
 - Policy control — Is this user and/or application allowed to make this reservation request?
 - Admission control — Are there enough bandwidth resources available on the relevant outgoing interface to accommodate this reservation request?
8. In this case, we assume that both policy and admission control are successful on 10.50.50.50, which means that the bandwidth provided by the Tspec in the path state for this session is reserved on the outgoing interface (in the same direction as the data flow, that is from Device 1 to Device 2), and a corresponding "reservation state" is created. Now router 10.50.50.50 can send a Resv message upstream by sending it as a unicast IP packet to the destination IP address stored in the P Hop for this session, which was 10.30.30.30. The N Hop object is also updated with the value of 10.50.50.50.
9. The Resv message now transits through the RSVP-unaware router identified as 10.40.40.40, which will route it toward its destination of 10.30.30.30 like any other IP packet. This mechanism allows RSVP signaling to work across a heterogeneous network where some nodes are not RSVP-enabled.
10. The RSVP-aware router identified as 10.30.30.30 receives the Resv message and processes it according to the mechanisms described in steps 7 and 8. Assuming policy and admission control are successful also at this hop, the bandwidth is reserved on the outgoing interface and a Resv message is sent to the previous hop, or 10.20.20.20 in this example.
11. After a similar process within the router identified as 10.20.20.20, the Resv finally reaches the RSVP sender, Device 1. This indicates to the requesting application that an end-to-end reservation has been established and that bandwidth has been set aside for this data flow in all RSVP-enabled routers across the network.

This example shows how the two main RSVP signaling messages, Path and Resv, travel across the network to establish reservations. Several other messages are defined in the RSVP standard to address error situations, reservation failures, and release of resources. In particular, the ResvErr message is used to signal failure to reserve the requested resources due to either policy control or admission control somewhere along the network. If, for example, admission control had failed at node 10.50.50.50 in [Figure 3-13](#), this node would have sent a ResvErr message back to Device 2, specifying the cause of the failure, and the application would have been notified.

Another important aspect of the RSVP protocol is that it adopts a soft-state approach, which means that for each session both the path state and the reservation state along the network need to be refreshed periodically by the application by sending identical Path and Resv messages. If a router does not receive refresh messages for a given session for a certain period of time, it deletes the corresponding state and releases the resources reserved. This allows RSVP to react dynamically to network topology changes or routing changes due to link failures. The reservations simply start flowing along the new routes based on the routing protocol decisions, and the reservations along the old routes time-out and are eventually deleted.

**Note**

This section focuses on providing an overview of RSVP principles and mechanisms. For more information on protocol behavior and extensions, complete message formats, and interactions with other protocols, refer to the numerous RFC documents related to RSVP, available at <http://www.ietf.org>.

RSVP and QoS in WAN Routers

RSVP has been supported in Cisco routers for many years, however most configurations recommended in this document are based on the RSVP Scalability Enhancements feature, which was first introduced in Cisco IOS Release 12.2(2)T.

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).

**Note**

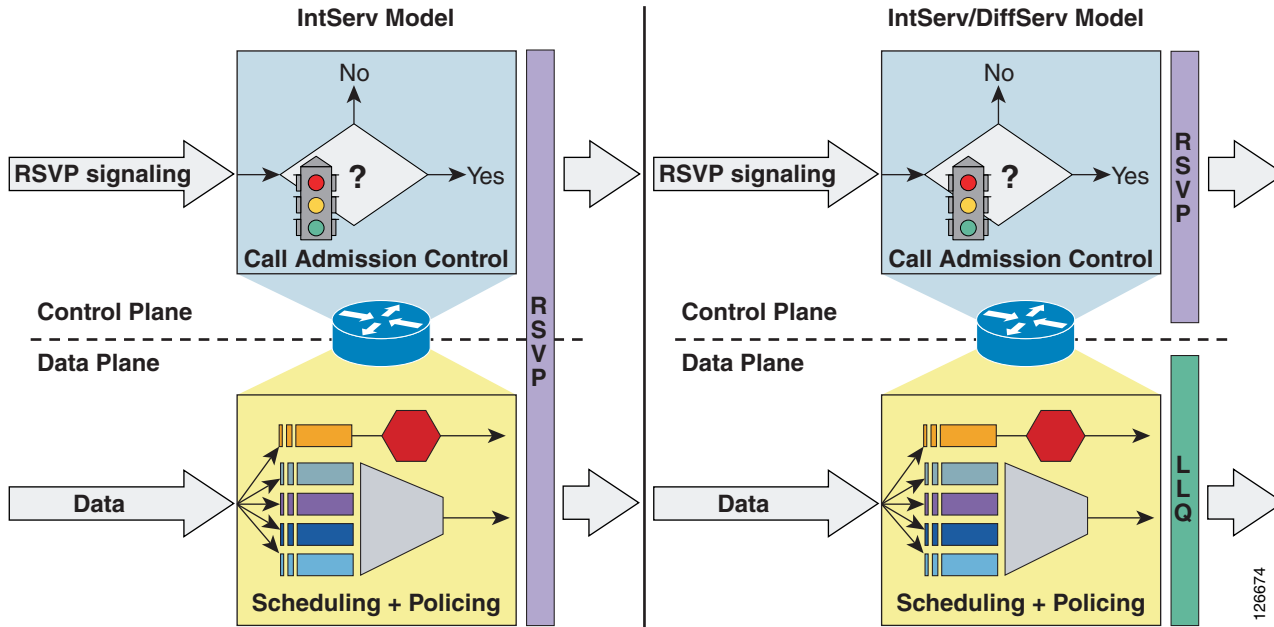
When RSVP is enabled on a router interface, all other interfaces in the router will drop RSVP messages unless they are also enabled for RSVP. To avoid dropping RSVP messages, enable RSVP on all interfaces through which you expect RSVP signaling to transit. If call admission control is not desired on an interface, set the bandwidth value to 75% of the interface bandwidth.

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 3-14](#) shows the difference between these two approaches from the perspective of a Cisco IOS router.

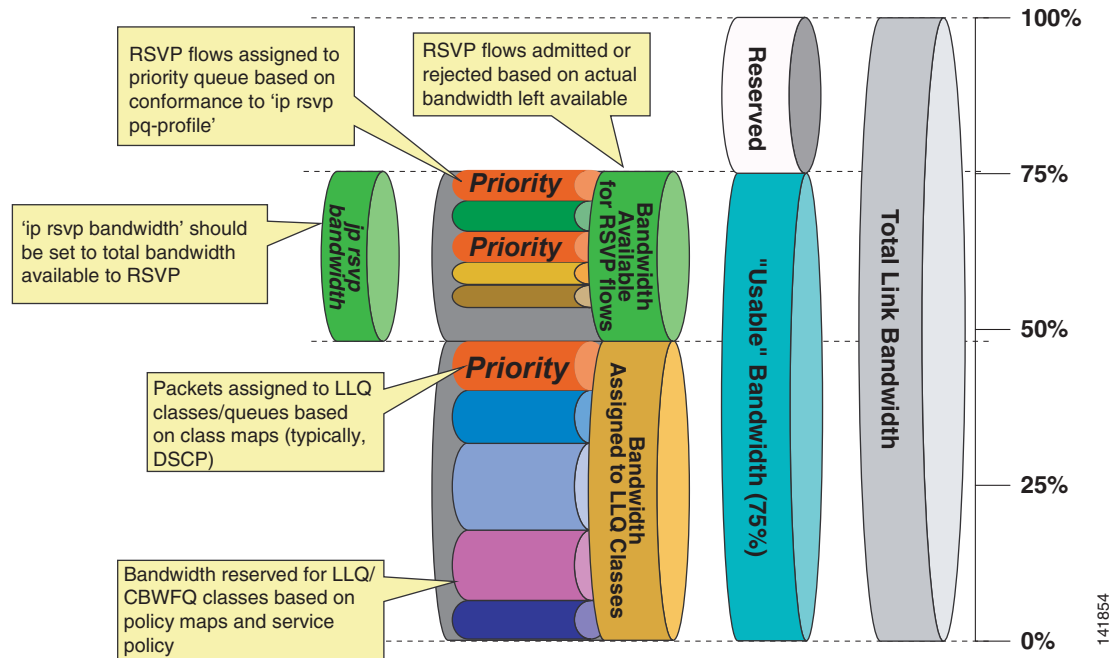
Figure 3-14 The Two RSVP Operation Models: IntServ and IntServ/DiffServ



The IntServ Model

As shown on the left side of [Figure 3-14](#), 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.

As shown in [Figure 3-15](#), when you combine the IntServ model with Low Latency Queuing (LLQ), the usable bandwidth is divided between RSVP and the predefined LLQ queues. RSVP controls the entrance criteria to the RSVP reserved bandwidth, while policy maps control the entrance criteria for the predefined queues.

Figure 3-15 Combining the IntServ Model with LLQ

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 the total maximum bandwidth that can be assigned to QoS mechanisms on a link is equal to 75% of the link speed, if you want to reserve 33% of the link bandwidth for RSVP-admitted flows, you have to make sure that the bandwidth assigned to LLQ classes does not exceed $(75 - 33) = 42\%$ of the link bandwidth.

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
- $p\text{-}to\text{-}r = 110\%$

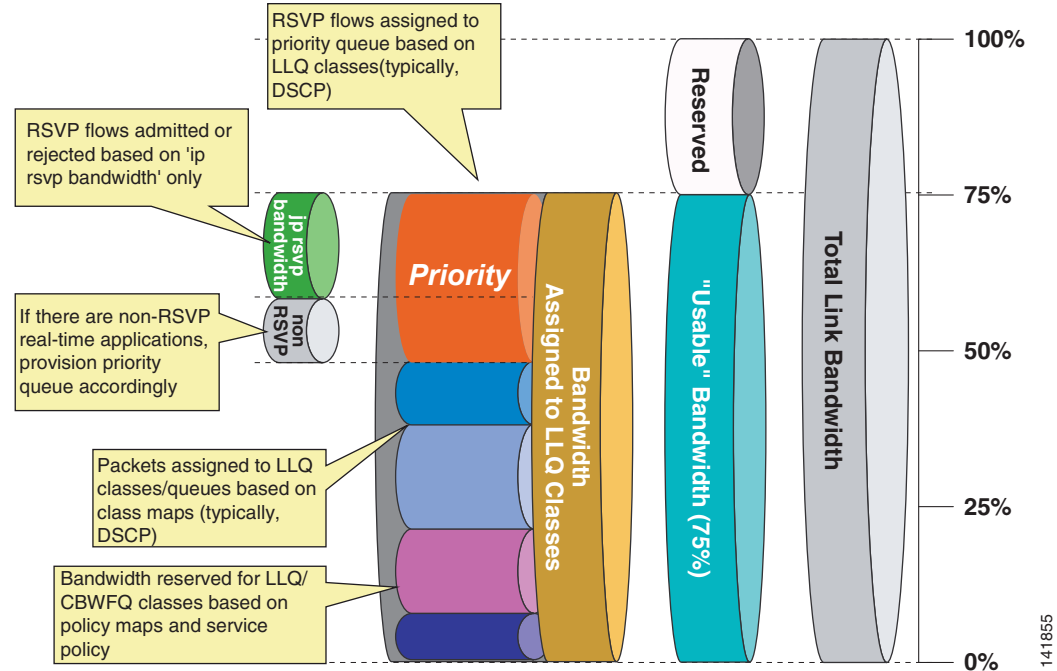
The IntServ/DiffServ Model

As shown on the right side of [Figure 3-14](#), 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.

The entire usable bandwidth (75% of the link speed) can be assigned to LLQ classes, as shown in [Figure 3-16](#), as it normally is today. The policy maps define the traffic that is admitted into each queue. RSVP is typically configured to admit flows up to the amount of bandwidth defined for priority traffic, but keep in mind that RSVP in this model does not adjust the scheduling, so any traffic admitted by RSVP in excess of the predefined priority queue may be dropped or remapped to other lower-priority queues.

If all applications that send priority traffic are RSVP-enabled, you may configure the RSVP bandwidth to match the size of the priority queue. If, on the other hand, there are non-RSVP applications that also need to send priority traffic (such as Unified CM static locations or a gatekeeper), as shown in [Figure 3-16](#), the priority queue is divided into priority traffic that is controlled by non-RSVP mechanisms and priority traffic that is controlled by RSVP. The combined non-RSVP and RSVP admission control mechanisms must not use more bandwidth than is allocated to ensure that the priority queue is never over-subscribed.

Figure 3-16 LLQ Bandwidth Allocation with RSVP

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.

RSVP Application ID

An application identity (app-id) is an RSVP object that can be inserted into the policy element of an RSVP message. This object is described in RFC 2872. This policy object helps to identify the application and associate it with the RSVP reservation request, thus allowing routers along the path to make appropriate decisions based on the application information.

The need for an app-id arises because RSVP is used to support multiple applications such as voice and video.

Without using an app-id, there is only one bandwidth value that is configurable per interface in RSVP. RSVP will admit requests until this bandwidth limit is reached. It does not differentiate between the requests and is not aware of the type of application for which the bandwidth is requested. As a result of this, it is quite possible for RSVP to exhaust the allowed bandwidth by admitting requests representing

just one type of application, thus causing all subsequent requests to be rejected due to unavailable bandwidth. In this way, a few video calls could prevent all or most of the voice calls from being admitted. For example, if an organization allocates 1000 units to RSVP, RSVP might exhaust a majority of this amount by admitting two 384-kbps video calls, thus leaving very little bandwidth for voice calls.

The solution to this problem is to configure separate bandwidth limits for individual applications or classes of traffic. Limiting bandwidth per application requires that an RSVP local policy matching the application bandwidth limit be applied to the router interface and that each reservation request flag the application to which it belongs so that it may be admitted against the appropriate bandwidth limit.

The app-id is not a single piece of information but multiple variable-length strings. As is described in RFC 2872, the object may include the following attributes:

- An identifier of the application (APP). This attribute is required.
- Global unique identifier (GUID). Optional.
- Version number of the application (VER). This attribute is required.
- Sub-application identifier (SAPP). An arbitrary number of sub-application elements can be included. Optional.

For example:

- APP = AudioStream
- GUID = CiscoSystems
- VER = 5.0.1.0
- SAPP = (not specified)

How Unified CM Uses the Application ID

To support RSVP Application ID functionality, Unified CM has two cluster-wide service parameters that define the Application ID used to tag audio and video call reservations using RSVP:

- RSVP Audio Application ID (Default is "AudioStream")
- RSVP Video Application ID (Default is "VideoStream")

It is possible to change these service parameters, but Cisco recommend that you leave them at their default values unless you require the ability to differentiate one cluster's reservations from another using the same link.

How Voice Calls are Tagged

When a voice call is made between locations with an RSVP policy, the resulting reservations for the audio stream will be tagged with the RSVP Audio Application ID.

How Video Calls are Tagged

When a video call is made between locations with an RSVP policy, the resulting reservations for the audio stream will be tagged with the RSVP Audio Application ID and the reservations for the video stream will be tagged with the RSVP Video Application ID.

Application ID Call Admission Control Model

As mentioned in the chapter on [Call Admission Control, page 9-1](#), the call admission control model supported by Application ID differs from the one supported by "static" locations. Because the audio stream of a video call is marked with the RSVP Audio Application ID, it is possible to guarantee a

minimum number of voice calls and allow them to take over the entire available bandwidth. Video calls are allowed up to a certain maximum bandwidth but may be denied when the entire bandwidth is consumed by voice calls that were established first.

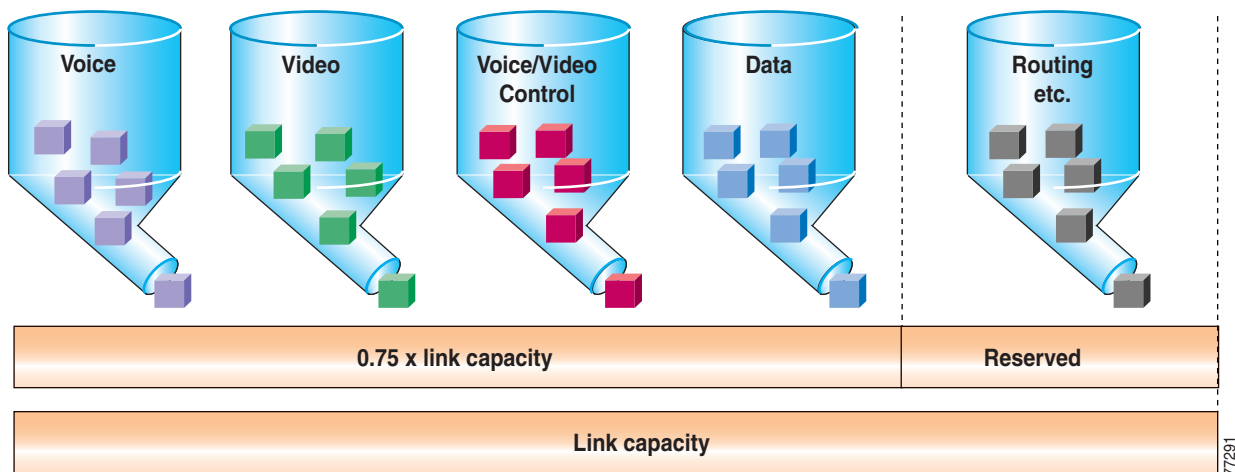
RSVP Design Best Practices

When deploying RSVP in the IP WAN in conjunction with Unified CM, observe the following design best practices:

- Cisco recommends that you use the IntServ/DiffServ model if either of the following statements is true:
 - The only traffic destined for the Priority Queue (PQ) in the IP WAN interfaces is RSVP-enabled traffic.
 - All the non-RSVP traffic destined for the PQ can be deterministically limited to a certain amount by an out-of-band call admission control mechanism (such as Unified CM locations or a Cisco IOS gatekeeper).
- If all the PQ traffic is RSVP-enabled, the value specified in the **ip rsvp bandwidth** command and the **priority** command should match once Layer 2 overhead is taken into account.
- If RSVP is enabled on one or more interfaces of a router, all interfaces through which you expect RSVP signaling to transit should also be enabled for RSVP to ensure that RSVP messages do not get dropped. If call admission control is not desired on an interface, set the bandwidth value to 75% of the interface bandwidth.
- 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.
- Enable RSVP Application ID support if you need to limit the maximum amount of bandwidth used by video calls. Application ID Support is introduced in Cisco IOS Release 12.4(6)T. For more information, see [RSVP Application ID, page 3-47](#).
- Enable RSVP at the edge of the network, including the router WAN interfaces on both sides of the WAN link.
- Enable RSVP at all possible WAN congestion points, including redundant links of different speeds.
- Ensure symmetric routing on load-balanced MPLS WAN links.
- Bundle Interfaces, including MLPPP, ATM-IMA, and FRF.16, should have the RSVP bandwidth configured to the size of one physical link.
- RSVP is currently not available on Tunnel Interfaces.
- RSVP is currently not available on the Catalyst Switching Platforms.

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 keep-alives. [Figure 3-17](#) illustrates this bandwidth provisioning process.

Figure 3-17 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 and video bearer streams, 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 bearer traffic but also the call control traffic. In fact, in multisite WAN deployments, the call control traffic (as well as the bearer traffic) 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 and video bearer traffic in all multisite WAN deployments (see [Provisioning for Bearer Traffic, page 3-51](#))
- Call control traffic in multisite WAN deployments with centralized call processing (see [Provisioning for Call Control Traffic with Centralized Call Processing, page 3-58](#))
- Call control traffic in multisite WAN deployments with distributed call processing (see [Provisioning for Call Control Traffic with Distributed Call Processing, page 3-63](#))

Provisioning for Bearer Traffic

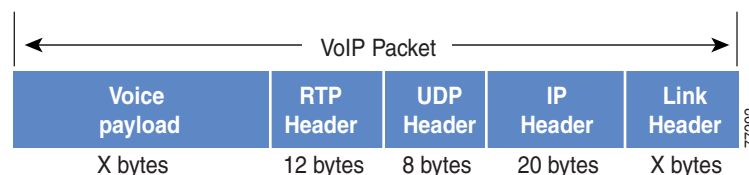
The section describes bandwidth provisioning for the following types of traffic:

- [Voice Bearer Traffic, page 3-51](#)
- [Video Bearer Traffic, page 3-53](#)

Voice Bearer Traffic

As illustrated in [Figure 3-18](#), 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 3-18 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 3-8](#) details the bandwidth per VoIP flow at a default packet rate of 50 packets per second (pps). [Table 3-8](#) 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 Unified CM Administration to adjust the packet rate.

[Table 3-8](#) 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 3-8 *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
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 3-9](#) lists the amount of bandwidth consumed by voice traffic when the Layer 2 headers are included in the calculations.

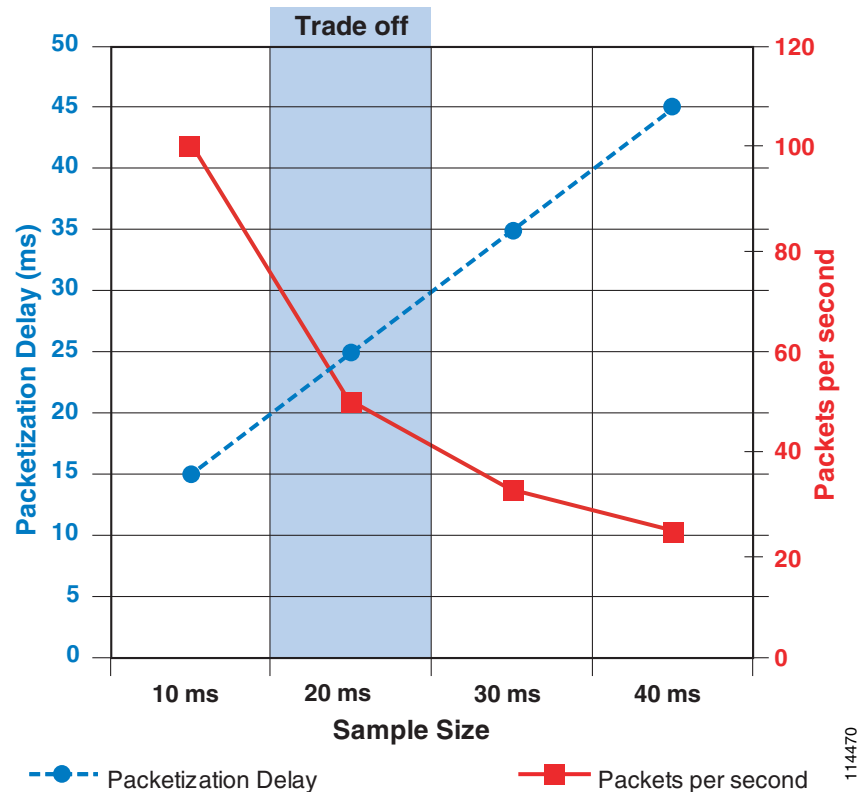
Table 3-9 *Bandwidth Consumption with Layer 2 Headers Included*

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 3-19](#), 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 3-19 Voice Sample Size: Packets per Second vs. Packetization Delay



Video Bearer Traffic

For audio, it is relatively easy to calculate a percentage of overhead per packet given the sample size of each packet. For video, however, it is nearly impossible to calculate an exact percentage of overhead because the payload varies depending upon how much motion is present in the video (that is, how many pixels changed since the last frame).

To resolve this inability to calculate the exact overhead ratio for video, Cisco recommends that you add 20% to the call speed regardless of which type of Layer-2 medium the packets are traversing. The additional 20% gives plenty of headroom to allow for the differences between Ethernet, ATM, Frame Relay, PPP, HDLC, and other transport protocols, as well as some cushion for the bursty nature of video traffic. (See [Table 3-10](#).)

Table 3-10 Recommended Bandwidth for Various Video Call Speeds

Call Speed Requested by Endpoint	Actual Layer-2 Bandwidth Required
128 kbps	153.6 kbps
256 kbps	307.2 kbps

Table 3-10 Recommended Bandwidth for Various Video Call Speeds (continued)

Call Speed Requested by Endpoint	Actual Layer-2 Bandwidth Required
384 kbps	460.8 kbps
512 kbps	614.4 kbps
768 kbps	921.6 kbps
1.5 Mbps	1.766 Mbps
2.048 Mbps	2.458 Mbps
7 Mbps	8.4 Mbps

Note that the values in [Table 3-10](#) represent the maximum burst speed of the call, with some additional amount for a cushion. The average speed of the call is typically much less than these values. The concepts of media channels and bandwidth usage are critical to understanding what values to use when configuring call admission control.

Additional Considerations for Bearer Traffic with RSVP

RSVP was built to support requesting a Quality of Service (QoS) for any traffic flow, not just voice or video, across a wide range of Layer 2 technologies. To accomplish this, RSVP must be able to describe in detail the traffic flow for which it is requesting QoS so that the intermediate routers can make admittance decisions correctly.

The RSVP Sender generates Path messages containing the stream's bandwidth requirements, in the form of a traffic specification (Tspec). The Tspec, as described in RFC 2210, contains the following attributes to describe the traffic flow in detail:

- Token Bucket Rate [r] — The average traffic rate in bytes per second
- Token Bucket Size [b] — The maximum burst of a flow in bytes
- Peak Data Rate [p] — The peak traffic rate in bytes per second
- Minimum Policed Unit [m]
- Maximum Packet Size [M]

Calculating RSVP Bandwidth Values for Use with Unified CM

At the time Unified CM instructs the Cisco RSVP Agent to make the initial reservation for the call flow, the endpoints that are involved in the call have not fully exchanged their codec capabilities. Without this information, Unified CM must rely on the region settings to determine how to describe the traffic flow. The size of the traffic flow is a function of two things, the codec bit-rate and the sampling rate (or packets per second). The region settings contain the codec but do not describe the sampling rate. The preferred sampling rates for G.729 and G.711 audio codecs are defined in the following clusterwide service parameters:

- Preferred G711 millisecond packet size: 20 ms by default
- Preferred G729 millisecond packet size: 20 ms by default

However, the codec sampling rate is negotiated for every call and might not be the preferred setting because it is not supported on one or more of the endpoints. To avoid having to increase the reservation size once the capabilities are fully exchanged, possibly causing a post-ring failure, this initial reservation

is for the worst-case scenario, or the smallest packet size, of that codec. Once media capabilities have been exchanged between the endpoints, then the reservation is revised to the correct bandwidth allocation. In most cases, the default sampling rate is used, resulting in the reservation being reduced.

**Note**

Unified CM does not include SRTP overhead or the Layer 2 overhead in the RSVP Reservation. When compared to the RSVP bandwidth value, the priority queue must be over-provisioned. (See [Table 3-9](#) and [Table 3-10](#).)

Voice Bearer Traffic

Inter-region call with audio codec set to G729:

- Initial request: 40 kbps, using a 10 ms worst-case scenario
- Updated request: 24 kbps, using the preferred sample size of 20 ms

Inter-region call with audio codec set to G711:

- Initial request: 96 kbps, using a 10 ms worst-case scenario
- Updated request: 80 kbps, using the preferred sample size of 20 ms

Video Bearer Traffic

As with the audio stream, the initial reservation for the video stream will rely on the region settings because the endpoint codec capabilities will not be fully negotiated at the time of the reservation. The region settings for video calls include the bandwidth for the audio stream. (See [IP Video Telephony, page 16-1](#), for more information.) Because the audio stream has its own reservation, the final reservation for the video stream will be the region setting minus the audio codec bit-rate. However, because these codecs have not been fully negotiated, the video stream reservation will be for the worst-case scenario, which assumes no audio stream. Once media capabilities have been exchanged between the endpoints, then the reservation will be revised to the correct bandwidth allocation.

Because video is inherently bursty, it is necessary to add some overhead to the stream requirements. (See [Video Bearer Traffic, page 3-53](#), for more information.) Unified CM uses the stream bandwidth to determine how to calculate the overhead, as follows:

- If the stream is < 256 kbps, then the overhead will be 20%
- If the stream is $\geq$ 256 kbps, then the overhead will be 7%

Inter-region video call, with G.729 audio codec and video setting of 384 kbps:

- Initial request: $384 * 1.07 = 410$ kbps
- Updated request: $(384 - 8) * 1.07 = 402$ kbps

Inter-region video call, with G.711 audio codec and video setting of 384 kbps:

- Initial request: $384 * 1.07 = 410$ kbps
- Updated request: $(384 - 64) * 1.07 = 342$ kbps

Configuration Recommendation

Because the initial reservation will be larger than the actual packet flow, over-provisioning the RSVP and LLQ bandwidth is required to ensure that the desired number of calls can complete.

When provisioning the RSVP bandwidth value for N calls, Cisco recommends that the Nth value be the worst-case bandwidth to ensure that the Nth call gets admitted.

For example:

- To provision four G.729 streams:
 $(3 * 24) + 40 = 112 \text{ kbps}$
- To provision four G.711 streams:
 $(3 * 80) + 96 = 336 \text{ kbps}$
- To provision four 384 kbps video streams (G.729 audio)
 $(3 * (384 - 8) + 384) * 1.07 = 1618 \text{ kbps}$
- To provision four 384 kbps video streams (G.711 audio)
 $(3 * (384 - 64) + 384) * 1.07 = 1438 \text{ kbps}$

Configuring Cisco IOS Application ID Support

RSVP Application ID feature support was introduced in Cisco IOS Release 12.4(6)T, and that is the minimum release required for the following examples.

Combined Priority Queue

To utilize the functionality allowed in Unified CM's implementation of Application ID support (that is, allowing voice calls to consume all the bandwidth available in the priority queue), we must modify the previous recommendations that voice and video priority queues be kept separate. (See [Application ID Call Admission Control Model, page 3-48.](#)) To use this functionality, you should combine both the voice and video match criteria into one class-map. Because the requirements are to match either voice traffic or video traffic, be sure to make the class-map match criteria **match-any** instead of **match-all**, as follows:

```
class-map match-any IPC-RTP
  match ip dscp ef
  match ip dscp af41 af42
```

Configure the priority queue to support both the voice and video traffic. The following example configuration allocates 33% of the link bandwidth to the priority queue:

```
policy-map Voice-Policy
  class IPC-RTP
    priority percent 33
```

Mapping Application ID to RSVP Policy Identities

The RSVP Local Policy provides the mechanism for controlling a reservation based on an Application ID. Application IDs are mapped to RSVP Local Policies through the **ip rsvp policy identity** command. RSVP Local Policy identities are defined globally and are available to each interface for policy enforcement. Each identity can have one policy locator defined to match an Application ID.

To give the user as much flexibility as possible in matching application policy locators to local policies, the RSVP local policy command line interface (CLI) accepts application ID match criteria in the form of Unix-style regular expressions for the policy locator. Regular expressions are already used in the CLI for existing Cisco IOS components such as Border Gateway Protocol (BGP). Refer to the follow documentation for more information on how regular expressions are used in Cisco IOS:

- *Access and Communication Servers Command Reference*
http://www.cisco.com/en/US/docs/ios/11_0/access/command/reference/arbook.html

- *Using Regular Expressions in BGP*
http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094a92.shtml
- *Regex Engine Performance Enhancement*
http://www.cisco.com/en/US/docs/ios/12_3t/12_3t4/feature/guide/gt_rexpe.html

RSVP Policy Identities for Matching the Default Unified CM Application IDs

```
ip rsvp policy identity rsvp-video policy-locator .*VideoStream.*
ip rsvp policy identity rsvp-voice policy-locator .*AudioStream.*
```

Interface RSVP Local Policies

Whether configuring Application ID support or not, for an interface to support RSVP, you must configure the **ip rsvp bandwidth <value>** command on that interface. This value cannot be exceeded by any one RSVP reservation or the sum of RSVP reservations on that interface, regardless of Application ID support. In fact, if a reservation passes the local policy check, it still must pass the interface RSVP bandwidth check before it is reserved.

Local policies based on Application ID are applied to an interface using the **ip rsvp policy local identity** command.

For reservations that match its policy locator value, a local policy has the ability to perform the following functions:

- Define the maximum amount of bandwidth the reservations can reserve as a group or as a single sender
- Forward or not forward RSVP messages
- Accept or not accept RSVP messages
- Define the maximum bandwidth the group or sender can reserve

For example, to limit the amount of video bandwidth to 384 kbps on a Serial T1, use the following commands:

```
interface Serial0/0/1:0
ip rsvp bandwidth 506
ip rsvp policy local identity rsvp-video
maximum bandwidth group 384
forward all
```

There is also a catch-all local policy called the default local policy. This local policy will match any RSVP reservation that did not match the other RSVP local policies configured on the link. The default local policy can be used to match reservations that are not tagged with an Application ID or reservations that are tagged with an Application ID that you want to treat as untagged traffic.

Example

The following example supports both voice and video calls using the model discussed in [How Unified CM Uses the Application ID, page 3-48](#). The voice calls are guaranteed 352 kbps of bandwidth while video calls are limited to 154 kbps of bandwidth. Voice calls can use all of the available RSVP bandwidth.

```
interface Serial0/0/1:0
ip address 10.2.101.5 255.255.255.252
service-policy output Voice-Policy
ip rsvp bandwidth 506
ip rsvp data-packet classification none
ip rsvp resource-provider none
ip rsvp policy local identity rsvp-voice
```

```
maximum bandwidth group 506
forward all
ip rsvp policy local identity rsvp-video
maximum bandwidth group 154
forward all
ip rsvp policy local default
no accept all ! Will not show in the configuration
no forward all! Will not show in the configuration
```

In this example, if an RSVP reservation is received that does not have an Application ID or its Application ID does not match the two configured options, the reservation will fail. This configuration works if RSVP traffic originates only from Cisco RSVP Agents controlled by Unified CM. However, if there is intercluster RSVP traffic via an IP-IP gateway or if RSVP messages from a controller other than Unified CM are traversing this link, then the default local policy should be configured to accept and forward the reservations and a maximum bandwidth value should be configured on the policy. Note that it is possible to oversubscribe the RSVP bandwidth via the use of multiple RSVP local policies (if the sum of the policies is greater than the RSVP interface bandwidth), but reservations then become first-come, first-serve.

Provisioning for Call Control Traffic with Centralized Call Processing

In a centralized call processing deployment, the Unified CM cluster and the applications (such as voicemail) are located at the central site, while several remote sites are connected through an IP WAN. The remote sites rely on the centralized Unified CMs 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 Unified CM 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), SIP (encrypted and non-encrypted), H.323, MGCP, and CTI-QBE. All the control traffic is exchanged between a Unified CM at the central site and endpoints or gateways at the remote branches.
- If RSVP is deployed within the cluster, the control traffic between the Unified CM cluster at the central site and the Cisco RSVP Agents at the remote sites uses the SCCP protocol.

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 endpoints (phones, gateways, and Cisco RSVP Agents) and Unified CM, regardless of call activity. This traffic is a function of the quantity of endpoints.
- Call-related traffic, which consists of signaling messages exchanged between the branch endpoints and the Unified CM at the central site when a call needs to be set up, torn down, forwarded, and so forth. This traffic is a function of the quantity of endpoints and their associated call volume.

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 3-60](#).

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

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

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

$$\text{Bandwidth (bps)} = 538 * (\text{Number of IP phones and gateways in the branch})$$

If a site features a mix of SCCP and SIP endpoints, the two equations above should be employed separately for the quantity of each type of phone used, and the results added.

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 this 25% 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 Unified CM and the endpoints. The following formula takes into account the impact of signaling encryption:

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

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

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

$$\text{Bandwidth with signaling encryption (bps)} = 619 * (\text{Number of IP phones and gateways in the branch})$$

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 3-11](#).

Table 3-11 Recommended Layer 3 Bandwidth for Call Control Traffic With and Without Signaling Encryption

Branch Office Size (Number of IP Phones and Gateways)	Recommended Bandwidth for SCCP Control Traffic (no encryption)	Recommended Bandwidth for SCCP Control Traffic (with encryption)	Recommended Bandwidth for SIP Control Traffic (no encryption)	Recommended Bandwidth for SIP Control Traffic (with encryption)
1 to 10	8 kbps	8 kbps	8 kbps	8 kbps
20	8 kbps	9 kbps	11 kbps	12 kbps
30	8 kbps	13 kbps	16 kbps	19 kbps
40	11 kbps	17 kbps	22 kbps	25 kbps
50	14 kbps	21 kbps	27 kbps	31 kbps
60	16 kbps	25 kbps	32 kbps	37 kbps
70	19 kbps	29 kbps	38 kbps	43 kbps
80	21 kbps	33 kbps	43 kbps	49 kbps
90	24 kbps	38 kbps	48 kbps	56 kbps
100	27 kbps	42 kbps	54 kbps	62 kbps
110	29 kbps	46 kbps	59 kbps	68 kbps
120	32 kbps	50 kbps	65 kbps	74 kbps
130	35 kbps	54 kbps	70 kbps	80 kbps
140	37 kbps	58 kbps	75 kbps	87 kbps
150	40 kbps	62 kbps	81 kbps	93 kbps

**Note**

Table 3-11 assumes 10 calls per hour per phone, and it does not include RSVP control traffic. To determine the RSVP-related bandwidth to add to the values in this table, see [Considerations for Calls Using RSVP](#), page 3-61.

**Note**

If an RSVP-based locations policy is used for inter-site calls, the values of Table 3-11 must be increased to compensate for the control traffic of the Cisco RSVP Agent. For example, if 10% of the calls go over the WAN, multiply the value from Table 3-11 by 1.1.

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 3A: Recommended Bandwidth Needed for SCCP Control Traffic 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 3B: Recommended Bandwidth Needed for SIP Control Traffic for a Branch with No Signaling Encryption.

$$\text{Bandwidth (bps)} = (138 + 40 * CH) * (\text{Number of IP phones and gateways in the branch})$$

Equation 4A: Recommended Bandwidth Needed for SCCP Control Traffic for a Remote Branch with Signaling Encryption.

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

Equation 4B: Recommended Bandwidth Needed for SIP Control Traffic for a Remote Branch with Signaling Encryption.

$$\text{Bandwidth with signaling encryption (bps)} = (159 + 46 * CH) * (\text{Number of IP phones and gateways in the branch})$$



Note

Equations 3A and 4A are based on the default SCCP keep-alive period of 30 seconds, while equations 3B and 4B are based on the default SIP keep-alive period of 120 seconds.

Considerations for Calls Using RSVP

In systems where call admission control uses RSVP, there is additional SCCP call control traffic between Unified CM and the Cisco RSVP Agents located at the branch when IP calls are placed across the WAN. To compute the associated bandwidth, use the following equation:

Equation 5: Recommended Bandwidth Needed for SCCP Control Traffic for Cisco RSVP Agents.

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

Where CHW represents the number of calls placed across the IP WAN per hour per phone, including calls between IP phones at different branches as well as calls made through gateways located in a different site. For example, in a site where 20 phones each make 10 calls per hour, if 20% of the calls are placed across the IP WAN, then CHW = 2. The equation thus yields: $(21 * 2) * 20 = 840$ bps.

The bandwidth calculated by Equation 5 should be added to the required bandwidth for phone call control.

Considerations for Shared Line Appearances

Calls placed to shared line appearances, or calls sent to line groups using the Broadcast distribution algorithm, have two net effects on the bandwidth consumed by the system:

- Because all the phones on which the line is configured ring simultaneously, they represent a load on the system corresponding to a much higher calls-per-hour (CH) value than the CH of the line. The corresponding bandwidth consumption is therefore increased. The network infrastructure's bandwidth provisioning requires adjustments when WAN-connected shared line functionality is deployed. The CH value employed for Equations 3 and 4 must be increased according to the following formula:

$$CHS = CHL * (\text{Number line appearances}) / (\text{Number of lines})$$

Where CHS is the shared-line calls per hour to be used in Equations 3 and 4, and CHL is the calls-per-hour rating of the line. For example, if a site is configured with 5 lines making an average of 6 calls per hour but 2 of those lines are shared across 4 different phones, then:

$$\text{Number of lines} = 5$$

$$\text{Number of line appearances} = (2 \text{ lines appear on } 4 \text{ phones, and } 3 \text{ lines appear on only one phone}) = (2 * 4) + 3 = 11 \text{ line appearances}$$

$$CHL = 6$$

$$\text{CHS} = 6 * (11 / 5) = 13.2$$

- Because each of the ringing phones requires a separate signaling control stream, the quantity of packets sent from Unified CM to the same branch is increased in linear proportion to the quantity of phones ringing. Because Unified CM is attached to the network through a 100 Mbps interface, it can instantaneously generate a very large quantity of packets that must be buffered while the queuing mechanism is servicing the signaling traffic. The servicing speed is limited by the WAN interface's effective information transfer speed, which is typically two orders of magnitude smaller than 100 Mbps.

This traffic may overwhelm the queue depth of the central site's WAN router. By default, the queue depth available for each of the classes of traffic in Cisco IOS is 64. In order to prevent any packets from being dropped before they are queued for the WAN interface, you must ensure that the signaling queue's depth is sized to hold all the packets from at least one full shared-line event for each shared-line phone. Avoiding drops is paramount in ensuring that the call does not create a race condition where dropped packets are retransmitted, causing system response times to suffer.

Therefore, the quantity of packets required to operate shared-line phones is as follows:

- SCCP protocol with Cisco Unified CM 4.x: 14 packets per shared-line phone
- SCCP protocol with Cisco Unified CM 5.0: 10 packets per shared-line phone
- SIP protocol with Cisco Unified CM 5.0: 6 packets per shared-line phone

For example, with Unified CM 4.1 and using 5 phones sharing the same line, the queue depth for the signaling class of traffic must be adjusted to a minimum of 70. [Table 3-12](#) provides recommended queue depths based on the quantity of shared line appearances within a branch site.

Table 3-12 Recommended Queue Depth per Branch Site

Number of Shared Line Appearances	SCCP		SIP
	Unified CM 4.x	Unified CM 5.0	Unified CM 5.0
	Queue Depth (Packets)	Queue Depth (Packets)	Queue Depth (Packets)
5	70	64	64
10	140	100	64
15	210	150	90
20	280	200	120
25	350	250	150

When using a Layer 2 WAN technology such as Frame Relay, this adjustment must be made on the circuit corresponding to the branch where the shared-line phones are located.

When using a Layer 3 WAN technology such as MPLS, there may be a single signaling queue servicing multiple branches. In this case, adjustment must be made for the total of all branches serviced. For example, with Unified CM 4.1, if 10 branches each require 5 phones sharing a line, the central site's WAN router must be adjusted to have a signaling queue depth of 700.

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 Unified CM 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 Unified CM clusters at each site, as well as between the Unified CM clusters themselves.

Therefore, bandwidth for control traffic must be provisioned on the WAN links between Unified CMs as well as between each Unified CM 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 Unified CM and the gatekeeper
- Call-related traffic, which in turn consists of two types of traffic:
 - Call admission control traffic, exchanged between the Unified CMs and the infrastructure control point (such as a gatekeeper or Cisco RSVP Agent) before a call can be set up and after it has been torn down.
 - Signaling traffic associated with a media stream, exchanged over an intercluster trunk 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.

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 Unified Wireless IP Phones 7920 and 7921G, 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 3-64](#)
- [WLAN Quality of Service \(QoS\), page 3-70](#)

For more information about WLAN and Voice over WLAN (VoWLAN) design, refer to the latest version of the *Enterprise Mobility Design Guide*, available at

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

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 3-64](#)
- [Wireless AP Configuration and Design, page 3-67](#)
- [Wireless Security, page 3-68](#)

Wireless Infrastructure Considerations

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

- [VLANs, page 3-64](#)
- [Roaming, page 3-65](#)
- [Wireless Channels, page 3-65](#)
- [Wireless Interference, page 3-66](#)
- [Multicast on the WLAN, page 3-66](#)

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 should 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

When devices roam at Layer 3, they move from one AP to another AP across native VLAN boundaries. When the WLAN network infrastructure consists of autonomous APs, the Cisco Catalyst 6500 Series Wireless LAN Services Module (WLSM) allows the Cisco Unified Wireless IP Phone to keep its IP address and roam at Layer 3 while still maintaining an active call. Seamless Layer 3 roaming occurs only when the client is roaming within the same mobility group. For details about the Cisco WLSM and Layer 3 roaming, refer to the Cisco WLSM product documentation available at

<http://www.cisco.com>

Seamless Layer 3 roaming for clients across a lightweight access point infrastructure is accomplished by WLAN controllers that use dynamic interface tunneling. Cisco Unified Wireless IP Phones that roam across WLAN controllers and VLANs can keep their IP address when using the same SSID and therefore can maintain an active call.

**Note**

In dual-band WLANs (those with 2.4 GHz and 5 GHz bands), it is possible to roam between 802.11b/g and 802.11a with the same SSID, provided the client is capable of supporting both bands. However, this can cause gaps in the voice path. In order to avoid these gaps, use only one band for voice.

**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 and 802.11g requires a minimum of five-channel separation between configured channels to prevent interference or overlap between channels. In North America, with allowable channels of 1 to 11, channels 1, 6, and 11 are the three usable non-overlapping channels for APs and wireless endpoint devices. However, in Europe where the allowable channels are 1 to 13, multiple combinations of five-channel separation are possible. Multiple combinations of five-channel separation are also possible in Japan, where the allowable channels are 1 to 14.

Optimal channel configuration for 5 GHz 802.11a requires a minimum of one-channel separation to prevent interference or overlap between channels. In North America, there are 20 possible non-overlapping channels: 36, 40, 44, 48, 52, 56, 60, 64, 100, 104, 108, 112, 116, 132, 136, 140, 149, 153, 157, and 161. In Europe, the same non-overlapping channels are allowed. However, many countries do not support the use of channel 40, so there are only 19 possible overlapping channels. In Japan, only 8 non-overlapping channels are supported: 36, 40, 44, 48, 52, 56, 60, and 64. Because of the larger set of non-overlapping channels, 802.11a allows for more densely deployed WLANs.

Note that the 802.11a band does require support for Dynamic Frequency Selection (DFS) and Transmit Power Control (TPC) on some channels in order to avoid interference with radar (military, satellite, and weather). Regulations require that channels 52 to 64, 100 to 116, and 132 to 140 support DFS and TPC. TPC ensures that transmissions on these channels are not powerful enough to cause interference. DFS monitors channels for radar pulses and, when it detects a radar pulse, DFS stops transmission on the channel and switches to a new channel.

AP coverage should be deployed so that no (or minimal) overlap occurs between APs configured with the same channel. Same channel overlap should typically occur at 19 dBm separation. However, proper AP deployment and coverage on non-overlapping channels require a minimum overlap of 20%. This amount of overlap ensures smooth roaming for wireless endpoints as they move between AP coverage cells. Overlap of less than 20% can result in slower roaming times and poor voice quality.

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. Both the 2.4 GHz and 5.0 GHz wave forms of 802.11 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.

**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

By design, multicast does not have the acknowledgement level of unicast. According to 802.11 specifications, the access point must buffer all multicast packets until the next Delivery Traffic Indicator Message (DTIM) period is met. The DTIM period is a multiple of the beacon period. If the beacon period is 100 ms (typical default) and the DTIM value is 2, then the access point must wait up to 200 ms before transmitting a single buffered multicast packet. The time period between beacons (as a product of the DTIM setting) is used by battery-powered devices to go into power save mode temporarily. This power save mode helps the device conserve battery power.

Multicast on WLAN presents a twofold problem in which administrators must weigh multicast traffic quality requirements against battery life requirements. First, delaying multicast packets will negatively affect multicast traffic quality, especially for applications that multicast real-time traffic such as voice. In order to limit the delay of multicast traffic, DTIM periods should typically be set to a value of 1 so that the amount of time multicast packets are buffered is low enough to eliminate any perceptible delay in multicast traffic delivery. However, by setting the DTIM period to a value of 1, the amount of time that battery-powered WLAN devices are able to go into power save mode is shortened, and therefore battery life is shortened. In order to conserve battery power and lengthen battery life, DTIM periods should typically be set to a value of 2 or more.

For WLAN networks with no multicast applications or traffic, the DTIM period should be set to a value of 2 or higher. For WLAN networks where multicast applications are present, the DTIM period should be set to a value of 2 whenever possible; however, if multicast traffic quality suffers or if unacceptable delay occurs, then the DTIM value should be lowered to 1. If the DTIM value is set to 1, administrators must keep in mind that battery life of battery-operated devices will be significantly shortened.

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

For additional considerations with multicast traffic, see [Music on Hold, page 7-1](#).

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 1000 Series Lightweight AP
- Aironet 1100 Series AP
- Aironet 1130AG Series AP
- Aironet 1200 Series AP
- Aironet 1230AG Series AP
- Aironet 1240AG Series AP
- Aironet 1300 Series AP
- Aironet 1500 Series AP

For these APs, Cisco IOS Release 12.3(4) JA or later is recommended.

AP Deployment

When you use Cisco access points (APs) for voice deployments, Cisco recommends that you do not associate more than 15 to 25 devices to a single 802.11b or 802.11b/g AP at any given time. For 802.11a or 802.11a/g APs, Cisco recommends that you do not associate more than 45 to 50 devices to a single AP. These numbers will vary depending on usage profiles and the enabled data rates. 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 devices than specified above can result in poor AP performance and slower response times for associated devices.

While there is no specific mechanism to ensure that only a limited number of devices are associated to a single AP, system administrators can manage device-to-AP ratios by conducting periodic site surveys and analyzing user and device traffic patterns. If additional devices and users are added to the network in a particular area, additional site surveys should be conducted to determine whether additional APs are required to handle the number of endpoints that need to access the network.

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.

- Enable Dynamic Transmit Power Control (DTPC) on the AP.

This ensures that the transmit power on the AP and on the voice endpoints match. Matching transmit power helps eliminate the possibility of one-way audio traffic. Voice endpoints adjust their transmit power based on the Limit Client Power (mW) setting of the AP to which they are associated.

- 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. Beginning with Cisco IOS Release 12.3(7)JA, the AP also provides 802.11e clear channel assessment (CCA) QBSS in beacons. The CCA-based QBSS values reflect true channel utilization.

- 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 3-71, 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:

- Extensible Authentication Protocol-Flexible Authentication via Secure Tunneling (EAP-FAST)

A standards-based security protocol that first requires the establishment of an authenticated tunnel between the wireless client and an Authentication, Authorization, and Accounting (AAA) server using a Protected Access Credential (PAC). Next the wireless endpoint authenticates across the tunnel using a user name and password to authenticate with the network via 802.1X. Once this authentication occurs, traffic to and from the wireless device is encrypted using TKIP or WEP. Using the 802.1X authentication 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.

- **Wi-Fi Protected Access (WPA)**

A standards-based security protocol requires the wireless endpoint to provide a user name and password to authenticate with the network. Once this authentication occurs using either 802.1X or WPA Pre-Shared Key (WPA-PSK), traffic to and from the wireless device is encrypted using Temporal Key Integrity Protocol (TKIP). Using the 802.1X authentication 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.

- **Wi-Fi Protected Access 2 (WPA2)**

This is the 802.11i enhanced version of WPA. It is a standards-based security protocol that requires the wireless endpoint to provide a user name and password to authenticate with the network. Once this authentication occurs using either 802.1X or Pre-Shared Key (WPA2-PSK), traffic to and from the wireless device is encrypted using Advanced Encryption Standards (AES). Using the 802.1X authentication 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**

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.

- **Static Wired 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.

Authentication and ACS Deployment Models

Extensible Authentication Protocol (EAP) 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 EAP-FAST, WPA, or 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 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 3-70](#)
- [Interface Queuing, page 3-71](#)
- [Bandwidth Provisioning, page 3-71](#)

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 3-5](#).

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 24 (or PHB CS3). 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 traffic 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.

For upstream queuing, devices that support Wi-Fi Multimedia (WMM) are able to take advantage of queueing mechanisms, including priority queueing. However, devices that do not support WMM are unable to take advantage of this queueing mechanism. The Cisco Wireless IP Phone 7921G does support WMM. The Cisco Wireless IP Phone 7920 does not support WMM; however, it can provide queuing upstream as the packets leave the device, but there is no mechanism in place to provide queuing among all clients on the wireless LAN because wireless networks are a shared medium.

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.

In order to set up this two-queue configuration for autonomous APs, 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 the packet should receive.

For lightweight APs, the WLAN controller has built-in QoS profiles that can provide the same queuing policy. Voice VLAN or voice traffic is configured to use the **Platinum** policy, which sets priority queueing for the voice queue. Data VLAN or data traffic is configured to use the **Silver** policy, which sets best-effort queueing for the Data queue. These policies are then assigned to the incoming and outgoing radio interfaces based on the VLAN.

The above configurations ensure that all voice media and signaling are given priority queueing treatment in a downstream direction.

Bandwidth Provisioning

Based on wireless voice network testing, Cisco has determined that an 802.11b-only AP with 802.11b clients and a data rate of 11 Mbps can support a maximum of seven active G.711 voice streams or eight G.729 streams. AP rates set lower than 11 Mbps will result in a lower call capacity per AP.

With 802.11a at a data rate of 54 Mbps, the maximum number of active voice streams increases to between 14 and 18 per AP.

For 802.11g environments with a data rate of 54 Mbps, in theory the maximum number of active voice streams also increases to between 14 and 18 per AP. However, because most 802.11g environments are mixed and include 802.11b clients (and therefore 11 Mbps data rates) as well as 802.11g clients, capacity is typically significantly lower, with a maximum of 8 to 12 active voice streams per AP.

**Note**

A call between two phones associated to the same AP counts as two active voice streams.

To prevent these limits from being exceeded, some form of call admission control is required. Cisco APs and wireless voice clients have two mechanisms that are used for call admission control:

- QoS Basic Service Set (QBSS)

QBSS is the beacon information element that enables the AP to communicate channel utilization information to the wireless IP phone. This QBSS value helps wireless phones to determine whether they 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 device should not roam to this AP. While this QBSS information is useful, it is not a true call admission control mechanism because it does not guarantee that calls will retain proper QoS or that there is enough bandwidth to handle the call. When a Cisco Unified Wireless 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 device. 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 oversubscription of the available bandwidth on that AP.

- Wi-Fi Multimedia Traffic Specification (WMM TSPEC)

WMM TSPEC is the QoS mechanism that enables WLAN clients to provide an indication of their bandwidth and QoS requirements so that APs can react to those requirements. When a client is preparing to make a call, it sends an Add Traffic Stream (ADDTS) message to the AP with which it is associated, indicating the TSPEC. The AP can then accept or reject the ADDTS request based on whether bandwidth and priority treatment are available. If the call is rejected, the phone will receive a Network Busy message. When roaming, mid-call clients supporting TSPEC will send a ADDTS message to the new AP as part of the association process to ensure that there is available bandwidth for priority treatment. If there is not enough bandwidth, the roam can be load-balanced to a neighboring AP if one is available.

The Cisco Unified Wireless IP Phone 7920 supports only QBSS, so this is the only mechanism that can be used for call admission control with these devices. However, the Cisco Unified Wireless IP Phone 7921G supports both QBSS and TSPEC. (TSPEC takes precedence over QBSS.) Therefore call admission control with the Cisco Unified Wireless IP Phone 7921G, when using TSPEC, is more accurate and eliminates the possibility of exceeding AP call capacities.

**Note**

Beginning with Cisco IOS Release 12.3(7)JA, the AP sends 802.11e CCA-based QBSS. These QBSS values represent true channel utilization for a particular AP.

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 3-67.](#))



CHAPTER 4

Gateways

Last revised on: October 30, 2008

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 gateway to provide the appropriate protocol and feature support for your IP Telephony network. The main topics discussed in this chapter include:

- [Traffic Patterns and Gateway Sizing, page 4-2](#)
- [Understanding Cisco Gateways, page 4-6](#)
- [Gateway Selection, page 4-7](#)
- [QSIG Support, page 4-24](#)
- [Fax and Modem Support, page 4-25](#)
- [Gateways for Video Telephony, page 4-36](#)

What's New in This Chapter

[Table 4-1](#) lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 4-1 *New or Changed Information Since the Previous Release of This Document*

New or Revised Topic	Described in:
Video gateways	Gateways for Video Telephony, page 4-36

Traffic Patterns and Gateway Sizing

This section presents a high-level discussion of the differences between various traffic models or patterns and how they can affect voice gateway selection. The emphasis is on traffic patterns and gateway sizing for traffic-intensive deployments.

Definitions and Terminology

This section uses the following terms and definitions:

- Simultaneous calls

The number of calls that are all active in the system at the same time.

- Maximum simultaneous calls

The maximum number of simultaneous calls in active (talk) state that the system can handle. The number of calls expected to be active simultaneously during the *busy hour* of the day should not exceed this number.

- Calls per second (cps)

The call arrival rate, described as the number of calls that arrive (that is, new call setup attempts) in one second. Call arrival rates are also often quoted in calls per hour, but this metric is looser in the sense that 100 calls arriving in the last five seconds of an hour provides an average call arrival rate of 100 calls per hour (which is an extremely low rate for a communications system), while it also provides an arrival rate of 20 calls per second (which is a high rate). Sustaining 20 calls per second for an entire hour would result in 72,000 calls per hour. Therefore calls-per-hour is not a very useful metric for ascertaining a system's ability to handle bursty call arrival traffic patterns.

- Busy Hour Call Attempts (BHCA)

The number of calls attempted during the busiest hour of the day (the peak hour). This is the same as the calls-per-second rating for the busiest hour of the day, but it is expressed over a period of an hour rather than a second. For example, 10 cps would be equal to 3600 calls per hour. There is also a metric for Busy Hour Call Completions (BHCC), which can be lower than the BHCA (call attempts) under the assumption that not all calls are successful (as when a blocking factor exists). This chapter assumes 100% call completions, so that BHCA = BHCC.

- Bursty traffic

Steady arrival means the call attempts are spaced more or less equally over a period of time. For example, 60 calls per hour at a steady arrival rate would present one call attempt roughly every minute (or approximately 0.02 cps). With bursty arrival, the calls arriving over a given period of time (such as an hour) are not spaced equally but are clumped together in one or more spikes. In the worst case, an arrival rate of 60 calls per hour could offer all 60 calls in a single second of the hour, thus averaging 0 cps for most of the hour with a peak of 60 cps for that one second. This kind of traffic is extremely stressful to communications systems.

- Hold time

This is the period of "talk time" on a voice call; that is, the period of time between call setup and call teardown when there is an open speech path between the two parties. A hold time of 3 minutes (180 seconds) is an industry average used for traffic engineering of voice systems. The shorter the hold time on the average call, the greater the percentage of system CPU time spent on setting up and tearing down calls compared to the CPU time spent on maintaining the speech path.

PSTN Traffic Patterns

Traffic, when used in the context of voice communication systems, refers to the volume of calls being sent and/or received. Of particular importance is the traffic carried by external circuits such as the public switched telephone network (PSTN). Traffic is measured in Erlangs, and an Erlang is defined as one call lasting for one hour. This section does not go into any further detail on Erlangs other than to say that there are mathematical tables (Erlang-B and Erlang-C) that are used to calculate how many circuits are required for a given amount of offered traffic.

The amount of traffic received and generated by your business determines the size of the external circuits required. However, many customers typically continue to use the same number of circuits for their IP-based communications system as they previously used for a TDM-based system. While this sizing method might work if no issues are encountered, the process of ongoing system traffic analysis should be part of any routine maintenance practices. Traffic analysis can show that the system is over-provisioned for the current levels of traffic (and, therefore, the customer is paying for circuits that are not needed) or, conversely, that the system is under-provisioned and may be suffering from occasional blocked and/or lost calls, in which case increasing the number of circuits will remedy the situation.

Normal Business Traffic Profile

Most customers have a normal traffic profile, which means that they typically have two *busy hours* per day, one occurring during the morning from 10:00 to 11:00 and the other in the afternoon from 14:00 to 15:00. These busy-hour patterns can often be attributed to such things as employees starting the work day or returning from a lunch break. The calls themselves tend to have longer hold times and they tend to arrive and leave in a steady manner. A generally accepted industry average holding time to use for traffic calculations is 3 minutes.

Assuming that the communications system is engineered with the busy-hour traffic in mind, no issues should arise. Engineering a system below these levels will result in blocked and/or lost calls, which can have a detrimental effect on business.

Contact Center Traffic Profile

Contact centers present somewhat different patterns of traffic in that these systems typically handle large volumes of calls for the given number of agents or interactive voice response (IVR) systems available to service them. Contact centers want to get the most out of their resources, therefore their agents, trunks, and IVR systems are kept busy all the while they are in operation, which usually is 24 hours a day. Call queuing is typical (when incoming call traffic exceeds agent capacity, calls wait in queue for the next available agent), and the agents are usually dedicated during their work shifts to taking contact center calls.

Call holding times in contact centers are often of a shorter average duration than normal business calls. Contributing to the shorter average call holding time is the fact that many calls interact only with the IVR system and never need to speak to a human agent (also termed self-service calls). A representative holding time for self-service calls is about 30 seconds, while a call that talks to an agent has an average holding time of 3 minutes (the same as normal business traffic), making the overall average holding time in the contact center shorter than for normal business traffic.

The goal of contact centers to optimize resource use (including IVR ports, PSTN trunks, and human agents), combined with the fact that contact centers are systems dedicated to taking telephone calls, also presents the system with higher call arrival rates than in a typical business environment. These call arrival rates can also peak at different times of day and for different reasons (not the usual busy hour) than normal business traffic. For example, when a television advertisement runs for a particular holiday

package with a 1-800 number, the call arrival rate for the system where those calls are received will experience a peak of traffic for about 15 minutes after the ad airs. This call arrival rate can exceed the average call arrival rate of the contact center by an order of magnitude.

Gateway Sizing for Contact Center Traffic

Short call durations as well as bursty call arrival rates impact the PSTN gateway's ability to process the traffic. Under these circumstances the gateway needs more resources to process all calls in a timely manner, as compared to gateways that receive calls of longer duration that are presented more uniformly over time. Because gateways have varying capabilities to deal with these traffic patterns, careful consideration should be given to selecting the appropriate gateway for the environment in which it will operate. Some gateways support more T1/E1 ports than others, and some are more able than others to deal with multiple calls arriving at the same time.

For a traffic pattern with multiple calls arriving in close proximity to each other (that is, high or bursty call arrival rates), a gateway with a suitable rating of calls per second (cps) is the best fit. Under these conditions, using calls with 15-second hold times, the Cisco AS5400XM Universal Gateway can maintain 20 cps (with 310 calls active at once), the Cisco 3845 Integrated Services Router can maintain 17 cps (with 255 calls active at once), and the Cisco Catalyst 6500 Communication Media Module can maintain 7 cps (with 130 calls active at once). The performance of the Cisco AS5350XM Universal Gateway is identical to that of the AS5400XM in terms of calls per second.

For traffic patterns with a steady arrival rate, the maximum number of active calls that a gateway can handle is generally the more important consideration. Under these conditions, using calls with 180-second hold times, the Cisco AS5400XM Universal Gateway can maintain 630 simultaneously active calls (with a call arrival rate of up to 3.5 cps), the Cisco 3845 Integrated Services Router can maintain 504 simultaneously active calls (with a call arrival rate of up to 3 cps), and the Cisco Catalyst 6500 Communication Media Module can maintain 240 simultaneously active calls (with a call arrival rate of up to 1.3 cps).

These numbers assume that all of the following conditions apply:

- CPU utilization does not exceed 75%.
- PSTN gateway calls are made with ISDN PRI trunks using H.323.
- Real Time Control Protocol (RTCP) timer is set to the default value of 5 seconds.
- Voice Activity Detection (VAD) is off.
- G.711 uses 20 ms packetization.
- Cisco IOS Release 12.4.11T or later is used.
- Dedicated voice gateway configurations are used, with ethernet (GE) egress and no QoS features. (Using QoS-enabled egress interfaces or non-ethernet egress interfaces, or both, will consume additional CPU resources.)
- No supplementary call features or services are enabled – such as general security (for example, access control lists or firewalls), voice-specific security (TLS, IPSec and/or SRTP), AAA lookups, gatekeeper-assisted call setups, VoiceXML or TCL-enabled call flows, call admission control (RSVP), and SNMP polling/logging. Such extra call features will use additional CPU resources.

Voice Activity Detection (VAD)

VAD is a digital signal processing feature that suppresses the creation of most of the IP packets during times when the speech path in a particular direction of the call is perceived as being silent. Typically only one party on a call speaks at a time, so that packets need flow in only one direction, and packets in

the reverse (or silent) direction need not be sent except as an occasional keepalive measure. VAD can therefore provide significant savings in the number of IP packets sent for a VoIP call, and thereby save considerable CPU cycles on the gateway platform. While the actual packet savings that VAD can provide varies with the call flow, the application, and the nature of speaker interactions, it tends to use 10% to 30% fewer packets than would be sent for a call made using a VAD-off configuration.

VAD is most often turned off in endpoints and voice gateways deployed in Unified CM networks; VAD is most often turned on in voice gateways in other types of network deployments.

Codec

Both G.711 and G.729A use as their default configuration a 20 ms sampling time, which results in a 50 packets per second (pps) VoIP call in each direction. While a G.711 IP packet (200 bytes) is larger than a G.729A packet (60 bytes), this difference has not proven to have any significant effect on voice gateway CPU performance. Both G.711 and G.729 packets qualify as "small" IP packets to the router, therefore the packet rate is the salient codec parameter affecting CPU performance.

Performance Overload

Cisco IOS is designed to have some amount of CPU left over during peak processing, to handle interrupt-level events. The performance figures in this section are designed with the processor running at an average load of approximately 75%. If the load on a given Cisco IOS gateway continually exceeds this threshold, the following will result:

- The deployment will not be supported by Cisco Technical Assistance Center (TAC).
- The Cisco IOS Gateway will display anomalous behavior, including Q.921 timeouts, longer post-dial delay, and potentially interface flaps.

Cisco IOS Gateways are designed to handle a short burst of calls, but continual overloading of the recommended call rate (calls per second) is not supported.



Note

With any gateway, you might be tempted to assign unused hardware ports to other tasks, such as on a CMM gateway where traffic calculations have dictated that only a portion of the ports can be used for PSTN traffic. However, the remaining ports *must* remain unused, otherwise the CPU will be driven beyond supported levels.

Performance Tuning

The CPU utilization of a Cisco IOS Voice gateway is affected by every process that is enabled in a chassis. Some of the lowest level processes such as IP routing and memory defragmentation will occur even when there is no live traffic on the chassis.

Lowering the CPU utilization can help to increase the performance of a Cisco IOS Voice Gateway by ensuring that there are enough available CPU resources to process the real-time voice packets and the call setup instructions. Some of the techniques for decreasing CPU utilization are described in [Table 4-2](#).

Table 4-2 **Techniques for Reducing CPU Utilization**

Technique	CPU Savings	Description
Enable Voice Activity Detection (VAD)	Up to 20%	Enabling VAD can result in up to 45% fewer voice packets in typical conversations. The difficulty is that, in scenarios where voice recognition is used or there are long delays, a reduction in voice quality can occur. Voice appears to "pop" in at the beginning and "pop" out at the end of talk spurts.
Disable Real Time Control Protocol (RTCP)	Up to 5%	Disabling RTCP results in less out-of-band information being sent between the originating and terminating gateways. This results in lower quality of statistics displayed on the paired gateway. This can also result in the terminating gateway having a call "hang" for a longer period of time if RTCP packets are being used to determine if a call is no longer active.
Disable other non-essential functions such as: Authentication, Authorization, and Accounting (AAA); Simple Network Management Protocol (SNMP); and logging	Up to 2%	Any of these processes, when not required, can be disabled and will result in lower CPU utilization by freeing up the CPU to provide faster processing of real-time traffic.
Change call pattern to increase the length of the call (and reduce the number of calls per second)	Varies	This can be done by a variety of techniques such as including a long(er) introduction prompt played at the beginning of a call or adjusting the call script at the call center.

Additional Information

For more information on Cisco Voice Gateway capabilities and call center traffic analysis, refer to the following sources:

- *Cisco Voice Gateway Router Interoperability with Cisco Unified CallManager* data sheet (Table 7):
http://www.cisco.com/en/US/prod/collateral/routers/ps259/product_data_sheet0900aecd8057f2e0.pdf
- *Cisco AS5400XM Universal Gateway* data sheet (Table 9):
http://www.cisco.com/en/US/products/hw/univgate/ps505/products_data_sheet0900aecd802efc92.html
- *How to Order a Cisco AS5400XM Universal Gateway*:
http://www.cisco.com/en/US/products/hw/univgate/ps505/prod_brochure0900aecd802f6ece.html
- Various voice traffic calculators, including Erlang calculators:
<http://www.erlang.com/calculator/>

Understanding Cisco Gateways

Cisco access gateways enable Cisco Unified Communications Manager (Unified CM) 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 Unified CM 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 Unified CM 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 Unified CM 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 use to connect to certain legacy voice mail systems.

Gateway Selection

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

- [Core Feature Requirements, page 4-7](#)
- [Gateway Protocols, page 4-8](#)
- [Gateway Protocols and Core Feature Requirements, page 4-11](#)
- [Site-Specific Gateway Requirements, page 4-16](#)

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, MGCP, or SIP) 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 4-25](#)

- Unified CM redundancy support

Cisco Unified Communications is based on a distributed model for high availability. Unified CM clusters provide for Unified CM redundancy. The gateways must support the ability to “re-home” to a secondary Unified CM in the event that a primary Unified CM fails. Redundancy differs from call survivability in the event of a Unified CM 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 4-16](#)).

Gateway Protocols

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

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

Cisco Unified CM Release 4.0 and later supports Session Initiation Protocol (SIP) on the trunk side. The SIP trunk implementation has been enhanced in Cisco Unified CM Release 5.0 to support more features.

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

Protocol selection depends on site-specific requirements and the installed base of equipment. For example, most remote branch locations have Cisco 2600XM, 2800, 3700, or 3800 Series routers installed. These routers support SIP, H.323, and MGCP 0.1 with Cisco IOS Release 12.2.11(T) and Cisco Unified CM Release 3.1 or later. For gateway configuration, MGCP might be preferred over H.323 or SIP due to simpler configuration. On the other hand, H.323 or SIP 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 or SIP devices do not identify the specific line being used from a group of ports. Use of H.323 or SIP 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 Unified CM deployment model being used can influence gateway protocol selection. (Refer to the chapter on [Deployment Models, page 2-1](#).)

[Table 4-3](#) 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 4-11](#), describes how each protocol provides these feature requirements.

Table 4-3 Supported Gateway Protocols and Cisco Unified Communications Gateways

Cisco Gateway	MGCP 0.1	H.323	SCCP	SIP
Cisco 3800	Yes, beginning with Cisco IOS Release 12.3.11T Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes, beginning with Cisco IOS Release 12.3.11T	Yes for DSP resources, beginning with Cisco IOS Release 12.3.11T. For FXS, use Cisco IOS Release 12.4.9.T.	Yes, SIP trunk
Cisco 2800	Yes, beginning with Cisco IOS Release 12.3.8T4 Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes, beginning with Cisco IOS Release 12.3.8T4	Yes for DSP resources, beginning with Cisco IOS Release 12.3.11T. For FXS, use Cisco IOS Release 12.4.9.T.	Yes, SIP trunk
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	Yes, SIP trunk
Communication Media Module (CMM)	Yes Supported with: • T1 CAS FXS • T1/E1 PRI • FXS	Yes	No	Yes
Catalyst 6000 WS-X6608-x1 Gateway Module and FXS Module WS-X6624 ¹	Yes Supported with: • T1 CAS E&M • T1 CAS FXS • T1/E1 PRI • FXS with WS-6624	No	No	No
VG224	Yes, FXS only. Also supports conferencing and transcoding for VG224 beginning with Cisco IOS Release 12.3(T).	Yes, FXS only	Yes, beginning with Cisco IOS Release 12.4(2)T	Yes, SIP trunk
VG248	No	No	Yes ²	No

Table 4-3 *Supported Gateway Protocols and Cisco Unified Communications Gateways (continued)*

Cisco Gateway	MGCP 0.1	H.323	SCCP	SIP
Cisco ATA 188	Yes, FXS only	Yes, FXS only	Yes, FXS only	Yes, third-party SIP phone
Cisco AS5350XM Cisco AS5400XM	No	Yes	No	Yes, SIP trunk
Cisco AS5850 ¹	No	Yes	No	Yes, SIP trunk
Cisco 5300 ¹	No	Yes	No	Yes, SIP trunk
Cisco 3640 and 3660 ¹	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	Yes, SIP trunk
Cisco 2600 and 2600XM ^{1 3}	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	Yes, SIP trunk
Cisco 1751 and 1760 ¹	Yes	Yes	Yes, conferencing and transcoding	Yes, SIP trunk
VG200 ¹	Yes Supported with: • Analog FXS/FXO • T1 CAS (E&M Wink Start; Delay Dial only) • T1/E1 PRI	Yes	Yes (DSP farm)	No
Cisco 7200	No	Yes	No	Yes, SIP trunk
Catalyst 4000 WS-X4604-GWY Gateway Module ¹	Yes	Yes	No	No
Cisco ICS7750-MRP ¹	No	Yes	No	No
Cisco ICS7750-ASI ¹	No	Yes	No	No
DE-30+, DT-24+ ¹	Yes	No	No	No
Cisco 827-V4 ¹	No	Yes, supported for FXS	No	No

1. These models have reached End of Sale (EOS).

2. The VG248 is not a true gateway in that it uses Skinny Client Control Protocol (SCCP) instead of H.323, MGCP, or SIP.
3. For IP Telephony applications, use Cisco 2800 Series Routers.

**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, MGCP, and SIP) supports the following gateway feature requirements:

- [DTMF Relay, page 4-11](#)
- [Supplementary Services, page 4-12](#)
- [Unified CM Redundancy, page 4-15](#)

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 3800 series products, can communicate with Unified CM 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 Unified CM 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. Unified CM 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 Unified CM MGCP gateway configuration interface.

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

SIP Gateway

The Cisco IOS-based VG224, 2600XM, 2800, 3700, 3800 platforms can use SIP for Unified CM communication. They support various methods for DTMF, but only the following two methods can be used to communicate with Unified CM:

- Named Telephony Events (NTE), or RFC 2833
- Unsolicited SIP Notify (UN)

The following example shows a configuration for NTE:

```
dial-peer voice 100 voip
destination-pattern 555...
session target ipv4:10.1.1.1
session protocol sipv2
dtmf-relay rtp-nte
```

The following example shows a configuration for UN:

```
dial-peer voice 100 voip
destination-pattern 555...
session target ipv4:10.1.1.1
session protocol sipv2
dtmf-relay sip-notify
```

For more details on DTMF method selection, see the chapter on [Media Resources, page 6-1](#).

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 VG224, VG248, and ATA 188 gateways provide full supplementary service support. The Cisco 2800 and 3800 Series gateways also support FXS SCCP ports with Cisco IOS Release 12.4.9T. The SCCP gateways use the Gateway-to-Unified CM signaling channel and SCCP to exchange call control parameters.

H.323 Gateways

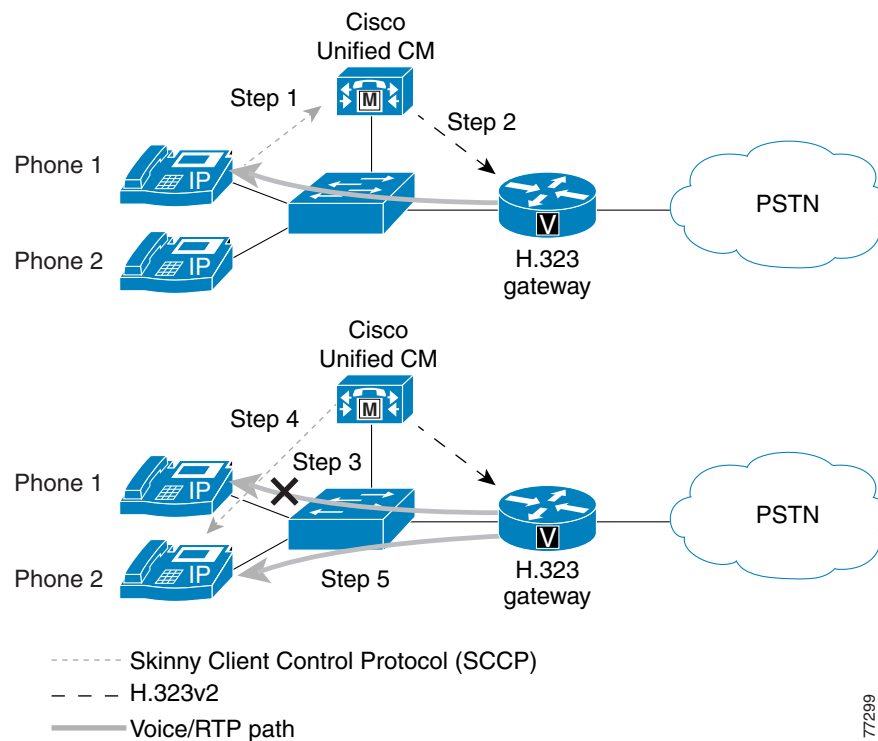
H.323v2 implements Open/Close LogicalChannel and the emptyCapabilitySet features. The use of H.323v2 by H.323 gateways, beginning in Cisco IOS Release 12.0(7)T and Cisco Unified CM Release 3.0 and later, eliminates the requirement for an MTP to provide supplementary services. With Unified CM Release 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 Unified CM 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 Unified CM using SCCP.
2. Unified CM 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. Unified CM issues a request to Phone 2, using SCCP, to set up an RTP connection to the Cisco IOS gateway. At the same time, Unified CM 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



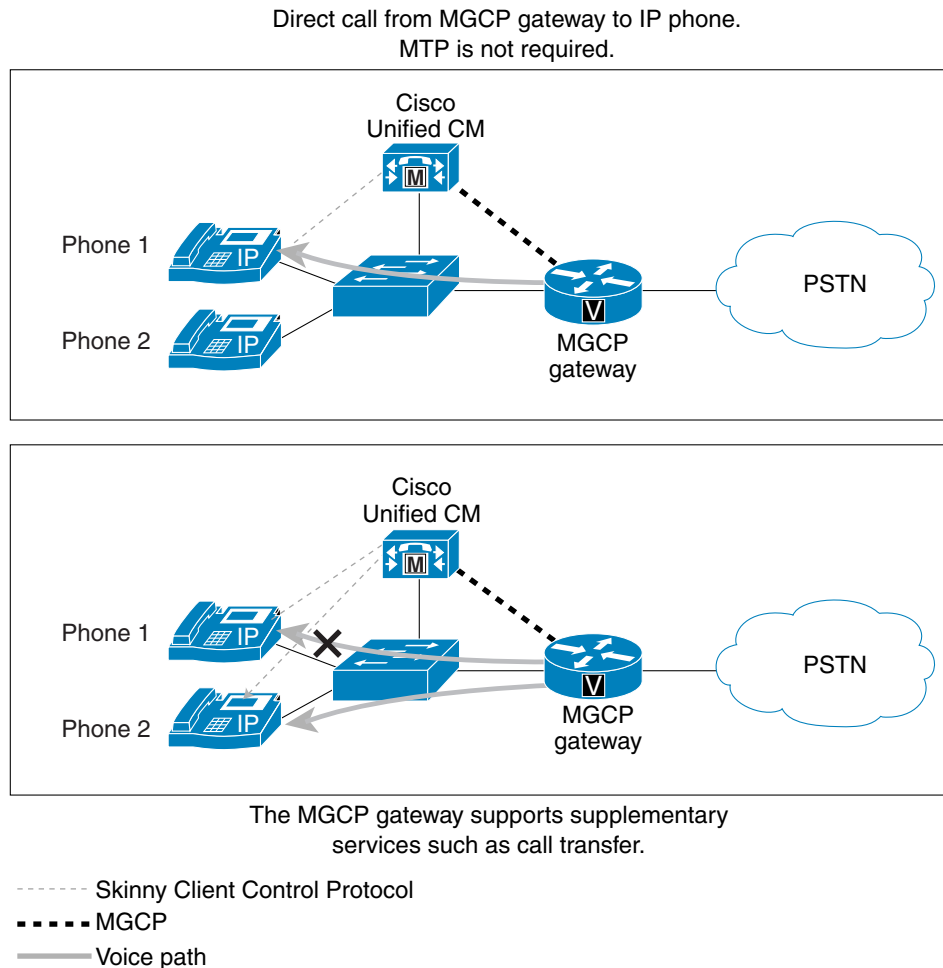
77299

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 Unified CM controlling all session intelligence, Unified CM 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 Unified CM using SCCP. Unified CM 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 Unified CM.

Figure 4-2 MGCP Gateway Supplementary Service Support



SIP Gateway

The Unified CM SIP trunk interface to Cisco IOS SIP gateways supports supplementary services such as hold, blind transfer, and attended transfer. The support for supplementary services is achieved via SIP methods such as INVITE and REFER. For more details, refer to the following documentation:

- *Cisco Unified Communications Manager System Guide*, available at <http://www.cisco.com>
- *Cisco IOS SIP Configuration Guide*, available at http://www.cisco.com/en/US/products/ps6441/products_installation_and_configuration_guides_list.html

Unified CM 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 Unified CMs. Even in its most simplistic form (a two-system cluster), a secondary Unified CM should be able to pick up control of all gateways initially managed by the primary Unified CM.

SCCP Gateways

Upon boot-up, the Cisco VG224, VG248, and ATA 188 gateways are provisioned with Unified CM server information. When these gateways initialize, a list of Unified CMs is downloaded to the gateways. This list is prioritized into a primary Unified CM and secondary Unified CM. In the event that the primary Unified CM becomes unreachable, the gateway registers with the secondary Unified CM.

H.323 VoIP Call Preservation Enhancements for WAN Link Failures

H.323 VoIP call preservation enhancements for WAN link failures sustain connectivity for H.323 topologies where signaling is handled by an entity that is different from the other endpoint, such as a gatekeeper that provides routed signaling or a call agent (such as the Cisco BTS 10200 Softswitch, Cisco PGW 2200 Softswitch, or Cisco Unified CM) that brokers signaling between the two connected parties.

Call preservation is useful when a gateway and the other endpoint (typically a Cisco Unified IP phone) are located at the same site but the call agent is remote and therefore more likely to experience connectivity failures.

Example of H.323 VoIP Call Preservation for All Calls

The following configuration example enables H.323 VoIP call preservation for all calls:

```
voice service voip
  h323
    call preserve
```

MGCP Gateway

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

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

If the primary Unified CM becomes available at a later time, the MGCP gateway can “re-home,” or switch back to the original Unified CM. 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>]
```

SIP Gateway

Redundancy with Cisco IOS SIP gateways can be achieved similarly to H.323. If the SIP gateway cannot establish a connection to the primary Unified CM, it tries a second Unified CM defined under another dial-peer statement with a higher preference.

By default the Cisco IOS SIP gateway transmits the SIP INVITE request 6 times to the Unified CM IP address configured under the dial-peer. If the SIP gateway does not receive a response from that Unified CM, it will try to contact the Unified CM configured under the other dial-peer with a higher preference.

Cisco IOS SIP gateways wait for the SIP 100 response to an INVITE for a period of 500 ms. By default, it can take up to 3 seconds for the Cisco IOS SIP gateway to reach the backup Unified CM. You can change the SIP INVITE retry attempts under the **sip-ua** configuration by using the command **retry invite <number>**. You can also change the period that the Cisco IOS SIP gateway waits for a SIP 100 response to a SIP INVITE request by using the command **timers trying <time>** under the **sip-ua** configuration.

One other way to speed up the failover to the backup Unified CM is to configure the command **monitor probe icmp-ping** under the **dial-peer** statement. If Unified CM does not respond to an Internet Control Message Protocol (ICMP) echo message (ping), the dial-peer will be shut down. This command is useful only when the Unified CM is not reachable. ICMP echo messages are sent every 10 seconds.

The following commands enable you to configure Unified CM redundancy on a Cisco IOS SIP gateway:

```
sip-ua
  retry invite <number>
  timers trying <time>

dial-peer voice 101 voip
  destination-pattern 2...
  session target ipv4:10.1.1.101
  preference 0
  monitor probe icmp-ping
  session protocol sipv2

dial-peer voice 102 voip
  destination-pattern 2...
  session target ipv4:10.1.1.102
  preference 1
  monitor probe icmp-ping
  session protocol sipv2
```

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 Unified Communications 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 multisite 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 Unified CM 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/en/US/products/sw/voicesw/ps556/products_device_support_tables_list.html

Cisco Analog Gateways

Table 4-4 lists supported interface types for Cisco analog gateways using H.323 or Session Initiation Protocol (SIP), and Table 4-5 lists supported interface types for Cisco analog gateways using Media Gateway Control Protocol (MGCP).

Table 4-4 Supported Analog H.323 and SIP Features

Cisco Gateway	Interface Type					
	FXS	FXO	E&M	FXO, Battery Reversal	Analog DID	CAMA 911
3800 Series	Yes	Yes	Yes	Yes	Yes	Yes
2800 Series	Yes	Yes	Yes	Yes	Yes	Yes
3700 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes
Communication Media Module (CMM) 24FXS	Yes	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
6608 and 6624 ¹	N/A	N/A	N/A	N/A	N/A	N/A
VG224	Yes	N/A	N/A	N/A	N/A	N/A
VG248	No	No	No	No	No	No
Analog Telephone Adapter (ATA) ¹	Yes	No	No	No	No	No
3600 Series ¹	Yes	Yes	Yes	Yes	Yes	12.2.11T
2600 Series ¹	Yes	Yes	Yes	Yes	Yes	12.2.11T
1751 and 1760 ¹	Yes	Yes	Yes	Yes	Yes	Yes
VG200 ¹	Yes	Yes	Yes	No	Yes	No
7x00 family	N/A	N/A	N/A	N/A	N/A	N/A
ICS 7750 ¹	Yes	Yes	Yes	Yes	Yes	No
Catalyst 4000 Access Gateway Module (AGM)	Yes	Yes	No	No	No	No
827-4V ¹	Yes	No	No	No	No	No

1. These models have reached End of Sale (EOS).

Table 4-5 Supported Analog MGCP Features

Cisco Gateway	Interface Type					
	FXS	FXO	E&M	FXO, Battery Reversal	Analog DID	CAMA 911
3800 Series	Yes	Yes	No	Yes	No	No
2800 Series	Yes	Yes	No	Yes	No	No
3700 Series ¹	Yes	Yes	No	Yes	No	No
Communication Media Module (CMM) 24FXS	Yes	N/A	N/A	N/A	N/A	N/A

Table 4-5 **Supported Analog MGCP 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
6608 and 6624 ¹	Yes	No	No	No	No	No
VG224	Yes	No	No	No	No	No
VG248	No	No	No	No	No	No
Analog Telephone Adapter (ATA) ¹	Yes	N/A	N/A	N/A	N/A	N/A
3600 Series ¹	Yes	Yes	No	Yes	No	No
2600 Series ¹	Yes	Yes	No	Yes	No	No
1751 and 1760 ¹	Yes	Yes	No	Yes	No	No
VG200 ¹	Yes	Yes	No	Yes	No	No
7x00 family	N/A	N/A	N/A	N/A	N/A	N/A
ICS 7750 ¹	Yes	Yes	No	No	No	No
Catalyst 4000 Access Gateway Module (AGM) ¹	Yes	Yes	No	No	No	No
827-4V ¹	No	No	N/A	N/A	N/A	N/A

1. These models have reached End of Sale (EOS).

Cisco Digital Gateways

Table 4-6 through Table 4-9 list supported interface types for Cisco digital gateways using H.323 or Session Initiation Protocol (SIP). Table 4-10 lists supported interface types for Cisco digital gateways using Media Gateway Control Protocol (MGCP).

Table 4-6 **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
3800 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
2800 Series	Yes	Yes	Yes	No	Yes	No	Yes	Yes
3700 Series ¹	Yes	Yes	Yes	No	Yes	No	Yes	Yes
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
6608 and 6624 ¹	N/A	N/A	N/A	N/A	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
VG248	No	No	No	No	No	No	No	No

Table 4-6 Supported Digital H.323 and SIP Features for BRI, T1 CAS, T1 FGB, T1 FGD, and T1 QSIG (continued)

Cisco Gateway	Interface Type							
	BRI (TE, User side)	BRI (NT, Network side)	BRI QSIG (Net3)	BRI Phones	T1 CAS (Robbed bit)	T1 FGB	T1 FGD	T1 QSIG
Analog Telephone Adapter (ATA) ¹	No	No	No	No	No	No	No	No
3600 Series ¹	Yes	Yes	Yes	No	Yes	No	Yes	Yes
2600 Series ¹	Yes	Yes	Yes	No	Yes	No	Yes	Yes
1751 and 1760 ¹	No	Yes	Yes	No	Yes	No	No	Yes
VG200 ¹	Yes	Yes	No	No	Yes	No	Yes	No
7x00 family	N/A	N/A	N/A	N/A	Yes	No	Yes	Yes
ICS 7750 ¹	Yes	Yes	No	No	Yes	No	Yes	No
Catalyst 4000 Access Gateway Module (AGM) ¹	Yes	No	Yes	No	Yes	No	Yes	Yes
827-4V ¹	No	No	No	No	No	No	No	No

1. These models have reached End of Sale (EOS).

Table 4-7 Supported Digital H.323 and SIP Features for T1 PRI SL-1, 4ESS, and 5ESS

Cisco Gateway	Interface Type					
	T1 PRI (User, DMS-100)	T1 PRI (Network, SL-1)	T1 PRI (User, 4ESS)	T1 PRI (Network, 4ESS)	T1 PRI (User, 5ESS)	T1 PRI (Network, 5ESS)
3800 Series	Yes	No	Yes	Yes	Yes	Yes
2800 Series	Yes	No	Yes	Yes	Yes	Yes
3700 Series ¹	Yes	No	Yes	No	Yes	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

Table 4-7 Supported Digital H.323 and SIP Features for T1 PRI SL-1, 4ESS, and 5ESS (continued)

Cisco Gateway	Interface Type					
	T1 PRI (User, DMS-100)	T1 PRI (Network, SL-1)	T1 PRI (User, 4ESS)	T1 PRI (Network, 4ESS)	T1 PRI (User, 5ESS)	T1 PRI (Network, 5ESS)
6608 and 6624 ¹	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A
VG248	No	No	No	No	No	No
Analog Telephone Adapter (ATA) ¹	No	No	No	No	No	No
3600 Series ¹	Yes	No	Yes	Yes	Yes	Yes
2600 Series ¹	Yes	No	Yes	Yes	Yes	Yes
1751 and 1760 ¹	Yes	No	Yes	No	Yes	No
VG200 ¹	Yes	No	Yes	No	Yes	No
7x00 family	Yes	No	Yes	No	Yes	No
ICS 7750 ¹	Yes	No	Yes	No	Yes	No
Catalyst 4000 Access Gateway Module (AGM) ¹	Yes	No	Yes	No	Yes	No
827-4V ¹	No	No	No	No	No	No

1. These models have reached End of Sale (EOS).

Table 4-8 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)
3800 Series	Yes	Yes	Yes	Yes	Yes	Yes
2800 Series	Yes	Yes	Yes	Yes	Yes	Yes
3700 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes
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	No
6608 and 6624 ¹	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A
VG248	No	No	No	No	No	No
Analog Telephone Adapter (ATA) ¹	No	No	No	No	No	No
3600 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes
2600 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes

Table 4-8 Supported Digital H.323 and SIP Features for T1 PRI NI2, NFAS, and Network Specific Facilities (NSF) Service (continued)

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)
1751 and 1760 ¹	Yes	Yes	No	No	No	No
VG200 ¹	Yes	Yes	No	No	No	No
7x00 family	Yes	Yes	No	No	No	No
ICS 7750 ¹	Yes	Yes	Yes	Yes	Yes	No
Catalyst 4000 Access Gateway Module (AGM) ¹	Yes	Yes	No	No	No	No
827-4V ¹	No	No	No	No	No	No

1. These models have reached End of Sale (EOS).

Table 4-9 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
3800 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2800 Series	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3700 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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
6608 and 6624 ¹	No	No	No	No	No	No	No
VG224	N/A	N/A	N/A	N/A	N/A	N/A	N/A
VG248	No	No	No	No	No	No	No
Analog Telephone Adapter (ATA) ¹	No	No	No	No	No	No	No
3600 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2600 Series ¹	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1751 and 1760 ¹	No	No	Yes	Yes	Yes	Yes	No
VG200 ¹	No	Yes	Yes	Yes	Yes	No	Yes
7x00 family	Yes	No	Yes	Yes	Yes	Yes	No
ICS 7750 ¹	No	No	Yes	Yes	Yes	No	No
Catalyst 4000 Access Gateway Module (AGM) ¹	No	No	Yes	Yes	Yes	Yes	No
827-4V ¹	No	No	No	No	No	No	No

1. These models have reached End of Sale (EOS).

Table 4-10 Supported Digital MGCP Features

Cisco Gateway	Interface Type						
	BRI ¹	T1 CAS (E&M)	T1 CAS Hookflash	T1 PRI	T1 QSIG	E1 PRI	E1 QSIG
3800 Series	12.4(2)T	Yes ³	Yes	Yes ³	Yes ³	Yes ³	Yes ³
2800 Series	12.4(2)T	Yes ³	Yes for 2811, 2821, 2851	Yes ³	Yes ³	Yes ³	Yes ³
3700 Series ²	12.4(2)T	Yes ³	Yes	Yes ³	Yes ³	Yes ³	Yes ³
Communication Media Module (CMM) 24FXS	N/A	N/A	Yes	N/A	N/A	N/A	N/A
CMM-6T1/E1	N/A	Yes	Yes	Yes	Yes	Yes	Yes
6608 ²	N/A	Yes	Yes – E&M only	Yes	Yes	Yes	Yes
6624 ²	N/A	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	N/A
VG248	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Analog Telephone Adapter (ATA) ²	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3600 Series ²	12.4(2)T	Yes ³	Yes	Yes ³	Yes ³	Yes ³	Yes ³
2600 Series ²	12.4(2)T	Yes ³	Yes for 2600XM only	Yes ³	Yes ³	Yes ³	Yes ³
1751 and 1760 ²	12.3(14)T	Yes	N/A	Yes	Yes	Yes	Yes
VG200 ²	No	Yes	N/A	Yes	Yes	Yes	Yes
7x00 family	N/A	No	N/A	No	No	No	No
ICS 7750	12.3.7T	Yes	N/A	Yes	Yes	Yes	Yes
Catalyst 4000 Access Gateway Module (AGM) ²	No	Yes	N/A	Yes	Yes	Yes	Yes
827-4V ²	N/A	N/A	N/A	N/A	N/A	N/A	N/A

1. Cisco IOS Release 12.4(2)T supports BRI MGCP with the following hardware: NM-HDV2, NM-HD-XX and on-board H-WIC slots. BRI MGCP is also supported on older Cisco IOS releases with NM-1V/2V hardware.
2. These models have reached End of Sale (EOS).
3. 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 4-11](#).

Table 4-11 QSIG Support on H.323 Gateways

Platform	Media	Minimum Cisco IOS Software Release Required
Cisco 3800	BRI and T1/E1 QSIG	12.3.11T
Cisco 2800 Series	BRI and T1/E1 QSIG	12.3.8T4
Cisco 3700 ¹	T1/E1 QSIG	12.2.8T
Cisco AS5350XM	T1/E1	12.2.2T
Cisco AS5400XM		
Cisco 5300 ¹	T1/E1	12.0.7T
Cisco 2600XM and 3600 Series ¹	BRI and T1/E1 QSIG	12.1.2T
Cisco 1751 and 1760 ¹	BRI	12.2(8)YH
	T1/E1 QSIG	12.2(4)YB
Cisco 7200	T1/E1 QSIG	12.1.2T

1. These models have reached End of Sale (EOS).

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

http://www.cisco.com/en/US/products/sw/iosswrel/ps1834/products_feature_guide09186a00800800b0.html#xtocid116542

Prior to Cisco Unified CM 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 Unified CM. 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 Unified CM.

For Unified CM to support QSIG functionality, QSIG must be back-hauled directly to Unified CM. This support is provided in Cisco Unified CM Release 3.3 and later, in conjunction with all MGCP T1/E1 ISDN gateways.

Fax and Modem Support

This section describes the fax and modem support available with Unified CM 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 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:

- **Fax Relay** — Cisco gateways support ITU standards-based T.38 fax relay as well as a proprietary method of Cisco fax relay. T.38 fax relay can be signaled using signaling protocols such as H.323, SIP, MGCP, or the RTP header. Cisco fax relay is always signaled using the RTP header.
- **Fax Pass-Through** — Fax pass-through enables fax tones to be passed with a G.711 codec. Other high-compression codecs can cause distortion of the fax tones. If high-compression codecs are being used for the voice call, upon detection of the fax tones, the terminating gateway will signal the originating gateway to upspeed to G.711. This signaling can be done using signalling protocols such as H.323, SIP, MGCP, or the RTP stream.

Fax relay mode is the preferred method to transmit fax traffic. However, if a specific gateway does not support 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 Unified Communications network, refer to the *Enterprise QoS Solution Reference Network Design Guide*, available at

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

- 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.

- T.38 fax relay provides the best fax performance based on network considerations. Both the originating and terminating sides must be configured for the same fax relay method. When the signaling protocols between terminating and originating sides are different, Cisco recommends using T.38 signaling via Named Service Event (NSE) in the RTP stream.

The only non-IOS gateway that does not support 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 and Modem Relay

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

- Modem pass-through
- Modem relay

Currently, both modem relay and modem pass-through are 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 terminating gateway will upspeed to G.711 upon detection of the modem tone. The configuration for modem pass-through and fax pass-through is done with the same command, **modem passthrough**. 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 *Enterprise QoS Solution Reference Network Design Guide*, available at

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

- 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.

Super-Group 3 Fax Support

Cisco IOS gateways with Cisco IOS Release 12.4.4T support Super-Group 3 fax; however, only Group 3 speeds are negotiated.

Supported Platforms and Features

The following Cisco platforms support fax and modem features:

Analog Gateways

Cisco IOS Gateways:

- 2600XM and 2691 (with FXS) – End of Sale (EOS)
- 2800 (with FXS)
- 3725 and 3745 (with FXS) – End of Sale (EOS)
- 3800 (with FXS)
- VG200 (with FXS) – End of Sale (EOS)
- VG224
- 1751 and 1760 – End of Sale (EOS)
- Communication Media Module (CMM) FXS card

Non-IOS Gateways:

- VG248
- ATA 188 – End of Sale (EOS)
- 6624 – End of Sale (EOS)

Digital Gateways

Cisco IOS Gateways:

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

Non-IOS Gateways:

- 6608 – End of Sale (EOS)

**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, Session Initiation Protocol (SIP), 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.

Some of the common combinations of protocols in a network include: MGCP and H.323; SCCP and H.323; SCCP and SIP; MGCP and SIP; H.323 and SIP; and SCCP and MGCP. Common voice gateways included the Cisco VG224, VG248, 2600XM, 2800, 3700, 3800, and Catalyst 6000.

Table 4-12 lists the protocol combinations that currently support fax and modem interoperability.

Table 4-12 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
Unified CM using MGCP combined with Unified CM using H.323 or SIP	Yes	Yes	No	Yes	Yes
Unified CM using MGCP combined with Unified CM using MGCP	Yes	Yes	No	Yes	Yes
SCCP combined with Unified CM using H.323 or SIP	Yes	Yes	No	Yes	Yes
SCCP combined with Unified CM using MGCP	Yes	Yes	No	Yes	Yes
Unified CM using H.323 combined with H.323 or SIP	Yes	Yes	Yes	Yes	Yes
Unified CM using SIP combined with H.323 or SIP	Yes	Yes	Yes	Yes	Yes



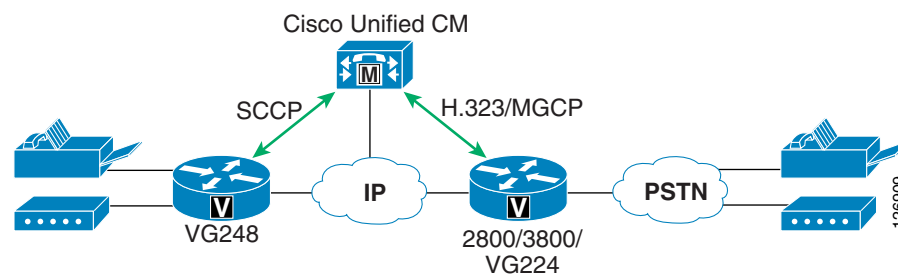
Note

Cisco ATA 188, VG248 and Catalyst 6000 platforms currently do not support T.38 fax relay. When these platforms connect to Cisco AS5350XM or AS5400XM gateways, only fax pass-through is supported for fax applications.

Gateway Combinations and Interoperability of Features

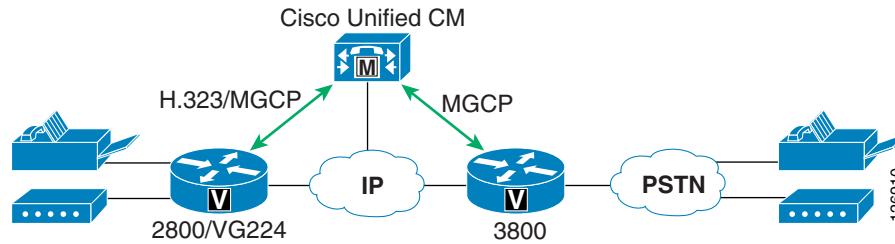
The most frequent questions about fax and modem interoperability arise from combining a Cisco IOS gateway (such as a Cisco 2800 or 3800) with a non-IOS gateway (such as a Cisco VG248), as illustrated in Figure 4-3.

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



Feature Support Between Similar Gateways

[Table 4-13](#) 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 4-13 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-manager 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) |
|-----|

```

Unified CM Configuration for Cisco IOS Gateways

Perform the following steps in Unified CM to configure it for the Cisco IOS gateways (such as the Cisco 6608 and 6624).

-
- Step 1** In Unified CM 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 Unified CM 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.
-

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 Series Routers. When operating in either H.323 or SIP mode, these platforms do not support MGCP.

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

- [Loose Gateway Controlled with Named Service Event \(NSE\), page 4-33](#)
- [Gateway Controlled with Capability Exchange Through H.245 or Session Description Protocol \(SDP\), page 4-33](#)
- [Call-Agent-Controlled T.38 with H.323 Annex D, page 4-35](#)

Loose Gateway Controlled with Named Service Event (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 Description Protocol (SDP)

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

- T.38 capability is exchanged between gateways. A Named Service Event (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.
- Unified CM cannot support this capability exchange with 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

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

- The call control agent (for example, Unified CM) 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 SIP.
- 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.
- The Cisco Voice Media Streaming Application does not support T.38, but Cisco IOS media termination points (MTPs) do. Therefore, ensure that Cisco IOS MTPs are correctly prioritized in the media resource group list (MRGL).

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-manager mgcp  
mgcp  
mgcp call-agent 10.10.10.1 service-type mgcp version 0.1  
!  
! T.38 fax relay is ON by default. HOWEVER, Unified CM 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  
!
```

Gateways for Video Telephony

Cisco offers voice gateway functionality in a variety of forms, such as standalone devices, modules that integrate into Cisco IOS Routers, or line cards that integrate into Cisco Catalyst Ethernet Switches. These gateways support multiple VoIP protocols (such as H.323, MGCP, SIP, and SCCP), multiple port interface types (such as FXS, FXO, E&M, T1/E1-CAS, T1/E1-PRI, ISDN BRI, and so on), and a myriad of advanced VoIP features. They also offer a rich set of management and troubleshooting interfaces.

Cisco IP/VC gateways are scalable and offer a robust video gateway solution for large networks. The IP/VC gateways have the following characteristics:

- They support only H.323 and H.320.
- They are standalone devices that cannot be integrated into Cisco IOS Routers or Cisco Catalyst Switches.
- They support only T1/E1-PRI, ISDN BRI, and V.35 interface types.
- They support only G.711, G.728, G.723, and G.722; they do not support G.729 audio.
- They support the H.245 Empty Capabilities Set (ECS).
- They do not support many of the manageability and troubleshooting capabilities inherent in Cisco Voice Gateways.

Cisco IOS routers also support H.320 protocol, enabling video gateway capability in addition to voice. The Cisco IOS video gateways now add support for the H.26x family of video codecs and bonding to provide channel aggregation for video conferencing.

Deploying Cisco IOS gateways for voice and video can have the following advantages:

- Single gateway for voice and video
- Single circuit for PSTN from service provider

Telephony customers can choose to deploy the Cisco IP/VC Video gateways or Cisco IOS gateways for video. Smaller locations with existing Cisco IOS voice gateways can add the additional video gateway functionality at these distributed locations, while the larger locations can use the IP/VC and dedicated voice gateways to better achieve scalability.

When using a single ISDN circuit for voice and H.320 videoconferencing, these lines are shared, as depicted in [Figure 4-5](#).

With separate Video gateways and voice gateways, PSTN lines are not shared as depicted in [Figure 4-6](#).

Figure 4-5 Traditional PBX Sharing PSTN Lines Between Voice and H.320 Videoconferencing

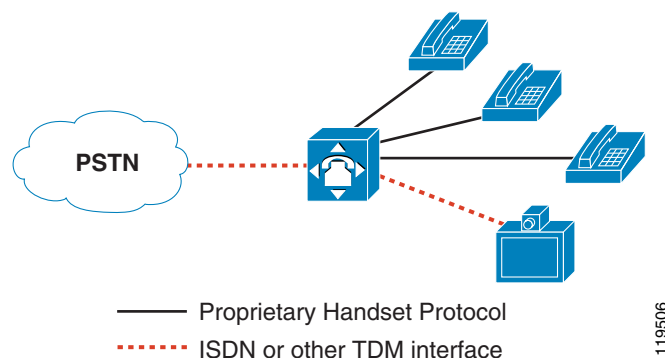
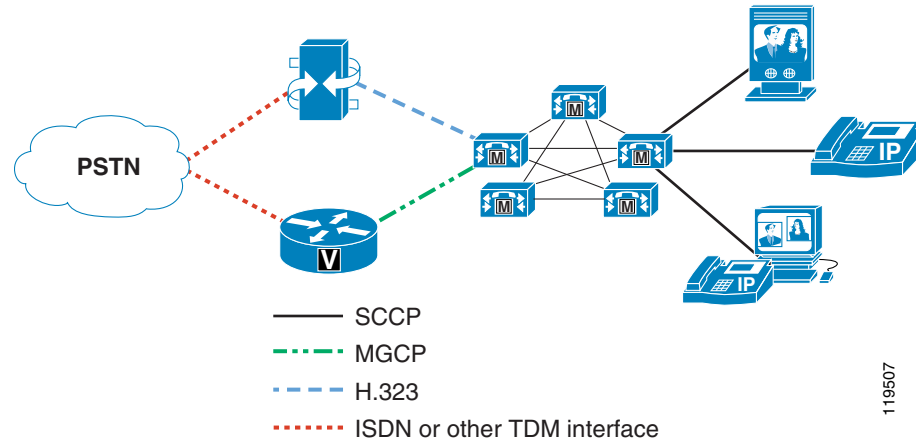


Figure 4-6 Unified CM System with Separate PSTN Lines for Voice and IP Video Telephony

With separate voice and video gateways, the route plans must also be separate for both inbound and outbound calls. For inbound calls, there is no way to have a single Direct Inward Dial (DID) extension for a user who wants to be able to receive both voice and video calls. Typically, each user will already have a DID for voice calls. When you introduce video into the scenario, users will have to be dialed some other way, such as via a second DID number or by dialing the main number of the video gateway and then entering the users video extension when prompted by the Interactive Voice Response (IVR). For outbound calls, there is no way to have a single PSTN access code for both voice and video calls. Typically, users will already have a well-known access code for voice (such as 9 in most US enterprises), but when you introduce video into the scenario, they will have to dial some other access code to place outbound video calls.

Another consideration for deploying two types of gateways is the placement of those gateways. Typically, enterprises have many PSTN gateway resources consolidated at their central site(s), and each branch office has some local gateway resources as well. For instance, Cisco Catalyst 6500 gateways may be deployed at the central site with several T1/E1 circuits connected to them, while Cisco Integrated Services Routers (ISRs) may be deployed at each branch office with either analog or digital trunks to the local CO. When video is introduced into this scenario, the customer must also determine the number of PSTN circuits they will need for video and where the video gateways will be placed. For instance, will they deploy only a few IP/VC 3500 Series gateways at the central site, or will they also deploy them at each branch office?

Finally, consider how calls will be routed across the IP network to a remote gateway for the purpose of providing toll bypass, and how calls will be re-routed over the PSTN in the event that the IP network is unavailable or does not have enough bandwidth to complete the call. More specifically, do you want to invoke automated alternate routing (AAR) for video calls?

Routing Inbound Calls from the PSTN

Use one of the following methods to route inbound calls from the PSTN:

- Assign at least two different directory numbers to each video-enabled device in the Unified CM cluster, with one line for audio and another line for video. With this method, the outside (PSTN) caller must dial the correct number to enable video.
- For video calls, have outside callers dial the main number of the video gateway. Cisco Unified Videoconferencing gateways offer an integrated IVR that prompts the caller to enter the extension number of the party they are trying to reach. Unified CM will then recognize that it is

a video call when ringing the destination device. This method relieves the caller from having to remember two different DID numbers for each called party, but it adds an extra step to dialing an inbound video call.



Note The outside video endpoints must support DTMF in order to enter the extension of the called party at the IVR prompt.

The following example illustrates the second method:

A user has a Cisco Unified IP Phone 7960 attached to a PC running Cisco Unified Video Advantage. The extension of the IP Phone is 51212, and the fully qualified DID number is 1-408-555-1212. To reach the user from the PSTN for a voice-only call, people simply dial the DID number. The CO sends calls to that DID number through T1-PRI circuit(s) connected to a Cisco Voice Gateway. When the call is received by the gateway, Unified CM knows that the gateway is capable of audio only, so it negotiates only a single audio channel for that call. Conversely, for people to reach the user from the PSTN for a video call, they must dial the main number of the video gateway and then enter the user's extension. For example, they might dial 1-408-555-1000. The CO would send calls to that number through the T1-PRI circuit(s) connected to a Cisco Unified Videoconferencing 3500 Series video gateway. When the call is received by the gateway, an IVR prompt asks the caller to enter the extension of the person they are trying to reach. When the caller enters the extension via DTMF tones, Unified CM knows that the gateway is capable of video, so it negotiates both audio and video channels for that call.

Gateway Digit Manipulation

The Cisco Unified Videoconferencing 3500 Series Gateways cannot manipulate digits for calls received from the PSTN. It takes the exact number of digits passed to it in the Q.931 Called Party Number field and sends them all to Unified CM. Therefore, Unified CM must manipulate the digits in order to match the directory number (DN) of the destination device. For instance, if the circuit from the CO switch to the gateway is configured to pass 10 digits but the extension of the called party is only five digits, Unified CM must strip off the leading five digits before attempting to find a matching DN. You can implement this digit manipulation in one of the following ways:

- By configuring the Significant Digits field on the H.323 gateway device or on the H.225 gatekeeper-controlled trunk that carries the incoming calls from the IP/VC gateway

This method enables you to instruct Unified CM to pay attention to only the least-significant N digits of the called number. For example, setting the Significant Digits to 5 will cause Unified CM to ignore all but the last 5 digits of the called number. This is the easiest approach, but it affects all calls received from that gateway. Thus, if you have variable-length extension numbers, this is not the recommended approach.

- By configuring a translation pattern and placing it in the calling search space of the H.323 gateway device or of the H.225 gatekeeper-controlled trunk that carries the incoming calls from the IP/VC gateway

This method enables Unified CM to match calls to the full number of digits received, to modify the called number, and then to continue performing digit analysis on the resulting modified number. This approach is slightly more complex than the preceding method, but it is more flexible and enables you to use a finer granularity for matching calls and for specifying how they will be modified.

Routing Outbound Calls to the PSTN

Use one of the following methods to route outbound calls to the PSTN:

- Assign different access codes (that is, different route patterns) for voice and video calls. For example, when the user dials 9 followed by the PSTN telephone number they are trying to reach, it could match a route pattern that directs the call out a voice gateway. Similarly, the digit 8 could be used for the route pattern that directs calls out a video gateway.
- Assign at least two different directory numbers on each video-enabled device in the Unified CM cluster, with one line for audio and another line for video. The two lines can then be given different calling search spaces. When users dial the access code (9, for example) on the first line, it could be directed out a voice gateway, while dialing the same access code on the second line could direct the call out a video gateway. This method alleviates the need for users to remember two different access codes but requires them to press the correct line on their phones when placing calls.

Gateway Service Prefixes

The Cisco Unified Videoconferencing Gateways use service prefixes to define the speed for outbound calls. When you configure a service prefix in the gateway, you must choose one of the following speeds:

- Voice-only
- 128 kbps
- 256 kbps
- 384 kbps
- 768 kbps
- Auto (dynamically determined; supports any call speed in the range of 128 kbps to 768 kbps)



Note

Each of the above speeds represents a multiple of 64 kbps. For 56-kbps dialing, there is a check-box on the service prefix configuration page to restrict each channel to 56 kbps. Therefore, a 128-kbps service with restricted mode enabled would result in a 112-kbps service; a 384 kbps service with restricted mode enabled would result in a 336-kbps service; and so on.

Calls from an IP endpoint toward the PSTN must include the service prefix at the beginning of the called number in order for the gateway to decide which service to use for the call. Optionally, you can configure the default prefix to be used for calls that do not include a service prefix at the beginning of the number. This method can become quite complex because users will have to remember which prefix to dial for the speed of the call they wish to make, and you would have to configure multiple route patterns in Unified CM (one for each speed). Fortunately, the Auto speed enables you to minimize this effort. If the majority of your calls are made using 64 kbps per channel (for example, 128 kbps, 384 kbps, 512 kbps, 768 kbps, and so on), you could use the Auto service in that case. You would then need to create only one other service for the rare case in which someone makes a call using 56 kbps per channel (for example, 112 kbps, 336 kbps, and so on).

Cisco recommends that you always use a # character in your service prefixes because the gateway recognizes the # as an end-of-dialing character. By placing this character in the service prefix, you block people from attempting to use the gateway for toll fraud by dialing the main number of the gateway, reaching the IVR, and then dialing out to an off-net number. The # can either be at the beginning (recommended) or the end of the service prefix. For example, if your access code to reach the PSTN is 8 for video calls, Cisco recommends that you configure the service prefix as #8 or 8#. Or, if you have two service prefixes as described above, you might use #80 for the Auto 64-kbps service and #81 for the Auto 56-kbps service.

The ramification of using a service prefix is that Unified CM must prepend the service prefix to the called number when sending calls to the IP/VC gateway. Because forcing users to dial the # would not be very user-friendly, Cisco recommends that you configure Unified CM to prepend the # to the dialed number. For example, if the access code to dial a video call to the PSTN is 8, you could configure a route pattern as 8.@ in Unified CM, and in the route pattern configuration you would configure the called number translation rule to prepend #8 whenever that route pattern is dialed. Or, if you have two service prefixes as described above, you might use 80.@ for the Auto 64-kbps service (prefixing # to the called number) and 81.@ for the Auto 56-kbps service (prefixing # to the called number).

Automated Alternate Routing (AAR)

When the IP network does not have enough bandwidth available to process a call, Unified CM uses its call admission control mechanism to determine what to do with the call. As described in the chapter on [IP Video Telephony, page 16-1](#), Unified CM performs one of the following actions with the call, depending on how you have configured it:

- Fail the call, playing busy tone to the caller and displaying a Bandwidth Unavailable message on the caller's screen
- Retry the video call as an audio-only call
- Use automated alternate routing (AAR) to re-route the call over an alternative path, such as a PSTN gateway

The first two options are covered in the chapter on [IP Video Telephony, page 16-1](#), and this section covers the AAR option.

To provide AAR for voice or video calls, you must configure the calling and called devices as members of an AAR group and configure an External Phone Number Mask for the called device. The External Phone Number Mask designates the fully qualified E.164 address for the user's extension, and the AAR group indicates what digits should be prepended to the External Phone Number Mask of the called device in order for the call to route successfully over the PSTN.

For example, assume that user A is in the San Jose AAR group and user B is in the San Francisco AAR group. User B's extension is 51212, and the External Phone Number Mask is 6505551212. The AAR groups are configured to prepend 91 for calls between the San Jose and San Francisco AAR groups. Thus, if user A dials 51212 and there is not enough bandwidth available to process the call over the IP WAN between those two sites, Unified CM will take user B's External Phone Number Mask of 6505551212, prepend 91 to it, and generate a new call to 916505551212 using the AAR calling search space for user A.

This same logic applies to video calls as well, with one additional step in the process. For video-capable devices, there is a field called *Retry Video Call as Audio*. As described in the chapter on [IP Video Telephony, page 16-1](#), if this option is enabled (checked), Unified CM does not perform AAR but retries the same call (that is, the call to 51212) as a voice-only call instead. If this option is disabled (unchecked), Unified CM performs AAR. By default, all video-capable devices in Unified CM have the *Retry Video Call as Audio* option enabled (checked). Therefore, to provide AAR for video calls, you must disable (uncheck) the *Retry Video Call as Audio* option. Additionally, if a call admission control policy based on Resource Reservation Protocol (RSVP) is being used between locations, the RSVP policy must be set to Mandatory for both the audio and video streams.

Furthermore, Unified CM looks at only the called device to determine whether the *Retry Video Call as Audio* option is enabled or disabled. So in the scenario above, user B's phone would have to have the *Retry Video Call as Audio* option disabled in order for the AAR process to take place.

Finally, devices can belong to only one AAR group. Because the AAR groups determine which digits to prepend, AAR groups also influence which gateway will be used for the rerouted call. Depending on your choice of configuration for outbound call routing to the PSTN, as discussed in the previous section, video calls that are rerouted by AAR might go out a voice gateway instead of a video gateway. Therefore, carefully construct the AAR groups and the AAR calling search spaces to ensure that the correct digits are prepended and that the correct calling search space is used for AAR calls.

While these considerations can make AAR quite complex to configure in a large enterprise environment, AAR is easier to implement when the endpoints are strictly of one type or the other (such as IP Phones for audio-only calls and systems such as the Tandberg T-1000 dedicated for video calls). When endpoints are capable of both audio and video calls (such as Cisco Unified Video Advantage or a Cisco IP Video Phone 7985G), the configuration of AAR can quickly become unwieldy. Therefore, Cisco recommends that large enterprise customers who have a mixture of voice and video endpoints give careful thought to the importance of AAR for each user, and use AAR only for select video devices such as dedicated videoconference rooms or executive video systems. [Table 4-14](#) lists scenarios when it is appropriate to use AAR with various device types.

Table 4-14 When to Use AAR with a Particular Device Type

Device Type	Device is used to call:	Enable AAR?	Comments
IP Phone	Other IP Phones and video-capable devices	Yes	Even when calling a video-capable device, the source device is capable of audio-only, thus AAR can be configured to route calls out a voice gateway.
IP Phone with Cisco Unified Video Advantage, or Cisco IP Video Phone 7985G	Other video-capable devices only	Yes	Because the device is used strictly for video calls, you can configure the AAR groups accordingly.
	IP Phones and other video-capable devices	No	It will be difficult to configure the AAR groups to route audio-only calls differently than video calls.
Sony or Tandberg SCCP endpoint	Other video-capable devices only	Yes	Because the device is used strictly for video calls, you can configure the AAR groups accordingly.
	IP Phones and other video-capable devices	No	It will be difficult to configure the AAR groups to route audio-only calls differently than video calls.
H.323 or SIP client	Other video-capable devices only	Yes	Because the device is used strictly for video calls, you can configure the AAR groups accordingly.
	IP Phones and other video-capable devices	No	It will be difficult to configure the AAR groups to route audio-only calls differently than video calls

Least-Cost Routing

Least-cost routing (LCR) and tail-end hop-off (TEHO) are very popular in VoIP networks and can be used successfully for video calls as well. In general, both terms refer to a way of configuring the call routing rules so that calls to a long-distance number are routed over the IP network to the gateway closest

to the destination, in an effort to reduce toll charges. (For Cisco Unified CM Release 4.1, LCR basically means the same thing as TEHO.) Unified CM supports this feature through its rich set of digit analysis and digit manipulation capabilities, including:

- Partitions and calling search spaces
- Translation patterns
- Route patterns and route filters
- Route lists and route groups

Configuring LCR for video calls is somewhat more complicated than for voice calls, for the following reasons:

- Video calls require their own dedicated gateways, as discussed previously in this chapter
- Video calls require much more bandwidth than voice calls

With respect to dedicated gateways, the logic behind why you might or might not decide to use LCR for video calls is very similar to that explained in the section on [Automated Alternate Routing \(AAR\)](#), [page 4-40](#). Due to the need to have different types of gateways for voice and video, it can become quite complex to configure all the necessary partitions, calling search spaces, translation patterns, route patterns, route filters, route lists, and route groups needed for LCR to route voice calls out one gateway and video calls out another.

With respect to bandwidth requirements, the decision to use LCR depends on whether or not you have enough available bandwidth on your IP network to support LCR for video calls to/from a given location. If the current bandwidth is not sufficient, then you have to determine whether the benefits of video calls are worth the cost of either upgrading your IP network to make room for video calls or deploying local gateways and routing calls over the PSTN. For example, suppose you have a central site with a branch office connected to it via a 1.544-Mbps T1 Frame Relay circuit. The branch office has twenty video-enabled users in it. A 1.544-Mbps T1 circuit can handle at most about four 384-kbps video calls. Would it really make sense in this case to route video calls up to the central site in order to save on toll charges? Depending on the number of calls you want to support, you might have to upgrade your 1.544-Mbps T1 circuit to something faster. Is video an important enough application to justify the additional monthly charges for this upgrade? If not, it might make more sense to deploy an IP/VC video gateway at the branch office and not bother with LCR. However, placing local IP/VC gateways at each branch office is not inexpensive either, so ultimately you must decide how important video-to-PSTN calls are to your business. If video is not critical, perhaps it is not worth upgrading the bandwidth or buying video gateways but, instead, using the Retry Video Call as Audio feature to reroute video calls as voice-only calls if they exceed the available bandwidth. Once a call is downgraded to voice-only, local gateway resources and bandwidth to perform LCR become more affordable and easier to configure.

ISDN B-Channel Binding, Rollover, and Busy Out

H.320 video uses multiple ISDN channels bound together to achieve the speeds needed to pass full-motion video. One of the problems with this bonding mechanism is that, when an inbound ISDN video call is received, the gateway does not know how many channels will be requested for that call until after it accepts the call and the source device indicates how many additional channels are required. If there are not enough B-Channels to satisfy the request, the call is disconnected. Therefore, careful traffic engineering is required to minimize the possibility that this situation will occur. Essentially, you want to ensure that there are always enough B-Channels available to handle the next call that might come in.

This B-Channel issue occurs in two cases:

- Inbound calls from the PSTN to the IP network
- Outbound calls from the IP network to the PSTN

Inbound Calls

For inbound calls, consider the following scenario:

A company has a Cisco 3526 IP/VC Gateway with an ISDN PRI circuit connecting it to a central office (CO) switch. The ISDN PRI circuit in this case offers 23 B-Channels. A video call is received from the PSTN at 384 kbps. This call takes six B-Channels, leaving 17 available. A second and third 384-kbps call are received on the line while the first one is still active. These each take an additional six channels, leaving five channels available. When the fourth 384-kbps call is received, the gateway will answer the call but, recognizing that it does not have enough B-Channels available (it only has five left but the call requires six), it will disconnect (by sending a Q.931 RELEASE COMPLETE with "16: Normal Call Clearing" as the reason). The caller attempting to make the fourth call will not know why the call failed and might redial the number repeatedly, trying to make the call work.

On Cisco Unified Videoconferencing gateways, you can minimize your chances of running into this issue by configuring the gateway to send a request to the CO to busy-out the remaining B-Channels (in this example, five channels) whenever the gateway reaches a certain threshold of utilization (configured as a percentage of total bandwidth).

In addition, you can have the CO provision multiple ISDN circuits in a trunk group. When the first circuit reaches the busy-out threshold, calls will roll over to the next PRI in the group. The Cisco 3540 IP/VC Gateway offers two ISDN PRI connections and supports bonding channels across both ports. For example, port 1 might have only five channels available while port 2 is sitting idle and, therefore, has 23 channels available. By taking the five channels from port 1 and one channel from port 2 and bonding them together, the fourth 384-kbps call can succeed. This leaves 22 channels available on controller 2, and at some point additional inbound calls would reach the busy-out threshold again. At that point the remaining channels on port 2 will be busied out, and all further inbound calls will be rejected with cause code "Network Congestion." Cisco Unified Videoconferencing gateways cannot bond channels across different gateways or across different Cisco 3540 gateway models in the same Cisco 3544 chassis, so two ports is the maximum that you can bond together. The CO switch can still roll calls over to a third or fourth PRI in the trunk group (most COs support trunk groups of up to 6 circuits), but you cannot bond channels between PRI number one and PRI number three, for example, as you can between PRI number one and PRI number two.

The busy-out logic described above depends on the assumption that all calls take place at the same speed. Suppose, for example, that two 384-kbps calls are active on a port and a 128-kbps call came in. This call would take only two channels, using a total of 14 channels for the three calls ($6+6+2 = 14$) and leaving nine channels available on the circuit. However, if the busy-out threshold is set at 18 channels (assuming that all calls would take place at 384-kbps), only four channels are still available under this busy-out threshold. If another 384 kbps call comes in at this point, the call will fail because the remaining four channels are not enough to support the call. Also, because the busy-out threshold of 18 channels has not been reached yet (only 14 channels are used), the circuit is not busied out and calls will not roll over to the next circuit. This condition will persist until one of the existing calls is disconnected. To avoid such situations, it is important to try to standardize on a single call speed for all calls.

Outbound Calls

Outbound calls encounter the same potential situations as inbound calls, but the way in which the busy-out occurs is different. The Cisco 3500 Series IP/VC Gateways support messages called Resource Availability Indicator and Resource Availability Confirm (RAI/RAC). The RAI/RAC messages are defined under the H.225 RAS specification and are used by the gateways to tell the gatekeeper that they are full and to no longer route any more calls to them. When the gateway reaches the busy-out threshold, it sends an RAI message with a status of True to the gatekeeper. True means "Do not send me any more calls;" False means "I am available." The gateway sends an RAI=False as soon as it is no longer at its

busy-out threshold. The busy-out threshold for outbound calls is separate from the busy-out threshold for inbound calls, and you can configure them differently so that inbound calls will roll over to the next available circuit but outbound calls will still be accepted, or vice versa. For example, you could configure the RAI threshold to 12 channels but the ISDN busy-out threshold to 18 channels. When two 384 kbps are active, outbound calls will roll over to the next available gateway, but a third 384-kbps inbound call could still be received. An equally efficient method of achieving outbound call busy-out failover is to use Unified CM's route group and route list construct, as described in the following section, instead of the RAI/RAC method.

Configuring the Gateways in Unified CM

You can configure an IP/VC gateway in either of the following ways in Unified CM:

- Configure it as an H.323 gateway, and Unified CM will route calls directly to the gateway.
- Configure an H.225 gatekeeper-controlled trunk to the gatekeeper, and route calls to the gateway through the gatekeeper.

If you have only one gateway, it is probably easier to configure it directly in Unified CM instead of going through a trunk to get to it. If you have multiple gateways for load balancing and redundancy, you can either configure them all in Unified CM and place them into a route group(s) and route list, or configure an H.225 trunk to the gatekeeper and rely on RAI/RAC between the gateways and the gatekeeper to tell Unified CM which gateway it should send a given call to.

For inbound calls from the PSTN to Unified CM, the Cisco Unified Videoconferencing gateways can either register with a gatekeeper or be configured with the IP addresses of up to three Unified CM servers to which they should send all inbound call requests. This method is known as peer-to-peer mode. Either way, the goal is have all inbound calls received by the gateways sent to Unified CM so that Unified CM can decide how to route the calls. See [Gatekeepers, page 16-22](#), for more details on how to configure the gatekeeper to route calls from the gateways to Unified CM.

Call Signaling Port Numbers

By default, the Cisco Unified Videoconferencing Gateways listen on TCP port 2720 instead of the well-known port 1720. However, also by default, Unified CM sends H.323 calls to port 1720. You can change the port that the gateway listens on or you can change the port that Unified CM sends to in the H.323 gateway device configuration in Unified CM. Either way, both sides have to match in order for outbound calls to the gateway to succeed.

In the inbound direction, when configured to operate in peer-to-peer mode, the Cisco Unified Videoconferencing Gateways will send the call to Unified CM on port 1720. When configured to register with a gatekeeper, Unified CM uses a randomly generated port number for all gatekeeper-controlled trunks. This method enables Unified CM to have multiple trunks to the same gatekeeper. This port number is included in the Registration Request (RRQ) from Unified CM to the gatekeeper, so the inbound H.225 setup message from the gateway to Unified CM will be sent to this port number. However, if the gateway is configured directly in Unified CM as an H.323 gateway device, Unified CM will ignore the fact that the call came in on the TCP port of the H.225 trunk and will instead match the source IP address to the H.323 gateway device configured in its database. If it does not find a matching device, Unified CM will treat the call as if it came in on the trunk.

In the outbound direction, if Unified CM uses a gatekeeper-controlled H.225 trunk to reach the gateway, the gatekeeper will tell Unified CM which TCP port to use to reach the gateway. If the gateway is configured in Unified CM as an H.323 gateway device (that is, peer-to-peer mode), then Unified CM must be configured to send calls either to port 2720 (default) or to 1720 (if the listening port on the gateway has been modified).

Call Signaling Timers

Due to the delay inherent in H.320 bonding, video calls can take longer to complete than voice calls. Several timers in Unified CM are tuned, by default, to make voice calls process as fast as possible, and they can cause video calls to fail. Therefore, you must modify the following timers from their default values in order to support H.320 gateway calls:

- H.245TCSTimeout
- Media Exchange Interface Capability Timer
- Media Exchange Timer

Cisco recommends that you increase each of these timers to 25 by modifying them under the Service Parameters in Unified CM Administration. Note that these are cluster-wide service parameters, so they will affect calls to all types of H.323 devices, including voice calls to existing H.323 Cisco Voice Gateways.

Voice Gateways Bearer Capabilities

H.323 calls use the H.225/Q.931 Bearer Capabilities Information Element (bearer-caps) to indicate what type of call is being made. A voice-only call has its bearer-caps set to "speech" or "3.1 KHz Audio," while a video call has its bearer-caps set to "Unrestricted Digital Information." Cisco voice gateways, some legacy PBXs, and most cellular carriers do not support Unrestricted Digital Information bearer-caps. Therefore, calls to a voice gateway might fail if Unified CM attempts the call as a video call.

Unified CM decides which bearer-caps to set, based on the following factors:

- Whether the calling and/or called devices are video-capable
- Whether the region in Unified CM is configured to allow video for calls between those devices

For example, consider a network in which a video-capable device (such as a Cisco Unified IP Phone with a VT Advantage client associated to it) is configured in the same region as a Cisco voice gateway. When the user dials 9 to access an outside line, Unified CM determines that the calling device is video-capable and that the region is set to allow 384 kbps of video bandwidth. Unified CM then sets the bearer-caps to Unrestricted Digital Information for that call. But because the call is to a Cisco voice gateway, the gateway rejects the call with the cause code "Incompatible Destination." This problem will occur in any network that uses H.323 voice gateways and that has IP Phones associated with Cisco Unified Video Advantage. From the user's perspective, things will appear to work fine before installing Cisco Unified Video Advantage, but calls to the PSTN will fail as soon as the user plugs a PC running Cisco Unified Video Advantage into the IP Phone.

This situation exists only on calls to H.323 voice gateways. If the Cisco voice gateway uses MGCP to communicate with Unified CM, the problem will not occur because Unified CM does not support video on its MGCP protocol stack and because, in MGCP mode, Unified CM has complete control over the D-Channel signaling to the PSTN. Likewise, if the Cisco voice gateway uses SIP to communicate with Unified CM, the problem will not occur then either because Unified CM does not support video on its SIP protocol stack, but even if it did, the gateway would simply need to ignore the video capabilities passed in Unified CM's outgoing Session Description Protocol (SDP) advertisement.

To prevent this situation, configure the bearer-caps on all Cisco H.323 voice gateways by using the **bearer-caps** command under the **voice-port** configuration mode, as illustrated in the following example:

```
gateway#configure terminal
gateway(config)#voice-port 1/0:23
gateway(config-voiceport)#bearer-caps speech
```




CHAPTER 5

Cisco Unified CM Trunks

Last revised on: October 30, 2008

Cisco Unified Communications Manager (Unified CM, formerly Cisco Unified CallManager) Release 4.0 introduced support for Session Initiation Protocol (SIP) trunks; prior to Release 4.0, Unified CM supported only H.323 trunks. This chapter discusses design considerations for Cisco Unified CM releases 5.x, but much of the discussion also applies to Unified CM releases 4.x and 3.3.

What's New in This Chapter

[Table 5-1](#) lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 5-1 *New or Changed Information Since the Previous Release of This Document*

New or Revised Topic	Described in:
H.323 trunks do not support the iLBC codec.	Table 5-2
T.38 Fax on SIP trunks in Unified CM 5.x does not interoperate with T.38 Fax on MGCP.	Table 5-2

A Comparison of H.323 and SIP Trunks

Cisco Unified CM trunk connections support both H.323 and SIP. In many cases, the decision to use H.323 or SIP is driven by the unique feature(s) offered by each protocol. There are also a number of external factors that can affect the choice of trunk protocol, such as customer preference or the protocol's maturity and degree of interoperability offered between various vendors' products.

For trunk connections between Cisco devices, this decision is relatively straightforward. For trunk connections to other vendors' products and to service provider networks, it is important to understand which features are required by the customer and the extent of interoperability between any two vendors' products.

[Table 5-2](#) compares some of the features offered over H.323 and SIP trunks between Unified CM clusters.

Table 5-2 Comparison of H.323 and SIP Features on Cisco Unified CM Trunks

Feature	H.323	QSIG over H.323	SIP
Calling Line (Number) Identification Presentation	Yes	Yes	Yes
Calling Line (Number) Identification Restriction	Yes	Yes	Yes
Calling Name Identification Presentation	Yes	Yes	Yes
Calling Name Identification Restriction	Yes	Yes	Yes
Connected Line (Number) Identification Presentation	Yes	Yes	Yes
Connected Line (Number) Identification Restriction	Yes	Yes	Yes
Connected Name Identification Presentation	Yes	Yes	Yes
Connected Name Identification Restriction	Yes	Yes	Yes
Alerting Name	No	Yes	Yes
Call Transfer (Blind/Attended)	Yes/Yes	Yes/Yes	Yes/Yes
Call Forward All	Yes	Yes	Yes
Call Forward Busy	Yes	Yes	Yes
Call Forward No Reply	Yes	Yes	Yes
Call Completion to Busy Subscriber	No	Yes	No
Call Completion No Reply	No	Yes	No
Subscribe/Notify, Publish – Presence	No	No	Yes
Message Waiting Indication (MWI: lamp ON, lamp OFF)	No	Yes	Yes
Path Replacement	No	Yes	No
Call Hold/Resume	Yes	Yes	Yes
Music On Hold	Yes	Yes	Yes
DTMF-relay	H.245 Out Of Band (OOB)	H.245 Out Of Band (OOB)	RFC 2833, KPML (OOB), Unsolicited Notify (OOB)
SIP Early Offer	N/A	N/A	Yes – Requires MTP for outbound calls
SIP Delayed Offer	N/A	N/A	Yes
H.323 Fast Start	Yes – MTP required for Outbound Fast Start	Yes – MTP required for Outbound Fast Start	N/A
H.323 Slow Start	Yes	Yes	N/A
CODECs	G.711, G.722, G.723, G.729, G722	G.711, G.722, G.723, G.729, G722	G.711, G.722, G.723, G.729, G722, iLBC, AAC
CODECs with MTP	G.711, G.722, G.723, G.729	G.711, G.722, G.723, G.729	G.711
Video	Yes	Yes	Yes
T.38 Fax ¹	Yes	Yes	Yes
Signalling Authentication	No	No	Digest, TLS

Table 5-2 Comparison of H.323 and SIP Features on Cisco Unified CM Trunks (continued)

Feature	H.323	QSIG over H.323	SIP
Signalling Encryption	No	No	TLS
Media Encryption	SRTP	SRTP	No

1. T.38 Fax on SIP trunks in Unified CM 5.x does not interoperate with T.38 Fax on MGCP.

Overview of H.323 Trunks

The H.323 trunk provides connectivity to other Unified CM clusters and other H.323 devices such as gateways. H.323 trunks support most of the audio and video codecs that Unified CM 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 Unified CM 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. (Unified CM does not support H.450 for supplementary call services.) In this case, MoH is an example of an ECS operation. H.323 trunks do not support multicast MoH, therefore 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 Unified CM and assigned to each trunk. A region limits the amount of bandwidth allocated for calls by specifying the audio codec type and video bandwidth per call 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.

Overview of SIP Trunks

SIP trunks provide connectivity to other SIP devices such as gateways, proxies, voicemail systems, and other Unified CM clusters. Cisco Unified CM 5.x and 6.0 introduced major enhancements for SIP trunks and removed the limitations in Cisco Unified CM 4.1 and 4.0, such as single codec support, lack of video support, and the mandatory media termination point (MTP) for RFC 2833 DTMF support.

The main enhancements to SIP trunks in Cisco Unified CM 6.0 are the support for the iLBC and AAC codecs and SIP PUBLISH. By providing improved performance, SIP PUBLISH provides the preferred mechanism for Cisco Unified CM 6.0 to send IP phone presence information to Cisco Unified Presence over a SIP trunk.

For the complete list of new enhancements for SIP trunks, refer to the Cisco Unified Communications Manager product release notes available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_release_notes_list.html

When used for intercluster trunking, SIP trunks do not support Secure Real-Time Transport Protocol (SRTP) or QSIG Tunneling using Annex M1.

IP PSTN and IP Trunks to Service Provider Networks

With support for both H.323 and SIP trunks in Cisco Unified CM, service providers are starting to offer non-TDM PSTN connections to enterprise customers. Apart from the obvious benefit of the cost savings from deploying non-TDM interfaces, in many cases these IP-based PSTN connections also offer additional voice features over traditional PSTN interfaces.

The choice of H.323 or SIP as the IP trunking protocol often depends on the service provider, although today the majority of IP-based PSTN connections are likely to use H.323 because of its relative maturity and greater number of proven deployments when compared with SIP. However, because SIP offers additional capabilities such as Presence and support for many multimedia applications (such as instant messaging), SIP will probably become the more widely deployed Unified Communications protocol in the long term.

Because both the SIP and H.323 standard are open to a degree of interpretation (there are many optional as well as mandatory requirements), the degree of interoperability between vendors is often something that improves over time as the protocols mature. To this end, the Cisco Unified Border Element (formerly the IP-to-IP Gateway) offers a number of interoperability features as well as a natural demarcation point as a Session Border Controller when connecting to other vendors' H.323 and SIP networks.

Cisco Unified Border Element

The Cisco Unified Border Element product provides a wide range of signalling and media functionality between the enterprise and service-provider Cisco Unified Communications networks. Cisco Unified Border Element provides a network-to-network interface point for:

- Address and port translations (privacy and topology hiding)
- Signaling interworking (H.323 and SIP)
- Media interworking (DTMF, fax, modem, and codec transcoding)
- QoS and bandwidth management (QoS marking using ToS, DSCP, and bandwidth enforcement using RSVP and codec filtering)
- Billing and CDR normalization

The Cisco Unified Border Element is a licensed Cisco IOS application that is available on the Cisco 2800 and 3800 Series Integrated Service Routers (ISR), the Cisco AS5350XM and AS5400XM Media Gateways, and the Cisco 7200VXR and Cisco 7301 Series router and gateway platforms.

Cisco recommends the use of Cisco Unified Border Element for any IP PSTN deployment.

H.323 Trunks

The following major types of H.323 trunks can be configured in a Unified CM:

- [Intercluster Trunk \(Non-Gatekeeper Controlled\)](#), page 5-5
- [Intercluster Trunk \(Gatekeeper Controlled\)](#), page 5-5
- [H.225 Trunk \(Gatekeeper Controlled\)](#), page 5-6

Intercluster Trunk (Non-Gatekeeper Controlled)

This trunk is the simplest and is used for connecting to other Unified CM 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 Unified CM if bandwidth control is required.

When defining this type of trunk, you may define up to three remote Unified CM servers in the same destination cluster. The trunk will automatically load-balance across all defined remote Unified CM servers. In the remote cluster, it is important to configure a corresponding intercluster trunk (non-gatekeeper controlled) that has a Unified CM Group containing the same servers that were defined as remote Unified CM servers in the first cluster. A similar configuration is required in each Unified CM 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 Unified CM 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 Unified CM 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 Unified CMs 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 Unified CM 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 Unified CM 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 Unified CM, page 5-12](#), for details of this discovery process.) This type of trunk is the recommended H.323 trunk if all Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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 Unified CM Group consisting of the two servers that make up the standard subscriber pair. (See [Call Processing Subscriber, page 8-7](#), for more information on subscriber redundancy.) For complete redundancy in each full cluster, four trunks would be needed, using four different device pools and resulting in eight subscribers registering with the gatekeeper. (The same result could be achieved with three trunks and larger Unified CM Groups.)

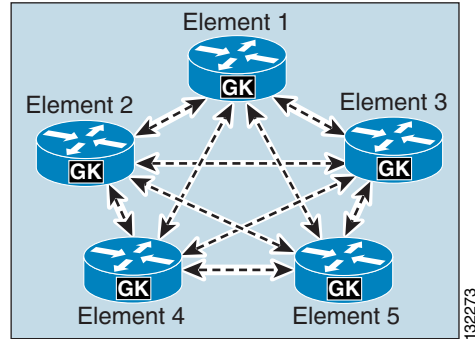
During registration, several parameters are passed between Unified CM and the gatekeeper. Unified CM uses an ephemeral User Datagram Protocol (UDP) port for gatekeeper Registration Admission Status (RAS) messages. This port would normally be UDP 1719. However, Unified CM must be able to identify precisely which H.323 daemon is the originator of a RAS message from a particular server; 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 Unified CM Group:

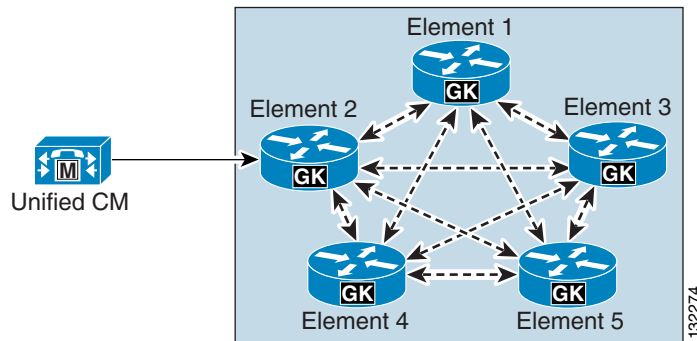
- H.225 call signaling port
- h323_id
- CanMapAlias support
- Technology prefix
- H.225 call signaling address

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.

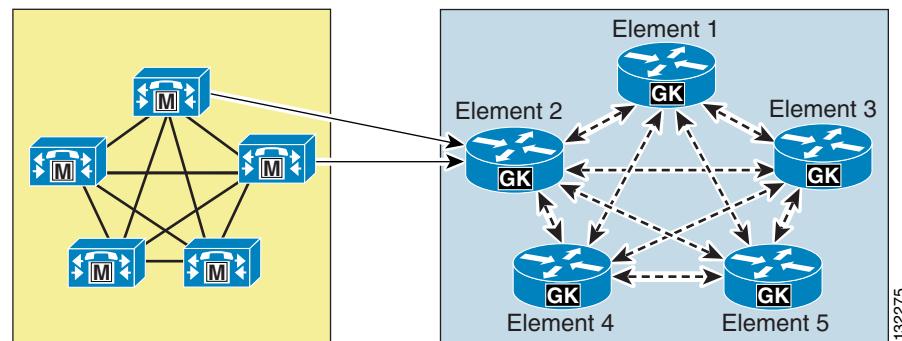
[Figure 5-1](#) shows a cluster of gatekeepers that use Gatekeeper Update Protocol (GUP) to communicate. (See the chapter on [Call Processing, page 8-1](#), for more information on gatekeepers.)

Figure 5-1 Gatekeeper Cluster

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

Figure 5-2 H.323 Trunk with a Single Unified CM Subscriber

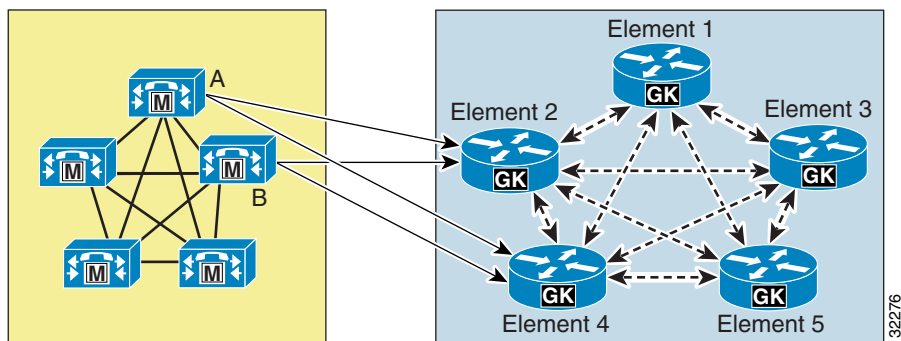
If there are multiple subscribers in the Unified CM Group associated with the trunk, additional connections will be established between the Unified CM cluster and the gatekeeper cluster, as illustrated in Figure 5-3.

Figure 5-3 H.323 Trunk with Multiple Unified CM 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 Unified CM. 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 Unified CM 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 Unified CM Group containing subscriber servers A and B

Using this approach, the Unified CM 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 Unified CM 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 Unified CM, and the “_2” and “_3” suffixes are appended automatically by the Unified CM 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 Unified CM in the Unified CM 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.

Load-Balancing Outbound Calls

In the majority of cases, the standard method of assigning Unified CM Groups to devices is sufficient to handle the call distribution of outbound calls over IP trunks from call processing subscribers. IP trunk calls might appear to originate randomly from call processing subscribers, but the trade-off for this random call origination is reduced call processing and reduced Intra-Cluster Communication Signaling (ICCS) traffic within the cluster. Load balancing of outbound IP trunk calls across call processing servers, as described below, can be counter-productive because the advantages gained from predictable call origination within the cluster can be outweighed by the increase in ICCS traffic created by calls from phones registered to one subscriber extending their communication to another server within the cluster to originate the outgoing IP trunk call.

For the initiation of outbound calls over an H.323 trunk, the following key factors within a Unified CM cluster determine which server is selected:

- Which Unified CM servers have an active H.323 daemon for the selected trunk
- Whether the phone originating the call is registered to a Unified CM server with an active H.323 daemon for the selected trunk

For IP trunks, the server selection process is not as intuitive as it is for H.323 connections to gateways where, for example, route lists and route groups can be used to select the gateway(s) used to originate the call to the PSTN and also to provide call distribution among multiple gateways. While the route list and route group process still operates for trunk selection and call distribution, a second process is also in operation to select the server from which the H.323 call will originate. In other words, after the route list and route group configuration selects the outgoing trunk, one server in the Unified CM Group for the selected trunk must be chosen to originate the H.323 call.

For all IP trunks, this process of server selection for outgoing calls is as follows:

- If there is an active H.323 daemon for the selected trunk on the Unified CM server to which the phone or device originating the call is registered (that is, if the server is one of those listed in the trunk's Unified CM Group), then use this Unified CM server to originate the H.323 call.
- If there is no active H.323 daemon for the selected trunk on the Unified CM server to which the phone or device originating the call is registered, then select a server on a round-robin basis from the Unified CM Group of the selected trunk.

**Note**

All servers defined in a trunk's Unified CM Group will run an active H.323 daemon for that trunk. On each server defined within a trunk's Unified CM Group, a unique H.323 daemon is created for each H.323 trunk.

To provide predictable and deterministic load balancing for outgoing calls over IP trunks, you must consider the trunk selection behavior described above.

Subscriber-based load balancing of outgoing IP trunk calls can be achieved as follows:

- To load-balance outbound trunk calls across only a subset of the call processing servers in the cluster, define multiple trunks and assign only a single subscriber to the Unified CM Group of each trunk.
- If you want subscriber redundancy within Unified CM Groups, the easiest way to provide predictable subscriber load balancing of outbound trunk calls is to ensure that the subscribers used to register phones are different from those initiating trunk calls.

For example, to spread outbound trunk calls across four subscribers in the cluster, perform the following tasks:

- Configure four H.323 trunks with four Unified CM Groups, all contained within a route group with circular call distribution.
- Define Unified CM Groups as follows:
 - Group A: Subscriber A
 - Group B: Subscriber B
 - Group C: Subscriber C
 - Group D: Subscriber D

With no backup subscribers defined, if the primary subscriber for the specified trunk fails, Unified CM will re-route outgoing calls to the next trunk in the route group.

In the example above, backup subscribers can be defined within each Unified CM Group for each trunk provided, using subscribers A, B, C, and D as follows:

- Group A: Subscriber A; Subscriber B
- Group B: Subscriber B; Subscriber C
- Group C: Subscriber C; Subscriber D
- Group D: Subscriber D; Subscriber A

However, to avoid subscriber selection for outgoing trunk calls based on phone registration and H.323 daemon subscriber co-location, all phones in the cluster should be registered to other call processing servers (for example, subscribers E, F, G, and H), with server redundancy as required using Unified CM Groups.

To spread outbound trunk calls across all eight subscribers in a cluster, perform the following tasks:

- Configure eight H.323 trunks with eight different Unified CM Groups, each containing only one subscriber and all contained within a circular route group.
- Define Unified CM Groups 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, that do not support the Empty Capabilities Set (ECS) for supplementary services, or that require H.323 FastStart.

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 Unified CM and the other device.

H323 Outbound FastStart Call Connections

Calls that are placed from IP phones over large WAN topologies can experience voice clipping when the called party goes off-hook to answer the call. When H.323 trunks or gateways are separated from the Unified CM server, significant delays can occur because of the many H.245 messages that are exchanged when a call is set up.

With the FastStart feature, information that is required to complete a media connection between two parties gets exchanged during the H.225 portion of call setup, and this exchange eliminates the need for H.245 messages. The connection experiences one round-trip WAN delay during call setup, and the calling party does not experience voice clipping when the called party answers the call.

Unified CM uses media termination points (MTP) for making an H.323 outbound FastStart call. Unified CM starts an outbound FastStart call by allocating an MTP and opening the receive channel. Next, the H.323 Fast Connect procedure sends the SETUP message with a FastStart element to the called endpoint. The FastStart element includes information about the receiving channel for the MTP.

Other MTP Uses

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 Unified CM 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 fix-up 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 Unified CM service parameter **Fail Call if MTP allocation fails** under the H.323 section to **True**.

H.323 Operation in Unified CM

This section provides information on how the H.323 protocol is used and implemented in Unified CM, 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 Unified CM 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. 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 Unified CM cluster but the Unified CM 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 Unified CM 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 Unified CM 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 use DHCP, they cannot be used with a Unified CM using DNS unless they support Dynamic DNS. With the RasAggregator, Unified CM 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 Unified CM in an ACF message.
- The RasAggregator also ensures that all calls by the H.323 clients are made through Unified CM and not directly between the clients themselves, thus ensuring that dialing rules and codec restrictions are enforced.

- H.323 gateway

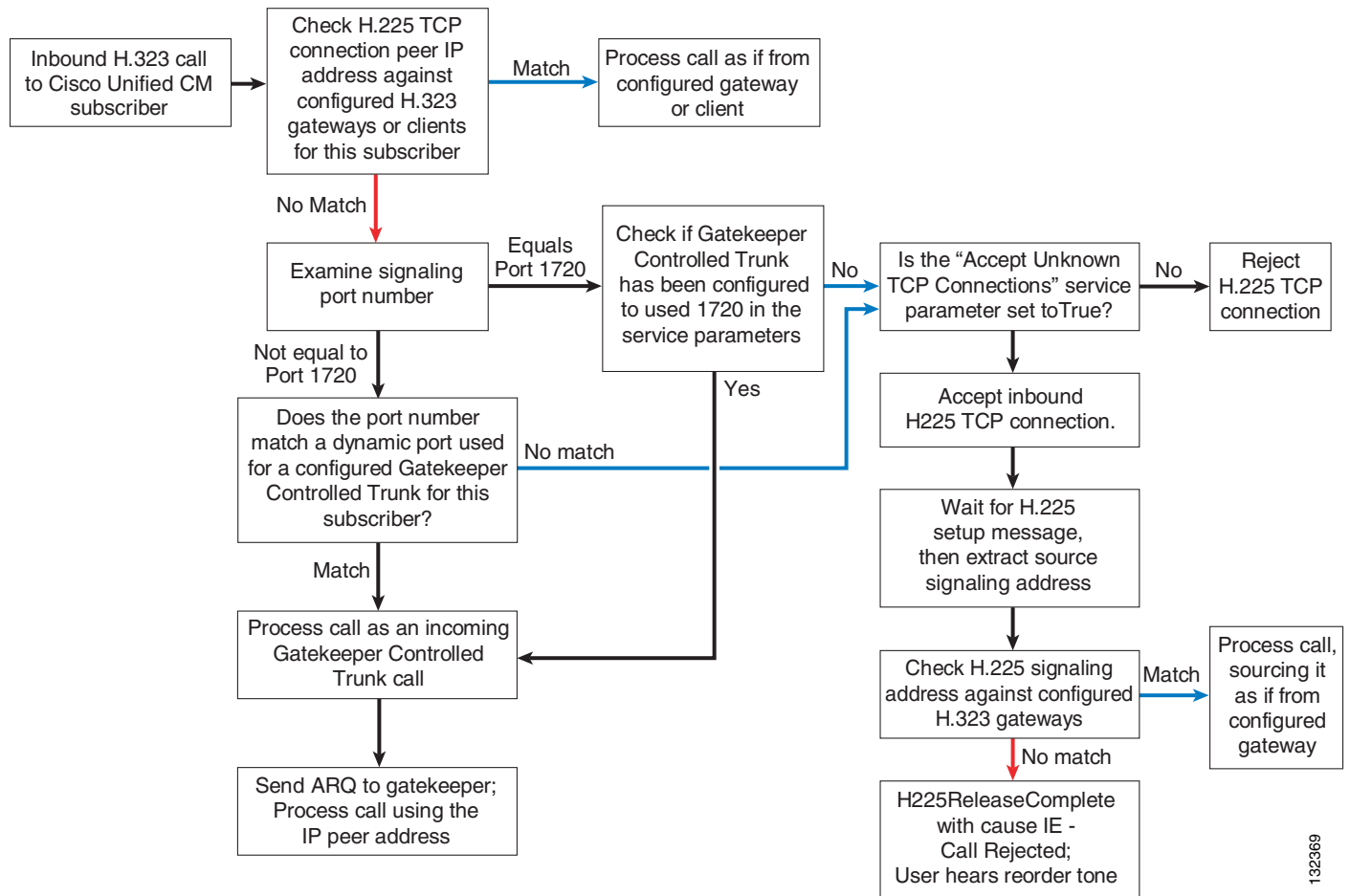
The H.225D is active on all subscribers in the Unified CM Group associated with the H.323 gateway.

- H.323 trunks

The H.225D is active on all subscribers in the Unified CM Group associated with the H.323 trunk. A RAS daemon registers the trunk with the gatekeeper from all subscribers in the associated Unified CM Group.

When an incoming H.323 call is made to a subscriber in a Unified CM 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 5-5 shows how this process works.

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



Unified CM 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 Unified CM Release 3.2 or later. Whenever a call is received, Unified CM looks for an H.225 User-to-User Information Element (UUIE) that indicates if the other end is another Unified CM. 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 Unified CM 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 Unified CM clusters.

- Tunneled Q.SIG or H.323 Annex M1

With the release of Cisco Unified CM 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 Unified CM clusters and other verified systems that also support H.323 Annex M1. These features include:

- Path replacement
- Message waiting indication (MWI)
- Callback

- Alternate Endpoints

When registering with a gatekeeper that supports this feature, such as a Cisco Multimedia Conference Manager (MCM) Gatekeeper, Unified CM 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 Unified CM 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), Unified CM 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 Unified CM 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.

SIP Trunks

As with H.323 trunks, there are several design considerations to be taken into account when deploying SIP trunks. This section describes those design considerations.

General Deployment Considerations:

For SIP trunk connections to third-party devices, such as SIP-based PBXs or service-provider IP PSTN connections, Cisco recommends the use of Delayed Offer for outbound calls from Unified CM and either Delayed Offer or Early Offer for inbound calls to Unified CM. Using Delayed Offer for outbound calls

removes the requirement to assign MTP resources to the SIP trunk, except in cases where a mismatch in DTMF transport types exists between the called and calling endpoints (in which case Unified CM will insert an MTP dynamically).

As previously mentioned, Cisco also recommends the deployment of the Cisco Unified Border Element on any IP PSTN SIP trunk connection from Unified CM to a voice service provider.

For DTMF over Unified CM SIP trunks, RFC 2833 is the recommended DTMF transport method.

SIP Delayed Offer, Early Offer, and DTMF are discussed in more detail in the subsequent sections.

DTMF Transport

There are several methods of transporting DTMF information between SIP endpoints. In general terms, these methods can be classified as out-of-band (OOB) and in-band signaling. In-band DTMF transport methods send either raw or signaled DTMF tones within the RTP stream, and they need to be handled and interpreted by the endpoints that generate and/or receive them. Out-of-band (OOB) signaling methods transport DTMF tones outside of the RTP path, either directly to and from the endpoints or via a call agent such as Cisco Unified CM, which interprets and/or forwards these tones as required.

Out-of-band (OOB) SIP DTMF signaling methods include Unsolicited Notify (UN), Information (INFO), and Key Press Markup Language (KPML). While KPML (RFC 4730) is the OOB signaling method preferred by Cisco, KPML is not widely used in the market place at this time. Currently, the only known products supporting KPML are Cisco Unified CM, Cisco IOS Gateways (Release 12.4 and later), and some models of Cisco IP Phones. Unsolicited Notify is a nonstandard method and is used only on Cisco IOS Gateways (Release 12.2 and later). INFO is not supported by Unified CM.

In-band DTMF transport methods send DTMF tones as either raw tones in the RTP media stream or as signaled tones in the RTP payload using RFC 2833. Among SIP product vendors, RFC 2833 has become the predominant method of sending and receiving DTMF tones and is supported by the majority of Cisco voice products.

Because in-band signaling methods send DTMF tones in the RTP media stream, the SIP endpoints in a session must either support the transport method used (for example, RFC 2833) or provide a method of intercepting this in-band signaling and converting it. If the two endpoints are using a back-to-back user agent (B2BUA) server for the call control (for example, Cisco Unified CM) and the endpoints negotiate different DTMF methods between each device and call control box, then the call control box determines how to handle the DTMF differences, either via MTP insertion or via OOB methods. With Unified CM, a DTMF transport mismatch (for example, in-band to out-of-band DTMF) is resolved by inserting a Media Termination Point (MTP), which terminates the RTP stream with in-band DTMF signaling (RFC 2833), extracts the DTMF tones from the RTP stream, and forwards these tones out-of-band to Unified CM, where they are then forwarded to the endpoint supporting out-of-band signaling. In this case, the MTP is always in the media path between the two endpoints because there is no MTP codec dependency for DTMF translation.

In-band DTMF tones can also be transported as raw (audible) tones in the RTP media stream. This transport method is not widely supported by Cisco products and, in general, is not recommended as an end-to-end DTMF transport mechanism. In-band audio DTMF tones can generally be reproduced reliably only when using G.711 a-law or mu-law codecs, and they are not suitable for use with low-bandwidth codecs. In cases where in-band audio is the only available DTMF transport mechanism, the Cisco Unified Border Element can be used to translate the in-band audio DTMF signaling into RFC 2833 signaling.

SIP Delayed Offer and Early Offer

Cisco Unified CM uses the SIP Offer/Answer model for establishing SIP sessions, as defined in RFC 3264. In this context, an Offer is contained in the Session Description Protocol (SDP) fields sent in the body of a SIP message. The Offer typically defines the media characteristics supported by the device (media streams, codecs, directional attributes, IP address, and ports to use). The device receiving the Offer sends an Answer in the SDP fields of its SIP response, with its corresponding matching media streams and codec, whether accepted or not, and the IP address and port on which it wants to receive the media streams. Unified CM uses this Offer/Answer model to establish SIP sessions as defined in the key SIP standard, RFC 3261.

RFC 3261 defines two ways that SDP messages can be sent in the Offer and Answer. These methods are commonly known as Delayed Offer and Early Offer, and support for both methods by User Agent Client/Servers is a mandatory requirement of the specification. In the simplest terms, an initial SIP Invite sent with SDP in the message body defines an Early Offer, whereas an initial SIP Invite without SDP in the message body defines a Delayed Offer.

In an Early Offer, the session initiator (calling device) sends its capabilities (for example, codecs supported) in the SDP contained in the initial Invite (thus allowing the called device to choose its preferred codec for the session). In a Delayed Offer, the session initiator does not send its capabilities in the initial Invite but waits for the called device to send its capabilities first (for example, the list of codecs supported by the called device, thus allowing the calling device to choose the codec to be used for the session).

Delayed Offer and Early Offer are the two media capabilities exchange options available to all standards-based SIP switches. Most vendors have a preference for either Delayed Offer or Early Offer, each of which has its own set of benefits and limitations.

**Note**

Unified CM can support Delayed Offer in one direction and Early Offer in the other direction. This capability can often be useful in situations where a SIP switch connecting to Unified CM via a SIP trunk wishes to control the codecs offered and selected for inbound and outbound calls (that is, where using Delayed Offer outbound from Unified CM and Early Offer inbound to Unified CM allows the service provider to send the Offer in all cases and, in doing so, to decide which codecs are offered for all calls.)

Early Media

In certain circumstances, a SIP session might require that a media path be set up prior to the finalization of the media capabilities exchange between the two SIP endpoints. To this end, the SIP protocol allows the establishment of Early Media after the initial Offer has been received by an endpoint. Some reasons for using Early Media include:

- The called device might want to establish an Early Media RTP path to reduce the effects of audio cut-through delay (clipping) for calls experiencing long signaling delays or to provide a network-based voice message to the caller.
- The calling device might want to establish an Early Media RTP path to access a DTMF or voice-driven IVR system.

Unified CM supports Early Media for both Early Offer and Delayed Offer calls.

**Note**

The terms *Early Offer* and *Early Media* are often confused; they are not the same.

Media Termination Points

Media Termination Points (MTPs) are generally not required for Delayed Offer calls from Unified CM SIP trunks. For this reason, Cisco recommends Delayed Offer as the call setup method for outbound calls from Unified CM SIP trunks. For outbound Early Offer calls from Unified CM, MTP resources are required (SIP Trunk **MTP required** box is checked) and remain in the media path for the duration of the call.

For calls inbound and outbound from Unified CM, endpoints can negotiate the use of RFC 2833 or an out-of-band DTMF method (for example, KPML) end-to-end. If a common DTMF method cannot be negotiated between the endpoints, Cisco Unified CM 5.x and later releases will insert an MTP dynamically. Cisco Unified CM 5.x and later releases support Delayed Offer (Invite without SDP) by default. Although Delayed Offer is a mandatory part of the SIP RFC 3261 specification, some SIP applications do not support it. In those cases, you must configure the SIP trunk to support Early Offer by pre-allocating an MTP under the SIP Trunk configuration.

MTPs are available in three forms:

- Software-based MTPs in Cisco IOS gateways — Available with any Cisco IOS T-train software release and scaling up to 500 sessions (calls) on the Cisco 3845 Integrated Services Router
- Hardware-based MTPs in Cisco IOS gateways — Available with any Cisco IOS T-train software release, hardware MTPs use on-board DSP resources and scale calls according to the number of DSPs supported on the Cisco router platform
- Software-based MTPs using the Cisco IP Voice Media Streaming Application on a Cisco Media Convergence Server (MCS)

The following example configuration is for a Cisco IOS software-based MTP:

```
!  
sccp local Vlan5  
sccp ccm 10.10.5.1 identifier 5 version 5.0.1  
! Communications Manager IP address (10.10.5.1)  
sccp  
!  
sccp ccm group 5  
  bind interface Vlan5  
  associate ccm 5 priority 1  
  associate profile 5 register MTP000E83783C50  
! MTP name (MTP000E83783C50) ... must match the Unified CM MTP name.  
!  
dspfarm profile 5 mtp  
  description software MTP  
  codec g711ulaw  
  codec pass-through  
  maximum sessions software 500  
  associate application SCCP
```

For more information on MTPs, see the chapter on [Media Resources](#), page 6-1.

SIP Trunk Transport Protocols

SIP trunks can use either TCP or UDP as a message transport protocol. As a reliable, connection-orientated protocol that maintains the connection state, TCP is preferred because failover to an alternative trunk is nearly instantaneous on the failure of the destination device at the far end of the SIP trunk. UDP is not connection-orientated and therefore relies on the SIP protocol stack to determine if the destination device at the end of the SIP trunk is unavailable. For UDP-based SIP trunks,

Unified CM uses the values for the SIP Invite Retry count and SIP Trying timers to detect and respond to far-end device failures. By default, the failover time is approximately 64 seconds per call, but these timers can be tuned to reduce failover times to acceptable values.

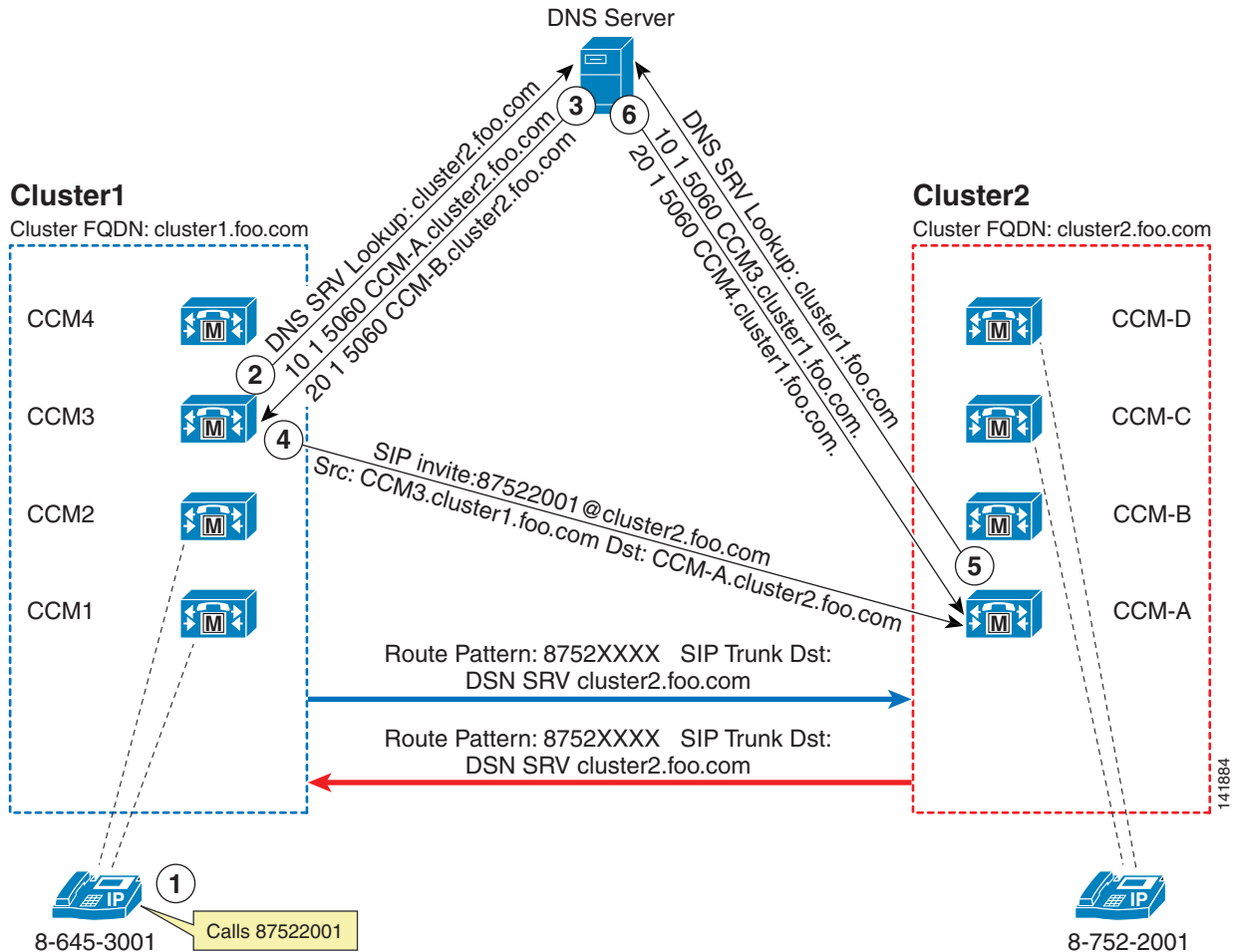
For more information on SIP trunk timer tuning, refer to the configuration example and technical notes at:

http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_configuration_example09186a008082d76a.shtml

SIP Intercluster Trunks

One of the main benefits of using SIP trunks for intercluster trunking is call survivability. However, as compared to H.323 trunks, SIP trunks do not support Secure Real-Time Transport Protocol (SRTP) or QSIG Tunneling using Annex M1. For outbound calls over SIP trunks using destination IP addresses rather than DNS, subscriber load balancing can be achieved in the same way as for H.323 trunk calls. (See the H.323 section on [Load-Balancing Outbound Calls](#), page 5-9.)

Unlike H.323 trunks in Cisco Unified CM Release 3.3 and above, SIP trunks can point to only a single IP address or DNS Server (SRV) record. To provide failover and load balancing for intercluster SIP trunks that are not capable of DNS SRV, configure multiple SIP trunks. In addition, these SIP trunks must be members of route groups and route lists. Also, it is important to note that Unified CM accepts calls only from a SIP device whose IP address matches one of the destination addresses of the configured SIP trunks. In addition, the incoming port number of the SIP messages has to match the port number configured for that SIP trunk. As a consequence, Cisco recommends that you configure as many SIP trunks with destination addresses as needed to match all IP addresses of any far-end SIP devices that can potentially place an inbound call. This method is not desirable for deployments with more than two Unified CM clusters, and it is better to use SIP trunks with DNS SRV if there are more than two clusters. [Figure 5-6](#) shows the call flow for an intercluster SIP trunk call using DNS SRV.

Figure 5-6 Call Flow for Intercluster SIP Trunk Using DNS SRV

Note: The DNS A Lookup has been removed from this call flow

Figure 5-6 illustrates the following steps in the call flow:

1. The IP phone in Cluster1 calls 87522001.
2. The call matches a route pattern of 8752XXXX that is pointing to SIP Trunk with DNS SRV of cluster2.foo.com. CCM3 in Cluster1 is the node handling this call because the SIP trunk is registered to it. CCM3 sends a DNS SRV lookup for cluster2.foo.com
3. The DNS server replies with two records: CCM-A.cluster2.foo.com and CCM-B.cluster2.foo.com. Because CCM-A.cluster2.foo.com has a higher priority, the call is attempted to this Unified CM. Before sending the SIP invite, another DNS lookup is done for CCM-A.cluster2.foo.com.
4. CCM3 sends a SIP invite to 87522001@cluster2.foo.com with destination address set to the IP address of CCM-A.
5. Unified CM interprets this call as a local call because the host portion of the uniform resource identifier (URI) matches the Cluster FQDN enterprise parameter. Cluster2 does not have any SIP trunk configured with a destination of CCM3, so it does a DNS SRV lookup for all domains configured under the SIP trunks with DNS SRV. In this case, the example shows a single trunk with a DNS SRV destination of cluster1.foo.com

6. The DNS server returns two entries, and one of them matches the source IP address of the invite. The cluster accepts the call and extends it to extension 87522001.



CHAPTER 6

Media Resources

Last revised on: October 30, 2008

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.

Use this chapter to determine which of the media resources described below are needed for your deployment. Also determine whether the resources that are required can be provided by software-based features or whether it is necessary to provision digital signal processors (DSPs) to implement the resources. Although the resources are discussed in individual sections, the same basic resources (DSPs and Cisco IP Voice Media Streaming Application) may be shared to implement higher-level functions.

This chapter focuses on the following features:

- [Voice Termination, page 6-2](#)
- [Audio Conferencing, page 6-8](#)
- [Transcoding, page 6-11](#)
- [Media Termination Point \(MTP\), page 6-13](#)
- [Annunciator, page 6-18](#)
- [Cisco RSVP Agent, page 6-19](#)
- [Cisco IP Voice Media Streaming Application, page 6-19](#)

For more details about the following features, refer to their respective sections:

- [Music on Hold, page 7-1](#)
- [Unified CM RSVP-Enabled Locations, page 9-16](#)

For details about the dependencies on hardware and software, see the section on [Hardware and Software Capacities, page 6-20](#).

You can control media resources on Cisco Unified Communications Manager (Unified CM) through the use of media resource groups and media resource group lists. You can create pools of resources to control the specific hardware or software that is used. Cisco recommends that you use pools to group resources based on physical location. For design guidelines based on the various call processing models, see the section on [General Design Guidelines, page 6-22](#).

What's New in This Chapter

[Table 6-1](#) lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 6-1 *New or Changed Information Since the Previous Release of This Document*

New or Revised Topic	Described in:
Dynamic allocation of media termination points (MTPs)	When are Media Termination Points Required for Named Telephony Events? , page 6-14

Voice Termination

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 [Audio Conferencing](#), page 6-8 and [Transcoding](#), page 6-11).

[Table 6-3](#) through [Table 6-7](#) 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 6-3](#) through [Table 6-7](#) 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 6-3](#) through [Table 6-7](#), 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 C5510 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 indicated in [Table 6-2](#), 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 6-3](#) through [Table 6-7](#) that contains the desired hardware. Check the medium and high complexity columns in [Table 6-2](#) 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 6-2](#). 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 6-2](#) 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 6-3](#) 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.

DSP Resources for Voice Termination

Table 6-3 through Table 6-7, organized by DSP chipset, provide information on DSP support by platform, DSP density, and the number of voice terminations (or calls) supported per DSP. Table 6-2 lists the codecs supported by the hardware modules in each complexity mode.

Table 6-2 Codecs Supported by Each Complexity Mode

Medium Complexity	High Complexity	Flex Mode
G.711 (a-law, mu-law)	G.711 (a-law, mu-law)	At 15 MIPS per call:
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	At 30 MIPS per call:
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)	At 40 MIPS per call:
	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 6-3.

Table 6-3 DSP Resources on Cisco IOS Hardware Platforms with C5510 Chipset

Hardware Module or Chassis	DSP Configuration	Maximum Number of Voice Terminations (Calls) per DSP and per Module		
		Medium Complexity (8 calls per DSP)	High Complexity (6 calls per DSP)	Flex Mode ¹ (240 MIPS per DSP)
VG-224	Fixed at 4 DSPs	N/A	24 calls per platform Supported codecs: • G.711 (a-law, mu-law) • G.729a	N/A
NM-HD-1V ²	Fixed at 1 DSP	4 calls per NM	4 calls per NM	240 MIPS per NM
NM-HD-2V	Fixed at 1 DSP	8 calls per NM	6 calls per NM	240 MIPS per NM
NM-HD-2VE	Fixed at 3 DSPs	24 calls per NM	18 calls per NM	720 MIPS per NM

Table 6-3 *DSP Resources on Cisco IOS Hardware Platforms with C5510 Chipset (continued)*

Hardware Module or Chassis	DSP Configuration	Maximum Number of Voice Terminations (Calls) per DSP and per Module		
		Medium Complexity (8 calls per DSP)	High Complexity (6 calls per DSP)	Flex Mode ¹ (240 MIPS per DSP)
NM-HDV2	1 to 4 of:	Calls per PVDM:	Calls per PVDM:	MIPS per PVDM:
NM-HDV2-2T1/E1	PVDM2-8 ³ (½ DSP)	4	3	120
NM-HDV2-1T1/E1	PVDM2-16 (1 DSP)	8	6	240
	PVDM2-32 (2 DSPs)	16	12	480
	PVDM2-48 (3 DSPs)	24	18	720
	PVDM2-64 (4 DSPs)	32	24	960
2801	1 to 2 of:	Calls per PVDM:	Calls per PVDM:	MIPS per PVDM:
2811	PVDM2-8 ³ (½ DSP)	4	3	120
	PVDM2-16 (1 DSP)	8	6	240
	PVDM2-32 (2 DSPs)	16	12	480
	PVDM2-48 (3 DSPs)	24	18	720
	PVDM2-64 (4 DSPs)	32	24	960
2821	1 to 3 of:	Calls per PVDM:	Calls per PVDM:	MIPS per PVDM:
2851	PVDM2-8 ³ (½ DSP)	4	3	120
	PVDM2-16 (1 DSP)	8	6	240
	PVDM2-32 (2 DSPs)	16	12	480
	PVDM2-48 (3 DSPs)	24	18	720
	PVDM2-64 (4 DSPs)	32	24	960
3825	1 to 4 of:	Calls per PVDM:	Calls per PVDM:	MIPS per PVDM:
3845	PVDM2-8 ³ (½ DSP)	4	3	120
	PVDM2-16 (1 DSP)	8	6	240
	PVDM2-32 (2 DSPs)	16	12	480
	PVDM2-48 (3 DSPs)	24	18	720
	PVDM2-64 (4 DSPs)	32	24	960

1. In flex mode, the maximum number of supported calls depends on the number of MIPS used per call (see [Table 6-2](#)).
2. With the NM-HD-1V module, the number of voice terminations (calls) is limited by the number of physical ports on the module.
3. The PVDM2-8 has a single half-capacity C5510.

Hardware that is based on the C5421 chipset may have DSPs configured as either medium complexity or high complexity. [Table 6-4](#) lists the call density per DSP, and [Table 6-2](#) lists the codecs supported in each complexity mode.

Table 6-4 *DSP Resources on Cisco IOS Hardware Platforms with C5421 Chipset*

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

Hardware that is based on the C549 chipset may have DSPs configured as either medium complexity or high complexity. [Table 6-5](#) lists the call density per DSP, and [Table 6-2](#) lists the codecs supported in each complexity mode.

Table 6-5 *DSP Resources on Cisco IOS Hardware Platforms with C549 Chipset*

Hardware Module	DSP Configuration	Maximum Number of Calls per DSP and per Module	
		Medium Complexity (4 calls per DSP)	High Complexity (2 calls per DSP)
NM-HDV NM-HDV-FARM	1 to 5 of PVDM-12 (3 DSPs per PVDM-12)	12, 24, 36, 48, or 60 calls per NM	6, 12, 18, 24, or 30 calls per NM
1751 ¹ 1760	1 to 2 of: PVDM-256K-4 (1 DSP) PVDM-256K-8 (2 DSPs) PVDM-256K-12 (3 DSPs) PVDM-256K-16HD (4 DSPs) PVDM-256K-20HD (5 DSPs)	Calls per NM: 4 or 8 8 or 16 12 or 24 16 or 32 20	Calls per NM: 2 or 4 4 or 8 6 or 12 8 or 16 10
PA-VXA-1TE1-24+ PA-VXA-1TE1-30+ PA-VXB-2TE1+ PA-VXC-2TE1+	Fixed at: 7 DSPs 8 DSPs 12 DSPs 30 DSPs	Calls per PA: 28 32 48 120	Calls per PA: 14 16 24 60
PA-MCX-2TE1 PA-MCX-4TE1 PA-MCX-8TE1	Fixed (No on-board DSPs)	Depends on PA-VX(<i>x</i>) ²	Depends on PA-VX(<i>x</i>) ²

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 6-6 lists the call density per DSP.

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

Hardware Module ¹	DSP Configuration	Maximum Number of Calls per DSP and per Module
NM-1V	Fixed at 2 DSPs	1 call per DSP 2 calls per NM
NM-2V	Fixed at 4 DSPs	1 call per DSP 4 calls per NM

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

Table 6-7 lists non-IOS hardware for DSP resources. All non-IOS hardware platforms have a fixed configuration of DSPs, as indicated in Table 6-7.

Table 6-7 DSP Resources on Non-IOS Hardware Platforms

Hardware Module or Platform	DSP Configuration	Maximum Number of Calls per DSP and per Module	Codecs Supported
WS-6608-T1 WS-6608-E1	Fixed at 64 of C549 (8 DSPs per port)	2 calls per DSP 256 calls per module ¹	G.711 a-law, mu-law G.729a
WS-6624-FXS	Fixed at 12 of C549	2 calls per DSP 24 calls per module	G.711 a-law, mu-law G.729a
VG-248	Fixed at 12 of C5409	4 calls per DSP 48 calls per platform	G.711 a-law, mu-law G.729a
WS-SVC-CMM-ACT	Fixed at 4 of Broadcom 1500	32 calls per DSP 128 calls per module	G.711 (10-30 ms) G.729 (10-60 ms) G.723 (30-60 ms)

Table 6-7 *DSP Resources on Non-IOS Hardware Platforms (continued)*

Hardware Module or Platform	DSP Configuration	Maximum Number of Calls per DSP and per Module	Codecs Supported
WS-SVC-CMM-6T1	Fixed at 12 of C5441	15 calls per DSP 144 calls per module	G.711 (10, 20, 30 ms) G.729 (10, 20, 30, 40, 50, 60 ms)
WS-SVC-CMM-6E1	Fixed at 12 of C5441	15 calls per DSP 180 calls per module	G.711 (10, 20, 30 ms) G.729 (10, 20, 30, 40, 50, 60 ms)
WS-SVC-CMM-24FXS	Fixed at 3 of C5441	15 calls per DSP 24 calls per module	G.711 a-law, mu-law G.729 G.729a
ATA-188 ²	Fixed at 1 of Komodo 3880	2 calls 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. If none of the DSPs are configured for T1 or E1, then a maximum of 256 DSP resources are available.
2. The ATA module does not define complexity. It supports G.711, G.729, and G.723 only.

Audio 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).

Audio Conferencing Resources

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 enables 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 Unified Communications Manager (Unified CM) use Skinny Client Control Protocol (SCCP) to communicate with Unified CM.

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

The number of individual conferences that may be supported by the resource varies, and the maximum number of participants in a single conference varies, depending on the resource.

The following types of conference bridge resources may be used on a Unified CM system:

- [Software Audio Conference Bridge \(Cisco IP Voice Media Streaming Application\)](#), page 6-9
- [Hardware Audio Conference Bridge \(Cisco NM-HDV2, NM-HD-1V/2V/2VE, 2800 Series, and 3800 Series Routers\)](#), page 6-9
- [Hardware Audio Conference Bridge \(Cisco WS-SVC-CMM-ACT\)](#), page 6-10
- [Hardware Audio Conference Bridge \(Cisco NM-HDV and 1700 Series Routers\)](#), page 6-10
- [Hardware Audio Conference Bridge \(Cisco Catalyst WS-X6608-T1 and WS-X6608-E1\)](#), page 6-10
- [Built-in Conference](#), page 6-11

Software Audio Conference Bridge (Cisco IP Voice Media Streaming Application)

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 conferences that can be supported on a given configuration depends on the server where the conference bridge software is running and on what other functionality has been enabled for the application. The Cisco IP Voice Media Streaming Application is a resource that can also be used for several functions, and the design must consider all functions together (see [Cisco IP Voice Media Streaming Application](#), page 6-19).

Hardware Audio Conference Bridge (Cisco NM-HDV2, NM-HD-1V/2V/2VE, 2800 Series, and 3800 Series Routers)

DSPs that are configured through Cisco IOS as conference resources will load firmware into the DSPs that is specific to conferencing functionality only, and these DSPs cannot be used for any other media feature.

The following guidelines and considerations apply to these DSP resources:

- Based on the C5510 DSP chipset, the NM-HDV2 and the router chassis use the PVDM2 modules for providing DSPs.
- DSPs on PVDM2 hardware are configured individually as either voice termination, conferencing, media termination, or transcoding, so that DSPs on a single PVDM may be used as different resource types. Allocate DSPs to voice termination first, then to other functionality as needed.
- The NM-HDV2 has 4 slots that will accept PVDM2 modules in any combination. The other network modules have fixed numbers of DSPs.
- A conference based on these DSPs allows a maximum of 8 participants. When a conference begins, all 8 positions are reserved at that time.
- The PVDM2-8 is listed as having $\frac{1}{2}$ a DSP because it has a DSP that has half the processing capacity of the PVDM2-16. For example, if the DSP on a PVDM2-8 is configured for G.711, it can provide $(0.5 * 8)$ bridges/DSP = 4 conference bridges.
- Use [Table 6-2](#) and [Table 6-3](#) to determine how many DSPs may be provisioned with specific hardware.
- A DSP farm configuration in Cisco IOS specifies which codecs may be accepted for the farm. A DSP farm that is configured for conferencing and G.711 provides 8 conferences. When configured to accept both G.711 and G.729 calls, a single DSP provides 2 conferences because it is also reserving its resources for performing transcoding of streams.
- The I/O of an NM-HDV2 is limited to 400 streams, so ensure that the number of conference resources allocated does not cause this limit to be exceeded. If G.711 conferences are configured, then no more than 6 DSPs (total of 48 conferences with 8 participants each) should be allocated per

NM because $(48 * 8)$ participants = 384 streams. If you configure all conferencing for both G.711 and G.729 codecs, then each DSP provides only 2 conferences of 8 participants each. In this case, it is possible to populate the NM fully and configure it with 16 DSPs so there would be 256 streams.

- Conferences cannot natively accept calls utilizing the GSM codec. A transcoder must be provided separately for these calls to participate in a conference.
- Any PVDM2-based hardware, such as the NM-HDV2, may be used simultaneously in a single chassis for voice termination but may not be used simultaneously for other media resource functionality. The DSPs based on PVDM-256K and PVDM2 have different DSP farm configurations, and only one may be configured in a router at a time.

Hardware Audio Conference Bridge (Cisco WS-SVC-CMM-ACT)

The following guidelines and considerations apply to this DSP resource:

- DSPs on this hardware are configured individually as either voice termination, conferencing, media termination, or transcoding, so that DSPs on a single module may be used as different resource types. Allocate DSPs to voice termination first.
- This Cisco Catalyst-based hardware provides DSP resources that may provide conference bridges of up to 32 participants per bridge.
- Each module contains 4 DSPs that are individually configurable. Each DSP can support 32 conference bridges.
- The G.711 and G.729 codecs are supported on these conference bridges without extra transcoder resources. However, transcoder resources would be necessary if other codecs are used.

Hardware Audio Conference Bridge (Cisco NM-HDV and 1700 Series Routers)

The following guidelines and considerations apply to these DSP resources:

- This hardware utilizes the PVDM-256K type modules that are based on the C549 DSP chipset.
- Conferences using this hardware provide bridges that allow up to 6 participants in a single bridge.
- The resources are configured on a per-DSP basis as conference bridges.
- The NM-HDV may have up to 5 PVDM-256K modules, while the Cisco 1700 Series Routers may have 1 or 2 PVDM-256K modules.
- Each DSP provides a single conference bridge that can accept G.711 or G.729 calls.
- The Cisco 1751 is limited to 5 conference calls per chassis, and the Cisco 1760 can support 20 conference calls per chassis.
- Any PVDM2-based hardware, such as the NM-HDV2, may be used simultaneously in a single chassis for voice termination but may not be used simultaneously for other media resource functionality. The DSPs based on PVDM-256K and PVDM2 have different DSP farm configurations, and only one may be configured in a router at a time.

Hardware Audio Conference Bridge (Cisco Catalyst WS-X6608-T1 and WS-X6608-E1)

The following guidelines and considerations apply to these DSP resources:

- This hardware has 8 DSPs that are physically associated to each port, and there are 8 ports per card. Configuration of the DSPs is at the port level, so that all DSPs associated to a port perform the same function.
- Conference bridges may have up to 32 participants, and each port supports 32 conference bridges.
- For conferences with G.711 or G.723 there may be 32 conferences per port. If G.729 calls are used, there may be 24 conferences per port.

Built-in Conference

Some phone models have a built-in conference resource that allows a three-way conference. This bridge is invoked only by the Barge feature and is not used as a general conferencing resource. For details on which phones have this bridge, see [IP Telephony Endpoints, page 20-1](#). This bridge accepts only G.711 calls.

Transcoding

A transcoder is a device that converts an input stream from one codec into an output stream that uses a different codec. It may also connect two streams that utilize the same codec but with a different sampling rate. In a Unified CM system, the typical use of a transcoder is to convert between a G.711 voice stream and the low bit-rate compressed voice stream G.729a. The following cases determine when transcoder resources are needed:

- Single codec for the entire system

When a single codec is configured for all calls in the system, then no transcoder resources are required. The G.711 codec is supported by all vendors. A single-site deployment usually has no need for conserving bandwidth, and a single codec can be used. In this scenario, G.711 is the most common choice.

- Multiple codecs in use in the system, and all endpoints are capable of all codec types

The most common reason for multiple codecs is to use G.711 for LAN calls to maximize the call quality and to use a low-bandwidth codec to maximize bandwidth efficiency for calls that traverse a WAN with limited bandwidth. Cisco recommends using G.729a as the low-bandwidth codec because it is supported on all Cisco Unified IP Phone models as well as most other Cisco Unified Communications devices, therefore it can eliminate the need for transcoding. Although Unified CM allows configuration of other low-bandwidth codecs between regions, the current phone models do not support those codecs and would require transcoders. They would require one transcoder for a call to a gateway and two transcoders if the call is to another IP phone. The use of transcoders is avoided if all devices support and are configured for both G.711 and G.729 because the devices will use the appropriate codec on a call-by-call basis.

- Multiple codecs in use in the system, and some endpoints support or are configured for G.711 only

This condition exists when G.729a is used in the system but there are devices that do not support this codec, or a device with G.729a support may be configured to not use it. In this case, a transcoder is also required. Devices from some third-party vendors may not support G.729. Another example where G.729 is supported but might not be configured is with Cisco Unity. Cisco Unity has support for accepting calls with G.729a, but the codec is implemented in software and is CPU-intensive. Because as few as 10 simultaneous calls can cause significant CPU utilization, many deployments choose to disable G.729 on Cisco Unity and off-load the transcoding function to a dedicated transcoding resource external to the Unity server. If your system includes Cisco Unity, determine whether Unity will accept G.729a calls or whether it will be configured for G.711 only.

**Note**

Cisco Unified MeetingPlace Express currently supports G.711 only. In an environment where G.729 is configured for a call into Cisco Unified MeetingPlace Express, transcoder resources are required.

To finalize the design, it is necessary to know how many transcoders are needed and where they will be placed. If multiple codecs are needed, it is necessary to know how many endpoints do not support all codecs, where those endpoints are located, what other groups will be accessing those resources, how many maximum simultaneous calls these device must support, and where those resources are located in the network.

Transcoding Resources

DSP resources are required to perform transcoding. Those DSP resources can be located in the voice modules and the hardware platforms for transcoding that are listed in the following sections.

Hardware Transcoder (Cisco NM-HDV2, NM-HD-1V/2V/2VE, 2800, and 3800 Series Routers)

The following guidelines and considerations apply to these DSP resources:

- Transcoding is available between G.711 mu-law or a-law and G.729a or G.729ab. Each DSP can support 8 sessions.
- Cisco Unified IP Phones use only the G.729a variants of the G.729 codecs. The default for a new DSP farm profile is G.729a/G.729ab/G.711u/G.711a. Because a single DSP can provide only one function at a time, the maximum sessions configured on the profile should be specified in multiples of 8 to prevent wasted resources.
- Transcoding is also available between G.711mu-law/G.711a-law and G.729/G729b, but is not usually used in a Unified CM system. Each DSP can support 6 sessions.
- Refer to the section on [DSP Resources for Voice Termination, page 6-4](#), to determine numbers of DSPs available on a particular platform or network module.

Hardware Transcoder (Cisco WS-SVC-CMM-ACT)

The following guidelines and considerations apply to this DSP resource:

- Transcoding is available between G.711 mu-law or a-law and G.729, G.729b, or G.723.
- There are 4 DSPs per ACT that may be allocated individually to DSP pools.
- The CCM-ACT can have 16 transcoded calls per DSP or 64 per ACT. The ACT reports resources as streams rather than calls, and a single transcoded call consists of two streams.

Hardware Transcoder (Cisco NM-HDV and 1700 Series Routers)

The following guidelines and considerations apply to these DSP resources:

- This hardware utilizes the PVDM-256K type modules that are based on the C549 DSP chipset.
- The NM-HDV may have up to 4 PVDM-256K modules, and the Cisco 1700 series routers may have 1 or 2 PVDM-256K modules.
- NM-HDV and NM-HDV2 modules may be used simultaneously in a single chassis for voice termination but may not be used simultaneously for other media resource functionality. Only one type of DSP farm configuration may be active at one time (either the NM-HDV or the HM-HDV2) for conferencing, MTP, or transcoding.
- Transcoding is supported from G.711 mu-law or a-law to any of G.729, G.729a, G.729b, G.729ab, or GSM codecs.
- Each DSP can provide 2 transcoding sessions.
- The Cisco 1751 has a chassis limit of 16 sessions, and the Cisco 1760 has a chassis limit of 20 sessions.

Hardware Transcoder (Cisco WS-X6608)

The following guidelines and considerations apply to this DSP resource:

- DSPs are allocated to functions at a port level. Each port can provide 24 transcoding sessions.
- There are 8 ports per blade.
- Transcoding is available between G.711 mu-law or a-law and G.729a, G.729ab, G.729, or G.729b.

A transcoder is also capable of performing the same functionality as a media termination point (MTP). In cases where transcoder functionality and MTP functionality are both needed, a transcoder is allocated by the system. If MTP functionality is required, Unified CM will allocate either a transcoder or an MTP from the resource pool, and the choice of resource will be determined by the media resource groups, as described in the section on [Media Resource Groups and Lists, page 6-23](#).

Media Termination Point (MTP)

A media termination point (MTP) is an entity that accepts two full-duplex 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. MTPs have many possible uses, as described in the following sections.

Re-Packetization of a Stream

An 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 sample sizes).

H.323 Supplementary Services

MTPs can be used to extend supplementary services to H.323 endpoints that do not support the H.323v2 OpenLogicalChannel and CloseLogicalChannel request features of the Empty Capabilities Set (ECS). This requirement occurs infrequently. Cisco H.323 endpoints support ECS, and most third party endpoints have support as well. 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.

When an MTP is required on an H.323 call and none is available, the call will proceed but will not be able to invoke supplementary services.

H.323 Outbound Fast Start

H.323 defines a feature called Fast Start, which reduces the number of packets exchanged during a call setup, thereby reducing the amount of time for media to be established. It is useful when two devices that are utilizing H.323 have high network latency between them because the time to establish media is dependant on that latency. Unified CM distinguishes between inbound and outbound fast start based on the direction of the call setup, and the distinction is important because the MTP requirements are not

equal. For inbound fast start, no MTP is required. Outbound calls on an H.323 trunk do require an MTP when fast start is enabled. Frequently it is only inbound calls that are problematic, and it is possible to use inbound fast start to solve the issue without also enabling outbound fast start.

Named Telephony Events (RFC 2833)

DTMF tones can be used during a call to signal to a far-end device for purposes of navigating a menu system, entering data, or other manipulation. They are processed differently than DTMF tones sent during a call setup as part of the call control.

Named Telephony Events (NTEs) defined by RFC 2833 are a method of sending DTMF from one endpoint to another after the call media has been established. The tones are sent as packet data using the already established RTP stream and are distinguished from the audio by the RTP packet type field. For example, the audio of a call can be sent on a session with an RTP packet type that identifies it as G.711 data, and the DTMF packets are sent with an RTP packet type that identifies them as NTEs. The consumer of the stream utilizes the G.711 packets and the NTE packets separately.

When are Media Termination Points Required for Named Telephony Events?

For Cisco Unified CM 4.x, only SIP trunks are supported, and they require that an MTP be allocated for all calls using the SIP trunk. The SIP trunk has a configuration parameter, "MTP required," which is selected by default and cannot be changed.

Cisco Unified CM 5.0 adds support for SIP on line devices and also removes the requirement that MTPs be allocated for all calls using SIP trunks. In Unified CM 5.0, an MTP is required when two endpoints do not have a method in common for sending DTMF between them or when the system configuration specifies that one must be allocated. The following description expands on the details and applies only to Unified CM 5.0.

Verify the types of endpoints planned for the system based on the following rules:

1. Two non-SIP endpoints do not require MTPs.

All Cisco Unified Communications endpoints other than SIP send DTMF to Unified CM via various signaling paths, and Unified CM forwards the DTMF between dissimilar endpoints. For example, an IP phone may use SCCP messages to Unified CM to send DTMF, which then gets sent to an H.323 gateway via H.245 signaling events. The two endpoints have a common method by sending DTMF to Unified CM. If no SIP endpoints are used, stop here.

2. Two Cisco SIP endpoints do not require MTPs.

All Cisco SIP endpoints support NTEs, therefore MTPs can be avoided entirely. DTMF is sent directly between the endpoints using NTE. If all endpoints are Cisco SIP devices, then no MTPs are required for DTMF conversion.

3. A combination of a SIP endpoint and a non-SIP endpoint might require MTPs.

To determine the support for NTE in your devices, see [Table 6-8](#). RFC 2833 is not limited to SIP and can be supported in devices with other call control protocols. For example, Cisco Unified IP Phones that run either an SCCP or SIP stack can support NTEs in both modes. Some devices support DTMF via multiple methods (for example, a Cisco Unified IP Phone 7960 with an SCCP stack can send NTEs to the other device or SCCP to Unified CM). Other devices such as the Cisco Wireless IP Phone 7920 can send only SCCP, and still others can send only NTE (such as the Cisco Unified IP Phone 7960 with a SIP stack). Unified CM has the ability to allocate MTPs dynamically on a call-by-call basis, based on the capabilities of the pair of endpoints. Use [Table 6-8](#) to determine whether MTPs should be provisioned.

4. Dynamic allocation of MTPs is only for outbound calls.

A SIP trunk that does not require MTPs will use dynamic allocation as described in rule 3. The SIP trunk is capable of accepting calls with or without Session Description Protocol (SDP) in the Invite, therefore it will not allocate an MTP for an inbound call. Only outbound calls may have an MTP allocated as described above. If a particular SIP trunk is used for inbound calls only, then it not necessary to allocate any MTP resources on the system for receiving calls on that trunk.

5. MTPs may be forced by the configuration.

In Cisco Unified CM 5.0, the SIP trunk parameter **MTP Required** is not selected by default, and the field is unlocked. When a SIP trunk is defined for a Cisco device (such as a SIP gateway), the device should be configured for NTE support. The default setting for MTP Required should normally be used, and MTPs are allocated only when needed.

There are other reasons for forcing an MTP. The establishment of a SIP dialog behaves differently when an MTP is allocated before call setup starts. SIP uses Session Description Protocol (SDP) for establishing the parameters of a session, and SDP is embedded into SIP messages. When an MTP is forced, the SDP is sent with the Invite message that initiates the call. When an MTP is not forced, it is allocated (if needed) after the Invite message. The Invite message does not contain SDP, and it is sent at a later point in the call setup. If the far-end device supports only Invite messages with embedded SDP, then the MTP Required parameter must be checked (enabled).

When this configuration parameter forces an MTP, the codec used on the call leg from the MTP to the far-end SIP device must be G.711 mu-law or a-law. There is an additional parameter on the trunk configuration that becomes unlocked when MTP Required is selected. This parameter allows a choice of which variant of G.711 to use. When an MTP is not forced by this configuration parameter, then the normal mechanisms for determining the codec apply. In these cases, an MTP may be inserted dynamically for DTMF relay or other purposes and will not impose any restriction on the codec negotiated.

Table 6-8 Endpoint Support for DTMF Methods

Endpoint Protocol Stack: Endpoints	DTMF Method		
	SCCP	NTE	KPML ¹
SCCP Stack: 12SP+, 30 VOIP, 7910, 7920, 7935, 7936 VG248, DPA-7610, DPA-7630, CTI Port, First Party Control	Yes	No	No
SCCP Stack: 7902, 7905, 7912, 7940, 7941, 7960, 7961, 7970 Future new phone models	Yes	Yes	No

Table 6-8 **Endpoint Support for DTMF Methods (continued)**

Endpoint Protocol Stack: Endpoints	DTMF Method		
	SCCP	NTE	KPML ¹
SIP Stack: 7905, 7911, 7912, 7940, 7941, 7960, 7961, 7970 Future new phone models	No	Yes	Yes for: 7911, 7941, 7961, 7970 No for: 7905, 7912, 7940, 7960

1. Key Press Markup Language (KPML)

**Note**

IP Phones play DTMF to the end user when DTMF is received via SCCP, but they do not play tones received by NTE. However, there is no requirement to send DTMF to another end user. It is necessary only to consider the endpoints that originate calls combined with endpoints that might need DTMF, such as PSTN gateways, application servers, and so forth.

Example 6-1 **Call Flow that Requires an MTP for NTE Conversion**

Assume the example system has CTI route points with first-party control (the CTI port terminates the media), which integrate to a system that uses DTMF to navigate an IVR menu. If all phones in the system are running SCCP, then no MTP is required. In this case Unified CM controls the CTI port and receives DTMF from the IP phones via SCCP. Unified CM provides DTMF conversion.

However, if there are phones running a SIP stack, an MTP is required. NTEs are part of the media stream, therefore Unified CM does not receive them. An MTP is invoked into the media stream and has one call leg that uses SCCP, and the second call leg uses NTEs. The MTP is under SCCP control of Unified CM and performs the NTE-to-SCCP conversion under the control of Unified CM.

Configuration of DTMF on SIP and H.323 Gateways

Configure either **sip-notify** or **rtp-nte** as a method under a SIP dial peer. The best configuration of a SIP dial peer will depend on the mix of endpoints that exist in the system (see [Table 6-8](#)). When Unsolicited Notify is used on the gateway, that dial peer may be used by endpoints that support SCCP without the need for MTP resources. Endpoints that support only NTE must invoke an MTP to use the gateway.

Conversely, if NTE is configured on the SIP dial peer, then any endpoint that can use NTE can send DTMF directly to the gateway without need for MTPs. An endpoint that supports only SCCP will need to invoke an MTP.

SIP gateways have the ability to send DTMF using NTEs, Unsolicited Notify, or audio tones in the media stream. Unsolicited Notify is a Cisco proprietary method that sends a SIP Notify message with an event that contains the DTMF tone. This method is also supported on Unified CM.

The following example lists a SIP gateway configuration for Named Telephony Events:

```
dial-peer voice 10 voip
dtmf-relay rtp-nte
```

H.323 gateways have support for H.245 Alphanumeric, H.245 Signal, NTE, and audio in the media stream. The NTE option must not be used because it is not supported on Unified CM for H.323 at this time. The preferred option is H.245 signal. MTPs are required for establishing calls to an H.323 gateway

if the other endpoint does not have signaling capability in common with Unified CM. For example, a Cisco Unified IP Phone 7960 running the SIP stack supports only NTEs, so an MTP is needed with an H.323 gateway.

The following configuration is recommended for DTMF methods on an H.323 gateway:

```
dial-peer voice 10 voip
dtmf-relay h.245-signal
```

CTI Route Points

CTI Route Points that have first-party control of a phone call will participate in the media stream of the call and require an MTP to be inserted. When the CTI has third-party control of a call such that the media passes through a device that is controlled by the CTI, then the requirement for an MTP is dependant on the capabilities of the controlled device.

MTP Resources

The following types of devices are available for use as an MTP:

Software MTP (Cisco IP Voice Media Streaming Application)

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 Unified CM node and informs Unified CM of how many MTP resources it supports. A software MTP device supports only G.711 streams. The IP Voice Media Streaming Application is a resource that may also be used for several functions, and the design guidance must consider all functions together (see [Cisco IP Voice Media Streaming Application, page 6-19](#)).

Software MTP (Based on Cisco IOS)

- The capability to provide a software-based MTP on the router is available beginning with Cisco IOS Release 12.3(11)T for the Cisco 3800 Series Routers and Release 12.3(8)T4 for other router models.
- This MTP allows configuration of any of the following codecs, but only one may be configured at a given time: G.711 mu-law and a-law, G.729a, G.729, G.729ab, G.729b, GSM, and pass-through. Some of these are not pertinent to a Unified CM implementation.
- The router configuration permits up to 500 individual streams, which support 250 transcoded sessions. This number of G.711 streams generates 5 Mbytes of traffic.

Hardware MTP (Cisco NM-HDV2, NM-HD-1V/2V/2VE, 2800 and 3800 Series Routers)

- This hardware uses the PVDM-2 modules for providing DSPs.
- Each DSP can provide 16 G.711 mu-law or a-law MTP sessions or 6 G.729, G.729b, or GSM MTP sessions.

Hardware MTP (Cisco WS-SVC-CMM-ACT)

- This module has 4 DSPs that may be configured individually.
- Each DSP can support 128 G.729, G.729b, or GSM MTP sessions or 256 G.711 mu-law or a-law MTP sessions.

Hardware MTP (Catalyst WS-X6608-T1 and WS-X6608-E1)

- Codec support is G.711 mu-law or a-law, G.729, G.729b, or GSM.
- Configuration is done at the port level. Eight ports are available per module.
- Each port configured as an MTP resources provides 24 sessions.

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 Unified CM cluster if it meets the performance requirements (see [Annunciator Performance, page 6-19](#)); 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 Unified CM at a time, as defined by its device pool. It will automatically fail over to a secondary Unified CM 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 (co-resident) with the Unified CM 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 RSVP Agent

In order to provide topology-aware call admission control, Unified CM invokes one or two RVSP Agents during the call setup to perform an RSVP reservation across the IP WAN. These agents are MTP or transcoder resources that have been configured to provide RSVP functionality. RSVP resources are treated the same way as regular MTPs or transcoders from the perspective of allocation of an MTP or transcoder resource by Unified CM.

The Cisco RSVP Agent feature was first introduced in Cisco IOS Release 12.4(6)T. For details on RSVP and Cisco RSVP Agents, refer to the chapter on [Call Admission Control, page 9-1](#).

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 limited amounts of 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).

The annunciator is the only media resource that is available only through the IP Voice Media Streaming Application. Conferencing, MTPs, and music on hold (MoH) can all reside on servers other than the Unified CM servers. Cisco highly recommends that you disable the MTP and conferencing resources on Unified CM and that you provide external dedicated resources for this functionality.

Cisco also strongly recommends that you run the IP Voice Media Streaming Application on a server other than the publisher or any Unified CM server that provides call processing. The increase in CPU load for media resources might adversely impact call processing performance, and security issues can arise because User Datagram Protocol (UDP) traffic must be received on the Unified CM server.

Hardware and Software Capacities

This section provides data on capacities of the network modules and chassis that carry DSPs, the capacities of the chassis to carry network modules, and software dependencies of the hardware.

PVDMs

[Table 6-9](#) and [Table 6-10](#) show the number of DSPs that can reside on the two models of PVDMs or on fixed-configuration network modules. The PVDM2-xx modules are newer than the PVDM-256K-xx modules, and the two types are not interchangeable.

Table 6-9 *Number of DSPs per PVDM-256K Module*

Module	Number of DSPs
PVDM-256K-4	1 DSP
PVDM-256K-8	2 DSPs
PVDM-256K-12	3 DSPs
PVDM-256K-16HD	4 DSPs
PVDM-256K-20HD	5 DSPs

Table 6-10 *Number of DSPs per PVDM2 Module or Fixed-Configuration Hardware*

Hardware Module or Chassis	Number of DSPs
PVDM2-8	½ DSP
PVDM2-16	1 DSP
PVDM2-32	2 DSPs
PVDM2-48	3 DSPs
PVDM2-64	4 DSPs
NM-HD-1V NM-HD-2V	1 DSP
NM-HD-2VE	3 DSPs

Table 6-11 specifies minimum versions of Cisco IOS software needed to support media resource functionality on these hardware platforms and network modules.

Table 6-11 PVDM2 Slots Available and Cisco IOS Versions Required for Media Support

Chassis or Network Module	Number of PVDM2 Slots	Minimum Cisco IOS Release for Media
2801	2	12.3(11)T
2811	2	12.3(8)T4
2821 or 2851	3	12.3(8)T4
3825 or 3845	4	12.3(11)T
NM-HDV2	4	

Cisco 2800 and 3800 Series Platforms

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 PVDM2 modules 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 6-12, up to the maximum numbers indicated.

All three families of modules in Table 6-11 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). In addition, 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. Not all chassis can be completely populated by these modules. Table 6-11 shows the maximum number of modules that each type of hardware platform can support.

Table 6-12 Number of Module Slots per Platform Type

Cisco IOS Platform	Number of Slots
2691, 2811, 2821, 2851	1
3620 ¹ , 3725, 3825	2
3640	3
3745, 3845	4
3660	6

1. Although it has two NM slots, the Cisco 3620 Router supports only a single NM-HDV module.

Table 6-13 Modules Supported by Platform Type

Cisco IOS Platform	Modules Supported by Platform		
	NM-HDV2	NM-HD-1V NM-HD-2V NM-HD-2VE	NM-HDV NM-HDV-FARM
VG2001 26002 36203 3640	No	No	Yes
3660	No	Yes	Yes
2600XM, 2691, 3725, 3745 2811, 2821, 2851 3825, 3845	Yes	Yes	Yes

**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.

Calculating DSP Requirements for the NM-HDV

There are specific situations where the NM-HDV may not be fully populated. The sampling rate is not normally changed from the system defaults. If you do not require the sampling rate to be changed, then this issue may be ignored.

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 PVDs, giving 60 usable DSP resources.

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:

$$100 \text{ pps (number of voice terminations)} + 600 \text{ pps (number of conferences)} + 200 \text{ pps (number of transcoding sessions)} < 6600 \text{ pps}$$

General Design Guidelines

The Unified CM constructs of media resource groups (MRGs) and media resource group lists (MRGLs) are used to control how the resources described in this chapter are organized and accessed. This section discusses considerations for how to utilize these constructs effectively. It also discusses specific considerations for the various Unified CM deployment models.

Media Resource Groups and Lists

Media resource groups and lists provide a method to control how resources are allocated that could include rights to resources, location of resources, or resource type for specific applications. This section assumes you have an understanding of media resource groups, and it highlights the following design considerations:

- The system defines a default media resource group that is not visible in the user interface and that all resources are members of when they are created. Consumers of media resources use resources first from any media resource group (MRG) or media resource group list (MRGL) that their configuration specifies. If the required resource is not available, the default MRG is searched for the resource. For simple deployments, the default MRG alone may be used.
- When using MRGs to control access to resources, it is necessary to move the resources out of the default MRG by explicitly configuring them in some other group. If the desired effect is for resources to be available only as a last resort for all calls, then the resources may remain in the default group. Also, if no control over resources is necessary, they may remain in the default group.
- An MRG may contain multiple types of resources, and the appropriate resource will be allocated from the group based on the feature needed. MTPs and transcoders are a special case because a transcoder may also be used as an MTP.
- One use of MRGs is to group resources of similar types. Conference bridge resources vary in the number of participants they support, and MRGs could be used to group the conference resources by conference bridge size.
- Use media resource groups (MRGs) and media resource group lists (MRGLs) to provide sharing of resources across multiple Unified CMs. If you do not use MRGs and MRGLs, the resources are available to a single Unified CM only.
- You can also use MRGs and MRGLs to separate resources based on geographical location, thereby conserving WAN bandwidth whenever possible.
- MRGLs will use MRGs in the order that they are listed in the configuration. If one MRG does not have the needed resource, the next MRG is searched. If all MRGs are searched and no resource is found, the search terminates.
- The algorithm for allocating like resources from an MRG attempts to spread the load across the like resources. When a resource has been used, a pointer for that MRG is incremented to the next device. A device may be present in more than one MRG, which will affect the pointers of all groups in which the device is a member. When MTPs are required and transcoders exist in the same group, the MTPs are always allocated until all MTPs are used, and then transcoders are used as MTPs. When resources of differing capacities exist in the same group, the load sharing attempts to allocate resources based on the capacity. The system will spread the load across resources, but because of the above factors, it frequently will not be round-robin in behavior.
- Unified CM Administration displays the devices in an MRG in alphabetical order, but they are allocated based on their order in the configuration database. You cannot change this order. If you want media resources to be allocated in a specific order, then create a separate MRG for each individual resource and use MRGLs to specify the order of allocation.
- Ensure that the media resources themselves have configurations that prevent further invocation of other media resources. For example, if an MTP is inserted into a call and the codec configured on that MTP does not match the one needed by Unified CM for the call, then a transcoder may also be invoked. A frequent mistake is to configure an MTP for G.729 or G.729b when Unified CM needs G.729a.

Deployment Models

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 6-24](#), consist of one or more call processing agents within a single site, with no voice traffic carried over the IP WAN.
- [Multisite WAN Deployments with Centralized Call Processing, page 6-24](#), consist of a single call processing agent that provides service for many sites connected through an IP WAN.
- [Multisite 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

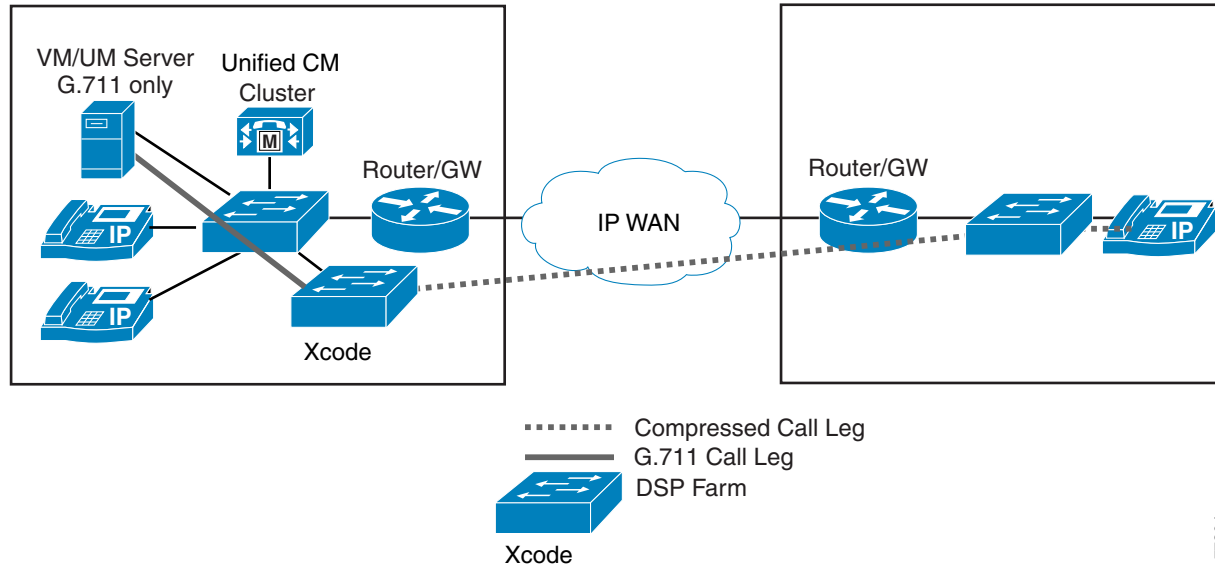
In a single-site deployment, there is no need for transcoding because there are no low-speed links to justify the use of a low bit-rate (LBR) codec. Some MTP resources might be required in the presence of a significant number of devices that are not compliant with H.323v2, such as older versions of Microsoft NetMeeting or certain video devices. MTP resources may be required for DTMF conversion if SIP endpoints are present (see [Named Telephony Events \(RFC 2833\), page 6-14.](#))

Multisite WAN Deployments with Centralized Call Processing

In a centralized call processing deployment, the Unified CM 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 Unified CMs 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 6-1.](#))

Voice compression between IP phones is easily configured through the use of *regions* and *locations* in Unified CM. 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.

Figure 6-1 Transcoding for the WAN with Centralized Call Processing

77904

Unified CM uses media resource groups (MRGs) to enable sharing of MTP and transcoding resources among the Unified CM servers within a cluster. In addition, when using an LBR codec (for example, G.729a) 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.

In [Figure 6-1](#), Unified CM knows that a transcoder is required and allocates one based on the MRGL and/or MRG of the device that is using the higher-bandwidth codec. In this case it is the VM/UM server that determines which transcoder device is used. This behavior of Unified CM is based on the assumption that the transcoder resources are actually located close to the higher-bandwidth device. If this system was designed so that the transcoder for the VM/UM server was located at the remote site, then G.711 would be sent across the WAN, which would defeat the intended design. As a result, if there are multiple sites with G.711-only devices, then each of these sites would need transcoder resources when an LBR is run on the WAN.

The placement of other resources is also important. For example, if a conference occurs with three phones at a remote site and the conference resource is located in the central (call processing) site, then three media streams are carried over the WAN. If the conference resource were local, then the calls would not traverse the WAN. It is necessary to consider this factor when designing the bandwidth and call admission control for your WAN.

Multisite WAN Deployments with Distributed Call Processing

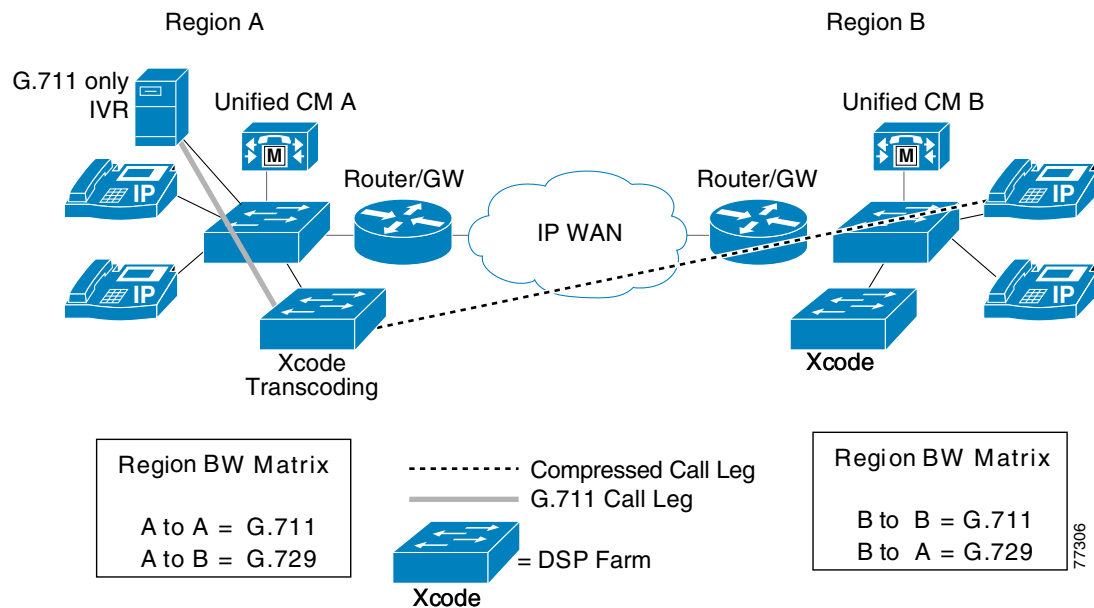
In distributed call processing deployments, several sites are connected through an IP WAN. Each site contains a Unified CM cluster that can, in turn, follow the single-site model or the centralized call processing model. A gatekeeper may be used for call admission control between sites.

Because WAN bandwidth is typically limited, calls between sites may be configured to use an LBR codec (such as G.729a) when traversing the WAN. H.323v2 intercluster trunks are used to connect Unified CM clusters. Unified CM also supports compressed voice call connections through the MTP service if a hardware MTP is used. (See [Figure 6-2](#).)

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 the use of transcoding resources. There might be specific instances where G.711 on a specific device cannot be avoided.
- Some endpoints (for example, video endpoints) do not support the H.323v2 features.

Figure 6-2 Intercluster Call Flow with Transcoding



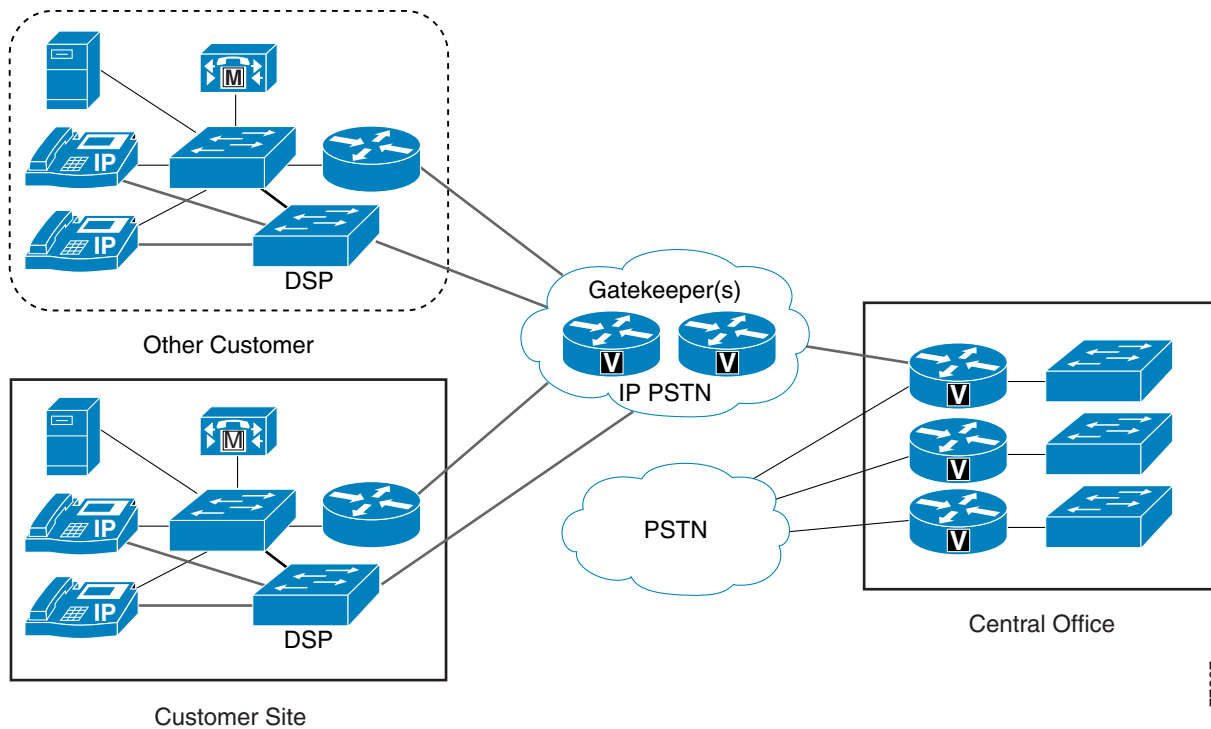
Unified CM uses media resource groups (MRGs) to enable sharing of MTP and transcoding resources among the Unified CM servers within a cluster. In addition, for calls across intercluster trunks, MTP and transcoding resources are used only when needed, thus eliminating the need to configure the MTP service 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

Another 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 6-3](#) illustrates this deployment model.

Figure 6-3 IP PSTN Access Example

Note that the customer site in [Figure 6-3](#) can use any of the previous three deployment models: single site, multisite WAN with centralized call processing, or multisite 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 Unified CM cluster to achieve more efficient use of the resources. This technique of utilizing MTPs for address hiding may be used for SIP trunks as well.



CHAPTER 7

Music on Hold

Last revised on: September 27, 2007

Music on hold (MoH) is an integral feature of the Cisco Unified Communications 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 Unified Communications Manager (Unified CM) 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.

Unified CM 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 7-20

Deployment Basics of MoH

For callers to hear music on hold, Unified CM 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
- Unified CM 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), Session Initiation Protocol (SIP), 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 Unified CM Administration interface. When an end device or feature places a call on hold, Unified CM connects the held device to an MoH media resource. Essentially, Unified CM 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

Unified CM 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 *IP Multicast SRND* document, available online at

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

Recommended Unicast/Multicast Gateways

The following recommended gateways support both unicast and multicast MoH:

- Cisco 6624 and 6608 gateway modules with MGCP and Cisco Unified CM Release 3.3(3) or later
- Cisco Communication Media Module (CMM) with MGCP or H.323 and Cisco Unified CM Release 4.0, Cisco IOS Release 12.2(13)ZP3 or later, and Catalyst OS Release 8.1(1) or later
- Cisco 2600, 2800, 3600, 3700, and 3800 Series Routers with MGCP or H.323 and Cisco IOS Release 12.2(8)T or later

Co-resident and Standalone MoH

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

- Co-resident deployment

In a co-resident deployment, the MoH feature runs on any server (either publisher or subscriber) in the cluster that is also running the Unified CM software.



Note The term *co-resident* refers to two or more services or applications running on the same server.

- Standalone deployment

A standalone deployment places the MoH feature on a dedicated server within the Unified CM cluster. The sole function of this dedicated server is to send MoH streams to devices within the network.

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 Unified CM 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
- G.729 Annex A
- Wideband

MoH audio files are generated automatically by Unified CM when .wav formatted audio files are uploaded to the MoH server using the upload file function on the MoH Audio File Management page (or on the Music On Hold Audio Source Configuration page). Unified CM then transcodes and formats audio source files into the appropriate MoH source files for the specified codec type(s). When an MoH event occurs, the MoH server streams the configured audio source file to the requesting device on hold.

**Note**

Before configuring a MoH audio source, the .wav formatted audio source file must be uploaded to every MoH server within the cluster. To upload the audio source file, the administrator must navigate to the Unified CM Administration interface on every MoH server within the cluster and use the upload file function on the MoH Audio File Management page. The administrator must perform this step for every audio source file. Only after the audio source file has been uploaded to all MoH servers in the cluster can the audio source be assigned to an MoH Audio Stream Number and configured as a MoH audio source.

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 connected to the audio input of 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 Unified CM 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 Cisco MoH USB audio sound card (MOH-USB-AUDIO=) must be used for connecting a fixed or live audio source to the MoH server. This USB sound card is compatible with all MCS platforms supporting Cisco Unified CM Release 5.x.

**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 Unified CM Cluster

The MoH feature requires that each MoH server must be part of a Unified CM cluster. All MoH servers must share their configurations with the publisher server and participate in the database replication schema. Specifically, the MoH server must share the following information (configured through Unified CM Administration) by means of the 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 Unified CM cluster and participates in the database replication automatically. To configure a standalone MoH server, start with a normal Unified CM 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 Unified CM, as well as typical call flow scenarios.

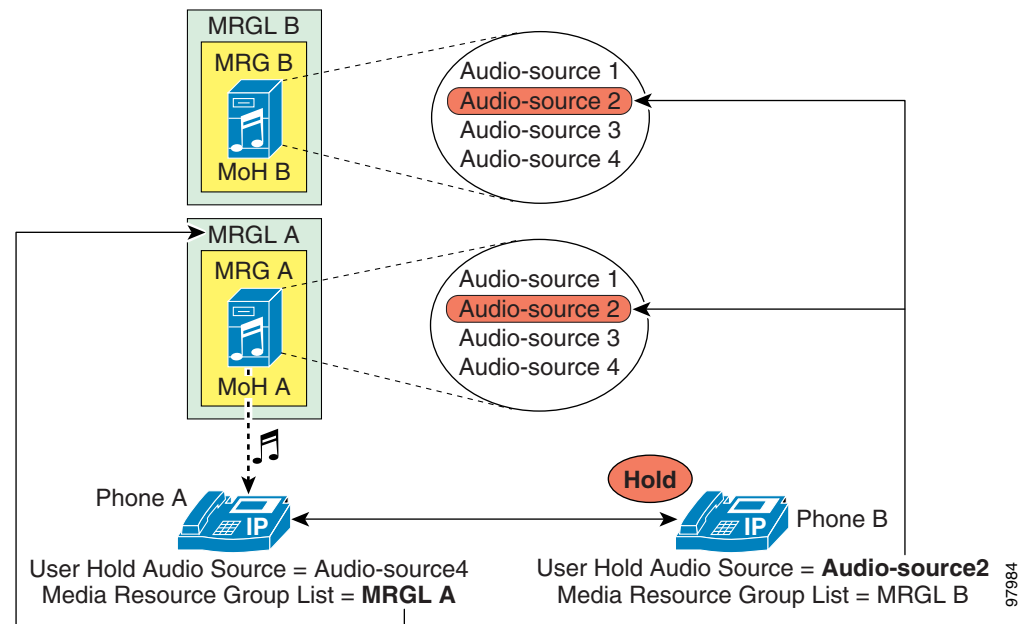
Basic MoH

The basic operation of MoH in a Cisco Unified Communications 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 Unified CM 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, Unified CM 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, Unified CM first looks at the User (or Network) Audio Source configured at the directory or line level. If this level is not defined, Unified CM looks at the User (or Network) Audio Source configured on the holder device. If this level is not defined, Unified CM looks at the User (or Network) Audio Source configured for the device pool of the holder device. If this level is not defined, then Unified CM looks at the cluster-wide default audio source ID configured under the Unified CM system parameters. (By default, this audio source ID is set to 1 for both User and Network Hold Audio Sources, which is the SampleAudioSource.)

Unified CM 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

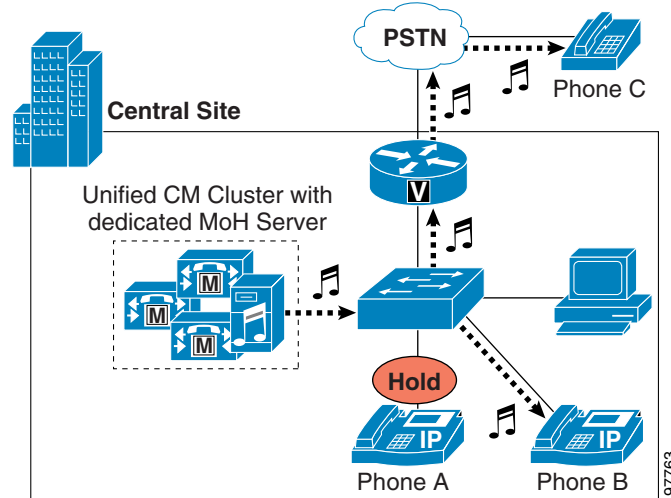
When attempting to determine the MRGL for a particular holdee, Unified CM looks at the MRGL configured at the device level. If this level is not defined, Unified CM looks at the MRGL configured for the device pool of the holdee device. If this level is not defined, then Unified CM uses the system default 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

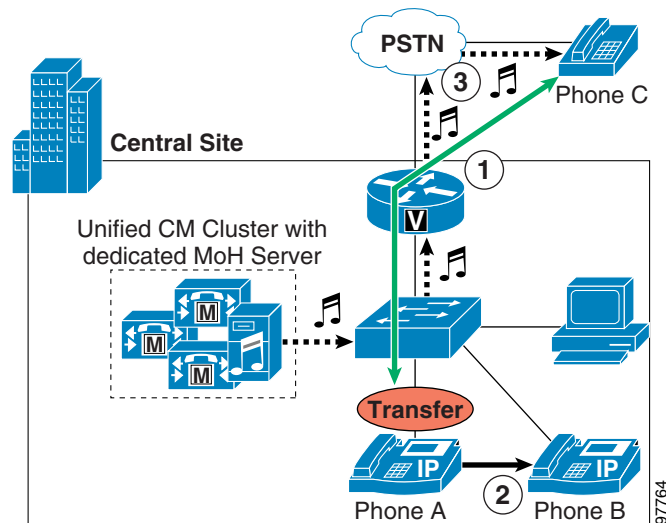
Figure 7-2 shows these two types of call flows. If phone A is in a call with phone B and phone A (holder) pushes the Hold softkey, then a music stream is sent from the MoH server to phone B (holdee). The music stream can be sent to holdees within the IP network or holdees on the PSTN, as is the case 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 an endpoint 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 Unified CM 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 Unified CM 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 7-20](#).

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 Unified CM 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

Configured audio sources are shared among *all* MoH servers in the Unified CM cluster, requiring each audio source file to be uploaded to every MoH server within the 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 methods of providing additional sources, refer to the sections on [Using Multiple Fixed \(Live\) Audio Sources](#), page 7-9, and [Multicast MoH from Branch Routers](#), page 7-16.

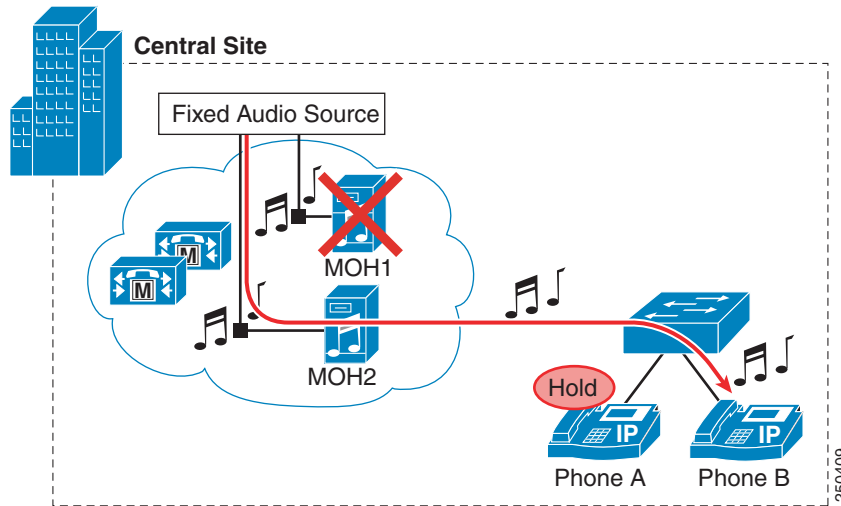
For those audio sources that will be used for multicast streaming, ensure that **Allow Multicasting** and **Play continuously (repeat)** are enabled. If continuous play of an audio source is not specified, only the first party placed on hold, not additional parties, will receive the MoH audio source.

Using Multiple Fixed (Live) Audio Sources

It is important to remember that only a single fixed audio source can be configured within Unified CM. However, each MoH server in the Unified CM cluster is capable of streaming a single fixed audio source via a Cisco MoH USB audio sound card (MOH-USB-AUDIO). When multiple fixed audio sources are needed, additional MoH servers can be added to provide these multiple sources. The audio supplied to each MoH server sound card can be the same or different, and the administrator can determine which MoH server is selected based on MRG and MRGL selections. When multiple audio sources are done in this manner, the holder's **User/Network Hold MoH Audio Source** should be configured for the fixed audio source (the single fixed audio source that is configured in Unified CM), and the MoH server to stream that fixed audio source to the device is then determined by the MRGL of the holdee.

In the case where the audio source is the same, this method also allows for redundancy of the fixed audio source. For example, in [Figure 7-4](#) there are two MoH servers, each with an MOH-USB-AUDIO sound card connected to an audio source streaming audio derived from a live radio station feed. Phone B's MRGL contains first an MRG that contains the MOH1 server and second an MRG that contains the MOH2 server. Assuming the User/Network Hold Audio Source at Phone A has been set to the fixed audio source, after a call is established between Phone A and Phone B, and Phone B is placed on hold by Phone A, Phone B will receive the live feed audio source from the MOH1 server. In the case where the MOH1 server is down (or has no available capacity) when Phone A puts Phone B on hold, Phone B will receive the live feed audio source from the MOH2 server.

Figure 7-4 Fixed Audio Source Redundancy Example

**Note**

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 Unified CM Cluster

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

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

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

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

When deploying separate MoH servers, configure one server without multicast enabled (unicast-only) and configure a second MoH server with multicast enabled. Assign the unicast-only MoH media resource and the multicast-enabled MoH media resource 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 whether the MRG is configured to use multicast and then on the server from which it is served.

When deploying a single MoH server for both unicast and multicast MoH, configure the server for multicast. Assign this same MoH media resource 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 MoH media 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 use 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. Lastly, configure a User Hold Audio Source and Network Hold Audio Source for each individual device or (in the case of phone devices) individual lines or directory numbers to assign 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.

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) value of 46 or a Per Hop Behavior (PHB) value 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.

Call signaling traffic between MoH servers and Unified CM servers is automatically marked with a DSCP value of 24 or a PHB value of CS3 (ToS of 0x60) by default. Therefore, as long as QoS is properly configured on the network, this call signalling traffic will be properly classified and queued within the network along with all other call signalling 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 7825	G.711 (A-law and mu-law) G.729a Wideband audio	Co-resident or standalone server: 250 MoH sessions ¹
MCS 7835 MCS 7845	G.711 (A-law and mu-law) G.729a Wideband audio	Co-resident or standalone server: 500 MoH sessions

1. You can configure a maximum of 51 unique audio sources per Unified CM cluster.

The following MoH Server Configuration parameters affect MoH server capacity:

- Maximum Half Duplex Streams**

This parameter determines the number of devices that can be placed on unicast MoH. By default this value is set to 250.

The Maximum Half Duplex Streams parameter should be set to the value derived from the following formula:

$$(\text{Server and deployment capacity}) - ((\text{Number of multicast MoH sources}) * (\text{Number of MoH codecs enabled}))$$

For example:

MCS-7835 standalone MoH server	Multicast MoH audio sources	MoH codecs enabled (G.711 mu-law and G.729)	Maximum half-duplex streams
500	- (12	* 2)	= 476

Therefore, in this example, the Maximum Half Duplex Streams parameter would be configured with a value of no more than 476.

The value of this parameter should never be set higher than the capacities indicated in Table 7-1, based on the platform and deployment type (co-resident or standalone).

- **Maximum Multicast Connections**

This parameter determines the number of devices that can be placed on multicast MoH. By default this value is set to 30.

The Maximum Multicast Connections parameter should be set to a number that ensures that all devices can be placed on multicast MoH if necessary. Although the MoH server can generate only a finite number of multicast streams (maximum of 204), large numbers of held devices can join each multicast stream, thus it is necessary to set this parameter to an artificially large number. Typically multicast traffic is accounted for based on the number of streams being generated; however, Unified CM maintains a count of the actual number of devices placed on multicast MoH or joined to each multicast MoH stream. This method is different than the way multicast traffic is normally tracked.



Note Because you can configure only 51 unique audio sources per Unified CM cluster and because there are only four possible codecs for MoH streams, the maximum number of multicast streams per MoH server is 204.

Failure to configure these parameters properly could lead to under-utilization of MoH 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 co-resident 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 co-resident (or standalone) MoH servers would be required to handle this load, as shown by the following calculation:

$$[(500 \text{ sessions per MCS 7835 or 7845 co-resident server}) * (1 \text{ co-resident server})] + [(250 \text{ sessions per MCS 7815 or 7825 co-resident server}) * (1 \text{ co-resident server})] > 600 \text{ sessions}$$

By comparison, a multicast-only MoH environment with 36 unique MoH audio streams, for example, would require only one co-resident MoH server (MCS 7815 or 7825), as shown by the following calculation:

$$(250 \text{ sessions per MCS 7815 or 7825 co-resident server}) * (1 \text{ co-resident server}) > 36 \text{ sessions}$$

These 36 unique multicast streams could 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.

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 [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 Multisite Deployments

Multisite 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 Unified CM 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 probable that voice media traffic might get delayed or dropped without appropriate call admission control. For additional information, see [Call Admission Control, page 9-1](#).

Call admission control for Unified CM (based on either static locations or RSVP-enabled locations) 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 ($4 * 24 \text{ kbps per G.729 audio stream} = 96 \text{ kbps}$).

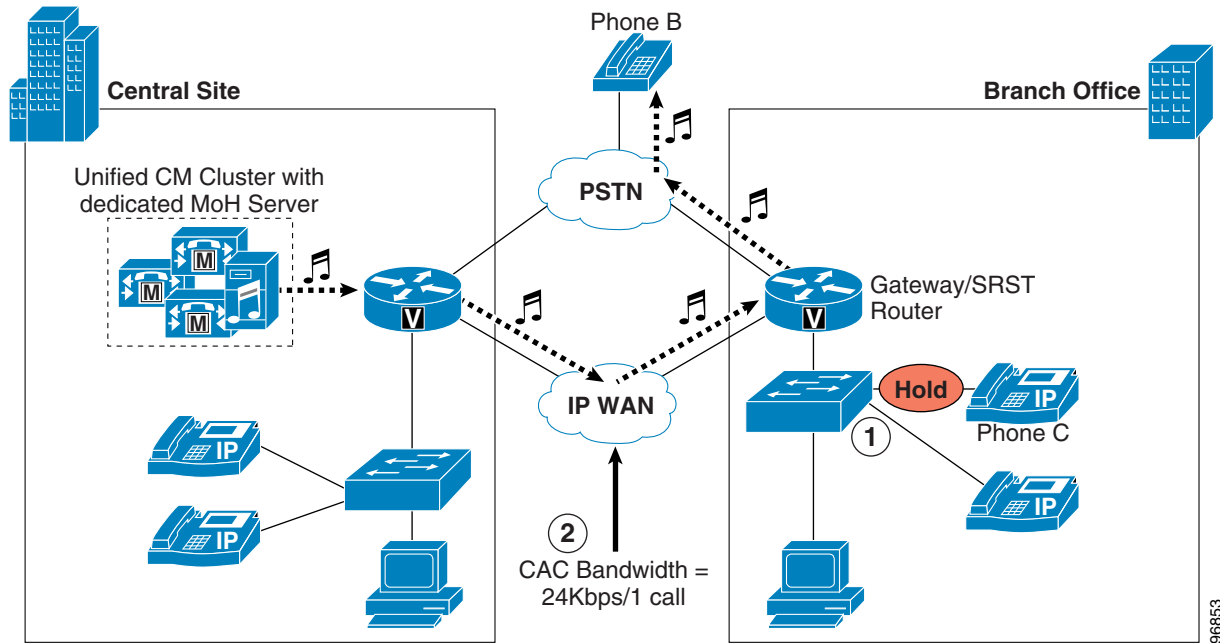
[Figure 7-5](#) shows an example of call admission control and MoH in a centralized multisite deployment. For this example, assume that IP phone C is in a call with a PSTN phone (phone B). At this point, no bandwidth has been consumed on the WAN. When phone C pushes the Hold softkey (step 1), phone B 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 Routers

Beginning with Cisco Unified Survivable Remote Site Telephony (SRST) Release 3.0, MoH can be multicast in a remote or branch site via the branch SRST router's flash. Beginning with SRST Release 3.2, MoH can be multicast in a branch site via a live feed connected to an analog port on the branch SRST router. Multicast MoH from a branch router via these two methods enhances the Cisco Unified Communications MoH feature in both of the following scenarios:

- Non-Fallback Mode

When the WAN is up and the phones are controlled by Unified CM, this configuration can eliminate the need to forward MoH across the WAN to remote branch sites by providing locally sourced MoH.

- Fallback Mode

When SRST is active and the branch devices have lost connectivity to the central-site Unified CM, the branch router can continue to provide multicast MoH.

When using the live feed option in either scenario, the SRST router provides redundancy by monitoring the live feed input, and it will revert to streaming MoH from a file in flash if the live feed is disconnected. You can use only a single multicast address and port number per SRST router to provide multicast MoH; therefore, the SRST router does not support streaming from both the live feed and the flash file at the same time. In addition, the SRST router can stream only a single audio file from flash.



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.

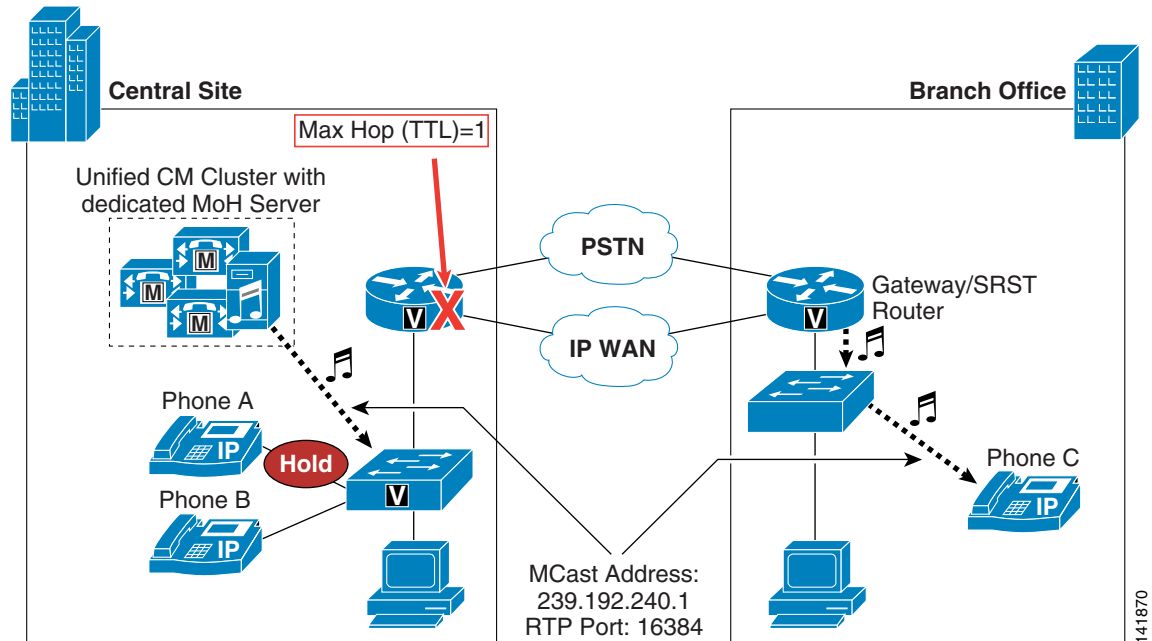
Non-Fallback Mode

During non-fallback mode (when the WAN is up and SRST is not active), the branch SRST router can provide multicast MoH to all local Cisco Unified Communications devices. To accomplish this, you must configure a Unified CM MoH 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 SRST router, 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 a branch router when it is not in fallback mode. After phone A places phone C on hold, phone C receives multicast MoH from the local SRST router. In this example, 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 either flash or a live feed. 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

When using this method for delivering multicast MoH, configure all devices within the Unified CM 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.

In networks with multiple branch routers configured to stream multicast MoH, this allows for more than 51 unique MoH audio sources within the Unified CM cluster. Each branch site router can multicast a unique audio stream, 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 audio, then the cluster actually contains 101 unique MoH audio sources (100 branch streams and one central-site stream). If you want more than 51 unique audio streams in the central site, see the methods described in [Using Multiple Fixed \(Live\) Audio Sources](#), page 7-9.

Fallback Mode

During fallback mode (when the WAN is down and SRST is active), the branch SRST router can stream multicast MoH to all analog and digital ports within the chassis, thereby providing MoH to analog phones and PSTN callers. Beginning with Cisco Unified SRST Release 4.0, SCCP IP phones in fallback mode can also receive the multicast MoH stream. At this time, SIP IP Phones will receive tone-on-hold.

The branch router's configuration for fallback mode multicast MoH is the same as the normal operation configuration. However, which multicast address you configure on the router depends on the intended operation. If you want the branch router to provide multicast MoH to devices only in fallback mode (for example, if MoH received by remote devices is to be sourced from the central-site MoH server during non-fallback mode), then the multicast address and port number configured on the SRST 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.

Note that, once the branch SRST/gateway router is configured to provide multicast MoH, the router will continue to multicast the MoH stream even when not in fallback mode.

Beginning with Cisco Unified SRST Release 4.0, it is also possible to configure the fallback mode to use Cisco Unified Communications Manager Express (Unified CME) in SRST mode. Fallback mode behavior is still the same, but the configuration commands are slightly different. SRST commands are entered under the Cisco IOS **call-manager-fallback** construct, while the commands for Unified CME in SRST mode are entered under **telephony-service**.

There are four methods of providing multicast MoH via SRST:

- SRST multicast MoH from branch router flash
- SRST multicast MoH from a live feed
- Unified CME in SRST mode with multicast MoH from branch router flash
- Unified CME in SRST mode with multicast MoH from a live feed

For more details on configuration of *Cisco SRST and Unified CME*, refer to the *Cisco Unified SRST System Administrator Guide* and the *Cisco Unified Communications Manager Express System Administrator Guide*, both available at

<http://www.cisco.com>

Distributed Multisite Deployments

Multisite 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 multisite 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 multisite deployment. Even when MoH servers are placed in a separate Unified CM region and the G.711 codec is configured between this region and the intercluster or SIP 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 Unified CM clusters (intercluster calls). Therefore, you must configure at least one unicast MoH resource in each Unified CM cluster if you want MoH on the intercluster or SIP trunk.

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 Unified CM 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 multisite 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

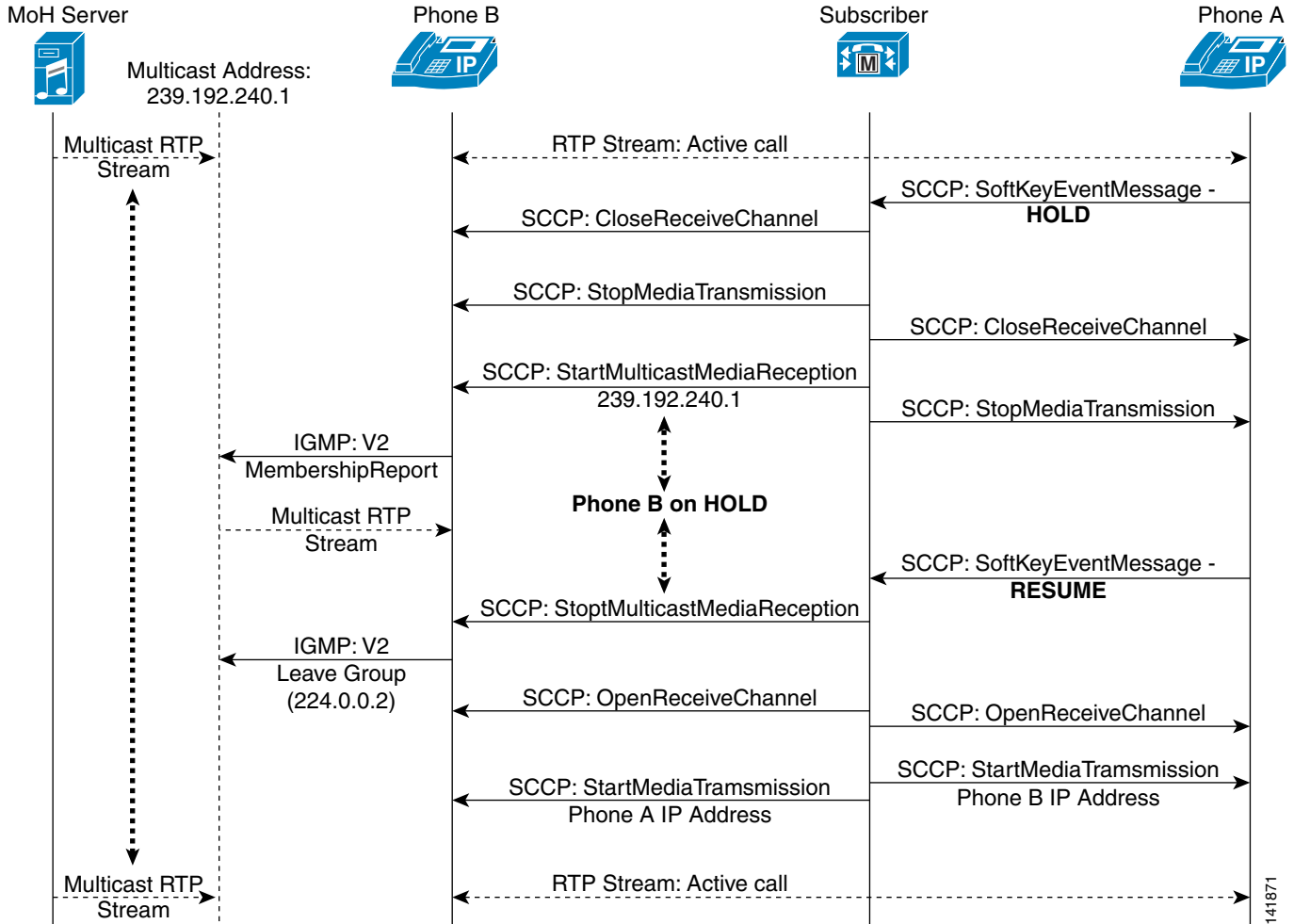
The following sections provide detailed illustrations and explanations of unicast and multicast MoH call flows for both SCCP and SIP endpoints.

SCCP Call Flows

This section describes the call flows for music on hold with Skinny Client Control Protocol (SCCP) endpoints.

SCCP Multicast Call Flow

[Figure 7-7](#) illustrates a typical SCCP multicast call flow. As shown in the diagram, when the Hold softkey is pressed at phone A, Unified CM 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, Unified CM 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) V2 Membership Report message indicating that it is joining this group.

Figure 7-7 Detailed SCCP Multicast MoH Call Flow

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, Unified CM instructs phone B to Stop Multicast Media Reception. Phone B then sends an IGMP V2 Leave Group message to 224.0.0.2 to indicate that the multicast stream is no longer needed. This effectively ends the MoH session. Next, Unified CM 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, Unified CM instructs both phones to Start Media Transmission to each other's IP addresses. The phones are once again connected via an RTP two-way audio stream.

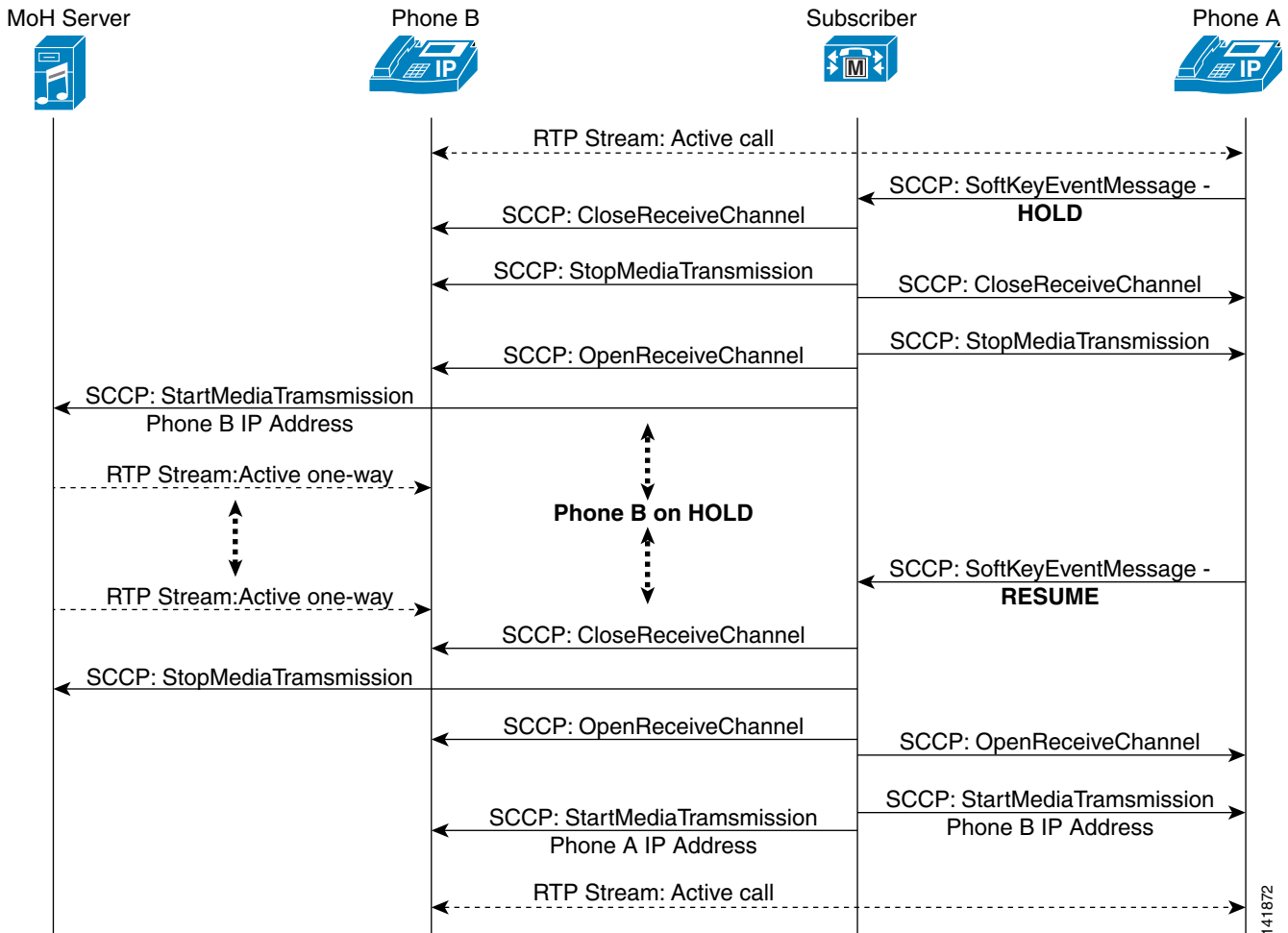
**Note**

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, keep-alives, 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.

SCCP Unicast Call Flow

Figure 7-8 depicts an SCCP unicast MoH call flow. In this call flow diagram, when the Hold softkey is pressed at phone A, Unified CM 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.

Figure 7-8 Detailed SCCP Unicast MoH Call Flow



Next, Unified CM tells phone B (the holdee) to Open Receive Channel. (This is quite different from the multicast case, where Unified CM tells the holdee to Start Multicast Media Reception.) Then Unified CM 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, Unified CM 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, Unified CM 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.

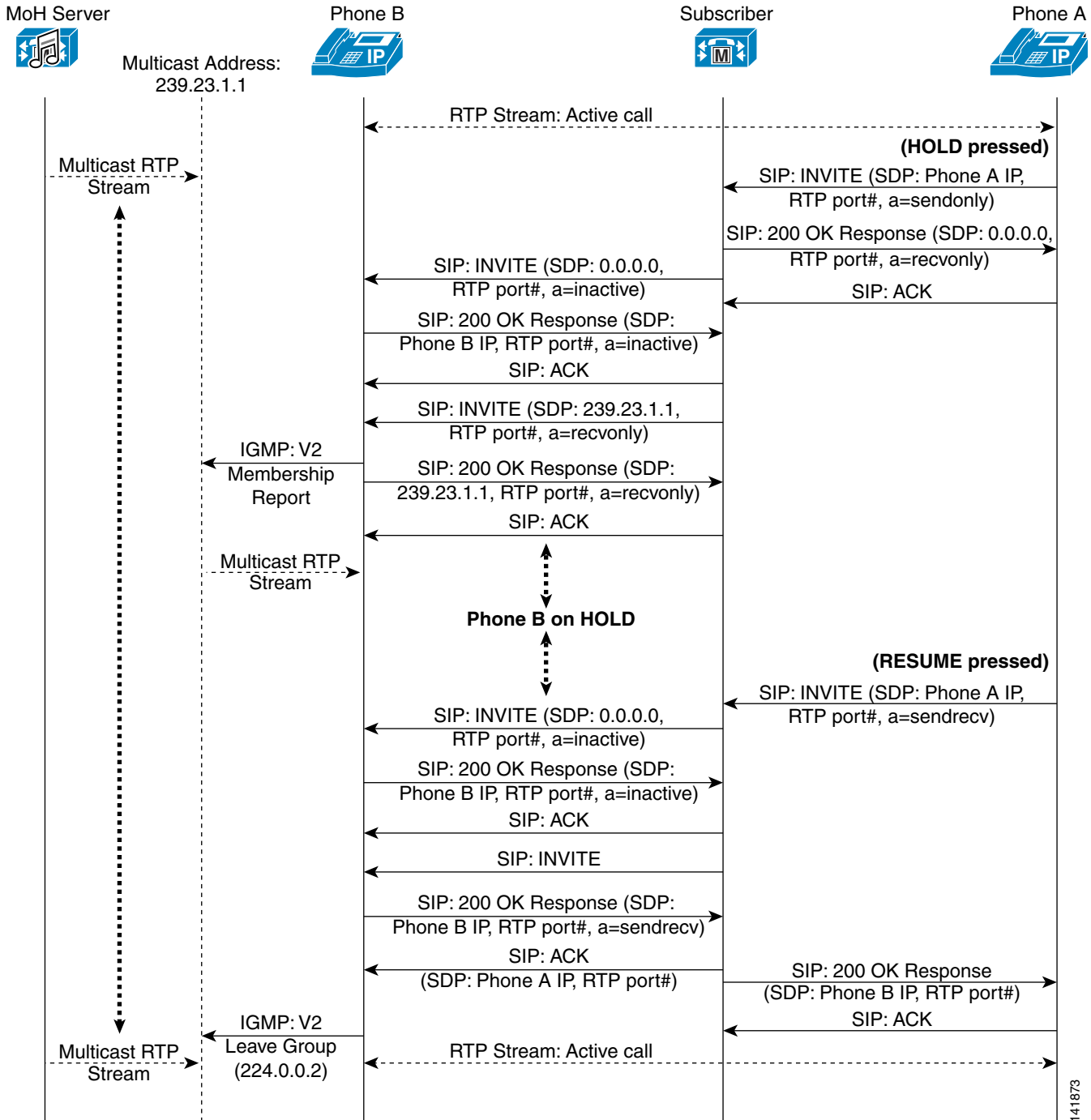
SIP Call Flows

This section describes the call flows for music on hold with Session Initiation Protocol (SIP) endpoints.

SIP Multicast Call Flow

[Figure 7-9](#) illustrates a typical SIP multicast call flow. As shown in the diagram, when the Hold softkey is pressed at phone A, phone A sends a SIP INVITE with a Session Description Protocol (SDP) connection information indication of phone A's IP address and a media attribute indication of sendonly. Unified CM then instructs phone A to disconnect the RTP stream via a SIP 200 OK Response with an SDP connection information indication of 0.0.0.0 and a media attribute indication of recvonly. Phone B is then told to disconnect the RTP stream via a SIP INVITE from Unified CM with an SDP connection information indication of 0.0.0.0 and a media attribute of inactive. After a SIP 200 OK Response is sent back from phone B to Unified CM indicating an SDP media attribute of inactive, Unified CM then sends a SIP INVITE to phone B with an SDP connection information indication of the MoH multicast group address (in this case 239.23.1.1) and a media attribute of recvonly.

Figure 7-9 Detailed SIP Multicast MoH Call Flow



141873

Next, phone B in Figure 7-9 issues an IGMP V2 Membership Report message indicating that it is joining this multicast group. In addition, phone B sends a SIP 200 OK Response back to Unified CM indicating an SDP media attribute of `recvonly` in response to the previous SIP INVITE. Meanwhile, the MoH server has been sourcing RTP audio to this MoH multicast group address and, upon joining the multicast group, phone B begins receiving the one-way MoH stream.

When the user at phone A presses the Resume softkey, phone A sends a SIP INVITE with an SDP connection information indication of phone A's IP address and media attribute indications of phone A's receiving RTP port and sendrecv. Unified CM then instructs phone B to disconnect from the multicast MoH stream via a SIP INVITE with an SDP connection information indication of 0.0.0.0 and a media attribute indication of inactive. A SIP 200 OK Response is sent back from phone B to Unified CM, indicating an SDP media attribute of inactive.

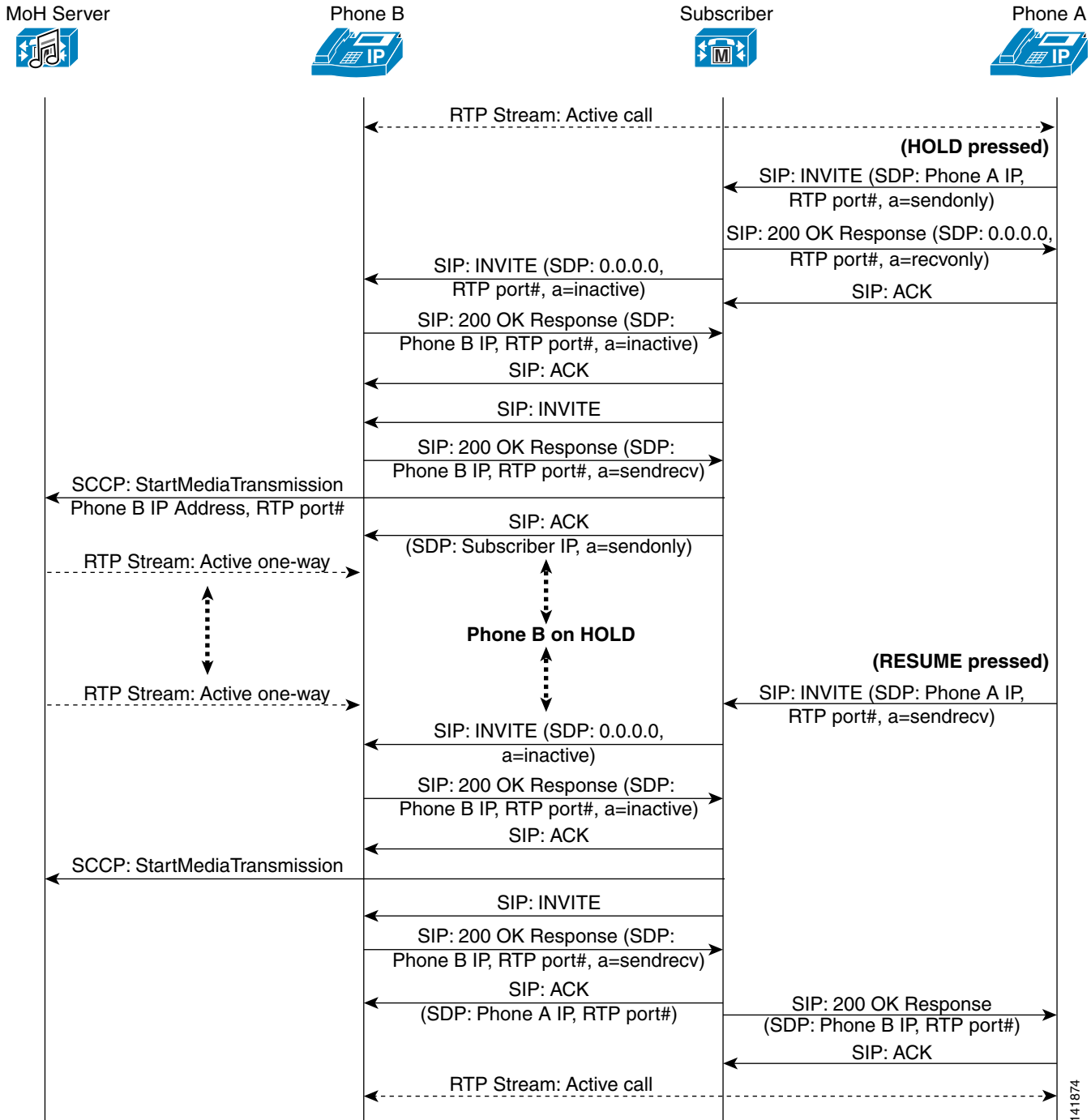
Next Unified CM sends a SIP INVITE to phone B, and phone B responds with a SIP 200 OK Response with an SDP connection information indication of phone B's IP address and media attribute indications of phone B's receiving RTP port and sendrecv. Unified CM responds by sending a SIP ACK to phone B with an SDP connection information indication of phone A's IP address and a media attribute of phone A's receiving RTP port number. Likewise, Unified CM forwards a SIP 200 OK Response to phone A's original resuming SIP INVITE, with an SDP connection information indication of phone B's IP address and a media attribute of phone B's receiving RTP port number. Phone B then sends an IGMP V2 Leave Group message to 224.0.0.2 to indicate that the multicast stream is no longer needed. Finally, the RTP two-way audio stream between phone A and B is reestablished.

**Note**

The call flow diagrams in [Figure 7-9](#), [Figure 7-10](#), and [Figure 7-11](#) 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, keep-alives, some acknowledgements, progression indications, 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.

SIP Unicast Call Flow

[Figure 7-10](#) depicts a SIP unicast MoH call flow. As shown in the diagram, when the Hold softkey is pressed at phone A, phone A sends a SIP INVITE with an SDP connection information indication of phone A's IP address and a media attribute indication of sendonly. Unified CM then instructs phone A to disconnect the RTP stream via a SIP 200 OK Response with an SDP connection information indication of 0.0.0.0 and a media attribute indication of recvonly. Phone B is then told to disconnect the RTP stream via a SIP INVITE from Unified CM, with an SDP connection information indication of 0.0.0.0 and a media attribute of inactive. Next a SIP 200 OK Response is sent back from phone B to Unified CM, indicating an SDP media attribute of inactive. Up to this point, unicast and multicast MoH call flows are exactly the same.

Figure 7-10 Detailed SIP Unicast MoH Call Flow

Unified CM then sends a SIP INVITE to phone B, and Phone B responds back with a SIP 200 OK Response indicating an SDP connection information indication with phone B's IP address and media attribute indications of phone B's receiving RTP port number and sendrecv. Unified CM then sends a SCCP StartMediaTransmission message to the MoH server, with phone B's address and receiving RTP

port number. This is followed by a SIP ACK from Unified CM to phone B indicating SDP connection information of the Unified CM IP address and a media attribute of sendonly. Meanwhile, the MoH server begins sourcing RTP audio to phone B, and phone B begins receiving the one-way MoH stream.

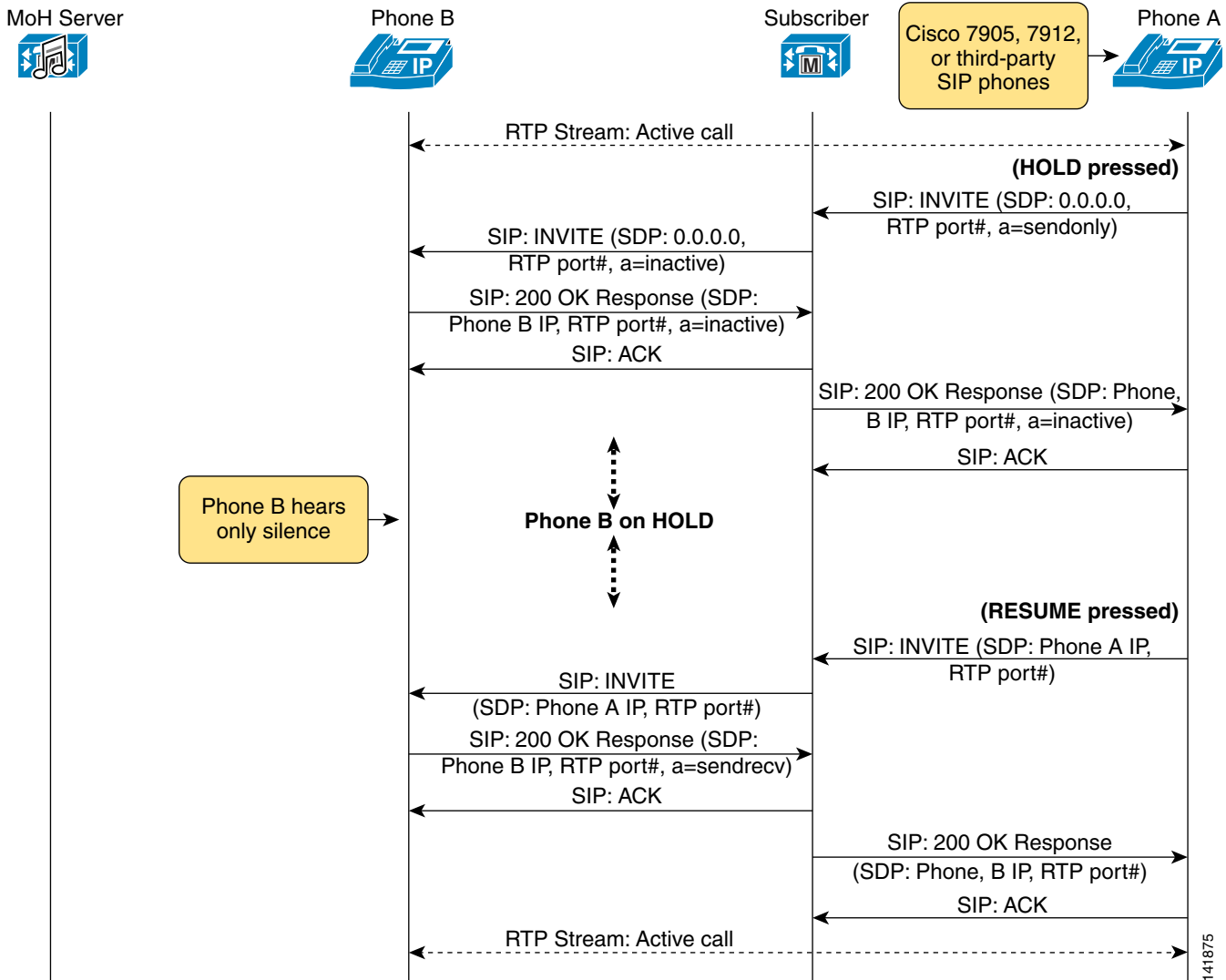
When the user at phone A presses the Resume softkey, phone A sends a SIP INVITE with an SDP connection information indication of phone A's IP address and media attribute indications of phone A's receiving RTP port and sendrecv. Unified CM then instructs phone B to disconnect from the multicast MoH stream via a SIP INVITE with an SDP connection information indication of 0.0.0.0 and a media attribute indication of inactive. A SIP 200 OK Response is sent back from phone B to Unified CM, indicating an SDP media attribute of inactive. Then Unified CM sends a SCCP StopMediaTransmission message to the MoH server, causing the MoH server to stop forwarding the MoH stream to phone B.

Next Unified CM sends a SIP INVITE to phone B, and phone B responds with a SIP 200 OK Response with an SDP connection information indication of phone B's IP address and media attribute indications of phone B's receiving RTP port and sendrecv. Unified CM responds by sending a SIP ACK to phone B, with an SDP connection information indication of phone A's IP address and a media attribute of phone A's receiving RTP port number. Likewise, Unified CM forwards a SIP 200 OK Response to phone A's original resuming SIP INVITE with an SDP connection information indication of phone B's IP address and a media attribute of phone B's receiving RTP port. Finally, the RTP two-way audio stream between phone A and B is reestablished.

SIP Media Hold Call Flow

[Figure 7-11](#) depicts an RFC 2543 media hold call flow. This call flow occurs only when the phone placing the call on hold (in this case phone A) is a Cisco Unified IP Phone 7905 or 7912, or any third-party SIP phone. As shown in the diagram, when the Hold softkey is pressed at phone A, phone A sends a SIP INVITE with an SDP connection information indication of 0.0.0.0 and a media attribute indication of sendonly. Phone B is then told to disconnect the RTP stream via a SIP INVITE from Unified CM, with an SDP connection information indication of 0.0.0.0 and a media attribute of inactive. Next a SIP 200 OK Response is sent back from phone B to Unified CM, with an SDP connection information indication of phone B's IP address and media attribute indications of phone B's receiving RTP port number and inactive. Next Unified CM forwards a SIP 200 OK Response back to phone A in response to phone A's original holding SIP INVITE, with an SDP connection information indication of phone B's IP address and media attribute indications of phone B's receiving RTP port and inactive.

At this point, phone B is on hold but receiving no MoH, so the phone B user hears only silence.

Figure 7-11 Detailed SIP Media Hold Call Flow

When the user at phone A presses the Resume softkey, phone A sends a SIP INVITE with an SDP connection information indication of phone A's IP address and a media attribute indication of phone A's receiving RTP port. Unified CM then sends a SIP INVITE to phone B, with an SDP connection information indication of phone A's IP address and a media attribute indication of phone A's receiving RTP port. Phone B in turn responds with a SIP 200 OK Response with an SDP connection information indication of phone B's IP address and media attribute indications of phone B's receiving RTP port and sendrecv. Likewise, Unified CM forwards a SIP 200 OK Response to phone A's original resuming SIP INVITE, with an SDP connection information indication of phone B's IP address and a media attribute indication of phone B's receiving RTP port. Finally, the RTP two-way audio stream between phone A and B is reestablished.

**Note**

Media hold occurs as described above when the Cisco Unified IP Phone 7905 or 7912 and third-party SIP phones place calls on hold. In addition, these phones will receive MoH if they are placed on hold by any other Cisco Unified IP Phone models, and the call flows for these scenarios look almost identical to the flows depicted in [Figure 7-9](#) and [Figure 7-10](#).



CHAPTER 8

Call Processing

Last revised on: September 27, 2007

This chapter provides design guidance for scalable and resilient call processing systems with Cisco Unified Communications Manager 5.0. This chapter also discusses how to choose the appropriate hardware and deployment scenario for Cisco Unified Communications Manager (Unified CM) 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

This chapter focuses on the following topics:

- [Unified CM Cluster Guidelines, page 8-2](#)

This section discusses the minimum hardware requirements for Unified CM. This section also explains the various feature services that can be enabled on the Unified CM server and the purpose of each.

- [Unified CM Platform Capacity Planning, page 8-15](#)

This section provides guidelines for using the Cisco Unified CM Capacity Tool, which must be used when planning an IP Telephony deployment. The Cisco Unified CM Capacity Tool provides guidance on resources used on a Unified CM server, based on certain deployment requirements.

- [Gatekeeper Design Considerations, page 8-20](#)

This section explains how gatekeepers can be used in an IP Telephony deployment. Cisco Gatekeeper may also be paired with other standby gatekeepers or may be clustered for higher performance and resilience. Gatekeepers may also be used for call routing and call admission control.

- [Interoperability of Unified CM and Unified CM Express, page 8-30](#)

This section explains the H.323 and SIP integration between Cisco Unified CM and Cisco Unified Communications Manager Express (Unified CME) in a distributed call processing deployment.

Unified CM Cluster Guidelines

The Unified CM 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 Unified CM 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.

This section describes the various functions performed by the servers that form a Unified CM cluster, and it provides guidelines for deploying the servers in ways that achieve the desired scale, performance, and resilience.

Hardware Platforms

Unified CM 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.

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 Unified CM Servers*

Server Type	Cisco Server Model	Characteristics
Standard server (not high availability)	MCS 7815 or equivalent	• Single processor • Single power supply • Non-RAID hard disk
High-availability standard server	MCS 7825 or equivalent	• Single processor • Multiple power supplies • Single SCSI RAID hard disk array
High-performance server	MCS 7835 and MCS 7845 or equivalent	• Multiple processors • Multiple power supplies • Multiple SCSI RAID hard disk arrays

Cisco Unified CM 5.0 is supported on specific Cisco MCS 7815, MCS 7825, MCS 7835, and MCS 7845 servers or on customer-provided HP and IBM servers that have been verified by Cisco to meet the following minimum requirements:

- Processor speed must be 2.0 GHz or greater
- Physical memory size must be 2 GB or greater
- Physical hard disk size must be 72 GB or larger

For a complete list of currently supported hardware configurations, refer to the documentation available at

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

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 Unified CM 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. Many servers also include the capability of using Gigabit Ethernet, which is also an option. Ensure that servers are connected to the network using full-duplex, which can be achieved with 10 Mbps and 100 Mbps by hard-coding the switch port and the server NIC. For 1000 Mbps, Cisco recommends using Auto/Auto for speed and duplex configuration on both the NIC and the switch port. The default for Cisco Unified CM 5.0 is Auto/Auto, and this setting is also the default following an upgrade from a previous Unified CM release.

**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, with the exception of Gigabit Ethernet ports which should be set to Auto/Auto.

NIC Teaming for Network Fault Tolerance

Hewlett-Packard (HP) server platforms with dual Ethernet network interface cards (NICs) can support NIC teaming for Network Fault Tolerance with Cisco Unified CM 5.0. This feature allows a server to be connected to the Ethernet via two NICs and, hence, two cables. NIC teaming prevents network downtime by transferring the workload from the failed port to the working port. NIC teaming cannot be used for load balancing or increasing the interface speed.

General Clustering Guidelines

The following guidelines apply to all Unified CM clusters:

**Note**

A cluster may contain a mix of server platforms, but all servers in the cluster must run the same Unified CM 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 2-17](#).

Unified CM Cluster Services

Within a Unified CM 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 database 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.

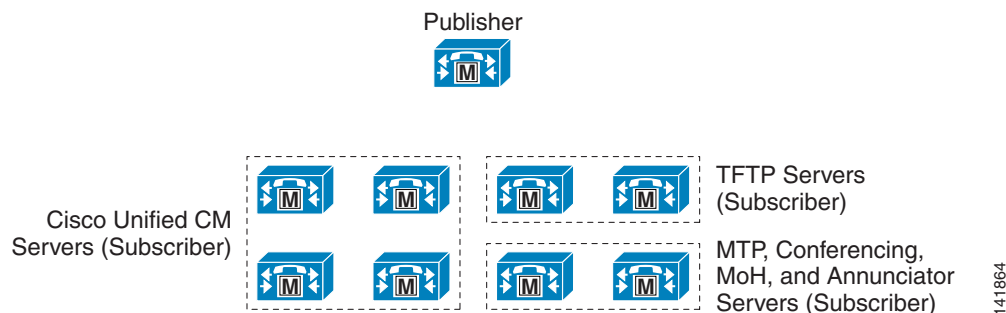
With the release of Cisco Unified CM 5.0, a cluster may contain as many as 20 servers, of which a maximum of eight may run the Cisco CallManager Service that provides call processing. The other servers may be configured as a dedicated database publisher, dedicated Trivial File Transfer Protocol (TFTP) 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 minimum 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. A maximum of 500 phones is supported in this configuration with Cisco Unified CM 5.1 on a Cisco MCS 7815 (or a maximum of 300 phones with Cisco Unified CM 5.0 on the MCS 7815). 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 and separate servers for primary and secondary call processing services is recommended, thus increasing the number of servers in a cluster.

It is also possible to deploy a single-server cluster with an MCS 7825 or greater servers. With an MCS 7825 or equivalent server, the limit is 500 users; with a higher-availability server, the single-server 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 Unified CM 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 Unified CM cluster.

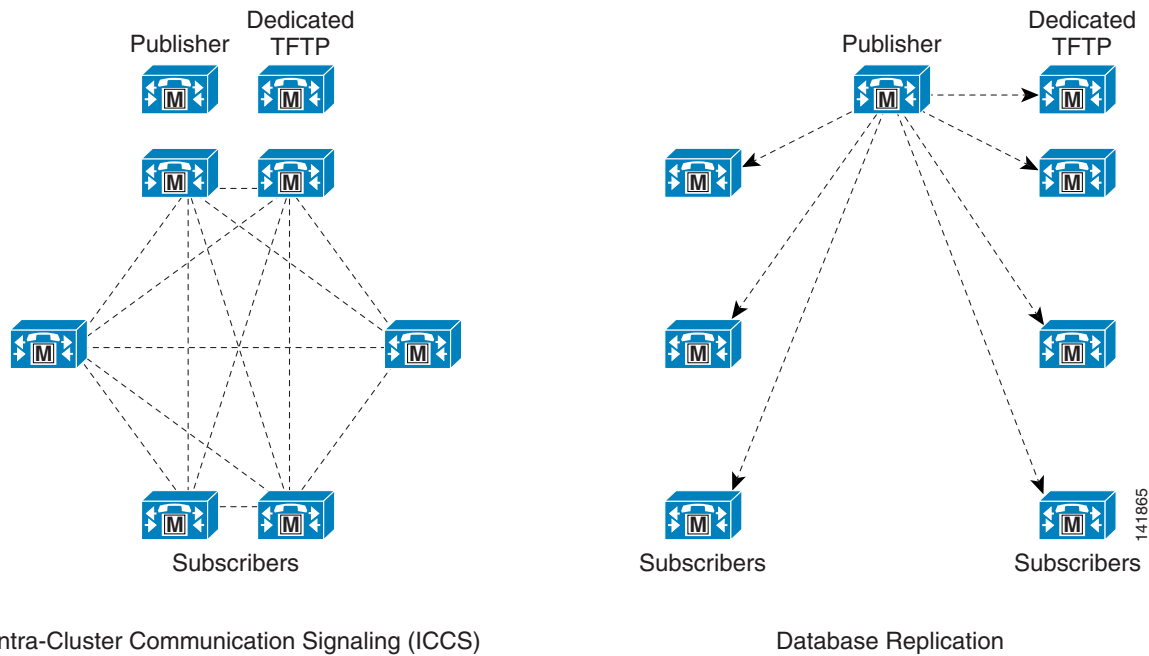
Figure 8-1 Typical Unified CM Cluster



Intracuster Communications

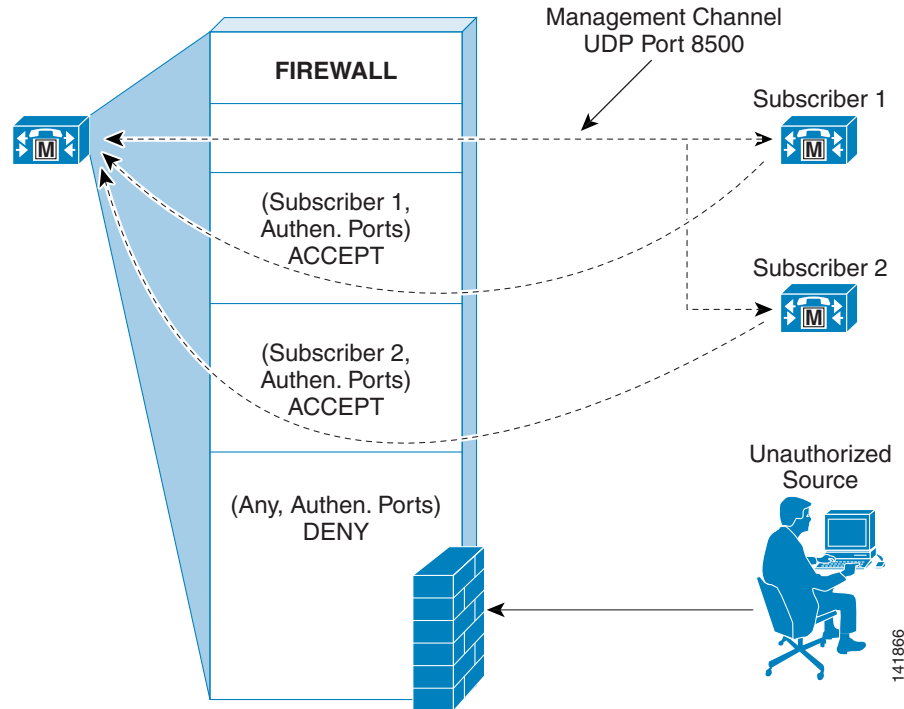
There are two primary kinds of communication within a Unified CM cluster (intracuster communications). (See Figure 8-2.) The first is a mechanism for distributing the database that contains all the device configuration information (see Database Replication in Figure 8-2). The configuration database is 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 (see ICCS in Figure 8-2). 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.

Figure 8-2 Intracluster Communications

Intracluster Security

With the release of Cisco Unified CM 5.0, a different architecture and mechanism are implemented for securing the Unified CM applications and communication links. Each server in a cluster runs an internal dynamic firewall. The application ports on Unified CM are protected using source IP filtering. The dynamic firewall opens these applications ports only to authenticated or trusted servers. (See [Figure 8-3.](#))

Figure 8-3 Intracluster Security

This security mechanism is applicable only between servers in a single Unified CM cluster. Unified CM subscribers are authenticated in a cluster before they can access the publisher's database. The Intra-cluster communication and database replication take place only between authenticated servers. During the installation process, a subscriber is authenticated to the publisher using a pre-shared key authentication mechanism. The authentication process involves the following steps:

1. Install the publisher server using a security password.
2. Configure the subscriber server on the publisher by using Unified CM Administration.
3. Install the subscriber server using the same security password used during publisher server installation.
4. After the subscriber is installed, the server attempts to establish connection to the publisher on a management channel using UDP 8500. The subscriber sends all the credentials to the publisher, such as hostname, IP address, and so forth. The credentials are authenticated using the security password used during the install process.
5. The publisher verifies the subscriber's credentials using its own security password.
6. The publisher adds the subscriber as a trusted source to its dynamic firewall table if the information is valid. The subscriber is allowed access to the database.
7. The subscriber gets a list of other subscriber servers from the publisher. All the subscribers establish a management channel with each other, thus creating a mesh topology.

Publisher

The publisher is a required server in all clusters, and there can currently be only one per cluster. This server is the first to be installed and provides the database 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 1250 users, Cisco recommends a dedicated publisher to prevent administrative operations from affecting the telephony users. A dedicated publisher does not have any call processing services or TFTP services running on the server. Other servers run the TFTP and Unified CM services.

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:

- Call forwarding changes
- Operations that requires licensing service
- Configuration 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 have 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 Unified CM 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, only the database and network services are enabled. All subscribers will subscribe to the publisher to obtain a read-only copy of the database information.

A call processing subscriber is a server that has the Cisco CallManager Service enabled. A single server license is required to enable this service on a subscriber. The Cisco CallManager Service cannot be enabled on a server if the publisher is not available because the publisher acts as a licensing server and distributes the licenses needed to activate the Cisco CallManager Service. 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. Cisco Unified CM 5.0 supports up to eight servers in a cluster with the Cisco CallManager Service enabled.

Depending on the redundancy scheme chosen (see [Call Processing Redundancy, page 8-8](#)), the call processing subscriber will be either a primary (active) subscriber or a backup (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 call processing service should not be enabled on the publisher and TFTP server. 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 [Unified CM Platform Capacity Planning, page 8-15](#), and redundancy is covered in the section on [Call Processing Redundancy, page 8-8](#).

Call Processing Redundancy

With Cisco Unified CM 5.0, 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 Skinny Client Control Protocol (SCCP) IP phones of approximately 125 registrations per second. The failover mechanism for Session Initiation Protocol (SIP) phones is approximately 40 registrations per second.

With the release of Cisco Unified CM 5.0, a cluster can be upgraded without impacting the services. Unified CM 5.0 allows two versions of Unified CM to be on the same server, one in the active partition and other in the inactive partition. All services and devices use the Unified CM version in the active partition for all Unified CM functionality. During the upgrade process, the cluster operations continue using its current release of Unified CM in the active partition, while the upgrade version gets installed in the inactive partition. Once the upgrade process is completed, the servers can be rebooted to switch the inactive partition to the active partition, thus running the new version of Unified CM.

Upgrading from Unified CM 5.0.x to Unified CM 5.0.x

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

-
Step 1	Install the new version of Unified CM on the publisher. Do not reboot.
Step 2	Install the new version of Unified CM on all subscribers simultaneously. Do not reboot.
Step 3	Reboot only the publisher. Switch to the new version of Unified CM and allow some time for the database to initialize.
Step 4	Reboot the TFTP server(s) one at a time. Switch to the new version of Unified CM and wait for the configuration files to be rebuilt before upgrading any further servers in the cluster.
Step 5	Reboot the dedicated music on hold (MoH) server(s) one at a time. Switch to the new version of Unified CM.
Step 6	Reboot the backup subscriber(s) one at a time. Switch to the new version of Unified CM. This step might impact some users if 50/50 load balancing is implemented.
Step 7	Fail-over the devices from the primary subscribers to their backups.
Step 8	Reboot the primary subscriber(s) one at a time. Switch to the new version of Unified CM.
-

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 Unified CM software.



Note

Once the upgrade process starts on the inactive partition, any changes made in the active partition of the publisher's database will not be migrated to the new version of the database.

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.

**Note**

Before you do an upgrade, Cisco recommends that you back up the Unified CM and Call Detail Record (CDR) database to an external network directory using the Disaster Recovery Framework. This practice will prevent any loss of data if the upgrade fails.

Call Processing Subscriber Redundancy

The following figures illustrate typical cluster configurations to provide call processing redundancy with Unified CM.

Figure 8-4 Basic Redundancy Schemes

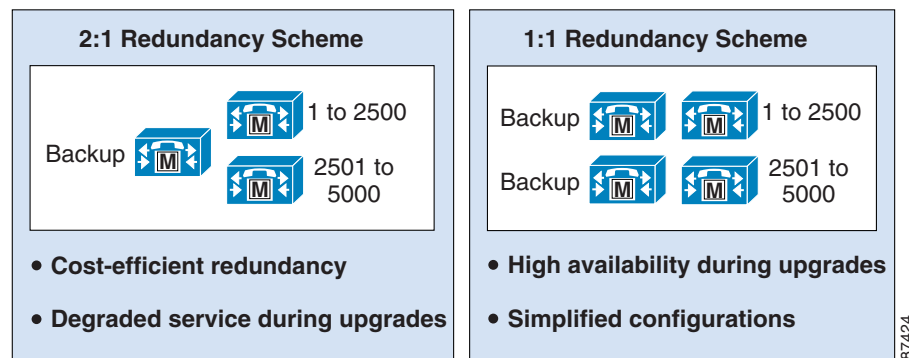
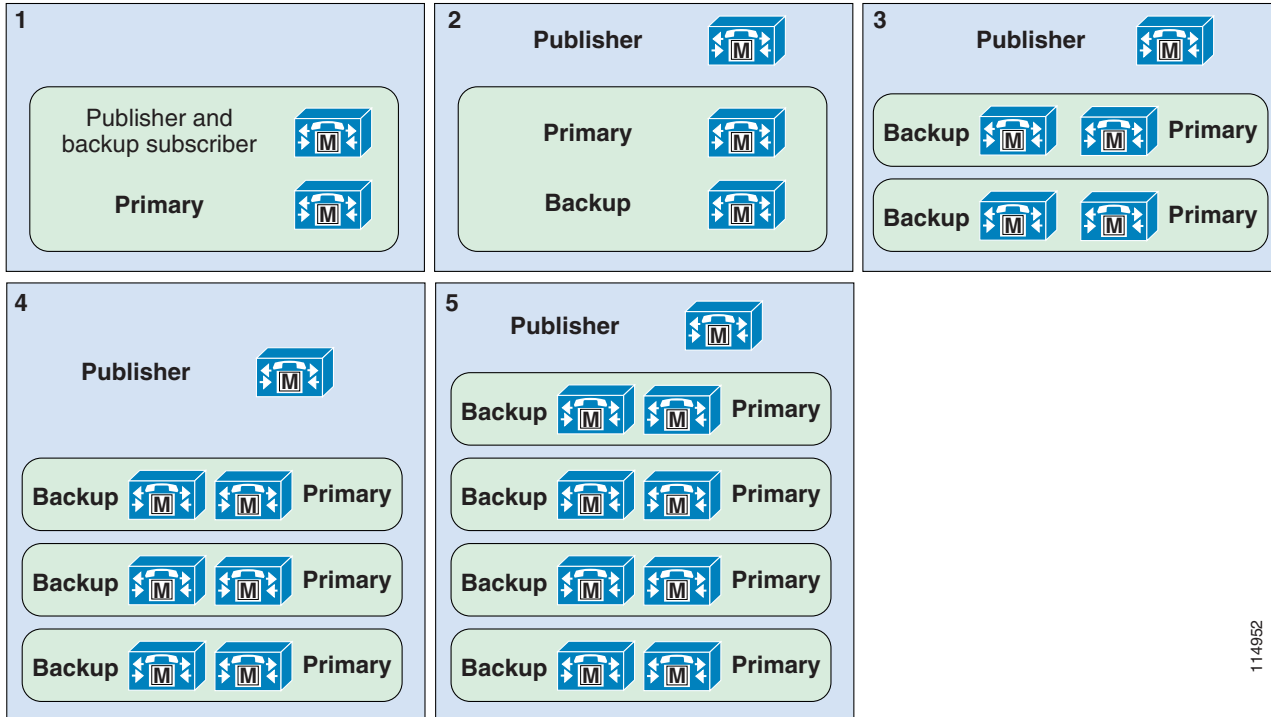


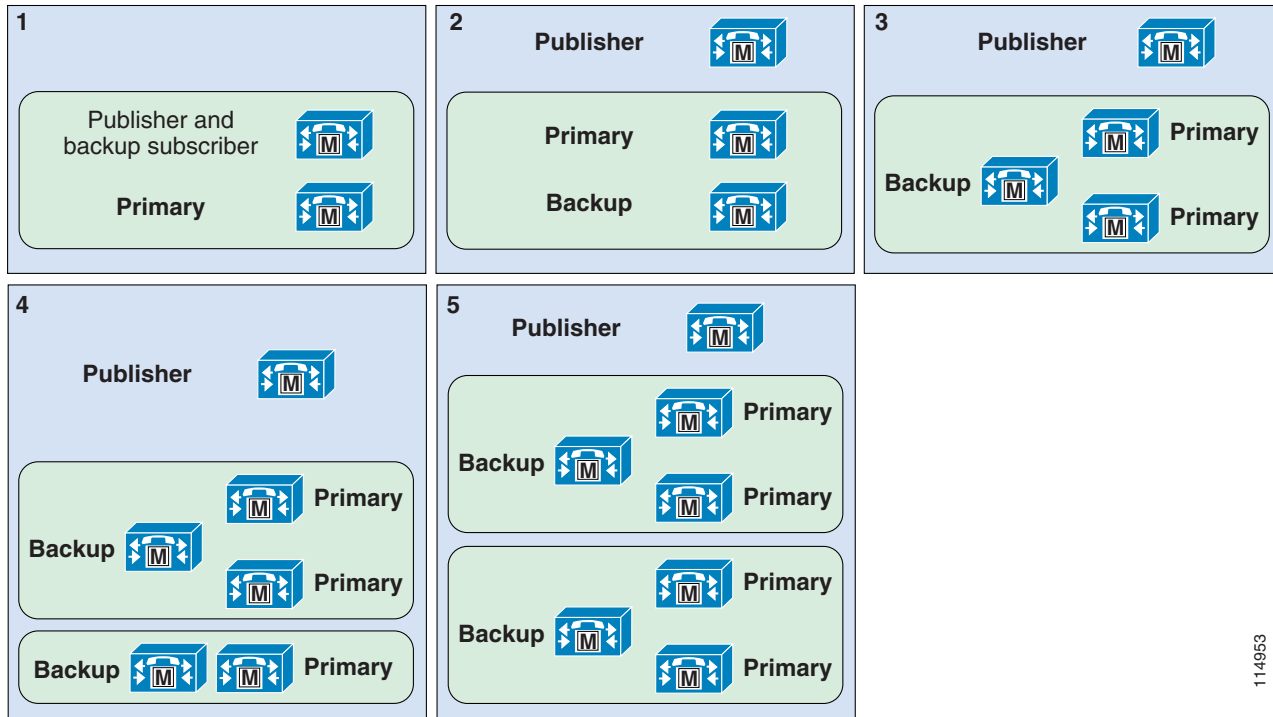
Figure 8-4 illustrates the two basic redundancy schemes available. In each case the backup server must be capable of handling the capacity of at least a single primary call processing server failure. In the 2:1 redundancy scheme, the backup might have to be capable of handling the failure of a single 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 choosing the hardware platforms is covered in the section on [Unified CM Platform Capacity Planning, page 8-15](#).

Figure 8-5 1:1 Redundancy Configuration Options

114952

In [Figure 8-5](#) 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 8-6 2:1 Redundancy Configuration Options

114953

Load Balancing

Normally a backup server has no devices registered to it unless its primary is unavailable. This model allows for:

- Easier troubleshooting — Because all call processing takes place on primary servers, obtaining traces and alert notifications becomes easier.
- Less configuration — Because all the devices are registered to the primary server, the need to define additional Unified CM redundancy groups or device pools for the various devices can be reduced by 50%.

The 1:1 redundancy scheme enables you to balance distribution of the devices over the primary and backup server pairs. With load balancing, you can move up to half of the device load from the primary to the secondary subscriber by using the Unified CM redundancy groups and device pool settings. This model allows for:

- Load sharing — The call processing load is distributed on multiple servers, which can provide faster response time.
- Faster failover and failback — Because all devices (such as IP phones, CTI ports, gateways, trunks, voicemail ports, and so forth) are distributed across all active subscribers, only some of the devices fail over to the secondary subscriber if the primary subscriber fails. 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 server, the total load on the primary and secondary subscribers should not exceed that of a single subscriber server.

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. There is no restriction on the number of servers that can have TFTP service enabled, however Cisco recommends deploying 2 TFTP servers for a large cluster, thus providing redundancy for TFTP service. More than 2 TFTP servers can be deployed in a cluster, but this can result in an extended period for rebuilding of all TFTP files on all TFTP servers. When configuring the TFTP options using DHCP or statically, you can normally define an IP address array (more than one IP address) for 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 Unified CM cluster, Cisco highly recommends that you upgrade the TFTP servers after the publisher and before any other server, also allowing additional time following the upgrade for the TFTP server to rebuild all the configuration files. Either use the typical Cisco TFTP - BuildDuration time or use the real-time monitoring tool to monitor the Cisco TFTP - DeviceBuildCount until it stops incrementing. 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 manually adding a specific binary or firmware load for a phone or gateway, be sure to copy the file to each TFTP server in the cluster.

Cisco Unified CM Release 5.0, by default, caches the configuration files in memory and does not store them on the hard drive of the TFTP server. This default setting can be changed to place the 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 in 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. Cisco recommends that you load-balance CTI applications across the various CTI Managers in the cluster to provide maximum resilience, performance, and redundancy.

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 Unified CM redundancy group of server A as the primary and server B as the backup (or secondary). The other two ports would have a Unified CM 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 Unified CM server failure.

IP Voice Media Streaming Application

Media resources such as conferencing and music on hold may be provided by IP Voice Media Streaming Application 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 6-18.](#))
- Conference bridges — Provide software-based conferencing for ad-hoc and meet-me conferences. (See [Audio 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 6-13.](#))

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 a dedicated media resource server is recommended for large-scale software-based conferencing and MTP unless those services are within the design guidelines 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 Unified CM service parameters, and you should disable it on H.323 and SIP dial peers by using the **no vad** command.

Unified CM Applications

Various types of applications can be enabled on Unified CM. This section covers only the scalability aspect of Unified CM applications. For detailed design guidance, see the chapter on [Cisco Unified CM Applications, page 22-1](#).

The Unified CM applications include the following:

Cisco Unified Communications Manager Assistant

The Cisco Unified CM Assistant application operates in conjunction with the CTI Manager Service. When Cisco Unified CM Assistant is used in a cluster, the expected capacity and performance can impact the choice of Unified CM and CTI subscriber hardware platform. If other CTI applications are used in a cluster, the number of CTI connections supported in a cluster might limit the maximum Unified CM Assistant configuration.

Currently supported limits for Unified CM Assistant in Cisco Unified CM Release 5.0 are as follows:

- Maximum of two Unified CM Assistant servers in a cluster
- Maximum of four CTI servers under Unified CM Assistant service parameters
- 1250 Assistants and 1250 Managers per cluster with Cisco MCS 7845 servers

For additional information on Unified CM Assistant capacity, refer to the Cisco Unified CM and Unified CM Assistant data sheets, documentation, and release notes available at

<http://www.cisco.com>

Cisco 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 activity takes place on the publisher.

Currently supported limits for Extension Mobility in Cisco Unified CM Release 5.0 are as follows:

- A Cisco MCS-7845 publisher server can support 50 sequential logins and/or logouts per minute.
- A Cisco MCS-7835 publisher server can support 30 sequential logins and/or logouts per minute.

For additional information on EM capacity, refer to the Cisco Unified CM data sheets, documentation, and release notes available at

<http://www.cisco.com>

Cisco Unified Communications Manager Attendant Console (AC)

The Cisco AC application interacts with the CTI Manager Service for line monitoring and phone control. When Cisco AC is used in a cluster, the expected capacity and performance can impact the choice of Unified CM and CTI subscriber hardware platform. If other CTI applications are used in a cluster, the number of CTI connections supported in a cluster might limit the maximum AC configuration.

Currently supported limits for AC in Cisco Unified CM Release 5.0 are as follows:

- An MCS 7845 can support a maximum of 1250 AC devices. The Attendant Console application can be used in any combination of up to 1250 attendant console devices. For example, you may configure Unified CM to have 125 hunt pilots with 10 members in each hunt pilot or 50 hunt pilots with 25 members in each hunt pilot.
- An MCS 7835 supports a maximum of 1000 AC devices, and an MCS 7825 supports a maximum of 750 AC devices.

Unified CM Platform Capacity Planning

Many types of devices can register with Unified CM; 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 Unified CM Capacity Tool, 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 co-resident 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, co-resident services, and other co-resident applications. All of these functions can consume additional resources within the Unified CM server.

To improve system 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.
- A Unified CM 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 can be modified to allow additional time for the configuration to initialize. For details on the service parameters, refer to the online help for Service Parameters in Unified CM Administration.

The following guidelines apply to Cisco Unified CM Release 5.0:

- 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 a maximum of 30,000 SCCP or SIP phones.
- Each cluster can support a maximum of 600 H.323 devices (gateways, trunks, and clients), digital MGCP devices, and SIP trunks

Capacity Calculations

The Cisco Unified Communications Manager Capacity Tool (Unified CMCT) for Cisco Unified CM Release 5.0 enables you to calculate the capacity of the system for various configurations. The capacity planning tool is currently available to all www.cisco.com users with a login account. 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 but you do not have access to the Cisco Unified CM Capacity Tool, please contact your Cisco Systems Engineer (SE) or the Cisco Technical Assistance Center (TAC).

The Cisco Unified CM Capacity Tool is available at

<http://www.cisco.com/partner/WWChannels/technologies/resources/CallManager/>

If your system meets the following requirements, you should not need to verify your configuration with the Cisco Unified CM 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 4 calls per hour
- Cluster security is not enabled
- 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 Unified CM 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 supported models)	7500	Yes	Yes
Cisco MCS-7835 (All supported models)	2500	Yes	No
Cisco MCS-7825 (All supported models)	1000	No	No
Cisco MCS-7815 (All supported models) ³	500 ⁴	No	No

1. A platform that is not a high-availability server can support a maximum of 500 IP Phones in a non-redundant installation.

2. A high-availability server supports redundancy for both the power supplies and the hard disks.

3. MCS-7815 servers support only N+1 redundancy and may not be a member of a cluster.

4. With Cisco Unified CM 5.0, the MCS 7815 server supports a maximum of 300 users.

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 Unified CM Capacity Tool

The Cisco Unified CM 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.

When you use the Cisco Unified CM Capacity Tool, the results will indicate the highest capacity of the various resources calculated for the configuration. The resources include hard limits, memory, processor utilization, and disk I/O activity. If you add more devices, it might look as if no more capacity is being used because the additional configuration is using resources that are below the current highest capacity indicated.

For example, if 2500 third-party controlled SCCP IP phones were added to the Cisco Unified CM Capacity Tool, it would show that a single MCS 7845 subscriber is required and is at 100% capacity. If an additional 1000 phones are added, the results still indicate a single MCS 7845 subscriber is required and is at 100% capacity. This result is due to the fact that CTI capacity for this server is at 100%, and additional IP phones do not add CTI requirements. Until the number of other devices exceeds one of the other limits, it will appear as if no additional capacity is being used. This is normal and expected behavior for the Cisco Unified CM Capacity Tool.

Use the guidelines in the following sections to provide information for the capacity calculator.

Phone Calculations

The main Telephony section (not the Contact Center section) of the Capacity Tool lists the following phone types:

- SCCP Phones (Non-secure)
- Secure SCCP Phones
- SIP Phones (non-secure)
- Secure SIP Phones

For each phone type, you can enter quantity, BHCA, utilization, and line appearance values, which are further explained in the following sections.

Quantity

This value is the total number of IP phones of each type that will be configured in the cluster. The quantity includes all Cisco 7900 Series IP Phones, VG248 ports, VG224, 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 of each type. If you have shared lines across multiple phones, the BHCA should include one call for each and every phone that shares that line. Shared lines across multiple phone types will affect the BHCA for each type. (That is, one call to a shared line will be calculated as multiple calls, one to each phone that rings.) 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 Unified CM Capacity Tool.

For example, assume there are two classes of user with the following characteristics:

- 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 to be 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 Unified CM 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. For example, if it is normal for a phone to have two calls, one active and the other on hold for various amounts of time, then the utilization will be higher for that device because two calls are actually connected.

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 Unified CM 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 Unified CM database. Generally, an entire module (WS-6624 or Unified CM) 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 Unified CM but are used via H.323 trunks.

**Note**

SIP gateways are added as 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 Unified CM database. A gatekeeper controlled trunk is not affected by the possible number of destinations and, therefore, 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. Each SIP gateway that is not connected via a SIP Proxy will require a SIP trunk to be defined in Unified CM.

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 or SIP trunk, then an MTP resource is required for every concurrent call on that trunk. RSVP agents should be included in the quantity and will also be based on the maximum number of sessions supported.

CTI Route Points, Ports, and Third-Party Controlled Lines

CTI route point calculations include the total quantity, average BHCA, and utilization values. Any assigned directory numbers should be added to the Dial Plan section of the tool.

CTI ports can be used in a variety of ways by the applications controlling them. For simplicity they have been divided into two groups:

- Simple call or redirect call

A simple call is a call either to or from the port to another device, with no supplementary service being invoked. It is the typical call scenario where one party calls another, the called party answers, the call proceeds, and the parties hang up.

A redirect is a call to a CTI port that is not actually answered. Instead, it is forwarded to another destination or redirected. This is even simpler than a simple call because no call is connected, so no media is connected to the CTI port.

- Transfer or conference

These call types are typical of an IVR or auto-attendant application.

A transfer type call is characterized as a call that is connected to a CTI port, then at some point the CTI port transfers the caller without consultation (or blind transfer) to another destination.

A conference is a variation of the transfer, whereby a call connected to the CTI port is either transferred with consultation (three-way call) or is conferenced with another party for the duration of the call.

Third-Party Controlled Lines

Each application that requires third-party control or monitoring of a device or line must be counted and entered in this field of the calculator. If more than one application is monitoring or controlling the same device, it will need to be counted more than once.

Gatekeeper Design Considerations

A single Cisco IOS gatekeeper can provide call routing and call admission control for up to 100 Unified CM clusters in a distributed call processing environment. Multiple gatekeepers can be configured to support thousands of Unified CM clusters. You can also implement a hybrid Unified CM and toll-bypass network by using Cisco IOS gatekeepers to provide communication and call admission control between the H.323 gateways and Unified CM.

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 Unified CM Release 3.3, the only method for providing gatekeeper redundancy was Hot Standby Router Protocol (HSRP); however, beginning with Cisco Unified CM 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 Unified CM 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 Unified CM environments prior to Release 3.3.

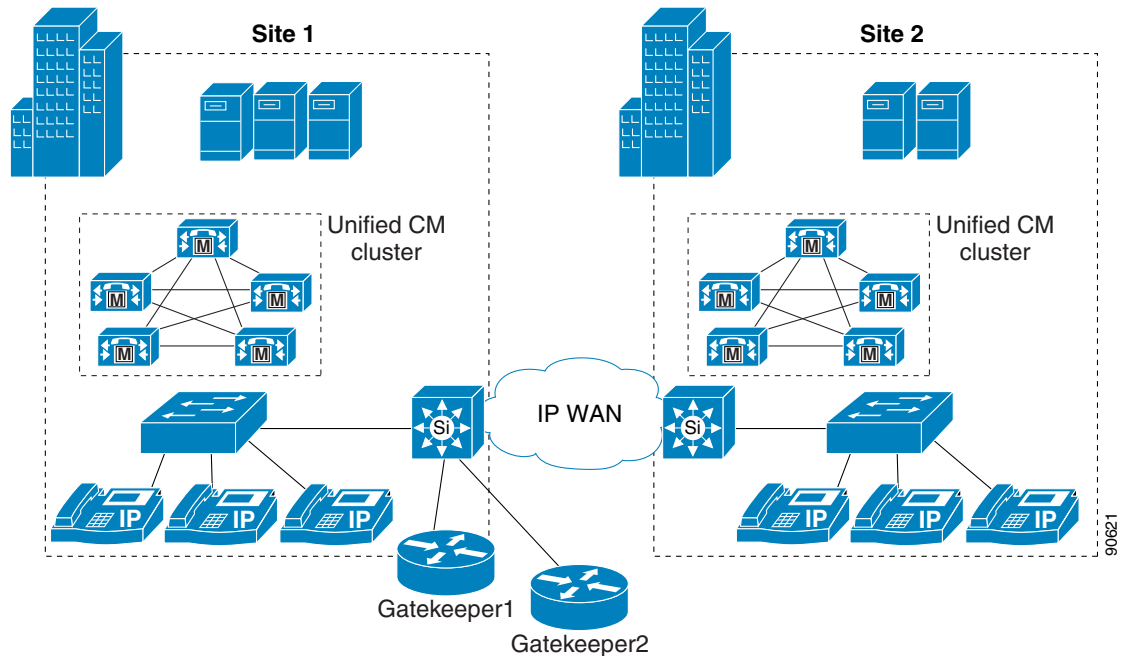
The following guidelines apply to HSRP:

- 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 8-7 show a network configuration using HSRP for gatekeeper redundancy.

Figure 8-7 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 8-7. 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 Unified CM cluster has a local zone configured on each router to support Unified CM 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** 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 Unified CM 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 Unified CM trunks.

For additional and advanced HSRP information, refer to the online documentation at the following locations:

- <http://www.cisco.com/en/US/docs/internetworking/case/studies/cs009.html>
- http://www.cisco.com/en/US/tech/tk648/tk362/technologies_q_and_a_item09186a00800a9679.shtml
- http://www.cisco.com/en/US/tech/tk648/tk362/technologies_tech_note09186a0080094afd.shtml

Gatekeeper Clustering (Alternate Gatekeeper)

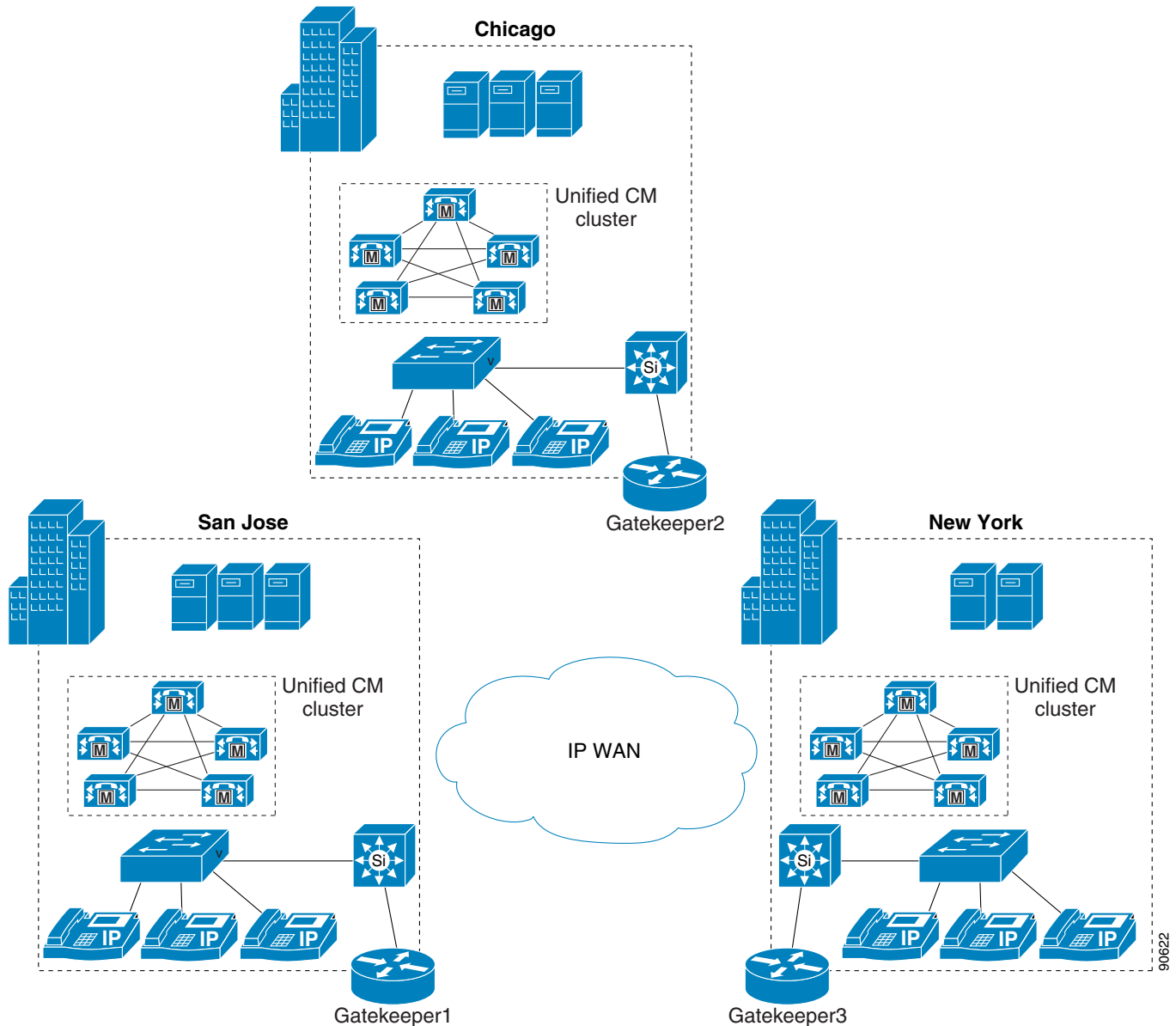
Gatekeeper clustering (alternate gatekeeper) enables the configuration of a "local" gatekeeper cluster, with each gatekeeper acting as primary for some Unified CM trunks and an alternate for others. Gatekeeper Update Protocol (GUP) is used to exchange state information between gatekeepers in a local 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 Unified CM 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 8-8 shows three sites with Unified CM distributed call processing and three distributed gatekeepers configured in a local cluster.

Figure 8-8 Gatekeeper Clustering

In [Figure 8-8](#), 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 Unified CM cluster has a local zone configured to support Unified CM 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** 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 Unified CM 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 Unified CM trunks.

All gatekeepers in the cluster display all Unified CM 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  RASSignalAddr  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 8-9 **Redundant Directory Gatekeepers**



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 8-9](#).

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 Unified CM cluster has a local zone configured to support Unified CM 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** 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 Unified CM trunks have been configured to register with a 1# prefix.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Unified CM trunks.

Interoperability of Unified CM and Unified CM Express

This section explains the requirements for interoperability and internetworking of Cisco Unified CM with Cisco Unified Communications Manager Express (Unified CME, formerly called Cisco IOS Telephony Services, or ITS) using H.323 or SIP protocol in a multisite IP telephony deployment. This section highlights the recommended deployments between phones controlled by Unified CM and phones controlled by Unified CME.

Cisco Unified CME 3.4 adds the ability to configure Cisco IP SIP Phones 7905, 7912, 7940, and 7960 in addition to the currently supported Cisco IP SCCP Phones 7902, 7905, 7910, 7912, 7920, 7935, 7936, 7940, 7960, 7970, 7971, and Cisco IP Communicator. If you use Unified CME with SIP phones, the WAN interface must be SIP. Unified CME SCCP phones are supported with either an H.323 or SIP WAN interface.

All call signaling is sent through Unified CME, regardless of the endpoint being used. However, with SCCP endpoints on the same Unified CME, the media can flow around Unified CME; whereas with SIP endpoints on the same Unified CME, media always flows through Unified CME.

The following sections provide guidelines for achieving interoperability between Unified CM and Unified CME:

- [Multisite IP Telephony Deployments Using Unified CM and Unified CME with SIP Trunks](#), page 8-31
- [Multisite IP Telephony Deployments Using Unified CM and Unified CME, with H.323 Trunks and an IP-to-IP Gateway](#), page 8-33

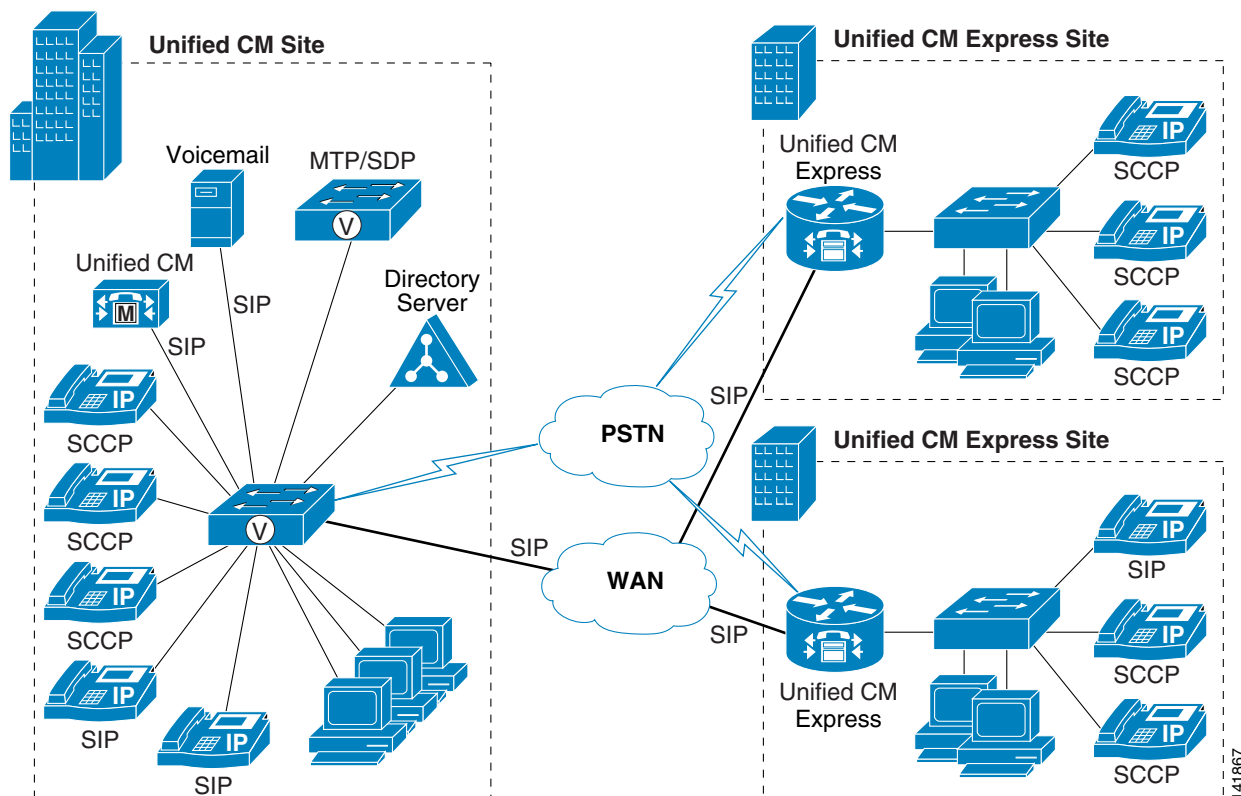
For more information about Unified CME, refer to the product documentation for Cisco Unified CME, available online at

<http://www.cisco.com>

Multisite IP Telephony Deployments Using Unified CM and Unified CME with SIP Trunks

Unified CM can communicate directly with Unified CME using a SIP interface. [Figure 8-10](#) shows an IP Telephony deployment with Unified CM networked directly with Cisco Unified CME using a SIP trunk WAN interface.

Figure 8-10 Multisite IP Telephony Deployment Using Unified CM and Unified CME with SIP Trunks



Best Practices

Follow these guidelines and best practices when using the deployment model illustrated in [Figure 8-10](#):

- Configure a SIP Trunk Security Profile with **Replaces** header acceptance.
- Configure a SIP trunk on Unified CM using the SIP Trunk Security Profile created, and also specify a ReRouting CSS. The ReRouting CSS is used to determine where a SIP user (transferor) can refer another user (transferee) to a third user (transfer target) and which features a SIP user can invoke using the SIP 3XX Redirection Response and INVITE with Replaces.
- For SIP trunks there is no need to enable the use of media termination points (MTPs) when using SCCP endpoints on Unified CME. However, SIP endpoints on Unified CME require the use of media termination points on Unified CM to be able to handle delayed offer/answer exchanges with the SIP protocol (that is, the reception of INVITEs with no Session Description Protocol).
- Use the Unified CM dial plan configuration (route patterns, route lists, and route groups) to send calls to the SIP trunk that connects to Unified CME.
- Use Unified CM device pools and regions to configure a G.711 codec within the site and the G.729 codec for remote Unified CME sites.
- Configure the **allow-connections sip to sip** command under **voice services voip** on Unified CME to allow SIP-to-SIP call connections.
- For SIP endpoints, configure the **mode cme** command under **voice register global**, and configure **dtmf-relay rtp-nte** under the **voice register pool** commands for each SIP phone on Unified CME.
- For SCCP endpoints, configure the **transfer-system full-consult** command and the **transfer-pattern .T** command under **telephony-service** on Unified CME.
- Configure the SIP WAN interface voip dial-peers to forward calls, destined for Unified CM, with **session protocol sipv2** and **dtmf-relay sip-notify rtp-nte** on Unified CME.



Note

When multiple PSTN connections exist (one for Unified CM and one for Unified CME), fully attended transfer between a Unified CM endpoint and a Unified CME endpoint to a PSTN endpoint will fail. The recommendation is to use blind transfer when using multiple PSTN connections, and it is configured under **telephony-service** as **transfer-system full-blind**.

[Example 8-11](#) lists a sample configuration for Unified CME using this SIP deployment model.

Example 8-11 Configuration for Cisco Unified CME 3.4 with SIP

```
voice service voip
  allow-connections sip to sip
  sip
  registrar server
dial-peer voice 1 voip      /* To Unified CM endpoints */
  destination-pattern xxxx
  session protocol sipv2
  session target ipv4:10.10.10.20
  session transport udp    /* tcp can be used here also */
  dtmf-relay rtp-nte
  codec g729r8             /* Voice class can also be used */
  no vad
voice register global
  mode cme
  source-address 10.10.10.21 port 5060
voice register pool 1
  id mac 0007.0E8B.5777
  type 7940
```

```
number 1 dn 1
  codec g729r8      /* Voice class can also be used */
  dtmf-relay rtp-nte
telephony-service
  ip source-address 10.10.10.22 port 2000
  create cnf-files
  keepalive 45
  max-conferences 8 gain -6
  moh music-on-hold.au
  transfer-system full-consult /* full-blind can also be used */
  transfer-pattern .T
```

Using the guidelines presented here ensures that basic calls can be completed between Unified CM and Unified CME phones. By using SIP trunk interfaces instead of H323, the need for MTPs is eliminated if you are using only SCCP endpoints on Unified CME. If SIP endpoints are used on Unified CME, the configuration of an MTP on Unified CM is required.

Multisite IP Telephony Deployments Using Unified CM and Unified CME, with H.323 Trunks and an IP-to-IP Gateway

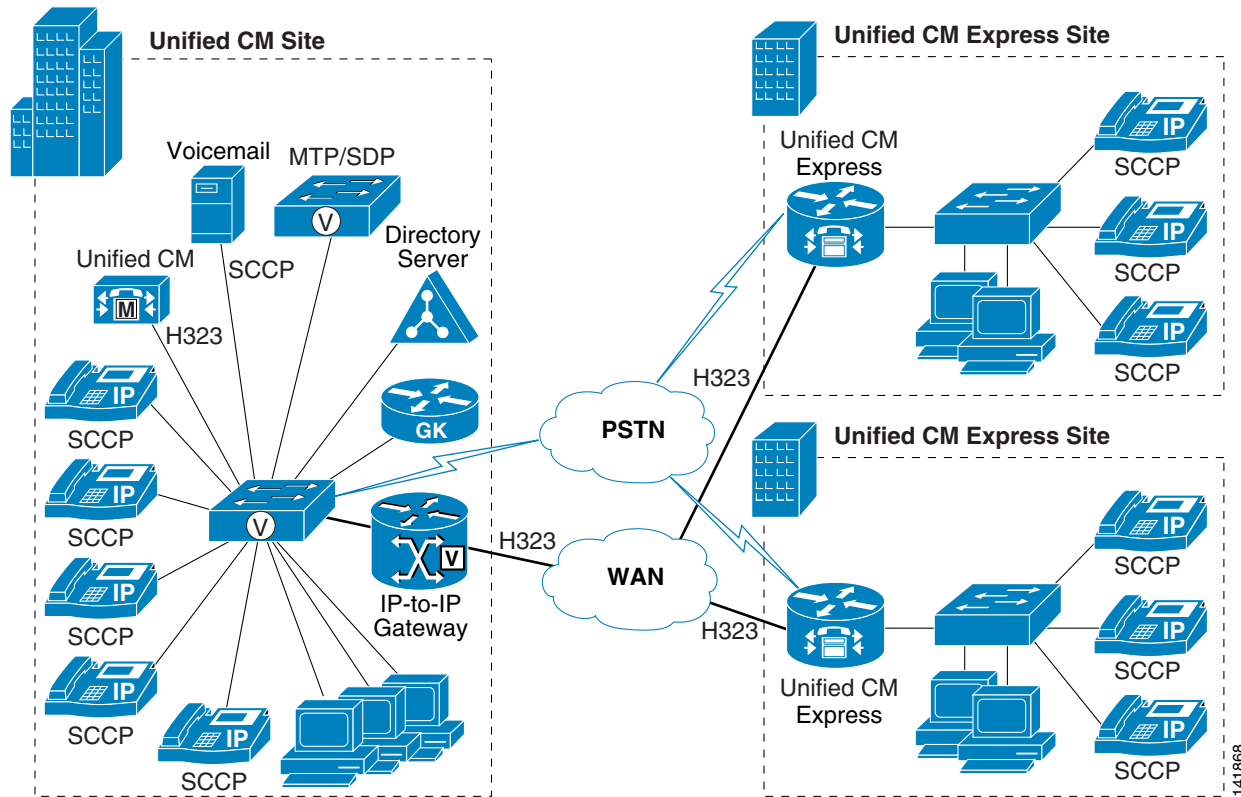
An IP-to-IP gateway is a separate router that provides a proxy (or front end) for a system that does not support H.450, such as Unified CM. The IP-to-IP gateway can be placed between Unified CM and a Unified CME router, and it provides H.323-to-H.323 call connections to terminate and re-originate calls for transfers and forwards to endpoints that do not support H.450.

Unified CME can be directly integrated with Unified CM using an H.323 interface without the IP-to-IP gateway. However, because Unified CM does not support H.450 specifications, supplementary services such call transfer and call forward initiated from Unified CM phones will require the media to be hair-pinned (over the WAN in some cases) through Unified CM even though the Unified CM phones might not be involved in the call. The IP-to-IP gateway ensure that this media hair-pinning does not occur over the WAN.

The IP-to-IP gateway can also act as the PSTN gateway for the remote Cisco Unified CME systems as well as for Unified CM. A separate PSTN gateway is not required in such cases.

The IP-to-IP gateway must run a Cisco IOS release that is compliant with Cisco Unified CME 3.4 and that supports H.450, such as Cisco IOS Release 12.4(6)T or later with the IP VOICE feature set.

[Figure 8-11](#) shows an IP Telephony deployment with Unified CM connected to Cisco Unified CME through an IP-to-IP gateway.

Figure 8-11 Multisite IP Telephony Deployment with Unified CM, Unified CME, and an IP-to-IP Gateway

Best Practices

Follow these guidelines and best practices when using the deployment model illustrated in [Figure 8-11](#):

- Configure a gatekeeper controlled H.225 trunk between Unified CM and the IP-to-IP gateway, with **Media Termination Point Required** checked and **Wait For Far End H.245 Terminal Capability Set** unchecked.
- In Unified CM, set the service parameter **Send H225 user info message** to **H225 info for Call Progress Tone**.
- Use the Unified CM dial plan configuration (route patterns, route lists, and route groups) to send calls to the H.225 trunk that connects to the IP-to-IP gateway.
- Register Cisco Unified CME and the IP-to-IP gateway as H.323 gateways on the gatekeeper.
- A media termination point (MTP) is required for H.450 to perform Empty Capability Set (ECS) signaling conversion. The MTP should be configured on the IP-to-IP gateway and registered to Unified CM. An MTP can cause media to be hair-pinned over the WAN in certain rare cases. For more details on media termination points, see the chapter on [Media Resources](#), [page 6-1](#).
- Configure the **allow-connection h323 to h323** command on the IP-to-IP gateway to allow H.323-to-H.323 call connections. Remote Unified CME routers need not have this command enabled.

**Note**

When using H.323 trunks between Unified CME and Unified CM, configure the **allow-connection h323 to h323** command on Unified CME to allow H.323-to-H.323 hair-pinned call connections. The Unified CME auto-detection feature (starting with Unified CME 3.1) disables all H.450-based communications on the interface connected to Unified CM. To complete the call transfer or forward between Unified CM and Unified CME phones, an H.323-to-H.323 connection is required.

- Define VoIP dial peers on the IP-to-IP gateway to route calls to the Unified CM and Unified CME endpoints.
- Define VoIP dial peers on Unified CME to forward calls, destined for Unified CM endpoints, to the IP-to-IP gateway.
- Supplementary services such as transfer and call forward will result in calls being media hair-pinned when the two endpoints reside in the same Unified CME branch location.

**Note**

When multiple PSTN connections exist (one for Unified CM and one for Unified CME), fully attended transfer between a Unified CM endpoint and a Unified CME endpoint to a PSTN endpoint will fail. The recommendation is to use blind transfer when using multiple PSTN connections, and it is configured under **telephony-service** as **transfer-system full-blind**.

[Example 8-12](#) lists a sample configuration for the IP-to-IP gateway.

Example 8-12 Configuration for IP-to-IP Gateway

```
voice service voip
  allow-connections h323 to h323
  supplementary-service h450.2
  supplementary-service h450.3
  supplementary-service h450.12
  h323
  emptycapability
  h225 id-passthru
  h225 connect-passthru
  h245 passthru tcsnonstd-passthru
dial-peer voice 1 voip          /* To Unified CM 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 Unified CM 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
```

[Example 8-13](#) lists a sample configuration for the in this H.323 deployment model.

Example 8-13 Configuration for Cisco Unified CME 3.4 with H.323

```
voice service voip
  h323
interface FastEthernet0/1
  ip address 10.10.10.23 255.255.255.0
```

```
h323-gateway voip interface
h323-gateway voip id cme ipaddr 10.10.10.30 1719
dial-peer voice 1 voip          /* To Unified CM endpoints */
destination-pattern xxxx
session target ras
session transport tcp
codec g729r8                   /* Voice class can also be used */
no vad
telephony-service
ip source-address 10.10.10.22 port 2000
create cnf-files
keepalive 45
max-conferences 8 gain -6
moh music-on-hold.au
transfer-system full-blind /* Used with multiple PSTN connections */
transfer-pattern .T
```



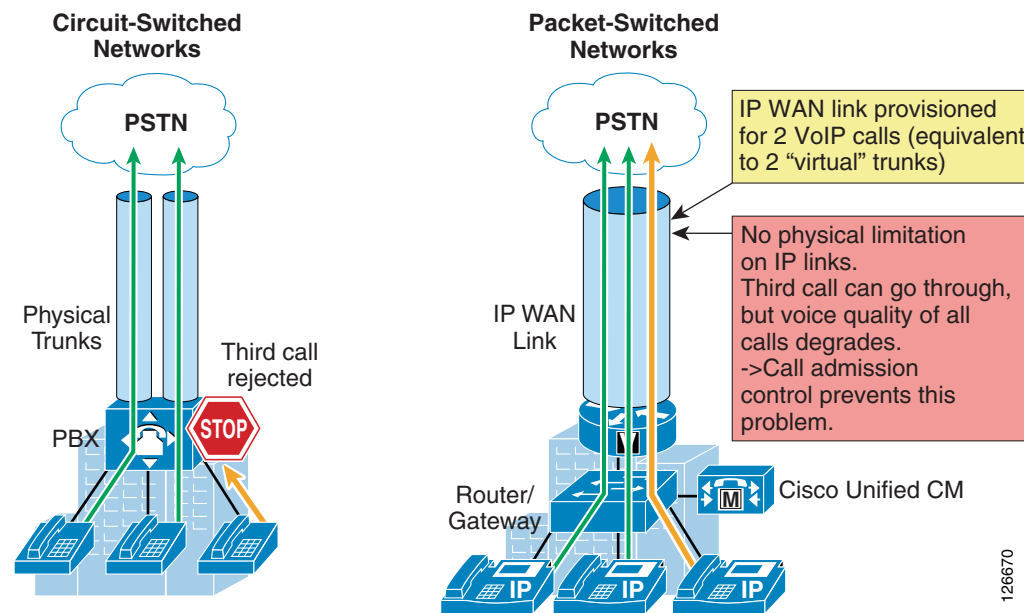
CHAPTER 9

Call Admission Control

Last revised on: September 27, 2007

The call admission control function is an essential component of any IP telephony 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 IP telephony 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 telephony 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 IP telephony 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 it is essential to guarantee good voice quality in a multisite deployment involving an IP WAN.

To preserve a satisfactory end-user experience, the call admission control function should always be performed during the call setup phase so that, if there are no network resources available, a message can be presented to the end-user or the call can be rerouted across a different network (such as the PSTN).

This chapter discusses the following main topics:

- [Best Practices Summary, page 9-2](#)

This section summarizes the key best practices, recommendations, and notes about call admission control for the readers who are already familiar with the principles and mechanisms described in the remainder of this chapter.

- [Call Admission Control Principles, page 9-3](#)

This section defines the two fundamental approaches to call admission control in an IP-based telephony system: topology-aware and topology-unaware call admission control.

- [Call Admission Control Elements, page 9-11](#)

This section describes the call admission control mechanisms available through the various components of a Cisco IP Communications system, such as Cisco Unified Communications Manager locations, Cisco IOS gatekeeper, RSVP, and the IP-to-IP gateway.

- [Call Admission Control Design, page 9-36](#)

This section shows how to apply and combine the mechanisms described in the previous sections, based on the IP WAN topology (simple hub-and-spoke, two-tier hub-and-spoke, MPLS, or other topologies) and also based on the Cisco Unified Communications Manager deployment model adopted.

Best Practices Summary

This section briefly summarizes the best practices for providing call admission control in various Cisco Unified Communications Manager (Unified CM) deployments. The remainder of this chapter explains these best practices in more detail.

The following recommendations apply to deployments with a single Unified CM cluster:

- For simple hub-and-spoke topologies with no dual links, use Unified CM static locations. Leave the hub site devices in the Hub_None location.
- For Multiprotocol Label Switching (MPLS) topologies with no dual links, use Unified CM static locations, with devices at every site (including the central site) assigned to a location.
- For any other topology, use Unified CM RSVP-enabled locations. Cisco recommends the **Mandatory** or **Mandatory (video desired)** policy as the default RSVP policy between sites. The Cisco RSVP Agent feature may reside on the IP WAN router in smaller sites or run on standalone platforms in larger sites.

The following recommendations apply to deployments with multiple Unified CM clusters:

- For simple hub-and-spoke topologies with no dual links, use Cisco IOS gatekeeper zones between sites where Unified CM clusters reside.

- For two-tier hub-and-spoke topologies with no dual links where Unified CM clusters are located at the first and second level hub sites, use Cisco IOS gatekeeper zones for the links between first- and second-level hub sites and use Unified CM static locations for the links between second-level hub sites and spoke sites.
- For MPLS topologies with no dual links, use Unified CM static locations, with every site in a location and with no gatekeeper zones. Leave intercluster trunks in the Hub_None location unless an MTP is required. You may use a gatekeeper for intercluster call routing, but it is not needed for call admission control.
- For any other topology and three or fewer clusters, use RSVP-enabled locations and the "remote agent" approach.
- For any other topology and more than three clusters, use RSVP-enabled locations within each cluster and gatekeepers with RSVP-enabled IP-to-IP gateways across clusters.

Call Admission Control Principles

As mentioned previously, call admission control is a function of the call processing agent in an IP-based telephony system, so in theory there could be as many call admission control mechanisms as there are IP-based telephony systems. However, most of the existing call admission control mechanisms fall into one of the following two main categories:

- Topology-unaware call admission control — Based on a static configuration within the call processing agent
- Topology-aware call admission control — Based on communication between the call processing agent and the network about the available resources

The remainder of this section first analyzes the principles of topology-unaware call admission control and its limitations, then it presents the principles of topology-aware call admission control.

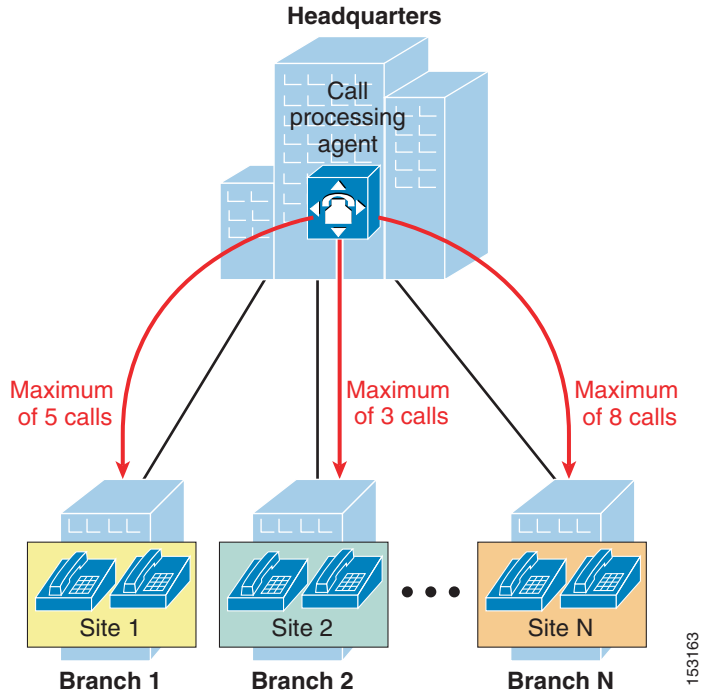
Topology-Unaware Call Admission Control

We define as topology-unaware call admission control any mechanism that is based on a static configuration within a call processing agent or IP-based PBX, aimed at limiting the number of simultaneous calls to or from a remote site connected via the IP WAN.

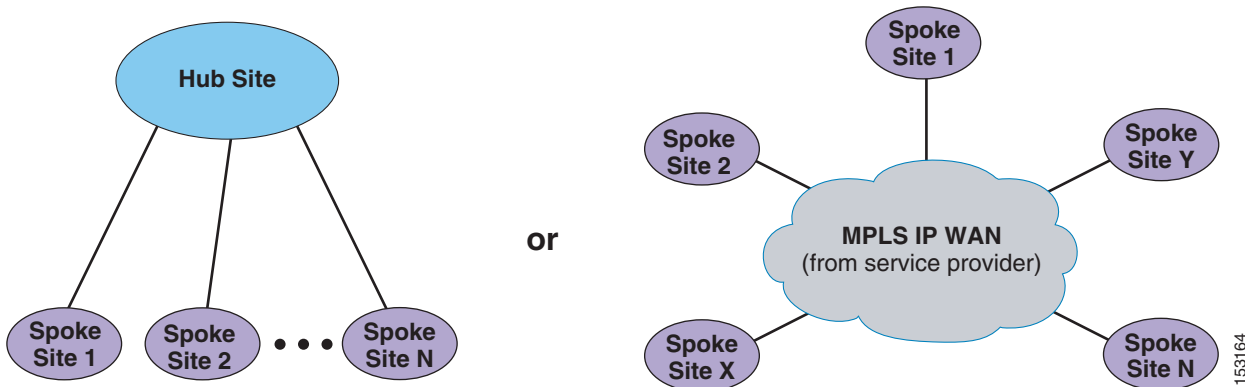
As shown in [Figure 9-2](#), most of these mechanisms rely on the definition of a logical "site" entity, which generally corresponds to a geographical branch office connected to the enterprise IP WAN.

After assigning all the devices located at each branch office to the corresponding site entity, the administrator usually configures a maximum number of calls (or a maximum amount of bandwidth) to be allowed in or out of that site.

Each time a new call needs to be established, the call processing agent checks the sites to which the originating and terminating endpoints belong, and verifies whether there are available resources to place the call (in terms of number of calls or amount of bandwidth for both sites involved). If the check succeeds, the call is established and the counters for both sites are decremented. If the check fails, the call processing agent can decide how to handle the call based on a pre-configured policy. For example, it could send a network-busy signal to the caller device, or it could attempt to reroute the call over a PSTN connection.

Figure 9-2 Principles of Topology-Unaware Call Admission Control

Because of their reliance on static configurations, topology-unaware call admission control mechanisms can generally be deployed only in networks with a relatively simple IP WAN topology. In fact, most of these mechanisms mandate a simple hub-and-spoke topology or a simple MPLS-based topology (where the MPLS service is provided by a service provider), as shown in Figure 9-3.

Figure 9-3 Domain of Applicability of Topology-Unaware Call Admission Control

In a hub-and-spoke network or MPLS-based network such as those shown in Figure 9-3, each spoke site is assigned to a "site" within the call processing agent, and the number of calls or amount of bandwidth for that "site" is configured to match the bandwidth available for voice (and/or video) on the IP WAN link that connects the spoke to the IP WAN.

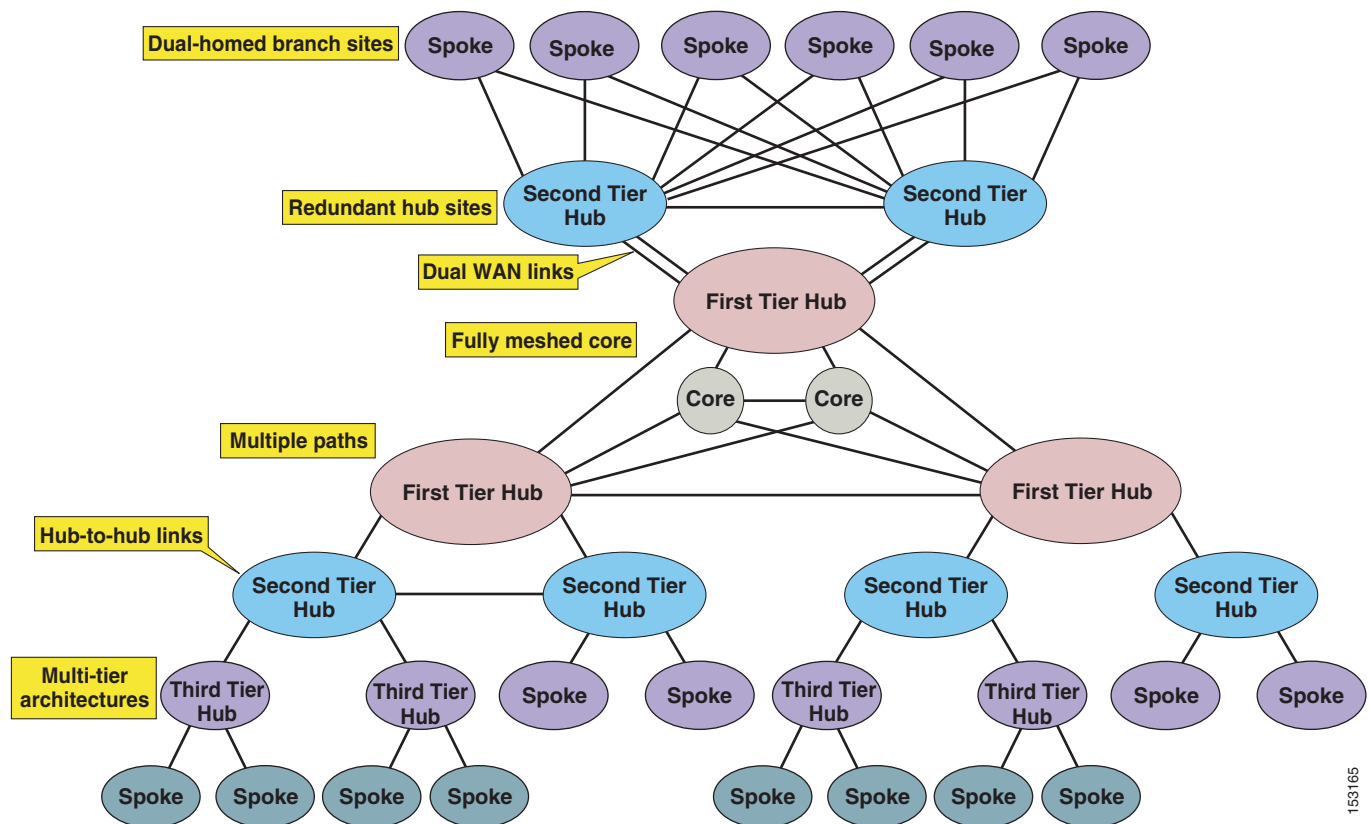
Notice the absence of redundant links from the spoke sites to the hub site and of links directly connecting two spoke sites. The next section explains why such links create problems for topology-unaware call admission control.

Limitations of Topology-Unaware Call Admission Control

In today's enterprise networks, high availability is a common requirement, and it often translates into a desire to provide redundancy for the IP WAN network connectivity.

When considering the IP WAN topology in a typical enterprise network, you are likely to encounter a number of characteristics that complicate the assumption of a pure hub-and-spoke topology. Figure 9-4 shows several of these network characteristics in a single diagram. Obviously, only the largest enterprise networks present all these characteristics at once, but it is highly likely that most IP WAN networks feature at least one of them.

Figure 9-4 Topology Characteristics of Typical Enterprise Networks

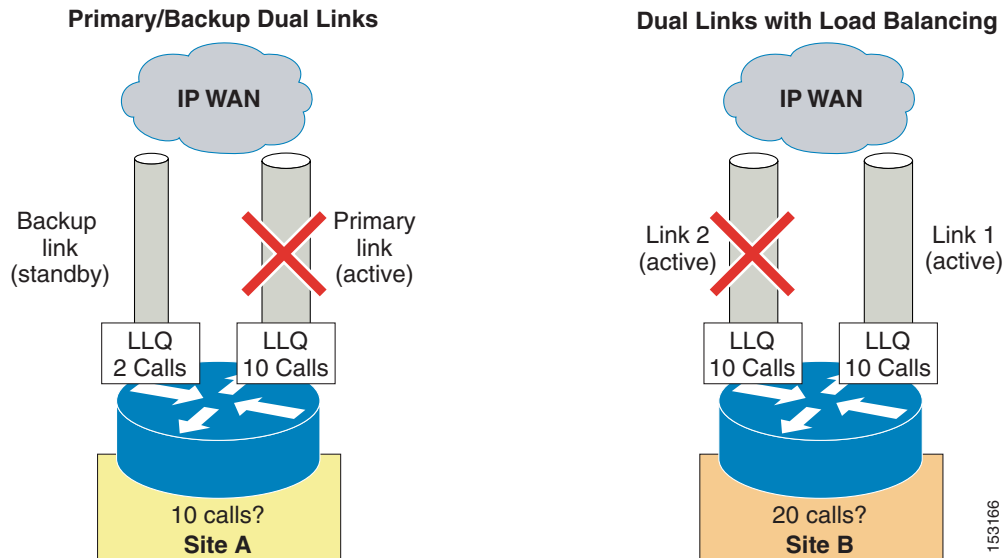


As explained in the section on [Call Admission Control Design, page 9-36](#), it is sometimes possible to adapt a topology-unaware call admission control mechanism to a complex network topology, but there are limitations in terms of when this approach can be used and what behavior can be achieved. For example, consider the simple case of a branch site connected to a hub site via the IP WAN, where redundancy is a network requirement. Typically, redundancy can be achieved in one of the following ways:

- A single router with a primary and a backup link to the IP WAN
- A single router with two active WAN links in a load-balancing configuration
- Two router platforms, each connected to the IP WAN, with load-balanced routing across them

The examples [Figure 9-5](#) attempt to apply a topology-unaware call admission control mechanism to the case of a single router with a primary and backup link and the case of a single router with two active load-balanced links. (The case of two router platforms has the same call admission control implications as the latter example.)

Figure 9-5 *Topology-Unaware Call Admission Control in Presence of Dual Links*



For the first example in [Figure 9-5](#), branch office A is normally connected to the IP WAN via a primary link, whose Low Latency Queuing (LLQ) bandwidth is provisioned to allow a maximum of 10 simultaneous calls. When this primary link fails, a smaller backup link becomes active and preserves the connectivity to the IP WAN. However, the LLQ bandwidth of this backup link is provisioned to allow only up to 2 simultaneous calls.

In order to deploy a topology-unaware call admission control mechanism for this branch office, we must define a "site" A in the call processing agent and configure it for a certain number of calls (or amount of bandwidth). If we choose to use 10 calls as the maximum for site A, the backup link can be overrun during failures of the primary link, thereby causing bad voice quality for all active calls. If, on the other hand, we choose 2 calls as the maximum, we will not be able to use the bandwidth provisioned for the remaining 8 calls when the primary link is active.

Now consider branch office B, which has two active links connecting it to the IP WAN. Each of these links is provisioned to allow a maximum of 10 simultaneous calls, and the routing protocol automatically performs load-balancing between them. When deploying a topology-unaware call admission control mechanism for this branch office, we must define a "site" B in the call processing agent and configure it for a certain number of calls (or amount of bandwidth). Similar to the case of branch office A, if we choose to add up the capacity of the two links and use 20 calls as the maximum for site B, there is a potential to overrun the LLQ on one of the two links during failures of the other one. For example, if link #2 fails, the system still allows 20 simultaneous calls to and from site B, which are now all routed via link #1, thus overrunning it and causing poor voice quality for all calls. On the other hand, if site B is configured for a maximum of 10 simultaneous calls, the available LLQ bandwidth is never fully utilized under normal conditions (when both links are operational).

These two simple examples show how IP WAN bandwidth provisioning in real enterprise networks is often too complex to be summarized in statically configured entries within the call processing agent. Deploying topology-unaware call admission control in such networks forces the administrator to make assumptions, develop workarounds, or accept sub-optimal use of network resources.

The optimal way to provide call admission control in the presence of a network topology that does not conform to a simple hub-and-spoke is to implement topology-aware call admission control, as described in the following section.

**Note**

Some IP telephony systems augment classic topology-unaware call admission control with a feedback mechanism based on observed congestion in the network, which forces calls through the PSTN when voice quality deteriorates. This approach is still not equivalent to true topology-aware call admission control because it is performed after the calls have already been established and because the call processing agent still does not have knowledge of exactly where congestion is occurring. As mentioned at the beginning of the chapter, in order to be effective, call admission control must be performed before the call is set up.

Topology-Aware Call Admission Control

We define as topology-aware call admission control any mechanism aimed at limiting the number of simultaneous calls across IP WAN links that can be applied to any network topology and can dynamically adjust to topology changes.

To accomplish these goals, topology-aware call admission control must rely on real-time communications about the availability of network resources between a call processing agent (or IP-based PBX) and the network. Because the network is a distributed entity, real-time communications require a signaling protocol.

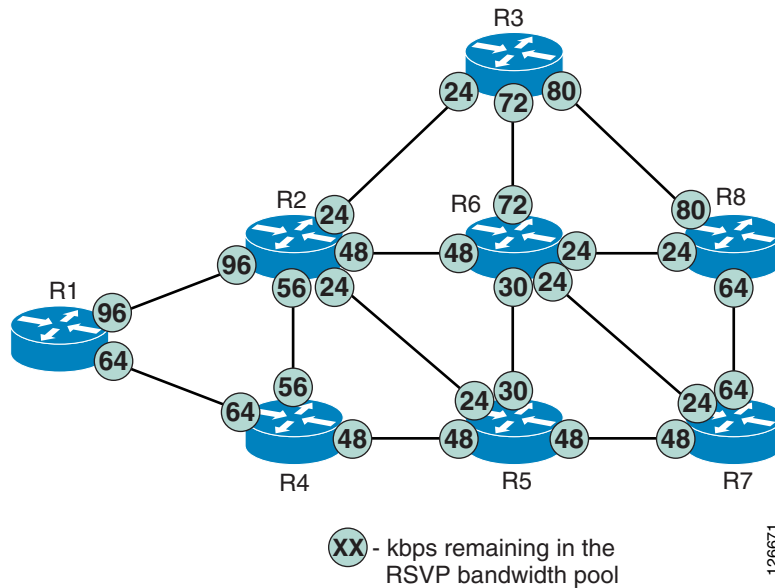
The Resource Reservation Protocol (RSVP) is the first significant industry-standard signaling protocol that enables an application to reserve bandwidth dynamically across an IP network. Using RSVP, applications can request a certain amount of bandwidth for a data flow across a network (for example, a voice call) and can receive an indication of the outcome of the reservation based on actual resource availability.

In the specific case of call admission control for voice or video calls, an IP-based PBX can synchronize the call setup process with RSVP reservations between the two remote sites and can make a routing decision based on the outcome of the reservations. 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.

To better understand the basic principles of how RSVP performs bandwidth reservation in a network, consider the simple example depicted in [Figure 9-6](#). This example does not analyze the exact message exchanges and protocol behaviors, but rather focus on the end results from a functionality perspective. For more information on the RSVP message exchanges, see [RSVP Principles, page 3-40](#).

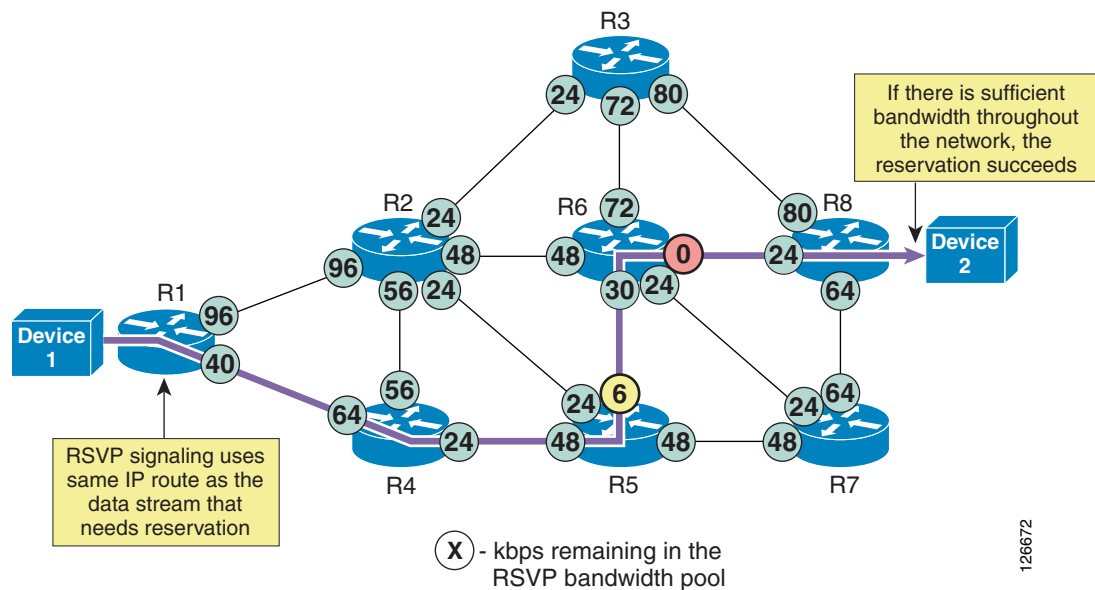
Assume that RSVP is enabled on each router interface in the network shown in [Figure 9-6](#) and that the numbers shown in the circles represent the amount of available RSVP bandwidth remaining on each interface.

Figure 9-6 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 9-7, which shows a particular data stream that requires 24 kbps of bandwidth from Device 1 to Device 2.

Figure 9-7 RSVP Signaling for a Successful Reservation

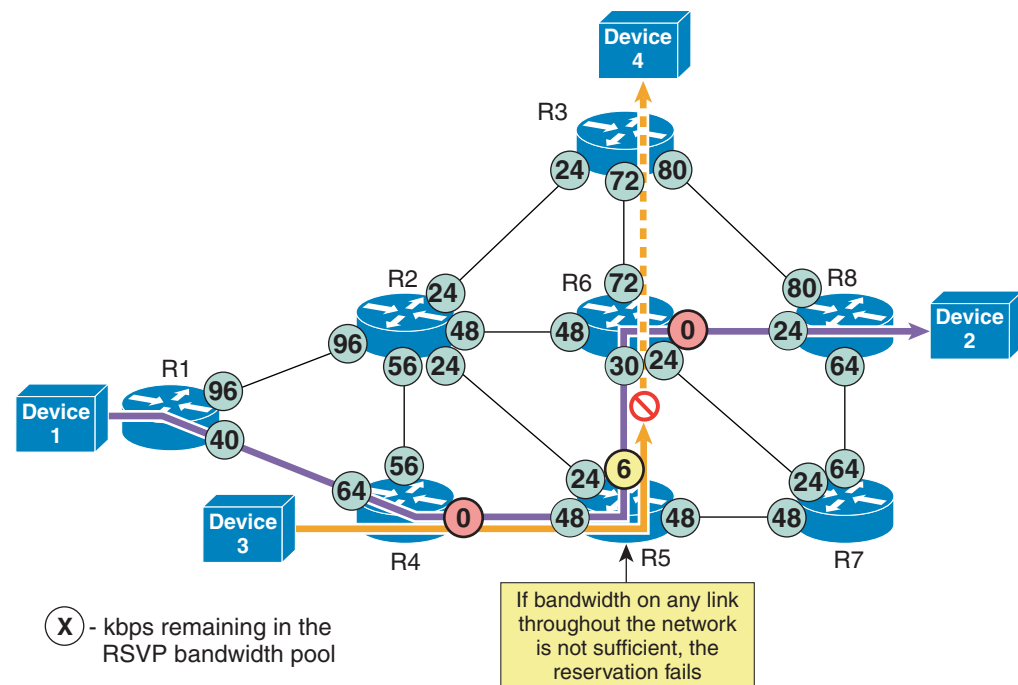


The following considerations apply to [Figure 9-7](#):

- 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 reservations 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 outbound router interfaces, as shown in [Figure 9-7](#).
- The available bandwidth on all outbound 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 pass them along like any other IP packet, and a reservation can still be established. (See [RSVP Principles, page 3-40](#), for details on protocol messages and behaviors.) 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 a 24-kbps reservation between Device 3 and Device 4, as depicted in [Figure 9-8](#).

Figure 9-8 *RSVP Signaling for an Unsuccessful Reservation*



The following considerations apply to [Figure 9-8](#):

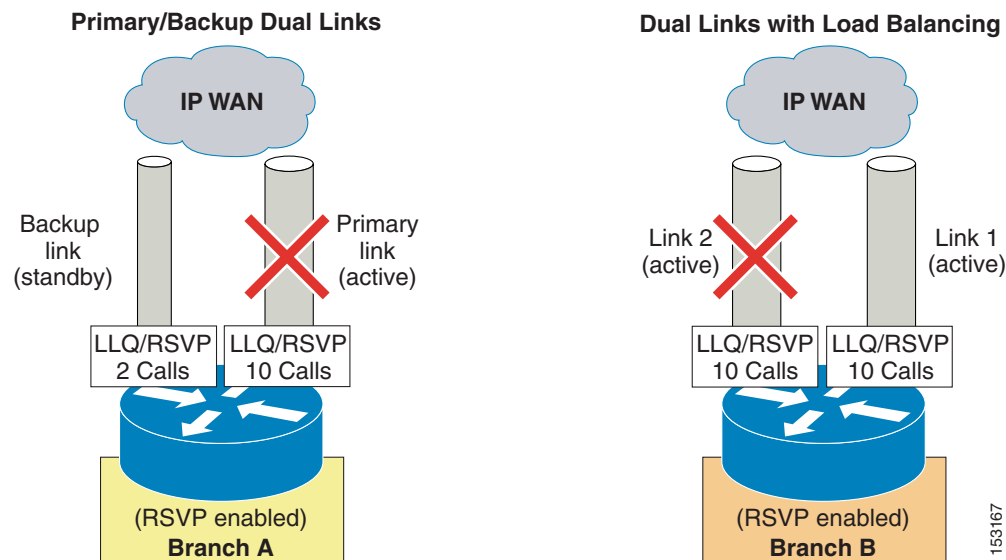
- The RSVP protocol attempts to establish an end-to-end reservation by checking for available bandwidth resources on all RSVP-enabled routers along the path from Device 3 to Device 4. As the RSVP messages progress through the network, the available RSVP bandwidth gets decremented by 24 kbps on the outbound router interfaces, as shown in [Figure 9-8](#).
- In this example, the available bandwidth on R5's outbound interface toward 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 outbound 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.

We can now apply the topology-aware call admission control approach based on RSVP to the examples of dual-connected branch offices A and B introduced in the previous section.

As shown in [Figure 9-9](#), branch office A has a primary link with an LLQ provisioned for 10 calls, while the backup link can accommodate only 2 calls. With this approach, RSVP is configured on both router interfaces so that the RSVP bandwidth matches the LLQ bandwidth. Branch A is also configured within the call processing agent to require RSVP reservations for all calls to or from other branches. Now calls are admitted or rejected based on the outcome of the RSVP reservations, which automatically follow the path determined by the routing protocol. Under normal conditions (when the primary link is active), up to 10 calls will be admitted; during failure of the primary link, only up to 2 calls will be admitted.

Policies can typically be set within the call processing agent to determine what to do in the case of a call admission control failure. For example, the call could be rejected, rerouted across the PSTN, or sent across the IP WAN as a best-effort call with a different DSCP marking.

Figure 9-9 Topology-Aware Call Admission Control for Dual Links



Similar considerations apply to branch B, connected to the IP WAN via two load-balanced links, as shown on the right side of [Figure 9-9](#). RSVP is enabled on each of the two router interfaces, with a bandwidth value that matches the LLQ configuration (in this case, enough bandwidth for 10 calls). Branch B is also configured within the call processing agent to request RSVP reservations for calls to or from other branches. Again, calls are admitted or rejected based on the actual bandwidth available along

the path chosen by the routing protocol. So in a case of perfectly even load-balancing across the two links, up to 20 calls could be admitted under normal conditions (when both links are operational); if one of the two links fails, only up to 10 calls would be admitted.

In the case that one of the two links failed while more than 10 calls were active, some calls would fail to re-establish a reservation on the new path. At this point, the call processing agent would be notified and could react based on the configured policy (for example, by dropping the extra calls or by remarking them as best-effort calls).

In conclusion, topology-aware call admission control allows administrators to protect call quality with any network topology, to automatically adjust to topology changes, and to make optimal use of the network resources under all circumstances.

Special Considerations for MPLS Networks

From the call admission control perspective, a network based on MPLS differs from one based on traditional Layer 2 WAN Services with respect to support for RSVP in the "hub" of the network. Hub sites of traditional Layer 2 wide-area networks consist, in most cases, of an enterprise-controlled router that can be enabled to participate in RSVP. Because the entire network (cloud) is the "hub site" in MPLS networks, there is no enterprise-controlled hub location to enable RSVP. (For more information, see [Simple MPLS Topologies, page 9-44](#).) Therefore, to provide topology-aware call admission control in an MPLS environment, the Customer Edge (CE) devices of the network must be configured for RSVP support.

Because RSVP must be enabled on the CE, control of this equipment is important. If this equipment is not under the control of the enterprise, you must work with your service provider to determine if they will enable RSVP on your WAN interface and if that implementation will support advanced features such as RSVP application ID.

RSVP messages will transparently pass across the RSVP-unaware MPLS cloud, so this does not pose a problem with end-to-end RSVP capability. Configuring RSVP on the CE WAN interface will ensure that its priority queue will not be overrun. Because RSVP reservations are unidirectional, the following rules must be observed to protect the priority queue on the Provider Edge (PE) router when RSVP is not enabled in the MPLS cloud:

- The media streams must be the same size in both directions.
- The media has to be symmetrically routed.

RSVP PATH messages record the egress IP address of the RSVP-aware routers they traverse. The information in the PATH message is used to send the RSVP RESV message back via the same route. Because of this mechanism, the WAN link between CE and PE must have routable IP addresses or the RSVP Reservations will fail.

If your MPLS network does not comply with these rules, contact your local Cisco account team for further assistance before implementing RSVP.

Call Admission Control Elements

There are several mechanisms that perform the call admission control function in a Cisco IP Communications system. This section provides design and configuration guidelines for all of these mechanisms, according to their category:

- Topology-unaware mechanisms
 - [Unified CM Static Locations, page 9-12](#)

- [Cisco IOS Gatekeeper Zones, page 9-14](#)
- Topology-aware mechanisms
 - [Unified CM RSVP-Enabled Locations, page 9-16](#)
 - [Cisco IOS Gatekeeper and IP-to-IP Gateway with RSVP, page 9-28](#)

**Note**

Cisco Unified CM 5.0 introduces topology-aware call admission control by extending the concept of *locations*, which already existed in previous releases. Therefore, this document refers to the former, topology-unaware mechanism as *static locations* and to the new, topology-aware mechanism as *RSVP-enabled locations*.

Unified CM Static Locations

Unified CM provides a simple mechanism known as *static locations* for implementing call admission control in the centralized call processing deployment. When you configure a device in Unified CM, the device can be assigned to a location. A certain amount of bandwidth will be allocated for calls to or from each location. The locations configured in Unified CM are virtual locations and not real, physical locations. Unified CM 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 on its location configuration so that Unified CM can correctly calculate bandwidth allocation for that device. Each device is in location Hub_None by default. Location Hub_None is a special location that is configured by default with unlimited audio and video bandwidth, and location Hub_None cannot be deleted. If the devices at a branch location are configured in the Hub_None location, none of the phone calls to or from that branch device will be subject to any call admission control.

Unified CM allows you to define a voice and video bandwidth pool for each location. If the location's audio and video bandwidth are configured as **Unlimited**, there will be unlimited bandwidth available for that location, and every audio or video call to or from that location will be permitted by Unified CM. On the other hand, if the bandwidth values are set to a finite number of kilobits per second (kbps), Unified CM 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. If the video bandwidth for the location is configured as **None**, every video call to or from that location is denied but the video calls within the same location are not affected.

For video calls, the video location bandwidth takes into account both the video and the audio portions of the call. Therefore, for a video call, no bandwidth is deducted from the audio bandwidth pool.

The devices that can specify membership in a location include:

- IP phones
- CTI ports
- H.323 clients
- CTI route points
- Conference bridges
- Music on hold (MoH) servers
- Gateways
- Trunks

The static locations call admission control mechanism also takes into account the mid-call changes of call type. For example, if an inter-site video call is established, Unified CM will subtract the appropriate amount of video bandwidth from the respective locations. If this video call changes to an audio-only call via a transfer to a device that is not capable of video, Unified CM 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.

Table 9-1 lists the amount of bandwidth requested by the static locations algorithm for various call speeds. For an audio call, Unified CM counts the media bit rates plus the Layer 3 overhead. For example, a G.711 audio call consumes 80 kbps allocated from the location's audio bandwidth pool. For a video call, Unified CM counts only the media bit rates for both the audio and video streams. For example, for a video call at a speed of 384 kbps, Unified CM will allocate 384 kbps from the video bandwidth pool.

Table 9-1 Amount of Bandwidth Requested by the Static Locations Algorithm

Call Speed	Static 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

Figure 9-10 shows the configuration for the location Branch 1, with 256 kbps of available audio bandwidth and 384 kbps of available video bandwidth. The Branch 1 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 different numbers of video calls depending on the video and audio codecs being used. For example, one video call requesting 384 kbps of bandwidth or three video calls with each requesting 128 kbps of bandwidth.

Figure 9-10 *Defining a Location in Cisco Unified CM*

Cisco Unified CallManager Administration For Cisco IP Telecommunication Solutions Logged in as: CCMAAdministrator

System ▾ Call Routing ▾ Media Resources ▾ Voice Mail ▾ Device ▾ Application ▾ User Management ▾ Bulk Administration ▾ Help ▾ **Log Off**

Location Configuration Related Links: [Back To Find/List](#) **Go**

Status
Status: Ready

Location Information
Name *

Audio Calls Information
Audio Bandwidth * ☐ Unlimited ☒ 256 kbps
If the audio quality is poor or choppy, lower the bandwidth setting. For ISDN, use multiples of 56 kbps or 64 kbps.

Video Calls Information
Video Bandwidth * ☐ None ☐ Unlimited ☒ 384 kbps

Location RSVP Settings
Location RSVP Setting
NOTE: Location(s) not displayed

Modify Setting(s) to Other Locations

Location	RSVP Setting
Branch 1	Use System Default
Branch 2	
Branch 3	
Hub_None	

Save Delete Copy Add New Resync Bandwidth

Call admission control does not apply to calls between devices within the same location.

When a call is placed from one location to the other, Unified CM deducts the appropriate amount of bandwidth from both locations. For example, a G.729 call between two locations causes Unified CM to deduct 24 kbps from the available bandwidth at both locations. When the call has completed, Unified CM returns the bandwidth to the affected locations. If there is not enough bandwidth at either branch location, the call is denied by Unified CM and the caller receives the network busy tone. If the calling device is an IP phone with a display, that device also displays the message "Not Enough Bandwidth."

When an inter-site call is denied by call admission control, Unified CM can automatically reroute the call to the destination via the PSTN connection by means of the Automated Alternate Routing (AAR) feature. For detailed information on the AAR feature, see [Automated Alternate Routing, page 10-27](#).

**Note**

AAR is invoked only when the locations-based call admission control denies the call due to a lack of network bandwidth. AAR is not invoked when the IP WAN is unavailable or other connectivity issues cause the called device to become unregistered with Unified CM. In such cases, the calls are redirected to the target specified in the Call Forward No Answer field of the called device.

Cisco IOS Gatekeeper Zones

A Cisco IOS gatekeeper can provide call routing and call admission control between devices such as Cisco Unified CM, Cisco Unified Communications Manager Express (Unified CME), 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. Refer to the section on [Call Admission Control Design, page 9-36](#), 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 section on [Gatekeeper Design Considerations, page 8-20](#). For call routing considerations, refer to [Call Routing in Cisco IOS with a Gatekeeper, page 10-41](#).

The call admission control capabilities of a Cisco IOS gatekeeper are based on 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 into or out of a given local zone.
- The **total** option controls the amount of bandwidth for all calls into, out of, or within 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.

The bandwidth value deducted by the gatekeeper for every active call 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

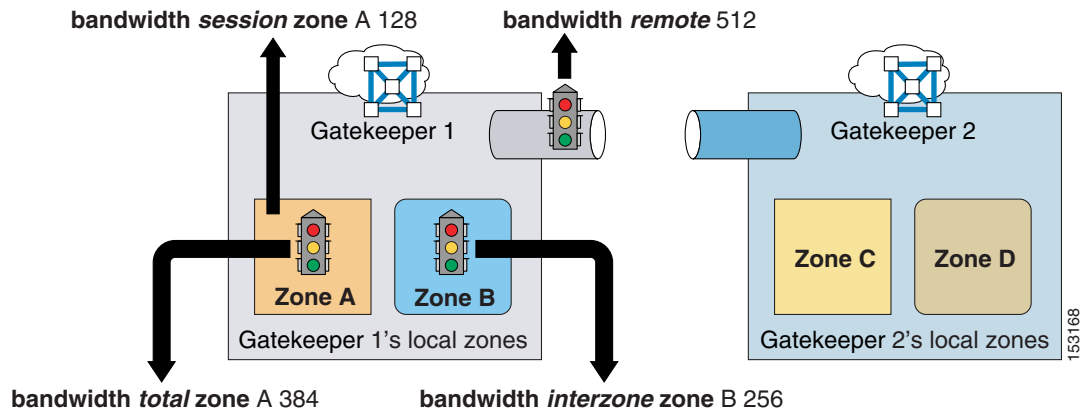
Call Speed	Gatekeeper Bandwidth Value
G.711 audio call (64 kbps)	128 kbps
G.729 audio call (8 kbps)	16 kbps
128-kbps video call	256 kbps
384-kbps video call	768 kbps
512-kbps video call	1024 kbps
768-kbps video call	1536 kbps

**Note**

Bandwidth calculations for the call Admission Request (ARQ) do not include compressed Real-Time Transport Protocol (cRTP) or any other transport overhead. See [Bandwidth Provisioning, page 3-49](#), for details on how to provision interface queues.

To better understand the application of the **bandwidth** commands in a real network, consider the example shown in [Figure 9-11](#).

Figure 9-11 Example of Cisco IOS Gatekeeper bandwidth Commands



Assuming that all calls are voice-only calls using the G.711 codec, and given the configuration commands shown in [Figure 9-11](#), the following statements hold true:

- The maximum amount of bandwidth requested by any device in zone A for a single call is 128 kbps, which means that calls trying to use codecs with a higher bit-rate than 64 kbps will be rejected.
- The maximum amount of bandwidth used by all calls involving devices in zone A (either within the zone or with other zones) is 384 kbps, which means that there can be at most three active calls involving devices in zone A.
- The maximum amount of bandwidth used by all calls between devices in zone B and devices in any other zone is 256 kbps, which means that there can be at most two active calls between devices in zone B and devices in zones A, C, and D.
- The maximum amount of bandwidth used by all calls between devices registered with gatekeeper GK 1 and devices registered with any other gatekeeper is 512 kbps, which means that there can be at most four active calls between devices in zones A and B and devices in zones C and D.

Unified CM RSVP-Enabled Locations

Cisco Unified CM Release 5.0 introduces a topology-aware call admission control mechanism based on the Resource Reservation Protocol (RSVP), which is applicable to any network topology and which eases the restriction of a traditional hub-and-spoke topology. The Cisco RSVP Agent is a Cisco IOS feature that enables Unified CM to perform the RSVP-based call admission control. The Cisco RSVP Agent feature has been introduced into Cisco IOS Release 12.4(6)T, and it is available on the Cisco 2600XM, 2691, 3700 Series, 2800 Series, and 3800 Series Integrated Services Routers platforms.

The Cisco RSVP Agent registers with Unified CM as either a media termination point (MTP) or a transcoder device with RSVP support. When an endpoint device makes a call in need of a bandwidth reservation, Unified CM invokes a Cisco RSVP Agent to act as a proxy for the endpoint to make the bandwidth reservation.

Figure 9-12 shows the signaling protocols used between Unified CM and various other devices, as well as the associated RTP streams for calls across the WAN in a given location. For any calls across the WAN, Unified CM directs the endpoint devices to send the media streams to their local Cisco RSVP Agent, which originates another call leg synchronized with an RSVP reservation to the Cisco RSVP Agent at the remote location. Figure 9-12 illustrates the following signaling protocols:

- Cisco RSVP Agents register to Unified CM via Skinny Client Control Protocol (SCCP).
- IP phones register with Unified CM via SCCP or Session Initiation Protocol (SIP).
- PSTN gateways register with Unified CM via Media Gateway Control Protocol (MGCP), SIP, or H.323 protocol.

Figure 9-12 Protocol Flows for Locations with RSVP

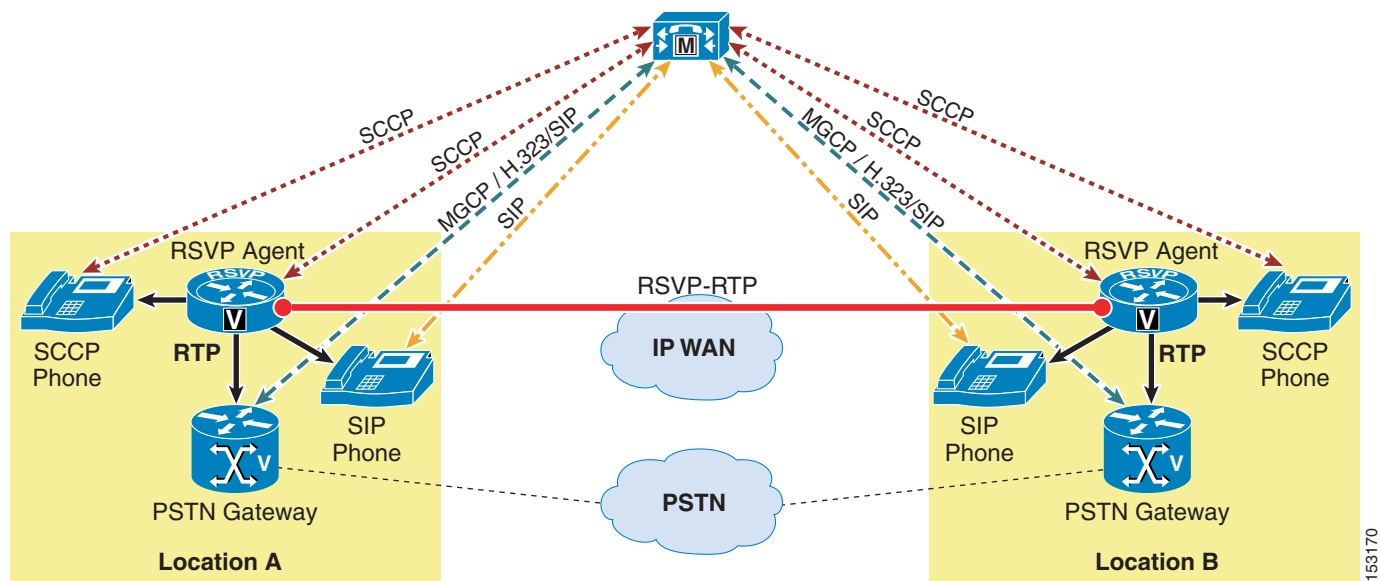


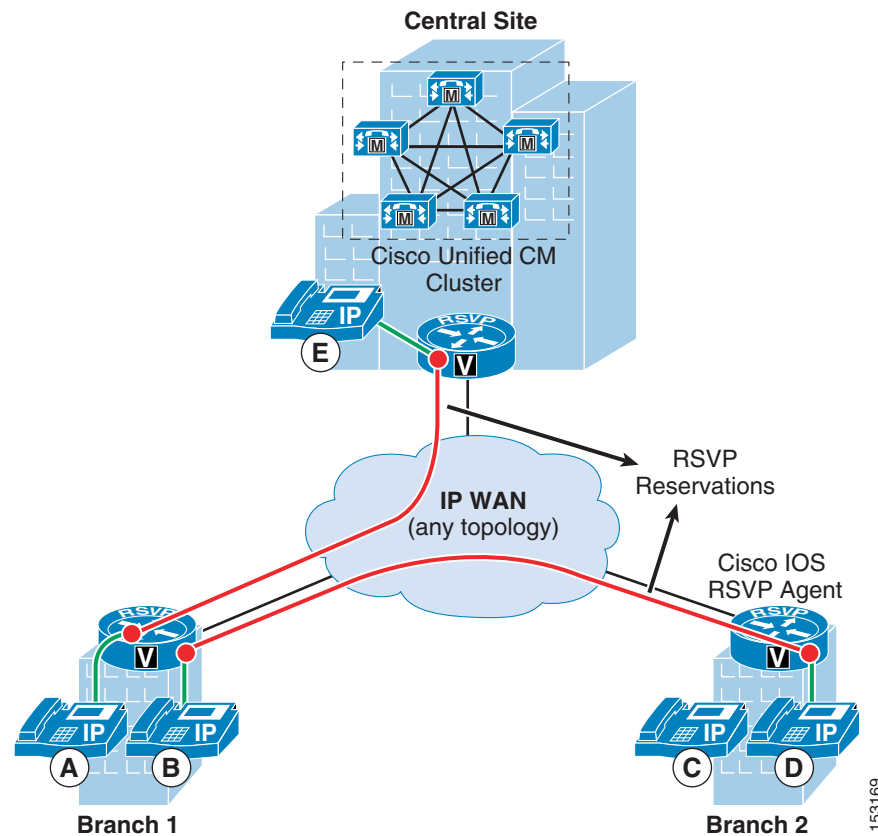
Figure 9-13 shows a typical Cisco RSVP Agent deployment within a Unified CM cluster, which includes three locations: Central Site, Branch 1, and Branch 2. The IP WAN connecting the three locations can be of any topology type and is not restricted to the hub-and-spoke topology. For any call between two locations that requires an RSVP reservation in the media path, a pair of Cisco RSVP Agents is invoked dynamically by Unified CM. The Cisco RSVP Agent acts as a proxy to make an RSVP reservation for the IP phone in the same location with the Cisco RSVP Agent. For example, when phone A in Branch 1 calls phone E in the Central Site, an RSVP reservation (illustrated as the red line in Figure 9-13) is established between Cisco RSVP Agents in the Branch 1 and Central Site locations.

There are three call legs for the media streams of this call. The first call leg is between phone A and the Branch 1 Cisco RSVP Agent, the second call leg is between the Branch 1 and Central Site Cisco RSVP Agents, and the third call leg is between the Central Site Cisco RSVP Agent and phone E. By the same token, when phone B in Branch 1 calls phone D in Branch 2, the RSVP reservation is established between the Branch 1 and Branch 2 Cisco RSVP Agents. Note that the media streams of a call between two branch locations are not sent through the Central Site in this case, which is different from a call made over the traditional hub-and-spoke topology using call admission control based on static locations.

**Note**

While RSVP-enabled locations and the use of Cisco RSVP Agent introduce support for arbitrary WAN topologies, they are based on static assignment of devices to a location, which means that every time a device is moved from one physical site to another, its configuration in Unified CM needs to be updated.

Figure 9-13 Cisco RSVP Agent Concept



Cisco RSVP Agent Provisioning

The Cisco RSVP Agent feature requires Cisco IOS Release 12.4(6)T and either of the following options:

- A Cisco Unified Survivable Remote Site Telephony license
- An Integrated Voice and Video Services image together with the Cisco Multiservice IP-to-IP Gateway feature license
- An Integrated Voice and Video Services image together with the Cisco IOS Advanced IP Services image

The capacity of Cisco RSVP Agent in terms of simultaneous calls (also referred to as sessions) depends on the following factors:

- For software-based MTP functionality, the session capacity is determined by the router platform and the relative CPU load (see [Table 9-3](#)).

- For hardware-based MTP and transcoder functionality, the session capacity is limited by the number of DSPs available. (See [Media Resources, page 6-1](#), for DSP sizing considerations.)

For software-based MTP functionality, [Table 9-3](#) provides guidelines for session capacity based on a router dedicated to the Cisco RSVP Agent and 75% CPU utilization. These numbers apply to Cisco IOS Release 12.4(6)T and should be considered as broad guidelines. Different combinations of specific services, configurations, traffic patterns, network topologies, routing tables, and other factors can significantly affect actual performance for a specific deployment and hence reduce the number of concurrent sessions supported. Cisco recommends careful planning and validation testing prior to deploying a multi-service router in a production environment.

Table 9-3 Session Capacity for Cisco RSVP Agent with Software-Based MTP Functionality

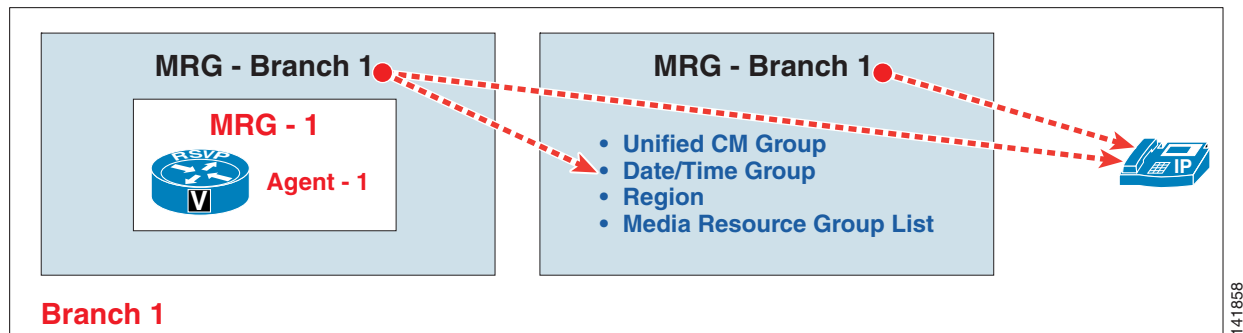
Cisco RSVP Agent Platform	Number of Sessions Supported
2611XM	40
2621XM	50
2651XM	65
2691	150
2801	130
2811	180
2821	240
2851	300
3725	250
3745	320
3825	400
3845	536



Note

When the Cisco RSVP Agent is deployed with a Cisco Unified Survivable Remote Site Telephony license, the supported number of sessions is determined by router performance (as indicated in [Table 9-3](#)) and license entitlement. When the Cisco RSVP Agent is deployed with the Cisco IOS Integrated Voice and Video Service image together with the Cisco Multiservice IP-to-IP Gateway feature license or the Cisco IOS Advanced IP Services image, then the supported number of sessions is determined only by router performance.

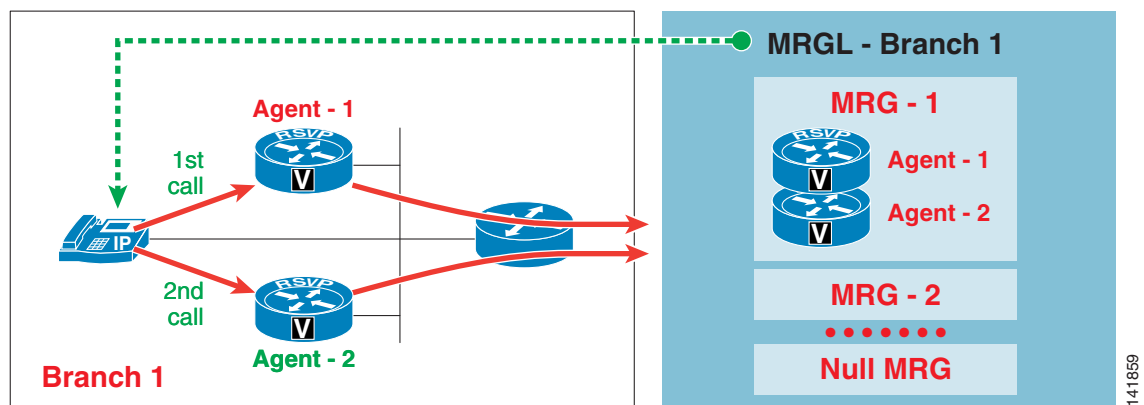
The Cisco RSVP Agent can be associated to the endpoint device by a combination of the device pool, media resource group (MRG), and media resource group list (MRGL) configurations. The Cisco RSVP Agent can be included in an MRG, and the MRG can be part of an MRGL. The MRGL can be assigned to the endpoint device either directly or via the device pool. As illustrated in [Figure 9-14](#), MRGL-Branch 1 can be associated to the IP phone either directly or via Device Pool-Branch 1. As a general rule, assign the MRGL directly to the endpoint device if the endpoint device requires an unique set of media resources; otherwise, assign the MRGL to the device pool in which the endpoint device is located.

Figure 9-14 Assigning a MRGL to an IP Phone

Unified CM allocates the Cisco RSVP Agent in the same way it allocates other conventional media resources including MTP, transcoder, conferencing resource, and annunciator.

Cisco recommends that you do not configure the Cisco RSVP Agent in the same MRG with other conventional media resources. Doing so could cause the Cisco RSVP Agent to be allocated for any call in need of an MTP device even though the call is not RSVP related.

Figure 9-15 shows how Cisco RSVP Agent load balancing is implemented via the MRG and MRGL configurations. For all the Cisco RSVP Agents within the same MRG, Unified CM load-balances and allocates the Cisco RSVP Agents in a round-robin fashion.

Figure 9-15 Cisco RSVP Agent Load Balancing

As Figure 9-15 illustrates, if both Cisco RSVP Agents in MRG-1 are available, Agent-1 will be selected for the first call and Agent-2 will be selected for the second call. If neither of the Cisco RSVP Agents in MRG-1 is available, Unified CM will try to search through MRG-2, MRG-3, and the rest of the MRGs until it finds a suitable Cisco RSVP Agent for the call. Any Cisco RSVP Agent that is not explicitly included in an MRG will be included in the Null MRG by default. Note that the Null MRG is always implicitly included as the last MRG in any MRGL configurations, but it is not displayed in Unified CM Administration. The Cisco RSVP Agent in the Null MRG is accessible by any endpoint devices in the Unified CM cluster. Therefore, Cisco recommends that you always configure a Cisco RSVP Agent in an MRG. For details on Unified CM's media resource allocation process and the associated best practices, see [Media Resources](#), page 6-1.

Cisco RSVP Agent Registration

The Cisco RSVP Agent registers with Unified CM as an MTP or transcoder device with RSVP support. The Cisco RSVP Agent does not have transcoding capability when registering as an MTP device. To have transcoding capability, the Cisco RSVP Agent must register with Unified CM as a transcoder device.

Registration Switchover and Switchback

If the primary Unified CM fails, the Cisco RSVP Agent switches over to the secondary Unified CM. When the primary Unified CM recovers from the failure, the Cisco RSVP Agent switches its registration back to the primary Unified CM. Use the following commands to configure the Cisco RSVP Agent registration switchover and switchback:

```
sccp ccm group
  switchover method immediate
  switchback method guard timeout 7200
!
gateway
  timer receive-rtp 180
```

- The **switchover method immediate** command specifies the immediate registration switchover to the secondary Unified CM server after failure of the primary Unified CM server is detected. The available DSP resources become available immediately for new calls after the switchover has completed.
- The **switchback method guard timeout 7200** command specifies the registration switchback mechanism after the primary Unified CM recovers from its failure. With this command configured, the Cisco RSVP Agent starts to switch its registration gracefully back to the primary Unified CM after the last active call disconnects. If the graceful registration switchback has not initiated by the time the guard timer expires, the Cisco RSVP Agent will use the immediate switchback mechanism and register with the primary Unified CM right away. The default value of the guard timer is 7200 seconds, and it can be configured statically in the range of 60 to 172800 seconds.
- The **timer receiver-rtp** command in the gateway configuration mode defines the RTP clean-up timer for RSVP reservations. If a failure occurs, the RSVP reservation for the existing call will stay in place until the RTP clean-up timer expires. The default value of this timer is 1200 seconds. Cisco recommends that you configure this timer with its lowest allowed value, which is 180 seconds.

Maximum Sessions Support

The Cisco RSVP Agent supports a maximum number of calls or sessions, based on the software-based (CPU) and hardware-based (DSP) resources equipped on the Cisco RSVP Agent router. The **maximum sessions** command in the **dspfarm profile** configuration mode specifies the maximum number of calls that the Cisco RSVP Agent is able to handle. The Cisco RSVP Agent notifies Unified CM of its session capacity based on this configuration. The maximum number of sessions decreases by one for every call going through the Cisco RSVP Agent. When the counter reaches zero, the Cisco RSVP Agent is regarded as having no resources available, and Unified CM skips that Cisco RSVP Agent for any subsequent calls.

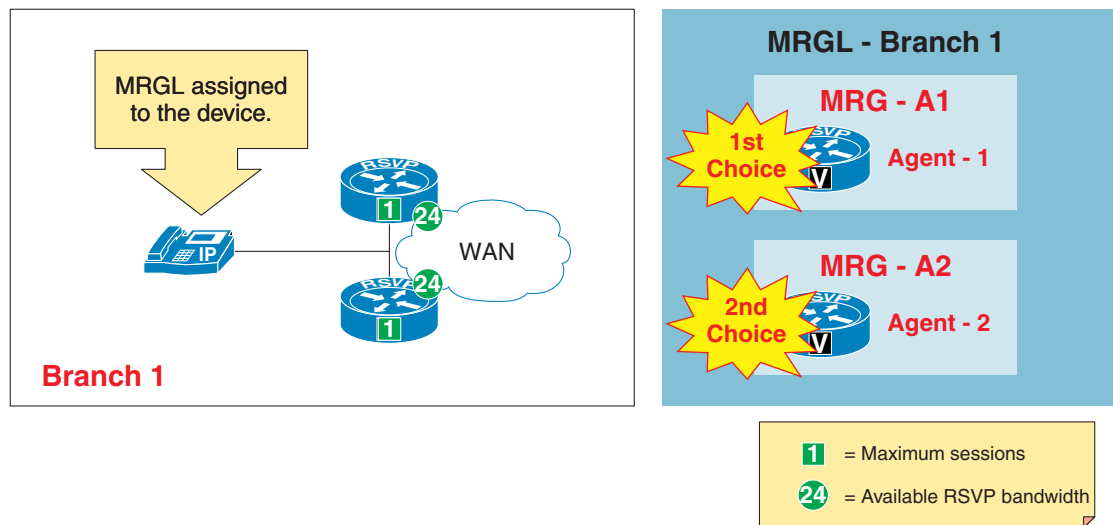
Figure 9-16 shows a branch site with dual Cisco RSVP Agents. The Cisco RSVP Agents are co-resident with the WAN routers, and Cisco RSVP Agent redundancy is achieved by assigning two Cisco RSVP Agents to different MRGs in the same MRGL. If Agent-1 in MRG-1 is unavailable or out of session capacity, Unified CM will try to allocate Agent-2 in MRG-2 for RSVP calls to or from Branch 1. To ensure that Agent-2 is selected when Agent-1's capacity is reached, Cisco recommends configuring the maximum number of sessions to match exactly the number of calls supported by the **ip rsvp bandwidth** configured on the WAN interface of the Cisco RSVP Agent. In this example, both Cisco RSVP Agents

need to be configured with **maximum sessions 1**. This recommendation is made on the assumption that all calls going across the WAN will use the same type of codec so that an accurate calculation of the number of calls across the WAN can be derived, which is done by dividing the total available RSVP bandwidth by the bandwidth requested per call.

**Note**

If the maximum number of sessions is higher than the number of calls supported by the **ip rsdp bandwidth** configuration, Unified CM will still send the call to the Cisco RSVP Agent but the RSVP reservation will fail because there is no bandwidth available, and Unified CM will follow the usual behavior for call admission control failure (that is, it will deny the call or invoke the AAR feature).

Figure 9-16 *Configuring Maximum Sessions on the Cisco RSVP Agent*



141860

Pass-Through Codec

The pass-through codec enables a Cisco IOS Enhanced MTP device to terminate an RTP media stream received from an endpoint without knowing the media encoding of the stream. That is, the UDP packets of the media stream flow through the MTP without being decoded. This method enables the MTP to support every audio, video, and data codec that is defined in Unified CM. Because the MTP does not decode the media stream, the pass-through codec can also be used with encrypted (SRTP) media streams. In fact, for video and SRTP media streams to use an MTP, it must support the pass-through codec. When configured with the pass-through codec, the Cisco RSVP Agent will substitute its own IP address for the source IP address in the IP/UDP headers of the packets and let them flow through.

The Cisco RSVP Agent will use the pass-through codec only if all of the following conditions are met:

- The two endpoint devices involved in the call have matching audio codec capability, and the region configuration permits the matching codec to be used for the call. In other words, no transcoder device needs to be inserted in the call.
- **MTP Required** is not configured for either endpoint device.
- All intermediate resource devices support the pass-through codec.

**Note**

If the Cisco RSVP Agent registers as an MTP device and a transcoder device needs to be inserted in the call, the codec configured in the Cisco RSVP Agent dspfarm MTP profile must match the inter-region codec configured in Unified CM Administration. For example, if the G.729 codec is the inter-region codec configured in Unified CM Administration, then the G.729 codec must also be configured in the dspfarm MTP profile.

The following example shows an Cisco RSVP Agent configuration on a Cisco 2800 IOS platform:

```
interface Loopback0
 ip address 10.11.1.100 255.255.255.255
!
sccp local Loopback0
sccp ccm 20.11.1.50 identifier 1 priority 1 version 5.0.1
sccp ccm 20.11.1.51 identifier 2 priority 2 version 5.0.1
sccp
!
sccp ccm group 1
 associate ccm 1 priority 1
 associate ccm 2 priority 2
 associate profile 1 register RSVPAgent
 switchover method immediate
 switchback method guard timeout 7200
!
dspfarm profile 1 mtp
 codec pass-through
 codec g729ar8
 rsvp
 maximum sessions software 100
 associate application SCCP
```

RSVP Policy

Unified CM can apply different RSVP policies to different location pairs. The RSVP policy can be configured in Unified CM Administration. The RSVP policy defines whether or not Unified CM will admit the call if the RSVP reservation attempt fails. The following RSVP policy settings can be configured between any two locations:

- **No Reservation**
No RSVP reservation attempt is made, and only static-locations call admission control is performed by Unified CM.
- **Mandatory**
Unified CM does not ring the terminating endpoint device until the RSVP reservation succeeds for the audio stream and, if the call is a video call, for the video stream as well.
- **Mandatory (Video Desired)**
A video call can proceed as an audio-only call if a reservation for the video stream cannot be reserved but the reservation for the audio stream succeeds.
- **Optional (Video Desired)**
A call can proceed as a best-effort audio-only call if it fails to obtain reservations for both its audio and video streams. The Cisco RSVP Agent re-marks the media packets as best-effort.

- Use System Default

The RSVP policy for the location pair matches the clusterwide RSVP policy. The default clusterwide RSVP policy is No Reservation. To change the default RSVP policy in Unified CM Administration, select **System > Service Parameters > Cisco CallManager Service.> Default Inter-location RSVP Policy**

**Note**

With the Optional (video desired) policy, IP WAN calls may proceed as best-effort not only if the RSVP reservation fails but also if the Cisco RSVP Agent is not available. In this case, Unified CM instructs SCCP and MGCP devices to re-mark their traffic as best-effort. However, this re-marking is not possible with H.323 and SIP devices, which will keep sending their traffic with the default QoS marking. To prevent over-subscribing the priority queue in the latter case, Cisco recommends configuring an access control list (ACL) on the IP WAN router to permit only packets marked with DSCP EF or AF41 if the source IP address is that of the Cisco RSVP Agent.

Figure 9-17 shows both the default and recommended configurations of the RSVP clusterwide parameters. Cisco recommends configuring the RSVP policy as **Mandatory** or **Mandatory (Video Desired)** because those settings guarantee the bandwidth reservation and the voice quality of the call. The most efficient method for setting the clusterwide RSVP policy is to configure the **Default Inter-location RSVP Policy** in the RSVP clusterwide parameters of the Cisco CallManager Service Service Parameter Configuration, and leave the RSVP configuration in the location configuration set to **Use System Default**.

Figure 9-17 **Setting RSVP Clusterwide Parameters**

Clusterwide Parameters (System - RSVP)			
Default inter-location RSVP Policy *	Mandatory		No Reservation
RSVP Retry Timer *	60		60
Mandatory RSVP Mid-call Retry Counter *	1		1
Mandatory RSVP mid call error handle option *	Call fails following retry counter exceeded		Call becomes best effort

141857

In the clusterwide RSVP parameters configuration, there is a service parameter named **Mandatory RSVP mid call error handle option**. If the RSVP policy is configured as **Mandatory** or **Mandatory (Video Desired)**, this parameter specifies how Unified CM will treat an existing RSVP call based on the failure of a mid-call RSVP reservation attempt. The mid-call RSVP reservation attempt can be triggered by (but is not limited to) a network convergence after a WAN failure or by an existing voice-only call becoming a video call. A network convergence makes the Cisco RSVP Agent not only start to send the media streams over the newly converged path but also to try to make a new RSVP reservation over the new path.

The default setting of the **Mandatory RSVP mid call error handle option** is **Call Becomes Best Effort**. With the default option configured, Unified CM will maintain the existing call even though the mid-call RSVP reservation attempt fails, but the RTP streams will be marked as best effort (DSCP 0). Cisco recommends configuring this parameter with the **Call Fails Following Retry Counter Exceeded** option. With this option configured, Unified CM will fail the call if the RSVP reservation attempt keeps failing after a certain number of retries. The default value of the retry counter is 1, which is defined by the **RSVP Mandatory mid-call retry counter** service parameter, and the default value of **RSVP retry timer** is 60 seconds. Cisco recommends having both the retry counter and the retry timer service parameters configured with their default values. With both set to their default values, Unified CM will wait for 60 seconds before it disconnects the call if the RSVP mid-call retry fails. During this period, users might experience degraded voice quality because no RSVP reservation is in place and the RTP streams are marked as best effort.

Migrating from Static Locations to RSVP Call Admission Control

The example in this section illustrates the best practices for migrating from the traditional static-locations call admission control to the RSVP-based call admission control mechanism.

Figure 9-18 shows a centralized call processing deployment with the static-locations call admission control mechanism. There are four locations in the Unified CM cluster, including the Hub_None location and three branches. For simplicity of illustration, the bandwidth used in this example refers only to the voice stream bandwidth. Table 9-4 and Table 9-5 show that every branch location is statically provisioned with 256 kbps of bandwidth, and the Hub_None location is provisioned with **Unlimited** bandwidth. The RSVP setting between any pair of locations is configured with **Use System Default**, and the clusterwide RSVP setting is configured with the default value **No Reservation**.

Figure 9-18 Configuration with Static Locations for Call Admission Control

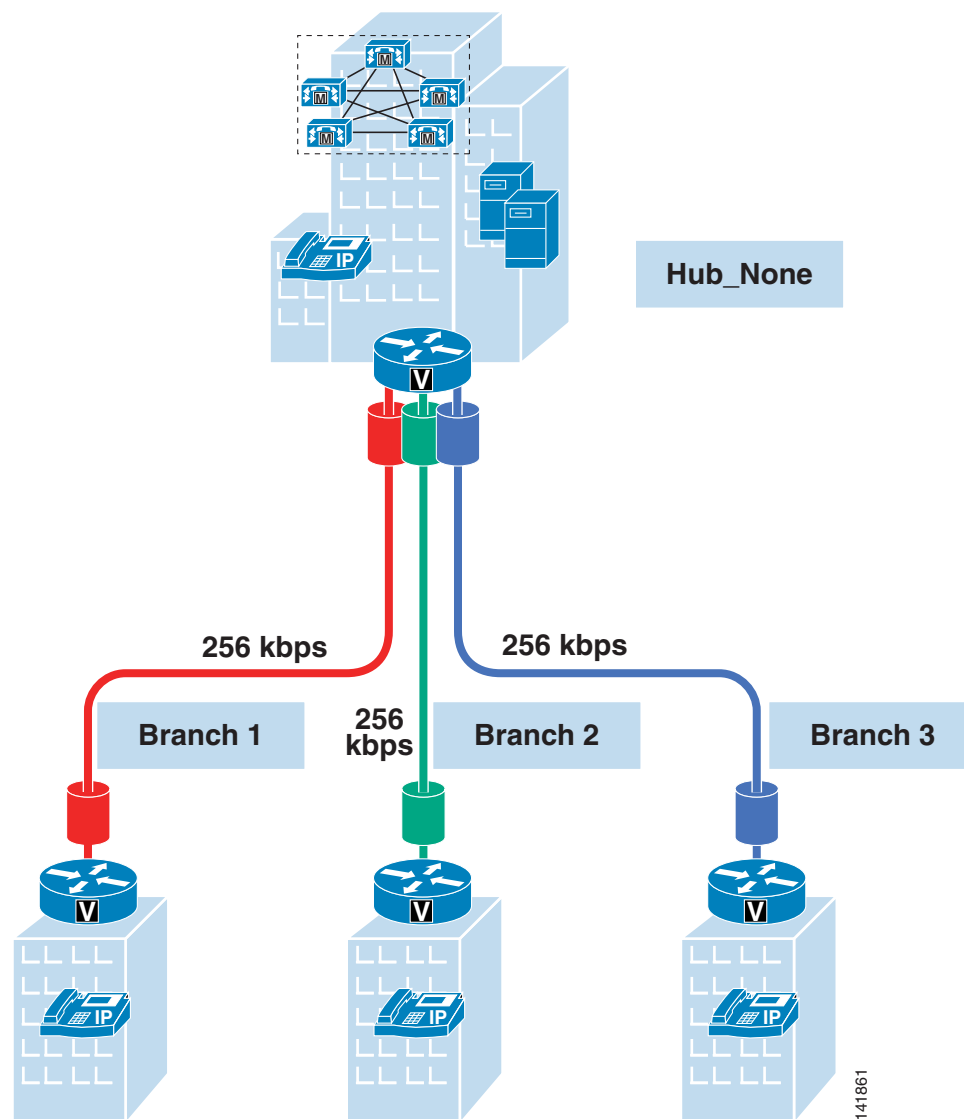


Table 9-4 *Locations and Bandwidth Settings for the Example in Figure 9-18*

Location Name	Bandwidth
Hub_None	Unlimited
Branch 1	256 kbps
Branch 2	256 kbps
Branch 3	256 kbps

Table 9-5 *RSVP Policy for the Example in Figure 9-18*

Locations Pair		Policy
Any	Any	No Reservation

To migrate to RSVP-based call admission control, Cisco recommends migrating one location at a time. For example, if Branch 1 is the first location to be migrated, observe the following procedures:

- Set up a Cisco RSVP Agent in the Branch 1 location and assign it to the Branch 1 MRG and MRGL to associate it with the Branch 1 IP phones.
- Set up another Cisco RSVP Agent in the Hub_None location and include this Cisco RSVP Agent in the MRG and MRGL that is associated to all the IP phones in the remaining three locations, including the Hub_None location. Note that this Cisco RSVP Agent should not be included in the Null MRG or Branch 1 MRG, otherwise it is possible that a Branch 1 phone will use the Cisco RSVP Agent at the Hub_None location to make RSVP reservations.
- Configure the Branch 1 bandwidth as **Unlimited**.
- Configure the RSVP setting between Branch 1 and any other location as **Mandatory**. For example, for a call between Branch 1 and Branch 2 phones, the voice stream will still be hair-pinned through the Hub_None location. For the first call leg between the Branch 1 and Hub_None locations, a RSVP reservation will be made between the Branch 1 and Hub_None Cisco RSVP Agents. For the second call leg between the Hub_None and Branch 2 locations, Unified CM will perform call admission control based on static locations by checking the bandwidth availability for the Branch 2 location.

Table 9-6 and Table 9-7 show the location bandwidth and RSVP policy configuration after the migration at Branch 1.

Table 9-6 *Locations and Bandwidth Settings After Migration of Branch 1*

Location Name	Bandwidth
Hub_None	Unlimited
Branch 1	Unlimited
Branch 2	256 kbps
Branch 3	256 kbps

Table 9-7 *RSVP Policy After Migration of Branch 1*

Locations Pair		Policy
Branch 1	Any	Mandatory
All other locations	All other locations	No Reservation

Table 9-8 and Table 9-9 show the location bandwidth and RSVP policy configurations after the clusterwide migration. With the clusterwide migration completed, any inter-site call will require a RSVP reservation to be made directly between two Cisco RSVP Agents, and the voice stream will be transported over the bandwidth reservation path.

The following procedures can be used to migrate Branch 2 and Branch 3 to RSVP call admission control:

- Set up a Cisco RSVP Agent in the Branch 2 location and assign it in the Branch 2 MRG and MRGL that is associated with the Branch 2 IP phones. Be sure to remove the Cisco RSVP Agent of the Hub_None location from the Branch 2 MRG so that the Cisco RSVP Agent in the Hub_None location will no longer be accessed by the IP phones in Branch 2.
- Configure the Branch 2 bandwidth as **Unlimited**.
- Configure the RSVP setting between Branch 2 and any other location as **Mandatory**.
- Set up a Cisco RSVP Agent in the Branch 3 location and assign it in the Branch 3 MRG and MRGL that is associated with the Branch 3 IP phones. Be sure to remove the Cisco RSVP Agent of the Hub_None location from the Branch 3 MRG so that the Cisco RSVP Agent of the Hub_None location will no longer be accessed by the IP phones in Branch 3.
- Configure the Branch 3 bandwidth as **Unlimited**.
- Configure the RSVP setting between Branch 3 and any other location as **Mandatory**.

Table 9-8 *Locations and Bandwidth Settings After Migration Is Complete*

Location Name	Bandwidth
Hub_None	Unlimited
Branch 1	Unlimited
Branch 2	Unlimited
Branch 3	Unlimited

Table 9-9 *RSVP Policy After Migration Is Complete*

Locations Pair		Policy
Any	Any	Mandatory

RSVP Application ID

The RSVP Application ID is a mechanism that enables Unified CM to add an identifier to both the voice and video traffic so that the Cisco RSVP Agent can set a separate bandwidth limit on either traffic based on the identifier received. To deploy the RSVP Application ID in the network, you must use Cisco IOS Release 12.4(6)T or later on the Cisco RSVP Agent router and Cisco Unified CM Release 5.0. The RSVP Application ID strings can be configured via two service parameters in the clusterwide RSVP parameter configuration: **RSVP Audio Application ID** and **RSVP Video Application ID**.

Unified CM uses SCCP to convey the RSVP Application ID to the Cisco RSVP Agent. The Cisco RSVP Agent also inserts the RSVP Application ID into the RSVP signaling messages (such as the RSVP Path and Resv messages) and sends those messages to the downstream or upstream RSVP routers.

The RSVP Application ID uses a model to separate the bandwidth of voice and video traffic that is different than the static locations model. In static locations, the voice and video streams of a video call are both deducted from the video bandwidth counter. When using the RSVP Application ID, the voice stream is deducted from the audio bandwidth pool while the video stream is deducted from the video bandwidth pool. With this change in the call admission control model, you can now reserve a certain amount of bandwidth for voice calls and allow them to use the entire available bandwidth in the priority queue, thus ensuring that all the available bandwidth can be used for voice calls if no video calls are in progress. If there is enough available bandwidth in the priority queue, calls can optionally be enabled for video. You can set limits on how much bandwidth the video-enabled calls can consume, but if voice calls are consuming all the available bandwidth, it might not be possible to place a video call at all. For detailed information on how to configure the RSVP Application ID, RSVP policy, and LLQ, see [RSVP Application ID](#), page 3-47.

Cisco IOS Gatekeeper and IP-to-IP Gateway with RSVP

The Cisco Multiservice IP-to-IP Gateway (also referred to as IP-IP gateway or IPIP GW) can be used to ease the restriction of hub-and-spoke topologies for the IP WAN connections between Unified CM 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 [Topology-Aware Call Admission Control](#), 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.

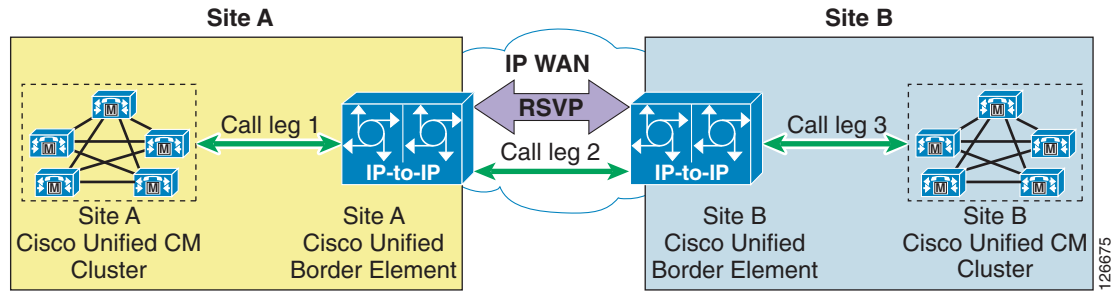


Note

For detailed information on Cisco support for deploying an IP-IP gateway with RSVP, refer to the *Cisco Multiservice IP-to-IP Gateway Application Guide* available at <http://www.cisco.com>.

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 9-19](#) shows a basic example where two sites, A and B, each have a Unified CM 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 Unified CM 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:

- Unified CM 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 Unified CM cluster. (This is call leg 3 in the figure.)

Figure 9-19 Simple Example of the IP-to-IP Gateway for RSVP Call Admission Control

The example in [Figure 9-19](#) is a simple scenario in which all calls between Unified CM 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 Unified CM clusters, H.323 gateways, H.323 videoconferencing endpoints, and IP-IP gateways.

**Note**

All scenarios involving IP-IP gateways described in this section apply to calls between different Unified CM clusters. Cisco does not recommend inserting IP-IP gateways for calls between endpoints registered to the same Unified CM cluster. For RSVP-based call admission control between endpoints registered to the same Unified CM cluster, see [Unified CM RSVP-Enabled Locations, page 9-16](#).

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 gatekeeper functionality and extend it with additional features. 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.

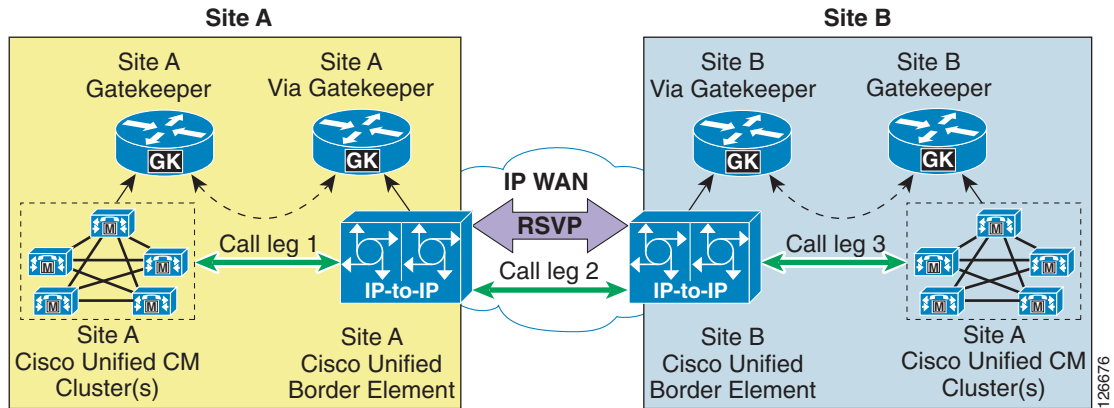
For ARQ messages, the gatekeeper determines if an **outvia** keyword is configured on the destination zone. If the **outvia** keyword is configured and the zone named with the **outvia** keyword is local to the gatekeeper, the call is directed to an IP-IP gateway in that zone by returning an ACF pointing 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 9-20](#) shows an example of how IP-IP gateways and via-zone gatekeepers can be used in conjunction with Unified CM clusters and legacy gatekeepers to provide call routing and call admission control. The following considerations apply to this scenario:

- Site A's Unified CM 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 Unified CM clusters, gatekeepers, and the IP-IP gateway at Site B are configured in a similar fashion to their counterparts at Site A.

Figure 9-20 IP-to-IP Gateway for RSVP Using Via-Zone Gatekeepers



Design Best Practices

When deploying IP-IP gateways in conjunction with Unified CM in order to enable RSVP call admission control in the IP WAN, observe these design best practices:

- When configuring trunks in Unified CM for either voice or video communication with other Unified CM clusters through one or more IP-IP gateways, use gatekeeper-controlled H.225 trunks. When using Cisco IOS Release 12.4(6)T or later and Cisco Unified CM Release 4.1 or later, MTP resources are no longer required to invoke supplementary services such as hold/resume, transfer, and conference across IP-IP gateways. In order to ensure interoperability, you must configure the following items:
 - On Unified CM Administration trunk configuration page, leave the **Media Termination Point required** field unchecked (default configuration), and also uncheck the field **Wait for Far End H.245 Terminal Capability Set**.
 - In Unified CM Administration, under the Advanced Service Parameters page for Unified CM, set the **Send H225 User Info Message** field to **H225 Info For Call Progress Tone**.
 - On the IP-IP gateways, configure the following Cisco IOS commands to ensure interoperability with Unified CM when invoking supplementary services:

```
voice service voip
  h323
    emptycapability
    h245 passthru tcsnonstd-passthru
```

- The MTP resources are preferred in some deployments to provide a proxy functionality and to terminate the signaling and media streams on behalf of endpoint devices. If the MTP resources are required, 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 an NM-HDV network module). Refer to the chapter on [Media Resources](#),

[page 6-1](#), for a complete list of available MTP resources. However, when 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. Note that video calls will not be established across clusters through IP-IP gateways if the **Media Termination Point required** option is checked on the H.225 trunks (because MTPs do not support video calls).

- Configure the IP-IP gateway as an H.323 gateway in Unified CM 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 Unified CM 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 Unified CM 3.3(2) or later on the IP-IP gateway is available on Cisco IOS Release 12.3(1) or later. Cisco recommends that you use Cisco IOS Release 12.4(6)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.
- You can run the via-zone gatekeeper function and the IP-IP gateway function on the same router platform (co-resident); however, be aware of the scalability requirements described in the section on [Redundancy, page 9-31](#).
- Do not use IP-IP gateways for calls between endpoints controlled by the same Unified CM cluster.
- Use RSVP-enabled locations to provide topology-aware call admission control between endpoints controlled by the same Unified CM cluster.
- Use the following options under the dial-peer configuration when enabling RSVP reservations for the IP-IP gateway:

```
req-qos guaranteed-delay audio
req-qos guaranteed-delay video
acc-qos guaranteed-delay audio
acc-qos guaranteed-delay video
```

This configuration ensures that for each voice or video call, the IP-IP gateway will request an RSVP reservation using the guaranteed delay service. The fact that both the requested QoS and the acceptable QoS specify this RSVP service means that the RSVP reservation is mandatory for the call to succeed (that is, if the reservation cannot be established, the call will fail). For more information on configuration details, see [Configuration Guidelines, page 9-32](#).

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
```

The above command specifies an aggregated session capacity of the IP-IP gateway, in terms of call legs. The default value is 1000 reserved call legs. Any call being handled by the IP-IP gateway costs two sessions from the available IP circuits, one session for the inbound call leg and another session for the outbound call leg. Like a regular H.323 gateway, the IP-IP gateway will automatically send its session capacity information to the via-zone gatekeeper using the H.323 version 4 protocol.

- On the gatekeeper:

```
endpoint resource-threshold onset onset-threshold abatement abatement-threshold
```

The above command makes the via-zone gatekeeper monitor the call volume in each of its gateways, including the IP-IP gateway. The via-zone gatekeeper does the active call counting when the gateway reports its session capacity information in an admission request (ARQ) or disengaged request (DRQ) message. If the active call capacity usage in a particular gateway is above the high-water mark (Range = 1 to 99; Default = 90), the via-zone gatekeeper will stop sending calls to that gateway. If the gateway's active call volume falls below the low-water mark (Range = 1 to 99; Default = 70), the via-zone gatekeeper will resume sending calls to the gateway. These thresholds are global values and affect all gateways registered with a given gatekeeper.

With both of the above commands configured, the via-zone gatekeeper is able to calculate the current session capacity utilization on the IP-IP gateway, which can prevent the via-zone gatekeeper from sending a call to the IP-IP gateway without enough capacity resources. Otherwise, the gatekeeper call admission control would fail, with an admission reject (ARJ) or location reject (LRJ) message being returned to the originating device.

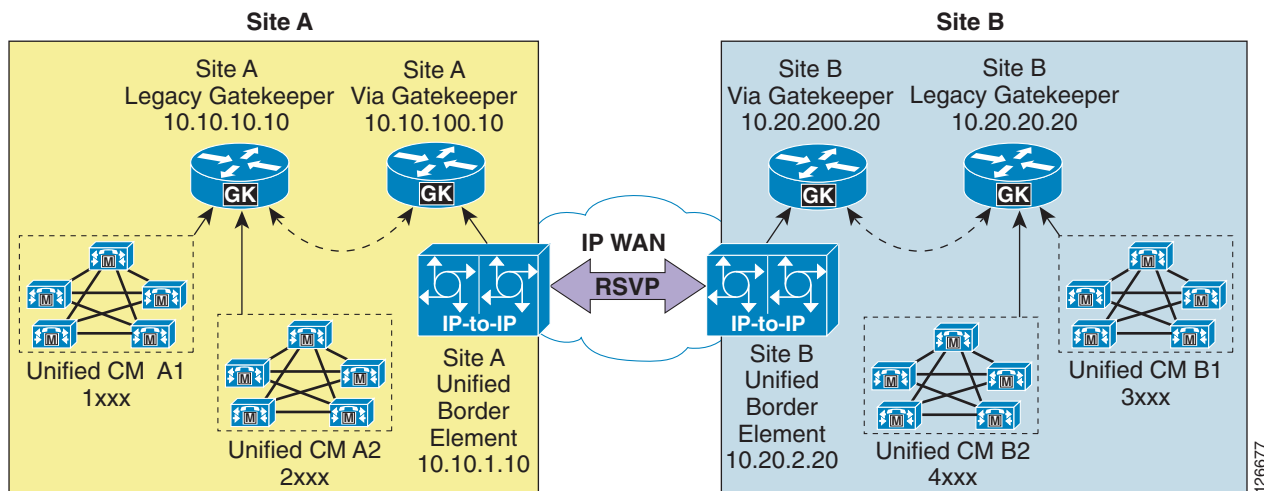
For more details on these commands, refer to the Cisco IOS command reference documentation available at

<http://www.cisco.com>

Configuration Guidelines

This section presents a simple configuration example based on the network diagram shown in [Figure 9-21](#). 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 9-21 Configuration Example for IP-IP Gateway with Via-Zone Gatekeeper

For the network shown in [Figure 9-21](#), assume that there are two Unified CM clusters at Site A: cluster A1, with phone extensions 1xxx, and cluster A2, with phone extensions 2xxx. There are also two Unified CM 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 Unified CM 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).

Unified CM

Both cluster A1 and cluster A2 use a gatekeeper-controlled intercluster trunk, which is an intercluster trunk (ICT) without MTP required and which points to Site A's legacy gatekeeper.

A [34]XXX route pattern points to the ICT through a route list and 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 the ICT through a route list and route group construct.

To support supplementary services with an IP-IP gateway and RSVP between clusters, the ICT must have the parameter "Calling Party Selection" set to "Last Redirect Number" or "First Redirect Number" for calls to complete.

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 Unified CM 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...
gw-type-prefix 1#* default-technology
arq reject-unknown-prefix
no shutdown

```

The following considerations apply to the configuration shown in [Example 9-2](#):

- The **invia** and **outvia** keywords in the command line related to the B-VIAGK remote zone trigger 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** and **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 Unified CM clusters (extensions 3xxx and 4xxx) but not for calls directed toward Site A's Unified CM 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
    emptycapability
    h245 passthru tcsnonstd-passthru
!
gateway
!

```

```
interface FastEthernet0/1
  ip address 10.10.1.10 255.255.255.0
  ip rsvp bandwidth 200
  ip rsvp data-packet classification none
  ip rsvp resource-provider none
  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
  h323-gateway voip tech-prefix 1#
!
dial-peer voice 5 voip
  session target ras
  incoming called-number [3-4]...
  codec g729r8
!
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 g729r8
!
dial-peer voice 15 voip
  session target ras
  incoming called-number [1-2]...
  req-qos guaranteed-delay audio
  req-qos guaranteed-delay video
  acc-qos guaranteed-delay audio
  acc-qos guaranteed-delay video
  codec g729r8
!
dial-peer voice 20 voip
  destination-pattern [1-2]...
  session target ras
  codec g729r8
```

The following considerations apply to the configuration shown in [Example 9-3](#):

- The **emptycapability** command enables the H.245 Empty Capabilities Set (ECS) between Unified CM and the IP-IP gateway to invoke supplementary services for the established 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 or 15.
- The **acc-qos guaranteed-delay [audio | video]** commands specify that the minimum 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, which equates to making the RSVP reservation mandatory. To configure the IP-IP gateway so that the RSVP reservation is optional (so that the call succeeds 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 Unified CM deployment models and to the following IP WAN topologies:

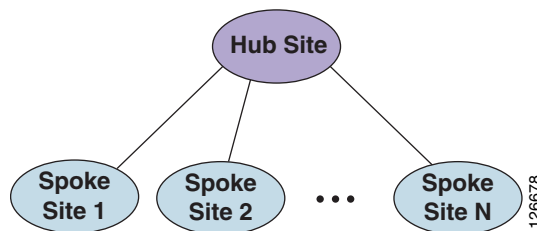
- [Simple Hub-and-Spoke Topologies, page 9-36](#)
- [Two-Tier Hub-and-Spoke Topologies, page 9-40](#)
- [Simple MPLS Topologies, page 9-44](#)
- [Generic Topologies, page 9-50](#)

For each topology, these sections present different sets of design considerations based on the Unified CM deployment model adopted.

Simple Hub-and-Spoke Topologies

[Figure 9-22](#) 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 via a single IP WAN link to a 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 9-22 A Simple Hub-and-Spoke Topology



The design considerations in this section apply to simple hub-and-spoke topologies that use 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 [Simple MPLS Topologies, page 9-44](#).

The remainder of this section contains design best practices for simple hub-and-spoke topologies according to the Unified CM deployment model adopted:

- [Centralized Unified CM Deployments, page 9-37](#)

One or more Unified CM clusters are located at the hub site, but only phones and gateways are located at the spoke sites.

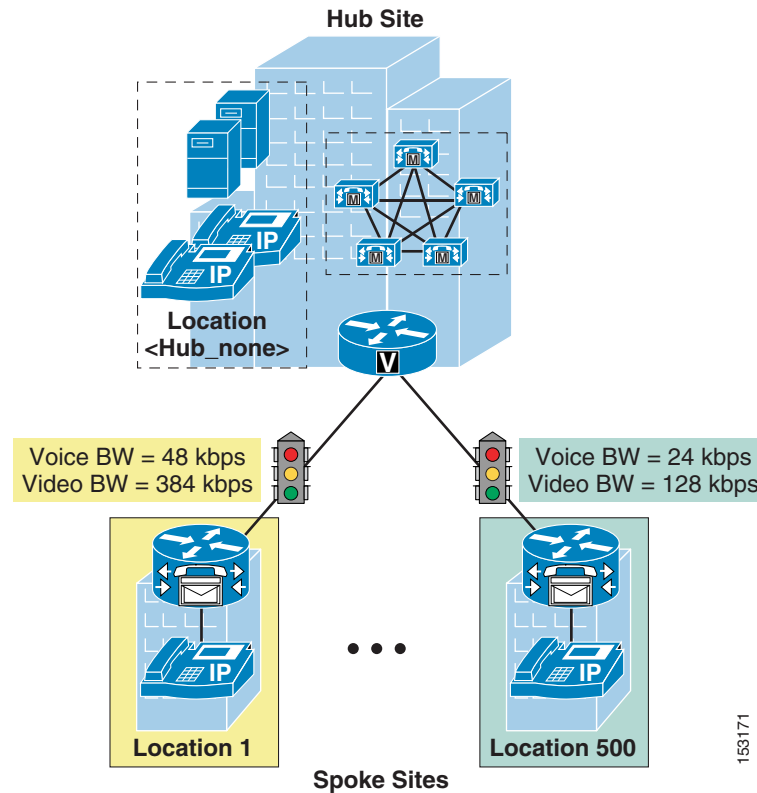
- [Distributed Unified CM Deployments, page 9-38](#)

A Unified CM cluster or Cisco Unified Communications Manager Express (Unified CME) is located at each site.

Centralized Unified CM Deployments

In multisite WAN deployments with centralized call processing in a simple hub-and-spoke topology, use Unified CM static *locations* for implementing call admission control. Figure 9-23 shows an example of how to apply this mechanism to such a topology.

Figure 9-23 Call Admission Control for Simple Hub-and-Spoke Topologies Using Static Locations



Follow these guidelines when using static locations for call admission control:

- Configure a separate location in Unified CM for each spoke site.
- Configure the appropriate bandwidth limits for voice and video calls for each site according to the types of codecs 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 Hub_None location.
- If you move a device to another location, change its location configuration as well.
- Unified CM 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 Unified CM. (See [Automated Alternate Routing](#), page 10-27.)
- If multiple Unified CM clusters are located at the same hub site, leave the intercluster trunk devices in the Hub_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.

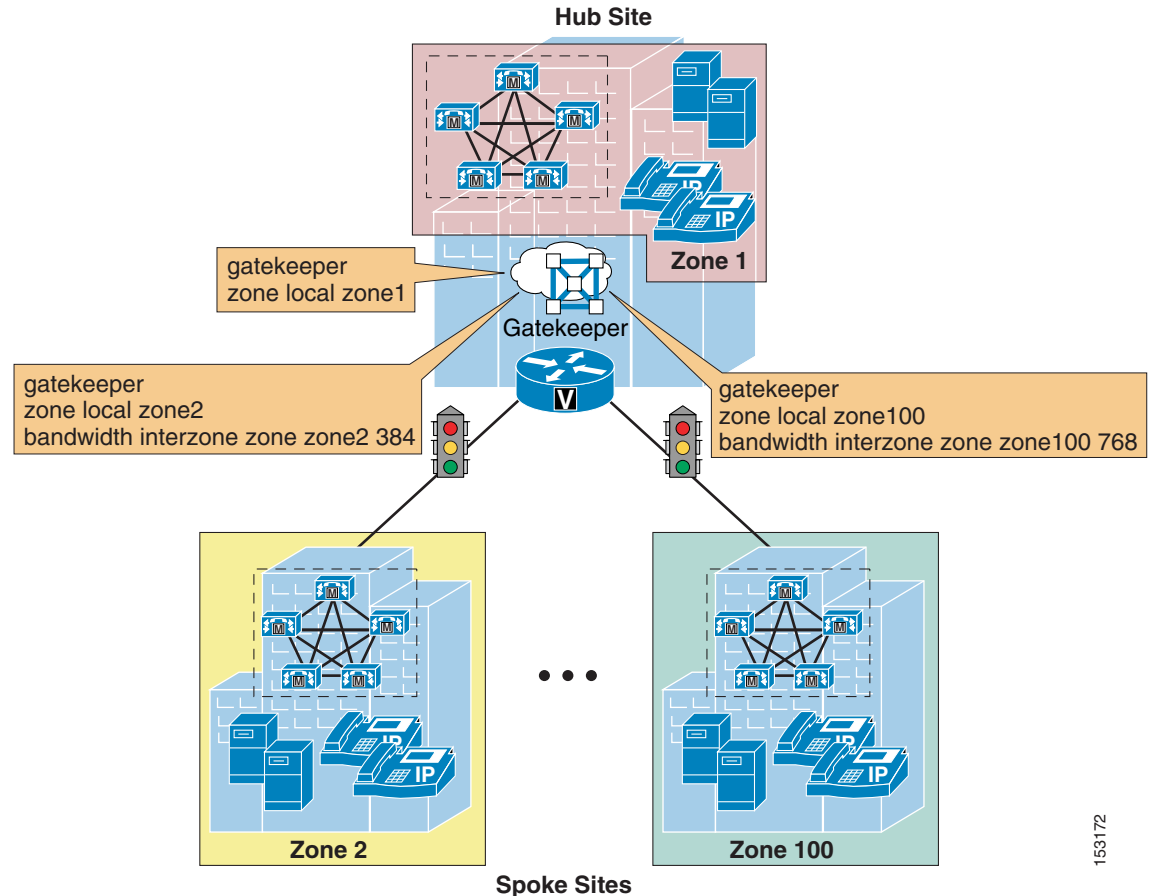
**Note**

If one or more sites have dual connectivity to the IP WAN and you want to take full advantage of the bandwidth available on both links, Cisco recommends that you deploy topology-aware call admission control, as described in the section on [Generic Topologies, page 9-50](#). See also [Limitations of Topology-Unaware Call Admission Control, page 9-5](#), for more information.

Distributed Unified CM 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 (which could be a Unified CM cluster, Cisco Unified Communications Manager Express (Unified CME), or an H.323 gateway) 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 9-24](#) 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.

Figure 9-24 Call Admission Control for Hub-and-Spoke Topologies Using a Gatekeeper

153172

Follow these guidelines when deploying call admission control with a gatekeeper:

- In Unified CM, configure an H.225 gatekeeper-controlled trunk if you have a mixed environment with Cisco Unified Communications Manager Express (Unified CME) and H.323 gateways.
- In Unified CM, configure an intercluster gatekeeper-controlled trunk if you have an environment exclusively based on Unified CM clusters.
- Ensure that the zone configured in Unified CM matches the correct gatekeeper zone for the site.
- Each Unified CM subscriber listed in the device pool's Unified CM Redundancy Group registers a gatekeeper-controlled trunk with the gatekeeper. (Maximum of three.)
- Calls are load-balanced across the registered trunks in the Unified CM cluster.
- Unified CM supports multiple gatekeepers and trunks.
- You can place the trunk in a route group and 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 Unified CMs, Unified CME, or H.323 gateways.
- Use the **bandwidth interzone** command on the gatekeeper to control bandwidth between Unified CM clusters, Unified CME 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.
- 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.

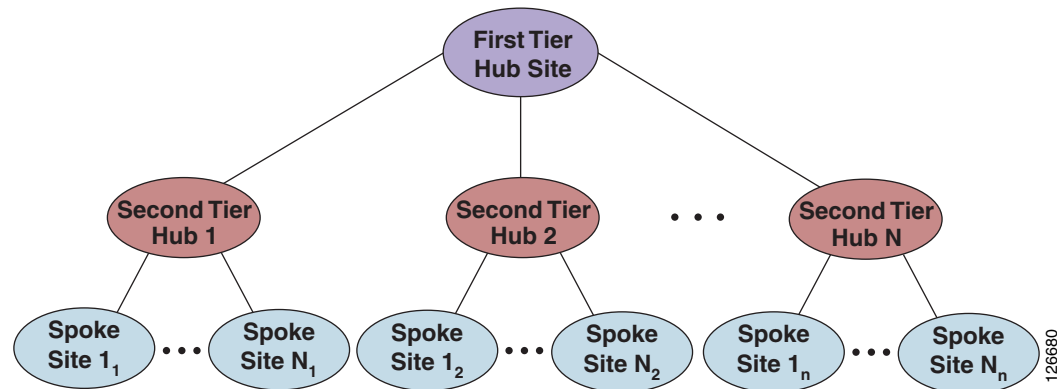
**Note**

If one or more sites have dual connectivity to the IP WAN and you want to take full advantage of the bandwidth available on both links, Cisco recommends that you deploy topology-aware call admission control, as described in the section on [Generic Topologies, page 9-50](#). See also [Limitations of Topology-Unaware Call Admission Control, page 9-5](#), for more information.

Two-Tier Hub-and-Spoke Topologies

Figure 9-25 depicts a two-tier hub-and-spoke topology. This type of network topology 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 communications 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 9-25 *A Two-Tier Hub-and-Spoke Topology*



The design considerations in this section apply to two-tier hub-and-spoke topologies that use 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 [Simple MPLS Topologies, page 9-44](#).

The remainder of this section contains design best practices for two-tier hub-and-spoke topologies according to the Unified CM deployment model adopted:

- [Centralized Unified CM Deployments, page 9-41](#)

One or more Unified CM 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.

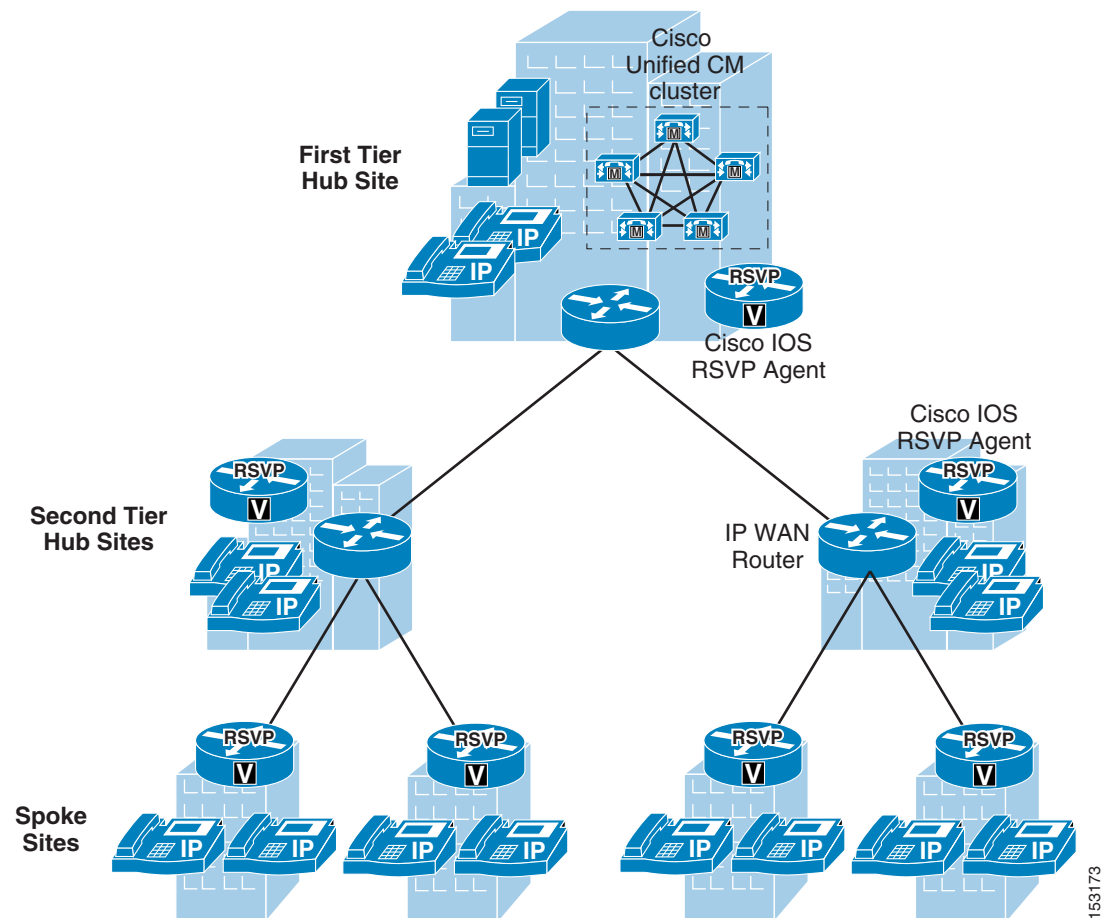
- [Distributed Unified CM Deployments, page 9-43](#)

Unified CM clusters are located at the first-tier hub site and at the second-tier hub sites, while only endpoints and gateways are located at the spoke sites.

Centralized Unified CM Deployments

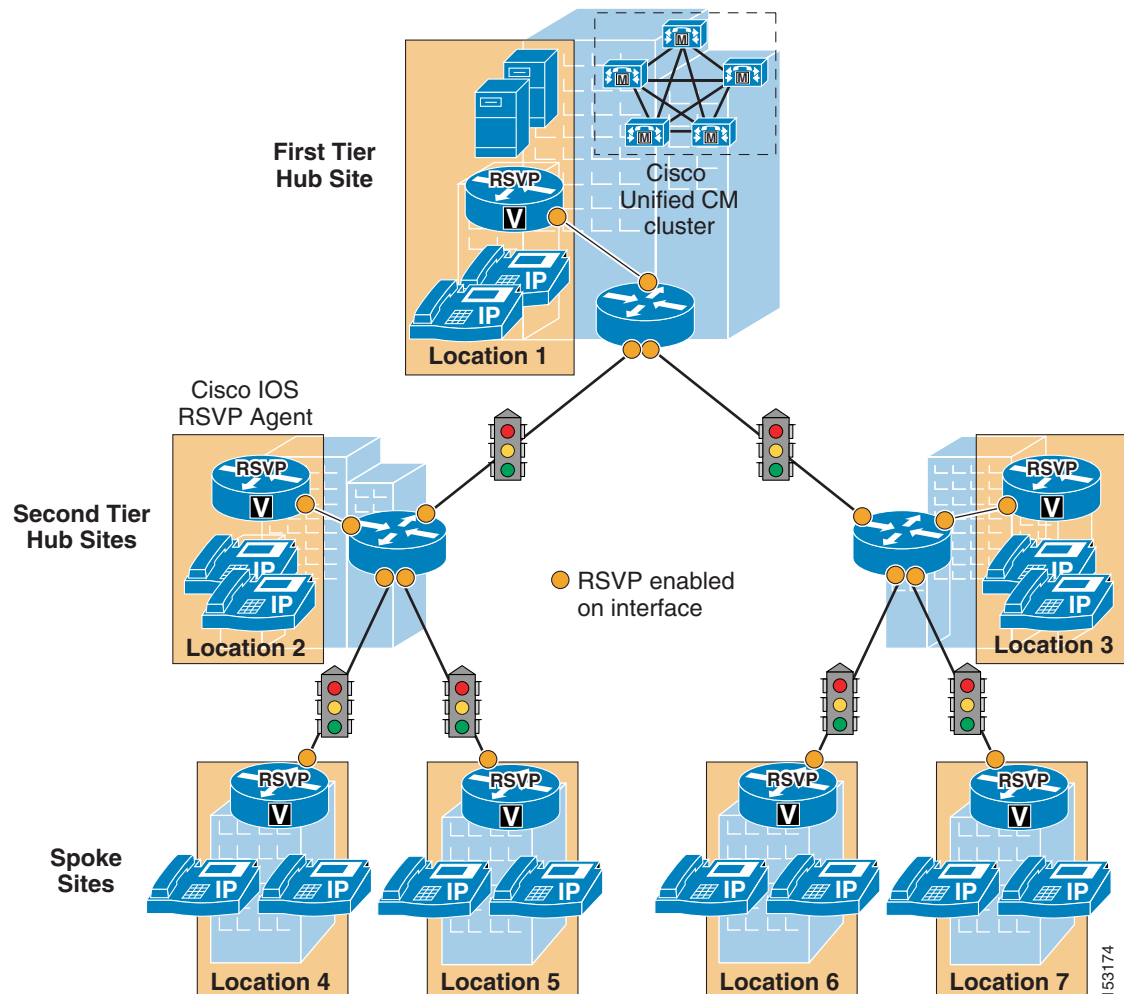
Figure 9-26 depicts a single centralized Unified CM cluster deployed in a two-tier hub-and-spoke IP WAN topology. In this scenario, the Unified CM cluster is located at the first-tier hub site, while all second-tier hub and spoke sites contain only endpoints and gateways.

Figure 9-26 Two-Tier Hub-and-Spoke Topology with Centralized Unified CM



This scenario requires that you deploy topology-aware call admission control, which for a single Unified CM cluster means using RSVP-enabled locations. Figure 9-27 shows how this mechanism can be deployed.

Figure 9-27 *Call Admission Control for Two-Tier Hub-and-Spoke Topologies Using Locations with RSVP*



The following guidelines apply to these deployments:

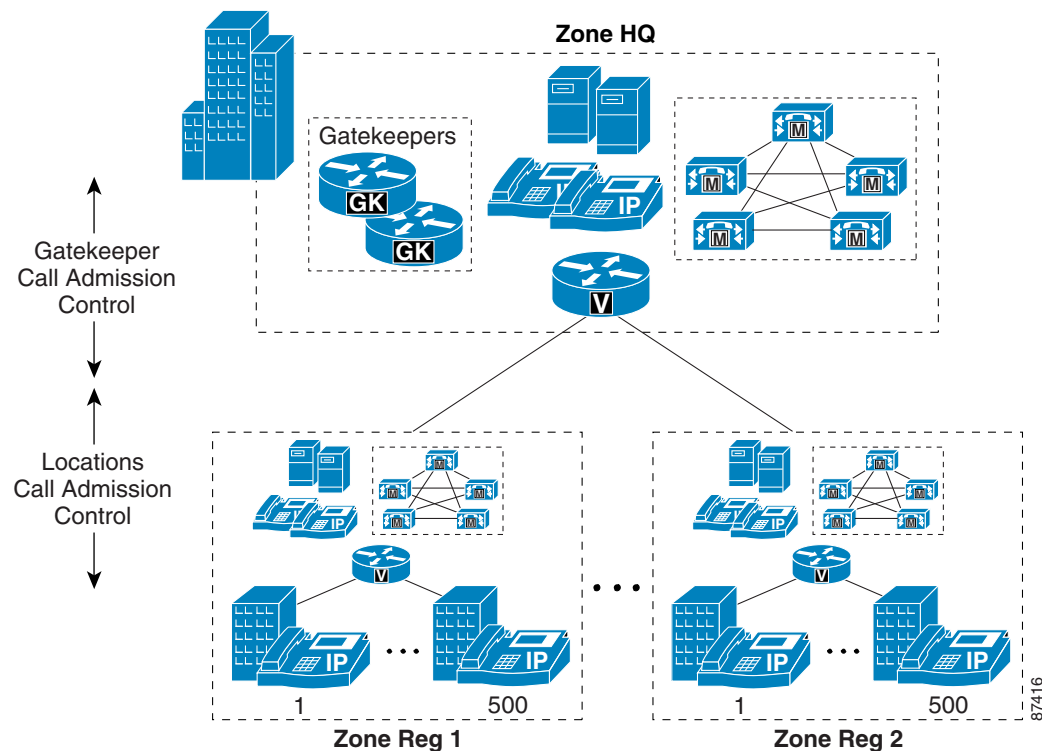
- Enable the Cisco IOS RSVP Agent feature on a Cisco IOS router at each site. At smaller sites, this router may coincide with the IP WAN router and PSTN gateway, while at larger sites they may be different platforms.
- In Unified CM, define a location for each site, and leave all bandwidth values as **Unlimited**.
- Assign all devices located at each site to the appropriate location (this includes endpoints, gateways, conferencing resources, and the Cisco RSVP Agent itself).
- Ensure that each Cisco RSVP Agent belongs to a media resource group (MRG) contained in the media resource group list (MRGL) of all devices at that site.
- In the Unified CM service parameters, set the **Default inter-location RSVP Policy** to **Mandatory** or **Mandatory (video desired)** and set the **Mandatory RSVP mid-call error handle option** to **Call fails following retry counter exceeded**.

- Enable RSVP on every WAN router interface in the network where congestion might occur, and configure the RSVP bandwidth based on the provisioning of the priority queue.
- If the Cisco RSVP Agent is not co-resident with the IP WAN router, enable RSVP on the LAN interfaces connecting the agent to the WAN router (as illustrated in [Figure 9-27](#)).

Distributed Unified CM Deployments

To provide call admission control in deployments that use a two-tier hub-and-spoke topology, with Unified CMs at the first-tier and second-tier hub sites, you can combine the static locations and gatekeeper zone mechanisms as illustrated in [Figure 9-28](#).

Figure 9-28 Combining the Locations and Gatekeeper Mechanisms for Call Admission Control



Follow these recommendations when combining gatekeeper zones with static locations for call admission control:

- Use call admission control based on static locations for sites with no local Unified CM (that is, the spoke sites).
- Use gatekeeper-based call admission control between Unified CM clusters (that is, between the first-tier hub site and the second-tier hub sites).
- For each site without a local Unified CM, configure a location for that site in the Unified CM cluster supporting the site.
- Configure the appropriate bandwidth limits for voice and video calls at 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 Unified CM to a location. If you move a device to another location, change its location configuration as well.

- Unified CM supports up to 500 locations.
- Each Unified CM cluster registers a gatekeeper-controlled trunk with the gatekeeper.
- On the gatekeeper, configure a zone for each Unified CM cluster, and use the **bandwidth interzone** command to control the number of calls to and from each cluster.

**Note**

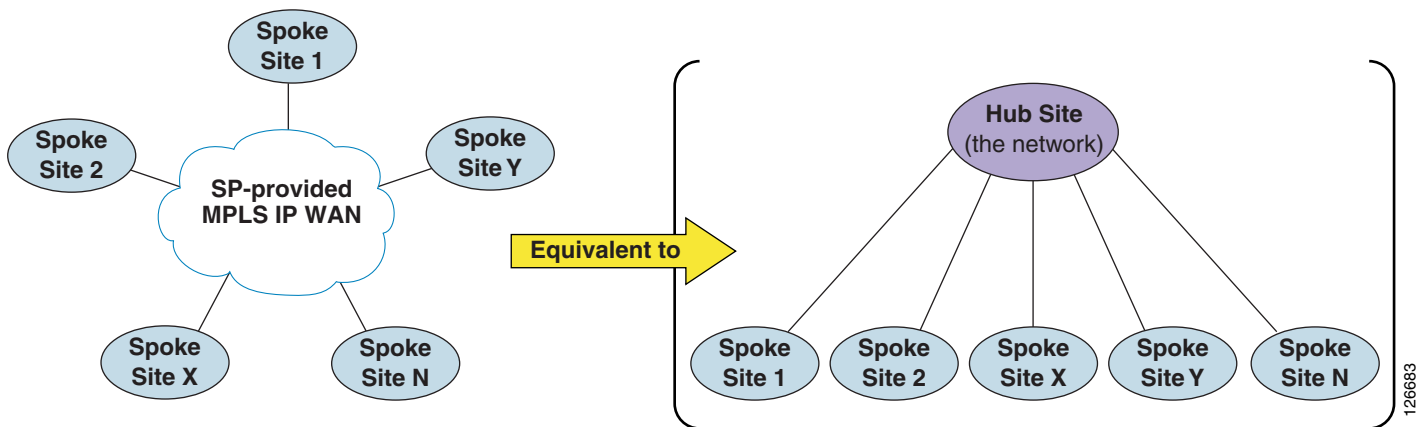
If one or more sites have dual connectivity to the IP WAN and you want to take full advantage of the bandwidth available on both links, Cisco recommends that you deploy topology-aware call admission control, as described in the section on [Generic Topologies, page 9-50](#). See also [Limitations of Topology-Unaware Call Admission Control, page 9-5](#), for more information.

Simple MPLS Topologies

Figure 9-29 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 9-29 MPLS IP WAN from a Service Provider, and Its Topology Equivalent



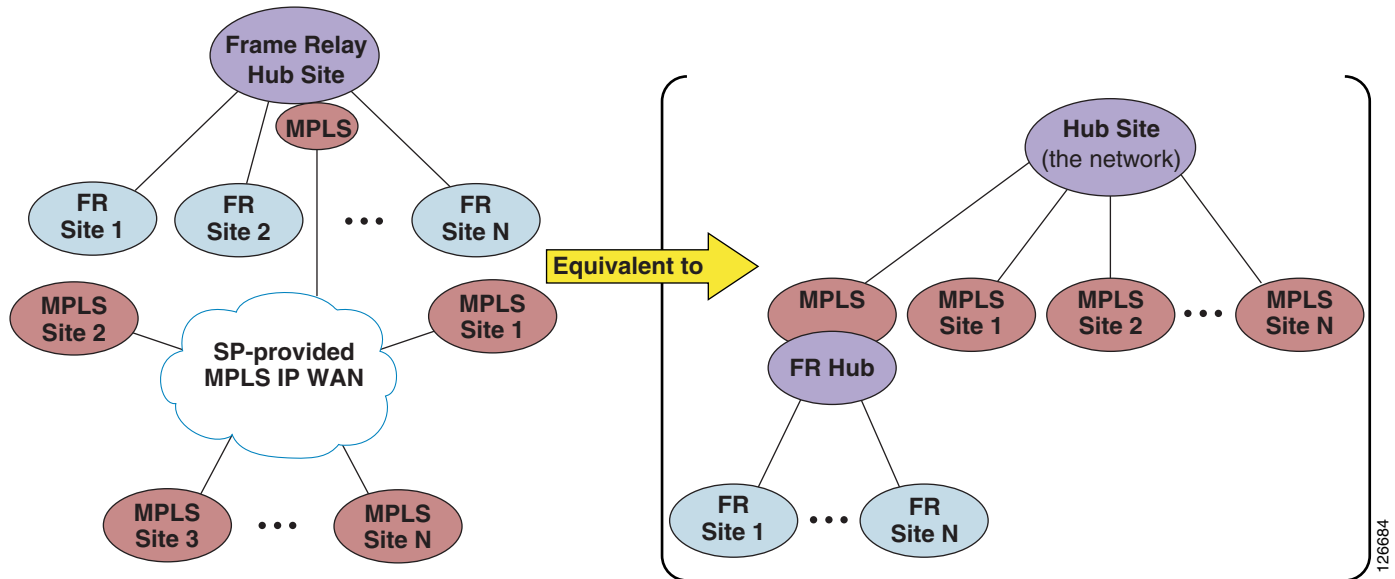
Based on these considerations, it is easy to see that, from a call admission control perspective, a service-provider IP WAN service based on MPLS is in reality equivalent to a hub-and-spoke topology without a hub site. (See Figure 9-29.) In fact, the network itself could be considered as the hub site, 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 multisite 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 9-30](#), 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 9-40](#).

Figure 9-30 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 Unified CM deployment model adopted:

- [Centralized Unified CM Deployments](#), [page 9-46](#)

One or more Unified CM clusters are located at only one site, while only endpoints and gateways are located at all other sites.

- [Distributed Unified CM Deployments](#), [page 9-48](#)

Unified CM clusters are located at multiple sites, while endpoints and gateways are located at all other 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.

**Note**

If one or more sites have dual connectivity to the IP WAN and you want to take full advantage of the bandwidth available on both links, Cisco recommends that you deploy topology-aware call admission control, as described in the section on [Generic Topologies, page 9-50](#). Particular care must be taken to guarantee symmetric routing in the presence of load-balanced links. See also [Limitations of Topology-Unaware Call Admission Control, page 9-5](#), and [Special Considerations for MPLS Networks, page 9-11](#), for more information, and contact your local Cisco account team for further assistance.

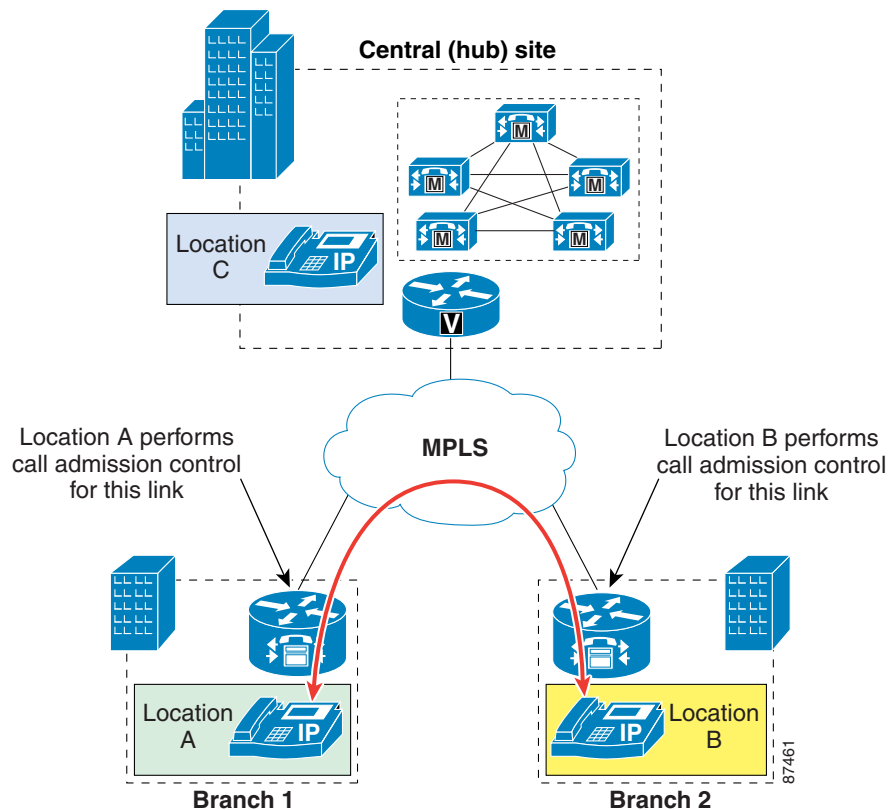
Centralized Unified CM Deployments

In multisite WAN deployments with centralized call processing in an MPLS topology, use Unified CM static *locations* for implementing call admission control.

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 Unified CM locations configuration, devices at the central site are left in the Hub_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 9-31](#) illustrates a spoke-to-spoke call between two branch sites in this type of deployment.

Figure 9-31 Spoke-to-Spoke Calls in an MPLS Deployment

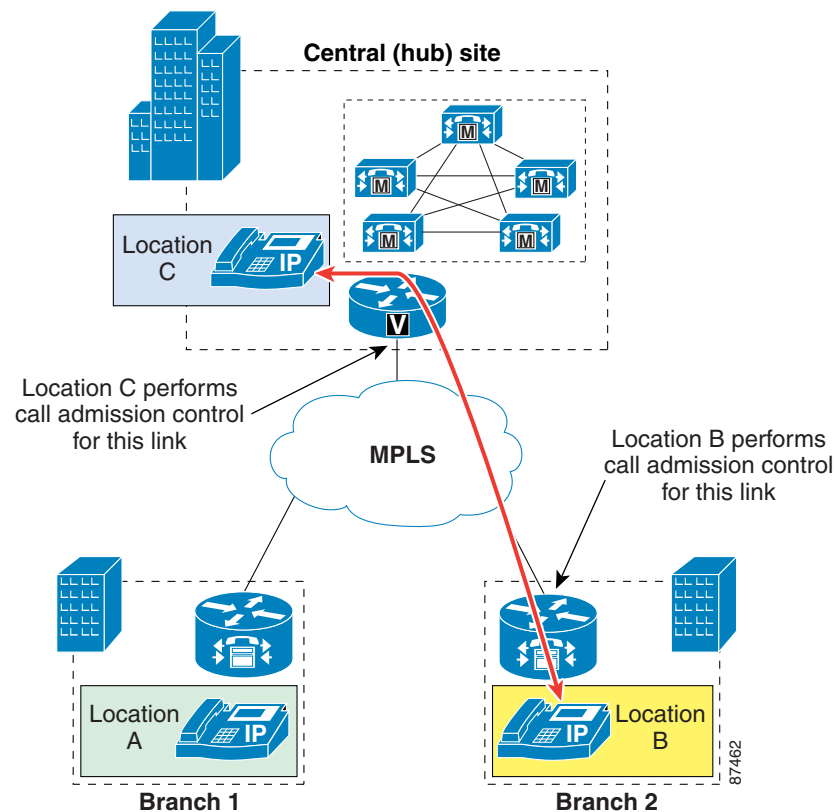


Also, in an MPLS WAN, the link connecting the central site to the WAN does not aggregate every branch's WAN link. By placing all the central site devices in their own call admission control location (that is, not in the Hub_None location), this configuration requires that call admission control be performed on the central site link independently of the branch links. (See [Figure 9-32](#).)

**Note**

Some devices such as trunks do not terminate media and are normally left in the Hub_None location. However, to avoid errors in call admission control when an MTP is required on a trunk, the trunk must be assigned to a location other than Hub_None and any MTP in the trunk's MRGL must be physically located at the site associated with that location. This configuration is required because an MTP cannot be assigned a location directly, so it inherits the location of the device that selected it.

Figure 9-32 Calls to and from the Hub in an MPLS Deployment

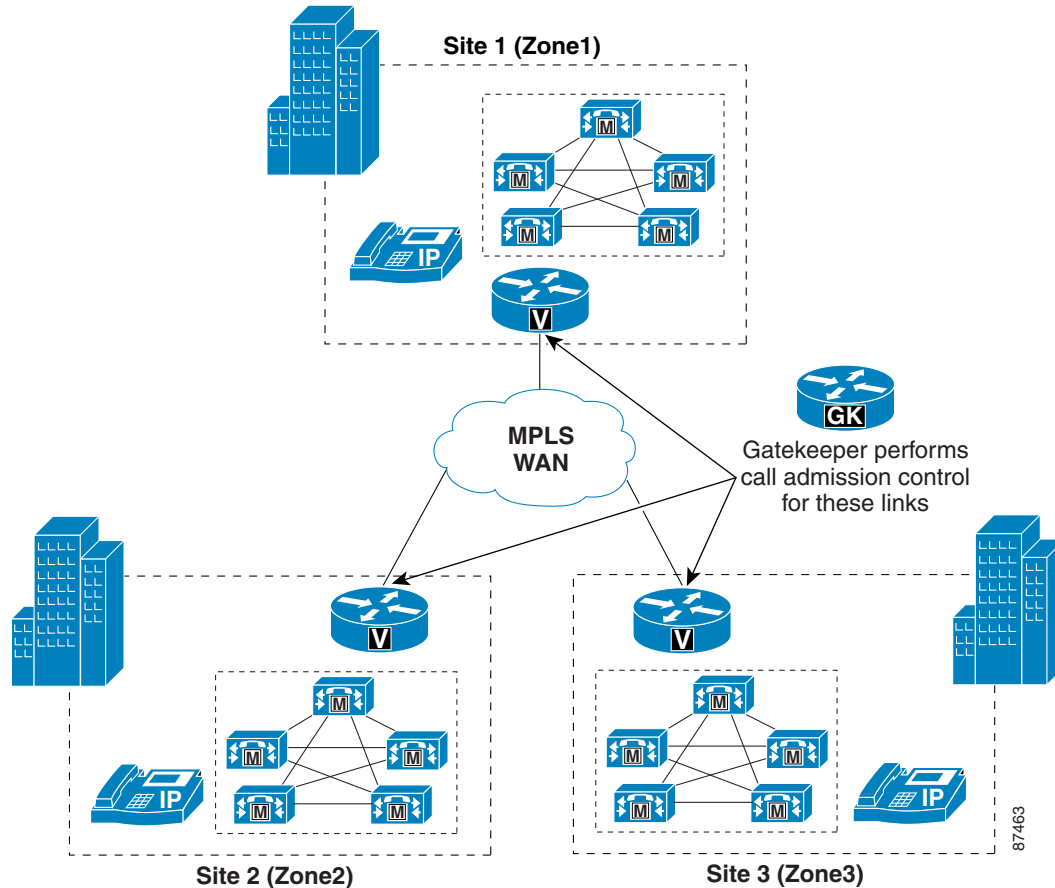


When all the available bandwidth for a particular site has been utilized, you can provide automatic failover to the PSTN by using the automated alternate routing (AAR) feature within Unified CM. (For more information on AAR, see [Automated Alternate Routing](#), page 10-27.)

Distributed Unified CM Deployments

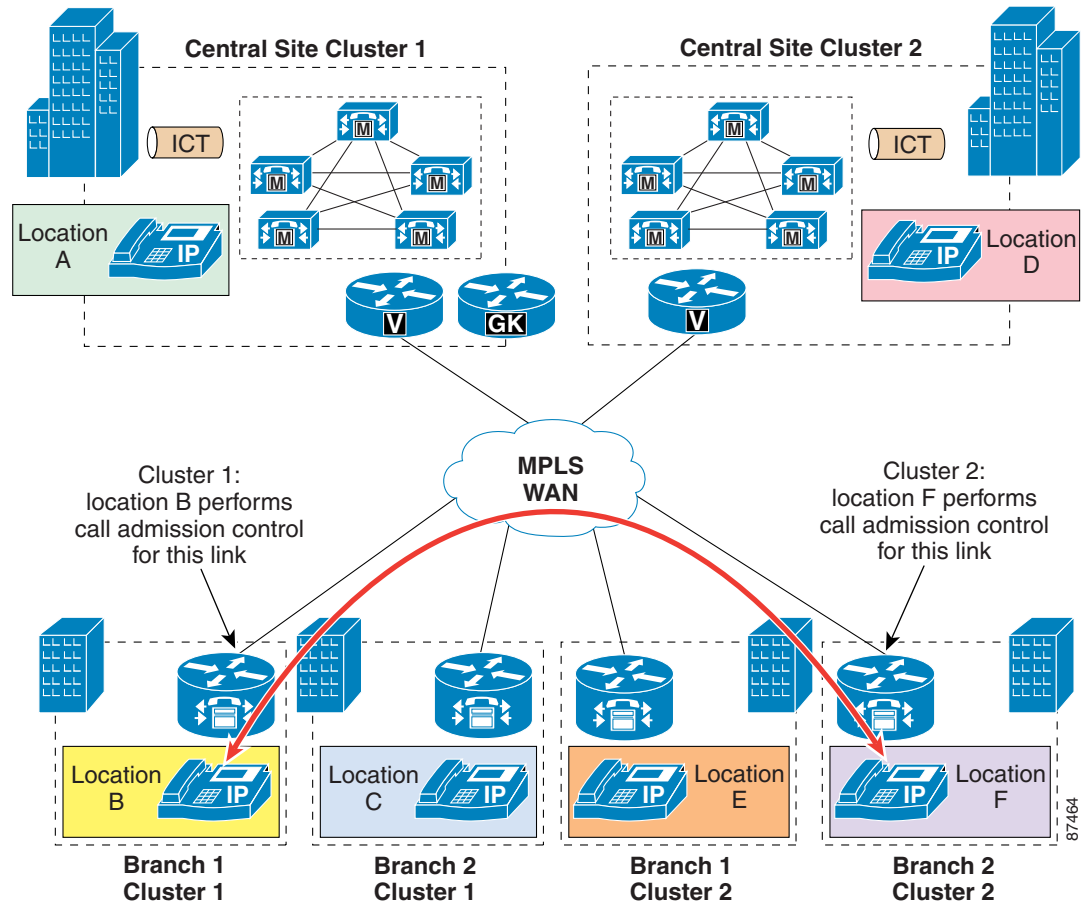
In multisite deployments where a Unified CM cluster is present at more than one site without any branch locations and the sites are linked through an MPLS WAN, a gatekeeper can provide dial-plan resolution as well as call admission control between the sites, with each site being placed in a different gatekeeper zone. This is the same mechanism adopted for hub-and-spoke topologies based on Layer 2 WAN technologies. (See [Figure 9-33](#).)

Figure 9-33 Gatekeeper Call Admission Control in a Distributed Deployment with MPLS



In deployments where branch sites are required, a gatekeeper can be used for dial-plan resolution between clusters, but a gatekeeper is not recommended for call admission control.

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 9-34](#).)

Figure 9-34 Multiple Clusters Connected by Intercluster Trunks (ICTs)

As in the centralized Unified CM 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 Hub_None. The exception is when the trunk requires an MTP, in which case the trunk and MTP should both be in the location of the site in which they reside.

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:

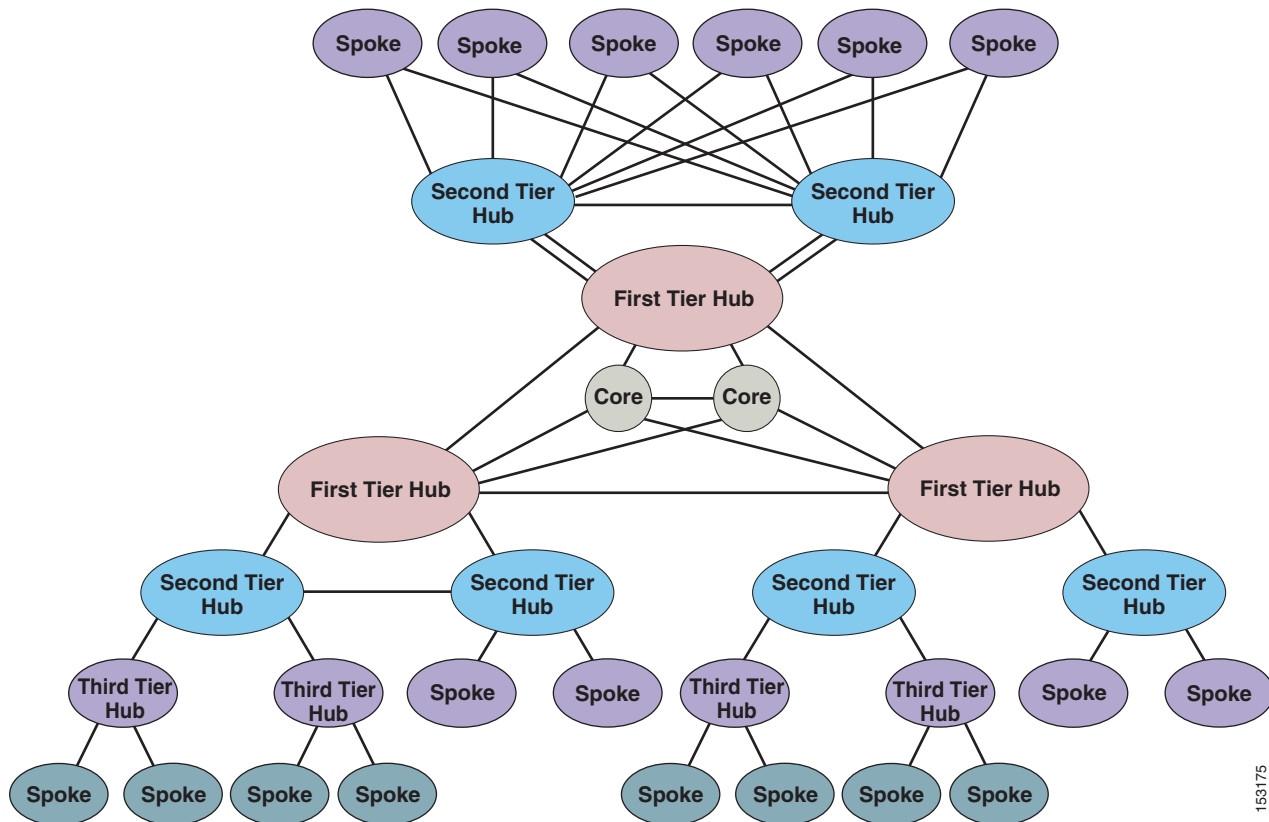
- The route list and route group construct for calls across multiple Unified CM clusters
- The automated alternate routing (AAR) feature for calls within a Unified CM cluster (For more information on AAR, see [Automated Alternate Routing, page 10-27.](#))

Generic Topologies

In the context of this chapter, a generic topology is a network topology that cannot be reduced to a simple or two-tier hub-and-spoke or to a simple MPLS-based network.

As [Figure 9-35](#) illustrates, a generic topology can present full-mesh features, hub-and-spoke features, partial-mesh features, or possibly all of them combined in a single network. It may also present dual connections between sites, as well as multiple paths from one site to another.

Figure 9-35 A Generic Topology



The complex nature of these networks requires the adoption of topology-aware call admission control mechanisms based on RSVP. In particular, these mechanisms can properly control bandwidth in presence of any of the following topology aspects:

- Remote sites dual-homed to different hub sites
- Multiple IP WAN links between any two sites, either in a primary/backup configuration or in an active/active load-balanced configuration
- Redundant hubs or data centers with a dedicated connection
- Fully-meshed core networks
- Multiple equal-cost IP paths between any two sites
- Multi-tiered architectures

The remainder of this section contains design best practices for generic network topologies according to the Unified CM deployment model adopted:

- [Centralized Unified CM Deployments, page 9-51](#)

One or more Unified CM clusters are located at a given site, but only endpoints and gateways are located at all other sites.

- [Distributed Unified CM Deployments, page 9-54](#)

Unified CM clusters are located at multiple sites, and endpoints and gateways are located at all other sites.

Centralized Unified CM Deployments

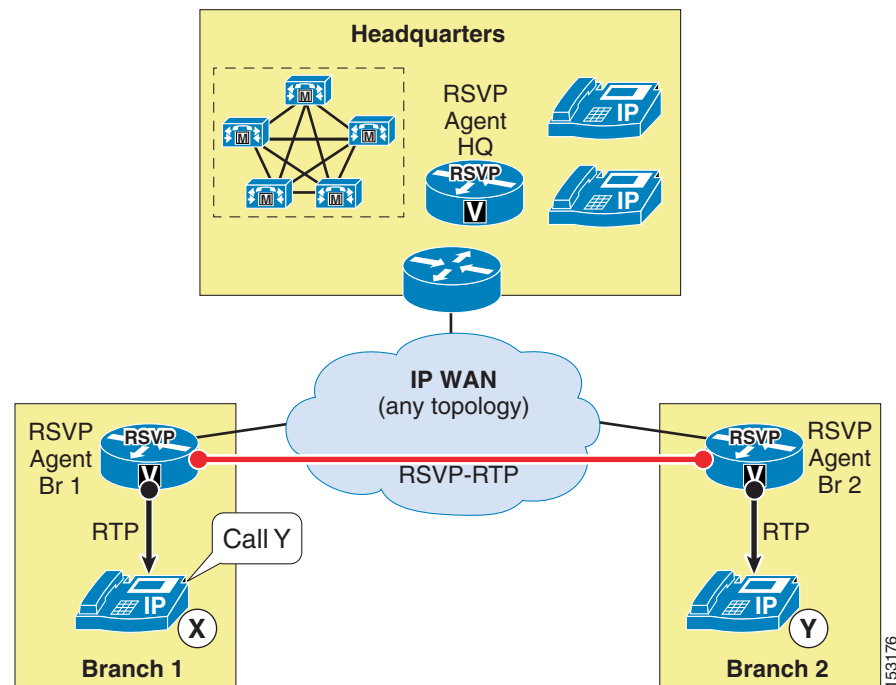
Centralized Unified CM deployments using a generic topology can be categorized into two sub-types:

- [Single Unified CM Cluster, page 9-51](#)
- [Co-Located Unified CM Clusters, page 9-52](#)

Single Unified CM Cluster

The recommendations in this section apply to a single Unified CM cluster deployed in a generic network topology, as illustrated in [Figure 9-36](#).

Figure 9-36 *A Single Unified CM Cluster in a Generic Topology*



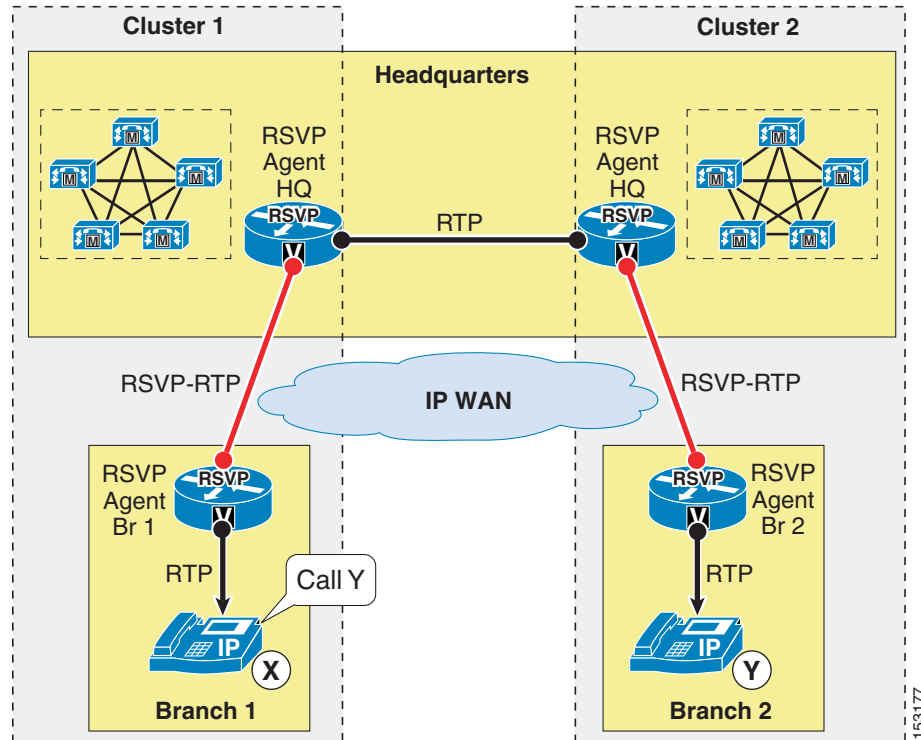
The following guidelines apply to this type of deployment:

- Enable the Cisco IOS RSVP Agent feature on a Cisco IOS router at each site, including the central site where Unified CM resides. At smaller sites, this router may coincide with the IP WAN router and PSTN gateway, while at larger sites they may be different platforms.
- In Unified CM, define a location for each site, and leave all bandwidth values as **Unlimited**.
- Assign all devices located at each site to the appropriate location (this includes endpoints, gateways, conferencing resources, and the Cisco RSVP Agents themselves).
- Ensure that each Cisco RSVP Agent belongs to a media resource group (MRG) contained in the media resource group list (MRGL) of all devices at that site.
- In the Unified CM service parameters, set the **Default inter-location RSVP Policy** to **Mandatory** or **Mandatory (video desired)** and set the **Mandatory RSVP mid-call error handle option** to **Call fails following retry counter exceeded**.
- Enable RSVP on every WAN router interface in the network where congestion might occur, and configure the RSVP bandwidth based on the provisioning of the priority queue. (See [Additional Considerations for Bearer Traffic with RSVP, page 3-54](#).)
- If you need to provision bandwidth separately for voice and video calls, also configure an RSVP application ID on the same WAN router interfaces.
- If the Cisco RSVP Agent is not co-resident with the IP WAN router, enable RSVP on the LAN interfaces connecting the agent to the WAN router.

Co-Located Unified CM Clusters

The recommendations in this section apply to deployments where multiple Unified CM clusters are located on the same LAN or MAN. However, the same considerations may also be valid if the sites where the Unified CM clusters reside are connected via a high-speed link, provided that no congestion occurs in the priority queue of the link and that bandwidth for voice and video can therefore be considered unlimited.

[Figure 9-37](#) illustrates a deployment with two Unified CM clusters located at a given site (HQ) and a number of remote sites with endpoints and gateways, which are controlled either by Cluster 1 (for example, Branch 1) or Cluster 2 (for example, Branch 2).

Figure 9-37 Co-located Unified CM Clusters in a Generic Topology

In addition to the guidelines listed in [Single Unified CM Cluster, page 9-51](#), observe the following best practices for this type of deployment:

- For each cluster, define an intercluster trunk to enable communications to the other clusters. A gatekeeper may be used for dial plan resolution, but it is not needed for call admission control.
- Assign the intercluster trunk to the same location used for all devices located at the central site (HQ in the example of [Figure 9-37](#)).
- Ensure that the intercluster trunk is assigned to a device pool that specifies an MRGL that in turn points to an MRG containing the Cisco RSVP Agent located at the central site (Cisco RSVP Agent HQ1 in the case of Cluster 1 in [Figure 9-37](#)).
- Use the AAR feature to provide automatic PSTN failover in case of call admission control failure within a cluster.
- Use the route list and route group constructs to provide automatic PSTN failover in case of call admission control failure across clusters.
- Both media and signaling traffic are hair-pinned via the central site for calls between two branch sites belonging to different clusters (as shown in [Figure 9-37](#), where the call between phone X in Branch 1 and phone Y in Branch 2 is hair-pinned via the HQ site).

Distributed Unified CM Deployments

In order to provide call admission control for distributed Unified CM deployments in a generic network topology, two approaches are possible, depending on the number of Unified CM clusters involved:

- [Remote Cisco RSVP Agent Approach, page 9-55](#)

This solution applies to deployments where three or fewer Unified CM clusters are located at different sites connected via an IP WAN with limited bandwidth.

- [IP-IP Gateway Approach, page 9-58](#)

This solution applies to deployments where any number of Unified CM clusters are located at different sites connected via an IP WAN with limited bandwidth.

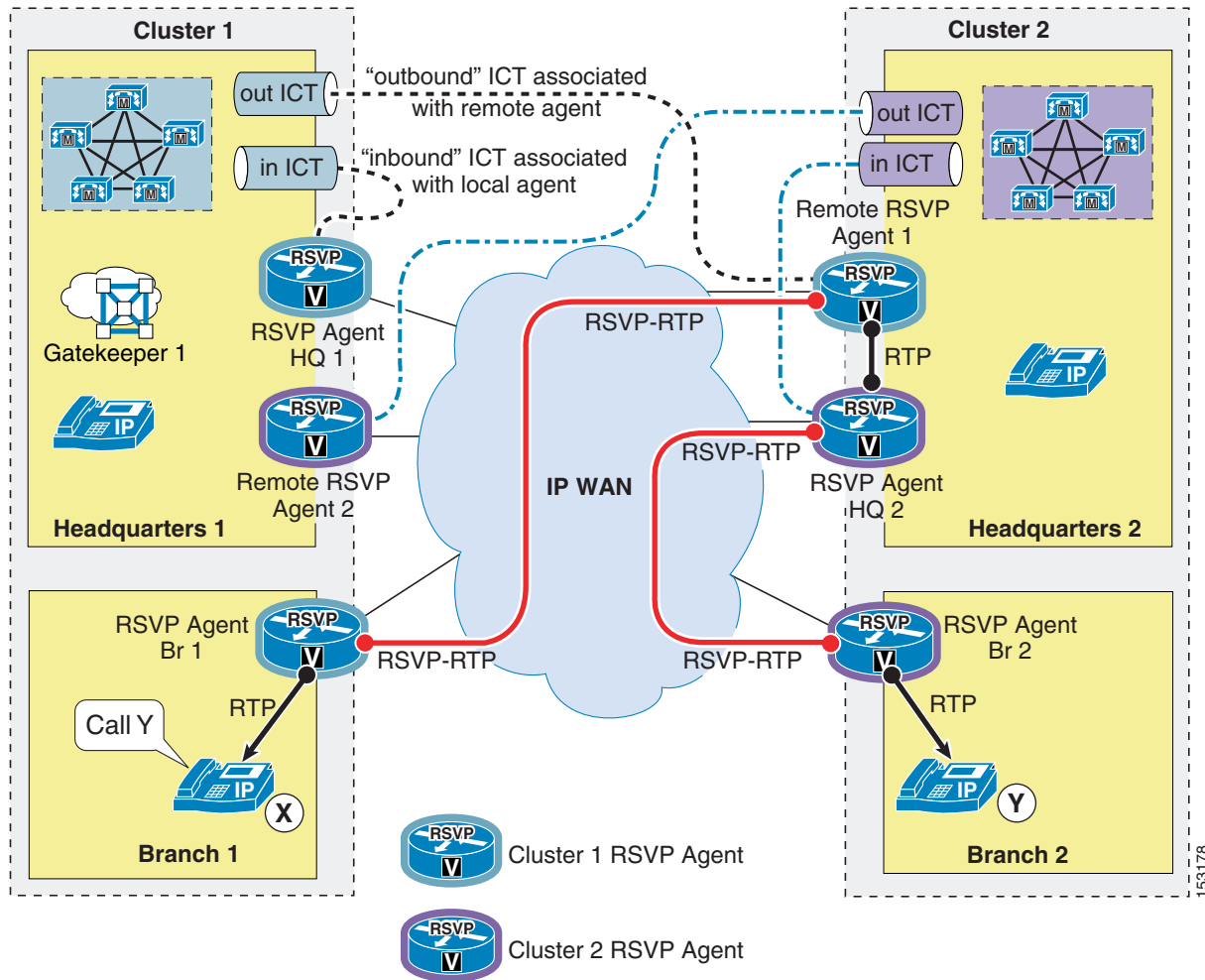
**Note**

In deployments where the Unified CM clusters are located at sites connected via a high-speed IP WAN link, it is possible to treat this scenario similar to that described in the section on [Co-Located Unified CM Clusters, page 9-52](#), provided that congestion does not occur in the priority queue of the IP WAN link.

Remote Cisco RSVP Agent Approach

To provide call admission control in generic topologies where three or fewer Unified CM clusters are located at different sites, you can extend the concept of RSVP-enabled locations to cover intercluster calls by defining "remote" Cisco RSVP Agents, as shown in [Figure 9-38](#).

Figure 9-38 Remote Cisco RSVP Agent Approach for Distributed Clusters in a Generic Topology



Note

For simplicity, the descriptions in this section are based on an example with two Unified CM clusters, as shown in [Figure 9-38](#). For deployments with three Unified CM clusters, some additional remarks are provided at the end of the section.

In addition to the guidelines listed in the section on [Single Unified CM Cluster, page 9-51](#), observe the following best practices for these deployments:

- For each cluster, define two intercluster trunks (ICTs) to enable communications to the other cluster: one "outbound" intercluster trunk and one "inbound" intercluster trunk.
- Configure a Cisco IOS gatekeeper for dial plan resolution (not for call admission control), defining one zone per Unified CM cluster. For example:

```
gatekeeper
  zone local cluster1 customer.com 10.10.10.10
  zone local cluster2 customer.com
```

- For each cluster, register the inbound trunk with the gatekeeper in the normal zone for that cluster. (For example, the inbound trunk of Cluster 1 registers in zone cluster1, while the inbound trunk of Cluster 2 registers with zone cluster2.)
- For each cluster, register the outbound trunk with the gatekeeper in specially created zones. For example:

```
gatekeeper
  zone local cluster1 customer.com 10.10.10.10
  zone local cluster2 customer.com
  zone local cluster1-to-cluster2 customer.com
  zone local cluster2-to-cluster1 customer.com
```

- Set up your Unified CM dial plan so that outbound calls to the other cluster use the outbound intercluster trunk. (For example, in Cluster 1, have a 2XXX route pattern that points to the outbound trunk via a route list and route group construct.)
- Set up your gatekeeper dial plan so that calls destined to a given cluster are routed to its inbound trunk. For example:

```
gatekeeper
  zone local cluster1 customer.com 10.10.10.10
  zone local cluster2 customer.com
  zone local cluster1-to-cluster2 customer.com
  zone local cluster2-to-cluster1 customer.com
  zone prefix cluster1 1...
  zone prefix cluster2 2...
```

- Assign the inbound trunk to the same location as all devices located at that site (for example, location HQ1 for Cluster 1's inbound trunk and location HQ2 for Cluster 2's inbound trunk).
- Assign the outbound trunk to a newly created location (for example, location remote-to-HQ2 for Cluster 1's outbound trunk toward Cluster 2, and location remote-to-HQ1 for Cluster 2's outbound trunk toward Cluster 1).
- At each of the two sites where Unified CM resides, have an instance of Cisco RSVP Agent registered with the local cluster (for example, Cisco RSVP Agent HQ1 at site HQ1 and Cisco RSVP Agent HQ2 at site HQ2).
- Assign these local Cisco RSVP Agents to the same location as all devices located at that site (for example, location HQ1 for Cisco RSVP Agent HQ1 and location HQ2 for Cisco RSVP Agent HQ2).
- For each cluster, assign the local Cisco RSVP Agent to an MRG contained in the MRGL of all devices located at the central site, including the MRGL used by the inbound trunk via its device pool.
- At each of the two sites where Unified CM clusters reside, add an instance of Cisco RSVP Agent registered with the other Unified CM cluster. (For example, Remote Cisco RSVP Agent 1 is registered with Cluster 1 and placed at site HQ2, where Cluster 2 resides, while Remote Cisco RSVP Agent 2 is registered with Cluster 2 and is placed at site HQ1, where Cluster 1 resides).

- Assign these remote Cisco RSVP Agents to the locations created for the outbound trunks (for example, location remote-to-HQ2 within Cluster 1 for Remote Cisco RSVP Agent 1 and location remote-to-HQ1 within Cluster 2 for Remote Cisco RSVP Agent 2).
- For each cluster, assign the remote Cisco RSVP Agent to an MRG contained in the MRGL used by the outbound trunk (via its device pool).

**Note**

While logically separate, the remote Cisco RSVP Agent instances may reside on the same router platform as the local Cisco RSVP Agent registered to the other cluster. For the example in [Figure 9-38](#), this means that Remote Cisco RSVP Agent 1 and Cisco RSVP Agent HQ2 may actually be located on the same router platform, and the same applies to Remote Cisco RSVP Agent 2 and Cisco RSVP Agent HQ1.

In [Figure 9-38](#), when phone X at Branch 1 places a call to phone Y at Branch 2, Unified CM Cluster 1 will route the call to the gatekeeper via the outbound trunk. Because phone X is assigned to location Branch 1 and the outbound trunk is associated to the location remote-to-HQ2, Unified CM Cluster 1 will initiate an RSVP reservation across the IP WAN between Cisco RSVP Agent Br 1 and Remote Cisco RSVP Agent 1. (The latter is located at HQ2, together with Cluster 2.)

The gatekeeper will then route the call to the inbound trunk of Cluster 2, based on its zone prefix configuration.

Unified CM Cluster 2 then receives a call from the inbound trunk (associated with location HQ1) and phone Y (associated with location Branch 2), so it will initiate another RSVP reservation across the IP WAN between Cisco RSVP Agent HQ2 and Cisco RSVP Agent Br 2.

The call is established across four call legs, two of which traverse the IP WAN and are RSVP-enabled.

**Note**

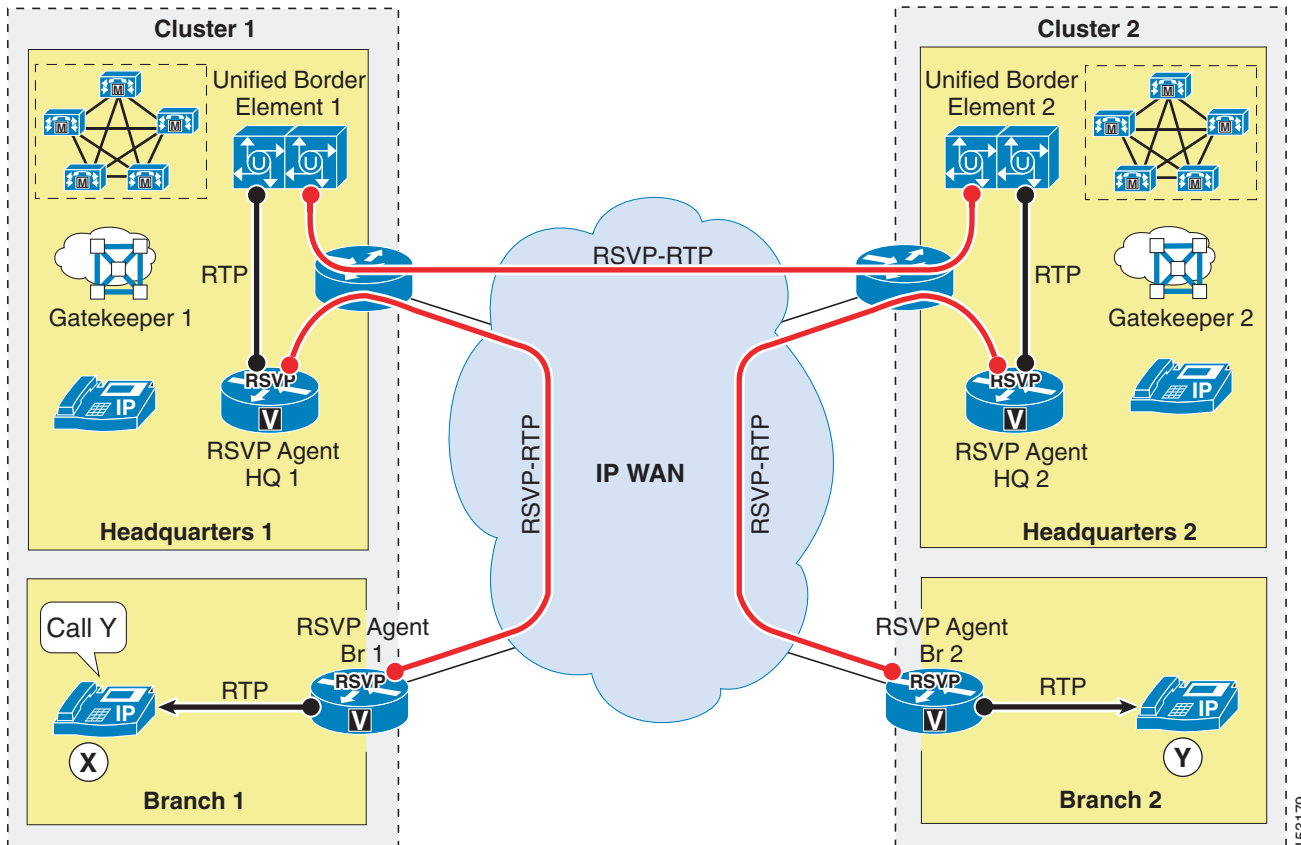
In presence of deployments involving three Unified CM clusters, the same considerations apply, except that instead of having a single outbound trunk per cluster, you need two of them, one for each of the other two clusters that will be called. Similarly, each cluster needs two remote Cisco RSVP Agents, each located with one of the other two clusters.

IP-IP Gateway Approach

The complexity of the remote Cisco RSVP Agent approach described in the previous section quickly increases with the number of Unified CM clusters, and it is therefore limited to a maximum of three clusters.

To provide call admission control in generic topologies where more than three Unified CM clusters are located at different sites, you can combine RSVP-enabled locations for calls within a cluster with RSVP-enabled IP-IP gateways for calls between clusters, as shown in Figure 9-39.

Figure 9-39 IP-IP Gateway Approach for Distributed Clusters in a Generic Topology



In addition to the guidelines listed in the section on [Single Unified CM Cluster](#), page 9-51, observe the following best practices for these deployments:

- For each cluster, define a gatekeeper-controlled intercluster trunk to enable communications to the other clusters. (Gatekeeper zones are used for dial plan resolution but are not needed for call admission control in this scenario.)
- Assign the intercluster trunk to the same location used for all devices located at the central site for that cluster.
- Ensure that the intercluster trunk is assigned to a device pool that specifies an MRGL that in turn points to an MRG containing the Cisco RSVP Agent located at the central site (for example, Cisco RSVP Agent HQ1 for Cluster 1 in Figure 9-39).
- For each cluster, place an IP-IP gateway at the central site for that cluster and enable it to use RSVP for VoIP calls across the IP WAN.

- For each cluster, configure a gatekeeper as a via-zone gatekeeper so that it invokes the local IP-IP gateway for all calls in or out of the respective zone. (Note that the gatekeeper can be co-resident with the IP-IP gateway.)
- Use the AAR feature to provide automatic PSTN failover in case of call admission control failure within a cluster.
- Use the route list and route group constructs to provide automatic PSTN failover in case of call admission control failure across clusters.
- For calls between two branch sites belonging to different clusters, both media and signaling traffic are hair-pinned via the central sites of the respective clusters (as shown in [Figure 9-39](#), where the call between phone X in Branch 1 and phone Y in Branch 2 is hair-pinned via the HQ1 and HQ2 sites).

**Note**

While logically separate, the Cisco RSVP Agents, gatekeeper, and IP-IP gateway functions may reside on the same router platform. For example, in the scenario shown in [Figure 9-39](#), IPIPGW 1, Gatekeeper 1, and Cisco RSVP Agent HQ1 may be resident on the same router platform, as may IPIPGW 2, Gatekeeper 2, and Cisco RSVP Agent HQ2.

In [Figure 9-39](#), when phone X at Branch 1 places a call to phone Y at Branch 2, Unified CM Cluster 1 will route the call to Gatekeeper 1 via the intercluster trunk. Because phone X is assigned to location Branch 1 and the intercluster trunk is associated to the location HQ1, Unified CM Cluster 1 will initiate an RSVP reservation across the IP WAN between Cisco RSVP Agent Br 1 and Cisco RSVP Agent HQ1.

Gatekeeper 1 then routes the call to IPIPGW 1 based on the via-zone configuration, and IPIPGW 1 establishes an RSVP reservation with IPIPGW 2 across the IP WAN. IPIPGW 2 in turn contacts Unified CM Cluster 2 via Gatekeeper 2.

Unified CM Cluster 2 then receives a call from the intercluster trunk associated with location HQ2 and directed to phone Y (associated with location Branch 2), so it will initiate another RSVP reservation across the IP WAN between Cisco RSVP Agent HQ2 and Cisco RSVP Agent Br 2.

The call is established across seven call legs, three of which traverse the IP WAN and are RSVP-enabled.



CHAPTER 10

Dial Plan

Last revised on: September 27, 2007

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 10-8](#)

This section provides detailed explanations of the elements of a Cisco Unified Communications dial plan. Covered topics include call routing logic, calling privileges, and digit manipulation techniques for various Cisco products. (Prerequisite: A working knowledge of Cisco Unified Communications Manager and Cisco IOS is recommended.)

- [Design Considerations, page 10-54](#)

This section contains design and deployment guidelines related to multisite IP Telephony networks, endpoint addressing methods, approaches to building classes of service, and call coverage functionality. (Prerequisite: A working knowledge of Cisco Unified Communications Manager and Cisco IOS is recommended.)

For more details, refer to the *Cisco Unified Communications Manager System Guide*, the *Cisco IOS Voice, Video, and Fax Configuration Guide, Release 12.2*, and other product documentation available at <http://www.cisco.com>

For purposes of this chapter only, we define two types of IP phones:

- Type-A
Includes the Cisco Unified IP Phone 7905, 7912, 7940, and 7960.
- Type-B
Includes the Cisco Unified IP Phone 7911, 7941, 7961, 7970, and 7971.

Type-A phones differ somewhat from Type-B phones in their behavior, and Type-B phones offer support for Key Press Markup Language (KPML) but Type-A phones do not. (See [User Input on Type-A SIP Phones, page 10-9](#) and [User Input on Type-B SIP Phones, page 10-11](#).)

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.

Dialed Pattern Recognition

Digit strings dialed by a user on a telephone generally follow patterns. For instance, many enterprises implement a five-digit abbreviated dialing pattern for calls made within the same office location. Also, many enterprises rely on a single-digit access code to represent outside dialing, followed by some quantity of digits to reach a local PSTN number or a long-distance PSTN number (for example, 9 followed by seven digits to reach a local number, or 9 followed by 1 and ten digits to reach a long-distance destination).

The system administrator must plan the system's recognition of these patterns to ensure that the system will act promptly upon detection of a string that corresponds to a predetermined pattern so that users experience no (or minimal) post-dialing delay.

For phones using the Skinny Client Control Protocol (SCCP) and for SIP phones using the Key Press Markup Language (KPML) during dialing, you can implement pattern recognition in Cisco Unified Communications Manager (Unified CM) by configuring route patterns, translation patterns, phone DNs, and so forth. With each digit dialed by the user, the phone sends a signaling message to Unified CM, which performs the incremental work of recognizing a matching pattern. As each key press from the user input is collected, Unified CM's digit analysis provides appropriate user feedback, such as:

- Playing dial tone when the phone first goes off-hook
- Stopping dial tone once a digit has been dialed
- Providing secondary dial tone if an appropriate sequence of digits has been dialed, such as when the off-net access code 9 is dialed

Once digit dialing is completed, Unified CM provides user feedback in the form of call progress tones, such as ringback tone if the destination is in the alerting stage or reorder tone if the destination is invalid.

IP phones running the Session Initiation Protocol (SIP) can be configured with pattern recognition instructions called SIP dial rules. When used, they accomplish the bulk of the task of pattern recognition within the phone. Once a pattern is recognized, the SIP phone sends an invitation to Unified CM to place a call to the number corresponding to the user's input. That action, called a SIP INVITE, is subjected to the Unified CM dial plan in the same way a call from an IP phone running the SCCP protocol would be, except that Unified CM's digit analysis is presented with a completed dial string (that is, all of the digits entered by the user are presented as a block to Unified CM for processing). In this mode of operation, user feedback during the dialing of the digit string is limited to what the phone can provide (see [SIP Dial Rules, page 10-13](#)). Once the string has been composed, user feedback can still be provided by Unified CM in the form of call progress tones.

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 is 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 Unified CM cluster but complex for large systems with multiple Unified CM 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 Unified Communications system:

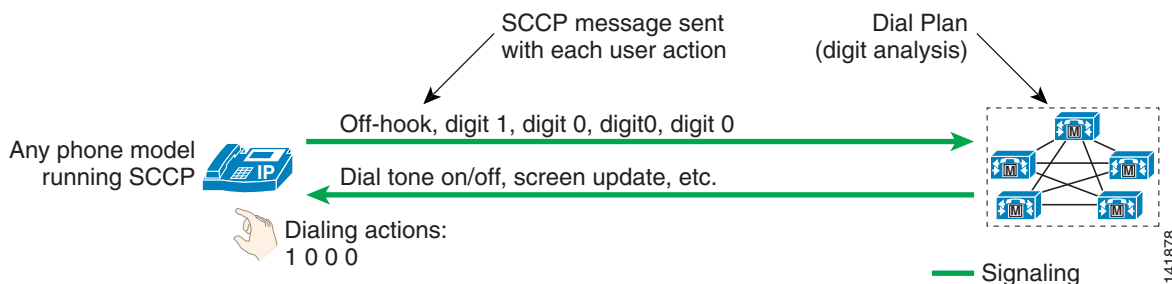
- [User Input on SCCP Phones, page 10-9](#)
- [User Input on Type-A SIP Phones, page 10-9](#)
- [User Input on Type-B SIP Phones, page 10-11](#)
- [SIP Dial Rules, page 10-13](#)
- [Call Routing in Unified CM, page 10-15](#)
- [Calling Privileges in Unified CM, page 10-21](#)
- [Digit Manipulation in Unified CM, page 10-26](#)
- [Automated Alternate Routing, page 10-27](#)
- [Extension Mobility, page 10-30](#)
- [Hunt Lists and Line Groups, page 10-33](#)
- [Call Routing in Cisco IOS with H.323 Dial Peers, page 10-38](#)
- [Calling Privileges in Cisco IOS with H.323 Dial Peers, page 10-50](#)
- [Digit Manipulation in Cisco IOS with H.323 Dial Peers, page 10-52](#)

User Input on SCCP Phones

IP phones using SCCP report every single user input event to Unified CM immediately. For instance, as soon as the user goes off-hook, a signaling message is sent from the phone to the Unified CM server with which it is registered. The phone can be considered to be a terminal, where all decisions resulting from the user input are made by the Unified CM server's configured dial plan.

As other user events are detected by the phone, they are relayed to Unified CM individually. A user who goes off-hook and then dials 1000 would trigger five individual signaling events from the phone to Unified CM. All the resulting feedback provided to the user, such as screen messages, playing dial tone, secondary dial tone, ring back, reorder, and so forth, are commands issued by Unified CM to the phone in response to the dial plan configuration. (See [Figure 10-1](#).)

Figure 10-1 User Input and Feedback for SCCP Phones



It is neither required nor possible to configure dial plan information on IP phones running SCCP. All dial plan functionality is contained in the Unified CM cluster, including the recognition of dialing patterns as user input is collected.

If the user dials a pattern that is denied by Unified CM, reorder tone is played to the user as soon as that pattern becomes the best match in Unified CM's digit analysis. For instance, if all calls to the pay-per-minute Numbering Plan Area (or area code) 976 are denied, reorder tone would be sent to the user's phone as soon as the user dials 91976.

User Input on Type-A SIP Phones

For purposes of this chapter only, we define Type-A phones to include the Cisco Unified IP Phone 7905, 7912, 7940, and 7960. Type-A phones differ somewhat from Type-B phones in their behavior, and Type-A phones do not offer support for Key Press Markup Language (KPML) as do Type-B phones. (See [User Input on Type-B SIP Phones](#), [page 10-11](#).)

Type-A IP phones using SIP offer two distinct modes of operation:

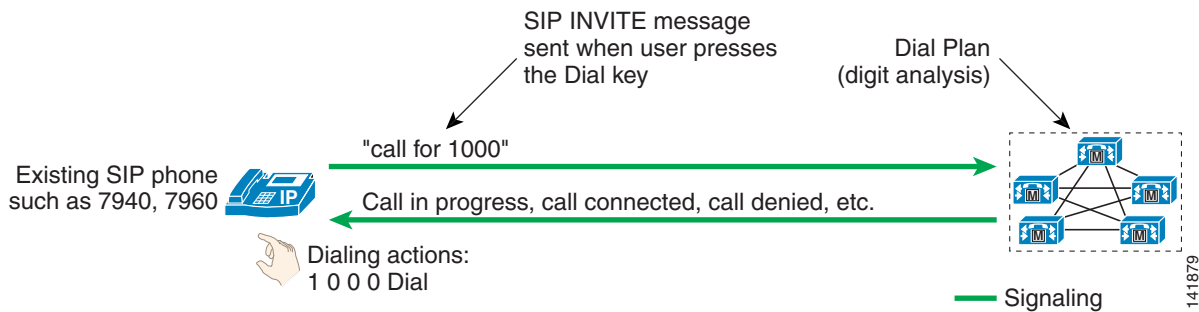
- [No SIP Dial Rules Configured on the Phone](#), [page 10-9](#)
- [SIP Dial Rules Configured on the Phone](#), [page 10-10](#)

No SIP Dial Rules Configured on the Phone

[Figure 10-2](#) illustrates the behavior of a SIP Type-A phone with no dial plan rules configured on the phone. In this mode of operation, the phone accumulates all user input events until the user presses either the # key or the Dial softkey. This function is similar to the "send" button used on many mobile phones. For example, a user making a call to extension 1000 would have to press 1, 0, 0, and 0 followed by the Dial softkey or the # key. The phone would then send a SIP INVITE message to Unified CM to indicate

that a call to extension 1000 is requested. As the call reaches Unified CM, it is subjected to the dial plan configuration for this phone, including all the class-of-service and call-routing logic implemented in Unified CM's dial plan.

Figure 10-2 User Input and Feedback for Type-A SIP Phones with No Dial Rules Configured



If the user dials digits but then does not press the Dial softkey or the # key, the phone will wait for inter-digit timeout (10 seconds by default) before sending a SIP INVITE message to Unified CM. For the example in [Figure 10-2](#), dialing 1, 0, 0, 0 and waiting for inter-digit timeout would result in the phone placing a call to extension 1000 after ten seconds.



Note

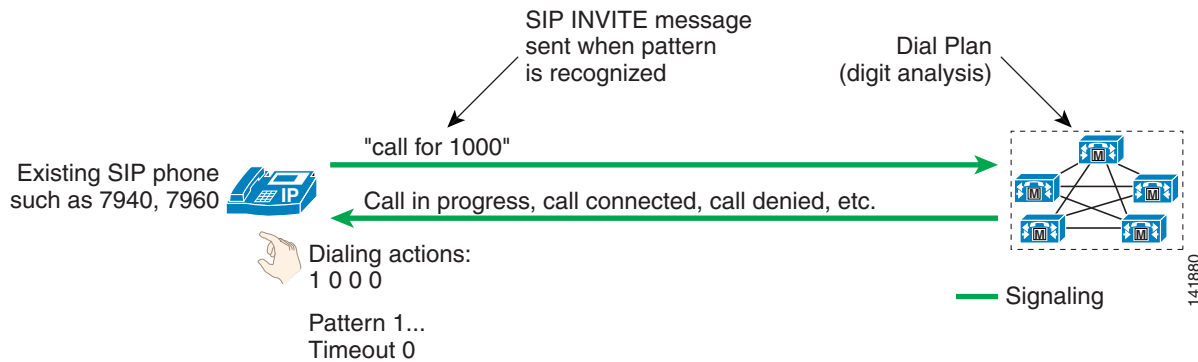
If the user presses the Redial softkey, and the action is immediate; the user does not have to press the Dial key or wait for inter-digit timeout.

If the user dials a pattern that is denied by Unified CM, the user must enter the entire pattern and press the Dial key, and the INVITE message must be sent to Unified CM, before any indication that the call is rejected (reorder tone) is sent to the caller. For instance, if all calls to NPA 976 are denied, the user would have to dial 919765551234 and press Dial before the reorder tone would be played.

SIP Dial Rules Configured on the Phone

SIP dial rules enable the phone to recognize patterns dialed by users. Once the recognition has occurred, the sending of the SIP INVITE message to Unified CM is automated and does not require the user to press the Dial key or wait for the inter-digit timeout. (For more details, see [SIP Dial Rules, page 10-13](#).)

For example, if a branch location of the enterprise requires that calls between phones within the same branch be dialed as four-digit extensions, the phone could be configured to recognize the four-digit patterns so that the user is not required to press the Dial key or wait for the inter-digit timeout. (See [Figure 10-3](#).)

Figure 10-3 User Input and Feedback for Type-A SIP Phones with Dial Rules Configured

In [Figure 10-3](#), the phone is configured to recognize all four-digit patterns beginning with 1 and has an associated timeout value of 0. All user input actions matching the pattern will trigger the sending of the SIP INVITE message to Unified CM immediately, without requiring the user to press the Dial key.

Type-A phones using SIP dial rules offer a way to dial patterns not explicitly configured on the phone. If a dialed pattern does not match a SIP dial rule, the user can press the Dial key or wait for inter-digit timeout.

If a particular pattern is recognized by the phone but blocked by Unified CM, the user must dial the entire dial string before receiving an indication that the call is rejected by the system. For instance, if a SIP dial rule is configured on the phone to recognize calls dialed in the form 919765551234 but such calls are blocked by the Unified CM dial plan, the user will receive reorder tone at the end of dialing (after pressing the final 4 key).

User Input on Type-B SIP Phones

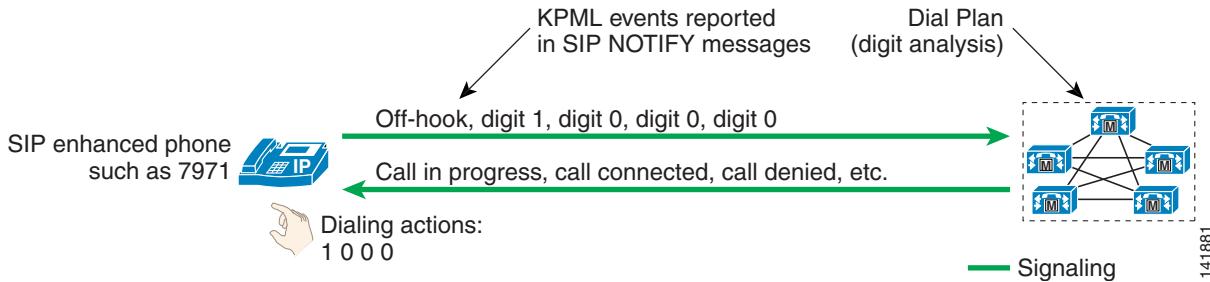
For purposes of this chapter only, we define Type-B phones to include the Cisco Unified IP Phone 7911, 7941, 7961, 7970, and 7971. Type-B phones differ somewhat from Type-A phones in their behavior, and Type-B phones offer support for Key Press Markup Language (KPML) but Type-A phones do not. (See [User Input on Type-A SIP Phones](#), page 10-9.)

Type-B IP phones (such as the Cisco Unified IP Phone 7911, 7941, 7961, 7970, and 7971) running SIP offer two distinct modes of operation:

- [No SIP Dial Rules Configured on the Phone](#), page 10-11
- [SIP Dial Rules Configured on the Phone](#), page 10-12

No SIP Dial Rules Configured on the Phone

Type-B IP telephones offer functionality based on the Key Press Markup Language (KPML) to report user key presses. Each one of the user input events will generate its own KPML-based message to Unified CM. From the standpoint of relaying each user action immediately to Unified CM, this mode of operation is very similar to that of phones running SCCP. (See [Figure 10-4](#).)

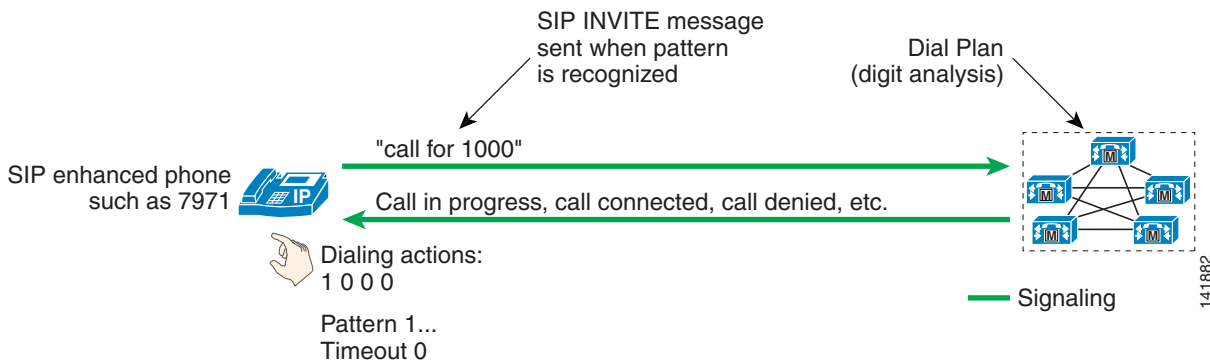
Figure 10-4 User Input and Feedback for Type-B SIP Phones with No Dial Rules Configured

Every user key press triggers a SIP NOTIFY message to Unified CM to report a KPML event corresponding to the key pressed by the user. This messaging enables Unified CM's digit analysis to recognize partial patterns as they are composed by the user and to provide the appropriate feedback, such as immediate reorder tone if an invalid number is being dialed.

In contrast to Type-A IP phones running SIP without dial rules, Type-B SIP phones have no Dial key to indicate the end of user input. In [Figure 10-4](#), a user dialing 1000 would be provided call progress indication (either ringback tone or reorder tone) after dialing the last 0 and without having to press the Dial key. This behavior is consistent with the user interface on phones running the SCCP protocol.

SIP Dial Rules Configured on the Phone

Type-B IP phones can be configured with SIP dial rules so that dialed pattern recognition is accomplished by the phone. (See [Figure 10-5](#).)

Figure 10-5 User Input and Feedback for Type-B SIP Phones with Dial Rules Configured

In [Figure 10-5](#), the phone is configured to recognize all four-digit patterns beginning with 1, and it has an associated timeout value of 0. All user input actions matching these criteria will trigger the sending of a SIP INVITE message to Unified CM.



Note

As soon as SIP dial rules are implemented on Type-B IP phones, KPML-based dialing is disabled. If a user dials a string of digits that do not match a SIP dial rule, none of the individual digit events will be relayed to Unified CM. Instead, the entire dialed string will be sent en-bloc to Unified CM once the dialing is complete (That is, once inter-digit timeout has occurred).

Type-B phones using SIP dial rules offer only one way to dial patterns not explicitly configured on the phone. If a dialed pattern does not match a SIP dial rule, the user has to wait for inter-digit timeout before the SIP NOTIFY message is sent to Unified CM. Unlike Type-A IP phones, Type-B IP phones do not have a Dial key to indicate the end of dialing, except when on-hook dialing is used. In the latter case, the user can press the “dial” key at any time to trigger the sending of all dialed digits to Unified CM.

**Note**

When a Type-B phone registers with the SRST router, the configured SIP dial rules are disabled.

If a particular pattern is recognized by the phone but blocked by Unified CM, the user must dial the entire dial string before receiving an indication that the call is rejected by the system. For instance, if a SIP dial rule is configured on the phone to recognize calls dialed in the form 919765551234 but such calls are blocked by the Unified CM dial plan, the user will receive reorder tone at the end of dialing (after pressing the 4 key).

SIP Dial Rules

Cisco Unified CM offers SIP dial rule functionality to allow phones to perform pattern recognition as user input is collected. For example, a phone can be configured to recognize the well established pattern 911 and to send a message to Unified CM to initiate an emergency call immediately, while at the same time allowing the user to enter patterns of variable length for international numbers.

It is important to note that pattern recognition configuration on the phone through the use of SIP dial rules does not supersede the Class of Service and Route Plan configurations of Unified CM. A phone might very well be configured to recognize long-distance patterns while Unified CM is configured to block such calls because the phone is assigned a class of service allowing only local calls.

There are two types of SIP dial rules, based on the phone model on which they will be deployed:

- 7905_7912 (Used for Cisco Unified IP Phones 7905 and 7912)
- 7940_7960_OTHER (Used for all other IP phone models))

There are four basic Dial Parameters that can be used as part of a dial rule:

- Pattern

This parameter is the actual numerical representation of the pattern. It can contain digits, wildcards, or instructions to play secondary dial tone. The following table provides a list of values and their effect for the two types of dial rules.

Pattern	Dial Rule Type	
	7905_7912	7940_7960_OTHER
Digits 0 through 9	Corresponding digit	Corresponding digit
.	Matches any digit (0 through 9)	Matches any character (0 through 9, *, #)
-	Indication that more digits can be entered. Must be at the end of an individual rule.	n/a

Pattern	Dial Rule Type	
	7905_7912	7940_7960_OTHER
#	Input termination key. Place the > character in the dial rule to indicate the character position after which the # key will be recognized as input termination. For instance, in 9>#..., the # character would be recognized any time after 9 has been pressed.	n/a
tn	Indicates a timeout value of <i>n</i> seconds. For example, 1...t3 would match 1000 and trigger the sending of an invite to Unified CM after 3 seconds.	n/a
rn	Repeats the last character <i>n</i> times. For example, 1.r3 is equivalent to 1....	n/a
S	When a pattern contains the modifier S, all other dial rules after this pattern are ignored. S effectively causes rule matching to cease.	n/a
*	Input termination key. Place the > character in the dial rule to indicate the character position after which the * key will be recognized as input termination.	Matches one or more characters. For instance, pattern 1* would match 10, 112, 123456, and so forth.
,	n/a	Cause the phone to play secondary dial tone. For instance, 8,... would cause the user to hear secondary dial tone after 8 is pressed.

- IButton

This parameter specifies the button to which the dial pattern applies. If the user is initiating a call on line button 1, only the dial patterns specified for Button 1 apply. If this optional parameter is not configured, the dial pattern applies to all lines on the phone. This parameter applies only to the Cisco SIP IP Phone models 7940, 7941, 7960, 7961, 7970, and 7971. The button number corresponds to the order of the buttons on the side of the screen, from top to bottom, with 1 being on top button.

- Timeout

This parameter specifies the time, in seconds, before the system times out and dials the number as entered by the user. To have the number dialed immediately, specify 0. This parameter is available only for 7940_7960_OTHER dial rules. If this parameter is omitted, the phone's default inter-digit timeout value is used (default of 10 seconds).

- User

This parameter represents the tag that automatically gets added to the dialed number. Valid values include **IP** (when SIP Call Agents other than Unified CM are deployed) and **Phone**. This parameter is available only for 7940_7960_OTHER dial rules. This parameter is optional, and it should be omitted in deployments where Unified CM is the only call agent.

**Note**

The Cisco Unified IP Phone 7905 and 7912 choose patterns in the order in which they have been created in the SIP dial rules, whereas all the other phone models choose the pattern offering the longest match. The following table shows which pattern would be chosen if a user dialed 95551212.

SIP Dial Rules	7905_7912	7940_7960_OTHER
..... 9.....	Chooses first matching pattern:	Chooses longest matching pattern: 9.....

Call Routing in Unified CM

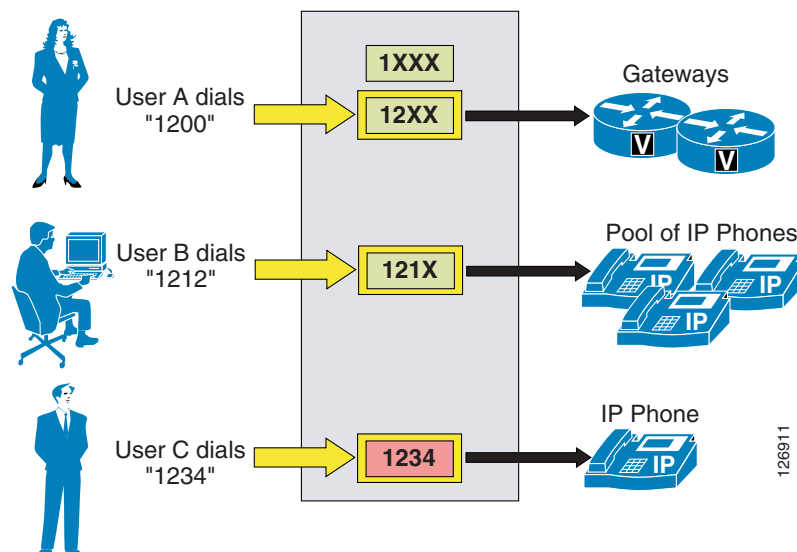
All dialing destinations configured in Unified CM are added to its internal call routing table as patterns. These destinations include IP phone lines, voicemail ports, route patterns, translation patterns, and CTI route points.

When a number is dialed, Unified CM 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 10-6](#), where the call routing table includes the patterns 1XXX, 12XX, and 1234.

Figure 10-6 Unified CM Call Routing Logic Example



When user A dials the string 1200, Unified CM 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.

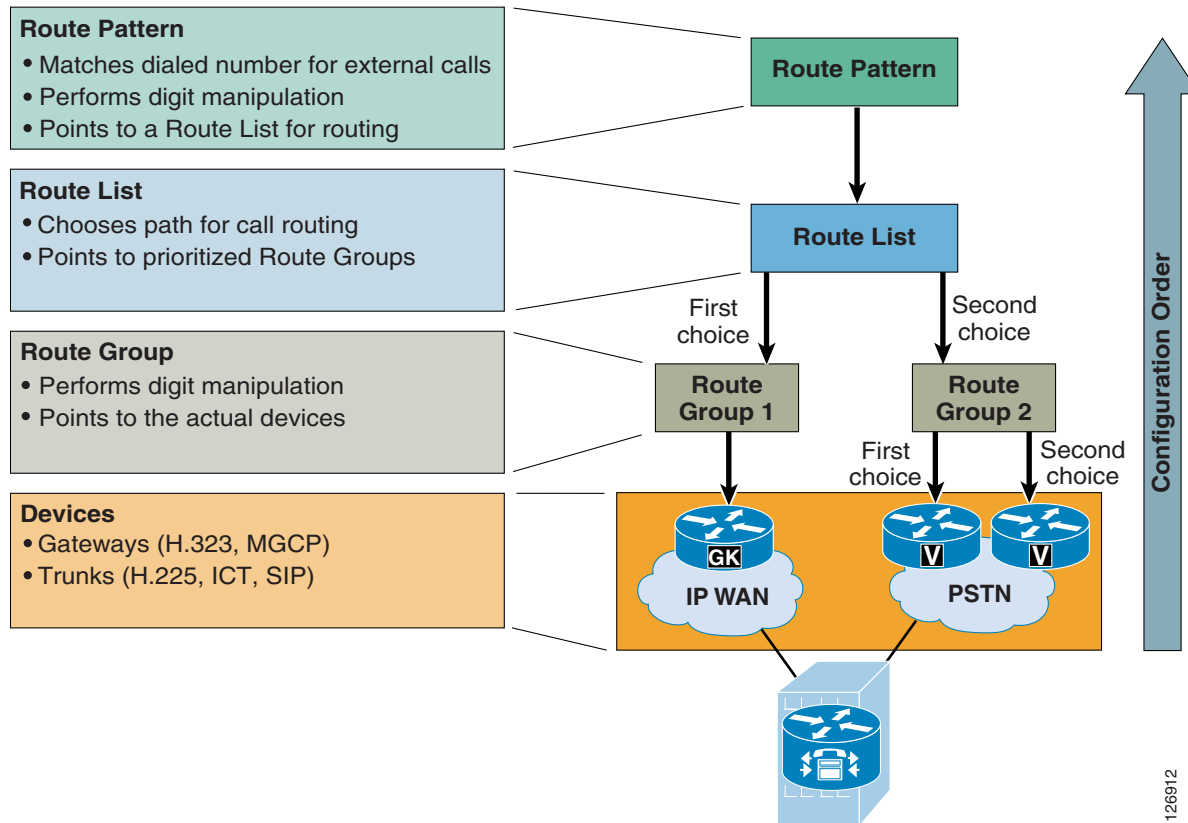
When user B dials the string 1212, there are three potentially matching patterns, 1XXX, 12XX and 121X. As mentioned above, 1XXX matches 1000 strings and 12XX matches 100 strings. However, 121X matches only 10 strings; therefore it is selected as the destination of the call.

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 Unified CM 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 Unified CM 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 No Answer field on primary patterns such as CTI route points because this field accepts wildcards in Cisco Unified CM Release 4.0 and later.

Unified CM automatically "knows" how to route calls to destinations within the same cluster. For external destinations such as PSTN gateways, H.323 gatekeepers, or other Unified CM 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. Unified CM 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 10-7](#) depicts the three-tiered architecture of the Unified CM external route construct.

Figure 10-7 External Route Pattern Architecture

The following sections describe the individual elements of the external route construct in Unified CM:

- [Route Patterns](#), page 10-17
- [Route Lists](#), page 10-20
- [Route Groups](#), page 10-21
- [Route Group Devices](#), page 10-21

Route Patterns

Route patterns are strings of digits and wildcards, such as 9.[2-9]XXXXXX, configured in Unified CM 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 Unified CM internal dial plan database.

- It is possible to configure Unified CM to accept other national numbering plans. Once this is done, the @ wildcard can be used for different numbering plans even within the same Unified CM cluster, depending on the value selected in the Numbering Plan field on the Route Pattern configuration page. For more information, please refer to the *Cisco Unified CallManager Dial Plan Deployment Guide*, available at http://www.cisco.com/en/US/products/sw/voicesw/ps5629/prod_maintenance_guides_list.html
- 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 10-18).

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 Unified CM 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, Unified CM 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 Unified CM for overlap sending and overlap receiving.

Overlap sending means that Unified CM 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 Unified CM Release 4.0 or later, check the Allow Overlap Sending box on the Route Pattern Configuration page. (In earlier Unified CM 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 Unified CM 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 Unified CM Release 3.3(3) or later, set the `OverlapReceivingFlagForPRI` service parameter to `True`. (In earlier Unified CM 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, Unified CM 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, Unified CM 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 forces the T302 timer to expire as soon as the configured pattern is the best 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 Unified CM, page 10-15](#), 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, Unified CM will not make its routing decision as soon as it receives the third digit because even though 1XX is an urgent pattern, it is not the best match (10 total patterns matched by 12! versus 100 patterns matched by 1XX). Unified CM will have to wait for inter-digit timeout before routing the call because the pattern 12! allows for more digits to be input by the user.

Consider another example, where pattern 12[2-5] is marked as urgent and 12! is configured as a regular pattern. If the user dials 123, the pattern 12[2-5] is the best match (4 total patterns matched by 12[2-5] versus 10 patterns matched by 12!). Because the urgent-priority pattern is the best match, the T302 timer is aborted and no further user input is expected. Unified CM routes the call using pattern 12[2-5].

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 Unified CM 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. Unified CM 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 Unified CM cluster, or SIP trunks to a SIP proxy. (In Cisco Unified CM 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.)

Unified CM sends calls to the devices according to the distribution algorithm assigned. Unified CM 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 Unified CMs. You can configure the following types of devices in Unified CM:

- 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 Unified CM cluster
- Intercluster trunk, gatekeeper controlled — trunk to other Unified CM clusters and/or H.323 gateways, via a gatekeeper
- SIP trunk — trunk to a SIP proxy (available with Cisco Unified CM 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 Unified CM and will select H.225 or Intercluster Trunk protocol accordingly.

Calling Privileges in Unified CM

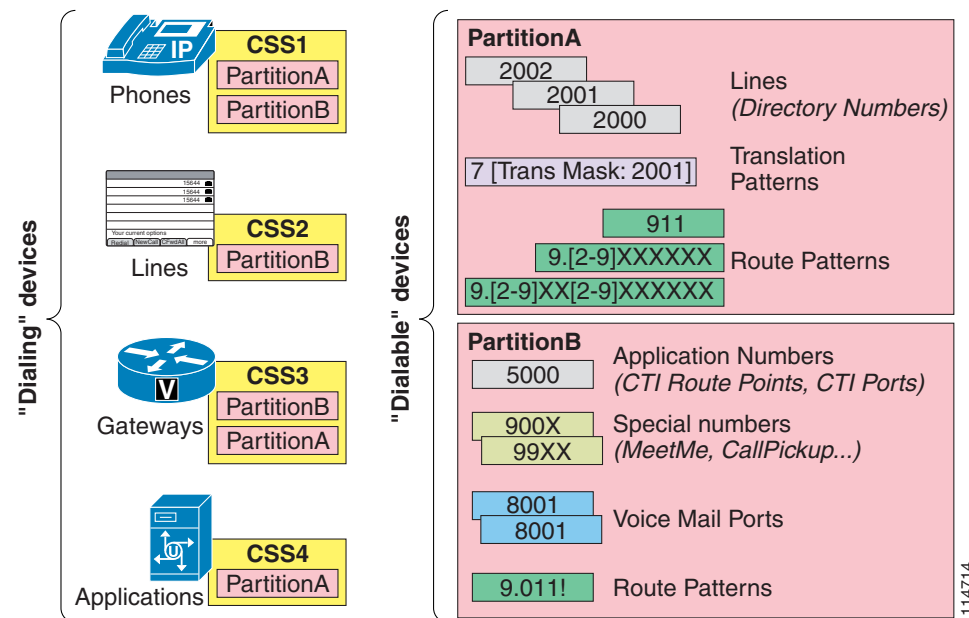
Dialing privileges are configured in order to control which types of calls are allowed (or prevented) for a particular endpoint (such as phones, gateways, or CTI applications). All calls handled by Unified CM are subjected to the dialing privileges implemented through the configuration of the following elements:

- [Partitions, page 10-22](#)
- [Calling Search Spaces, page 10-23](#)

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 10-8](#), 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).

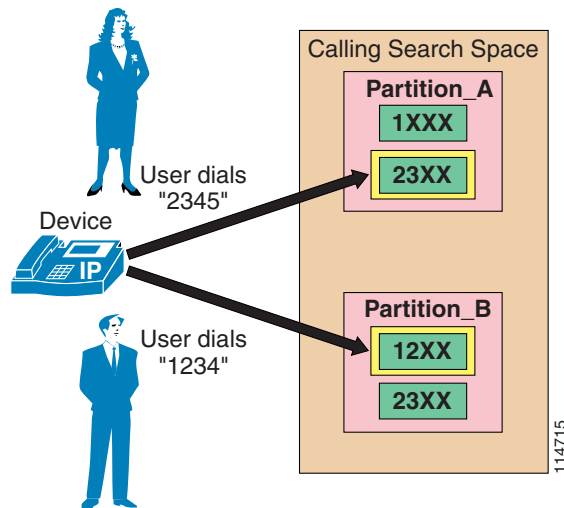
Figure 10-8 Partitions and Calling Search Spaces



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 Unified CM, page 10-15](#), if two or more dial plan entries (directory numbers, route patterns, or so forth) overlap, Unified CM 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, Unified CM selects the dial plan entry that appears first in the calling search space of the device making the call.

For example, consider [Figure 10-9](#), 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, Unified CM 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, Unified CM 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 10-9 *Impact of Partition Order on the Matching Logic***Note**

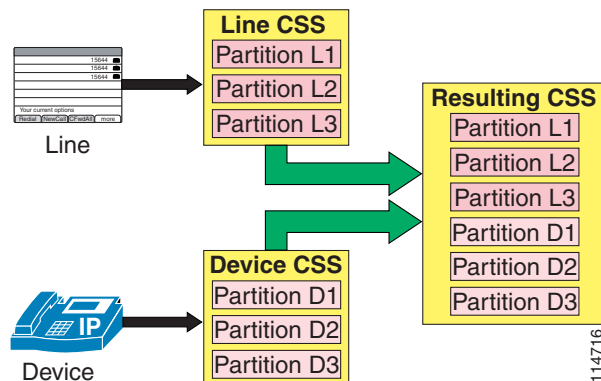
When multiple equal-precision matches occur in the same partition, Unified CM selects the entry that is listed first in its local dial plan database. Since you cannot configure the order in which the dial plan database lists dial plan entries, Cisco strongly recommends that you avoid any possibility of equal-precision matches coexisting within the same partition because the resulting dial plan logic is not predictable in such cases.

Beginning with Cisco Unified CM 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 Unified CM 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 Unified CM call routing engine. For more information on this feature, see [Time-of-Day Routing, page 10-37](#).

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, Unified CM 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 10-10](#).

Figure 10-10 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 10-22](#), Unified CM 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 Unified CM with the Traditional Approach, page 10-76](#), and [Building Classes of Service for Unified CM with the Line/Device Approach, page 10-80](#).

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 Unified Communications Manager 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.

**Note**

Cisco strongly recommends that you do not leave any calling search space defined as <None>. Doing so can introduce dial plan behavior that is difficult to predict.

Call-Forward Calling Search Spaces

For Unified CM versions other than releases 5.x, 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.

With Cisco Unified CM releases 5.x, the Call Forward All calling search space is concatenated with the Secondary Calling Search Space for Forward All available on the Directory Number configuration page, and the resulting calling search space applies to all calls when the Forward All action is initiated from the User page or from the Administrative pages.

These concatenated calling search spaces are also used when the Forward All action is invoked from a phone running SCCP or from an Type-B phone running SIP. For further details, see [Building Classes of Service for Unified CM with the Line/Device Approach](#), page 10-80.

On Type-A IP phones running SIP, if Call Forward All is invoked from the phone itself, the device's Rerouting Calling Search Space is used for forwarded calls. If Forward All actions are invoked from the Unified CM User page or the Unified CM Administrative page, then any Forward All action initiated from the phone is irrelevant.

For example, assume an Type-A IP phone running SIP is configured with Forward All to extension 3000 from the Unified CM User page. At the same time, the phone itself is configured to Forward All to extension 2000. All calls made to that phone will be forwarded to extension 3000.

**Note**

On Type-A IP phones running SIP, invoking Forward All from the Unified CM User or Administrative pages will not be reflected on the phone. The phone does not display any visual confirmation that calls are forwarded.

When Forward All is initiated from an IP phone running SCCP or from an Type-B IP phone running SIP, user input is simultaneously compared to the patterns allowed in the configured Forward All calling search space(s). If an invalid destination pattern is configured, the user will be presented with reorder tone. When Forward All is invoked from an Type-A IP phone running SIP, Forward All user input is stored locally on the phone and is not verified against any calling search space in Unified CM. If user input corresponds to an invalid destination, no notification is offered to the user. Calls made to that phone will be presented with reorder tone as the phone tries to initiate a SIP re-route action to an invalid destination number.

When the Forward All calling search space is left as <None>, the results are difficult to predict and depend on the Unified CM 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 DN's for the voicemail pilot and voicemail ports but not external PSTN numbers.
- Configure both the Call Forward All calling search space and the Secondary Calling Search Space for Forward All, 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.

Digit Manipulation in Unified CM

The following tools provide digit manipulation capability in Unified CM:

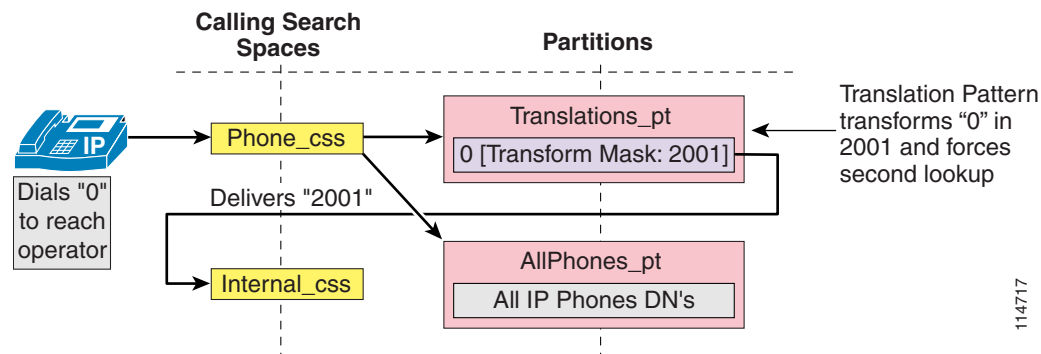
- The external route construct (route patterns, route lists, and route groups)
- Translation patterns

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 Unified CM, page 10-15](#).

Translation patterns are the most powerful tool in Unified CM 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, Unified CM 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 10-11](#).

Figure 10-11 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 Unified CM 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). However, when the digit manipulation occurs within a route list, the CDR will show the originally dialed number, not the translated one. The Placed Calls directory on the IP phone always shows the string as it was dialed by the user.

Automated Alternate Routing

The automated alternate routing (AAR) feature enables Unified CM to establish an alternate path for the voice media when the preferred path between two endpoints within the same cluster 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 10-30](#), for more details.

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

- [Establish the PSTN Number of the Destination, page 10-27](#)
- [Prefix the Required Access Codes, page 10-28](#)
- [Select the Proper Dial Plan and Route, page 10-28](#)

**Note**

Beginning with Cisco Unified CM 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 Unified CM 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 Unified CM 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 1234), the AAR group configuration prefixes 91 to the dialed string, yielding a completed string of 91 212 555 1234.

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 Unified CM 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 Multisite Deployments, page 10-55](#), for more details.

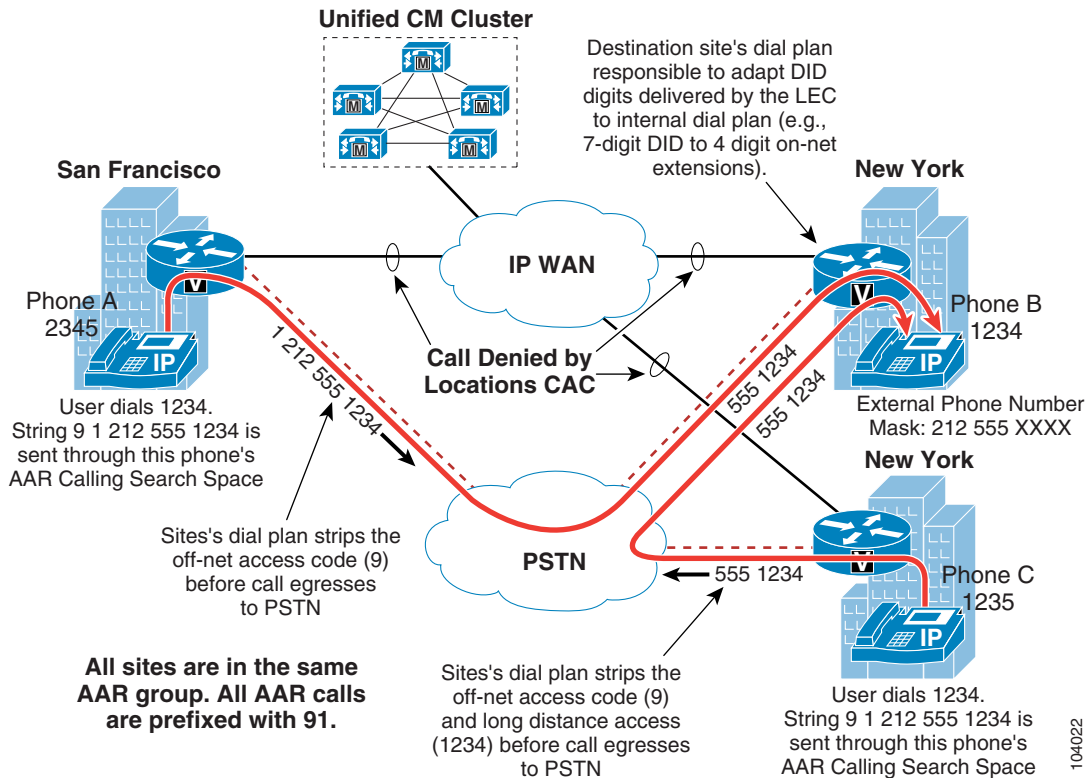
**Note**

To avoid denial of re-routed calls due to call admission control, AAR functionality requires the use of a LAN as the IP path between each endpoint and its associated gateway to the PSTN. Therefore, AAR dial plans cannot rely on centralized gateways for PSTN access.

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 10-12](#).) 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 10-12 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

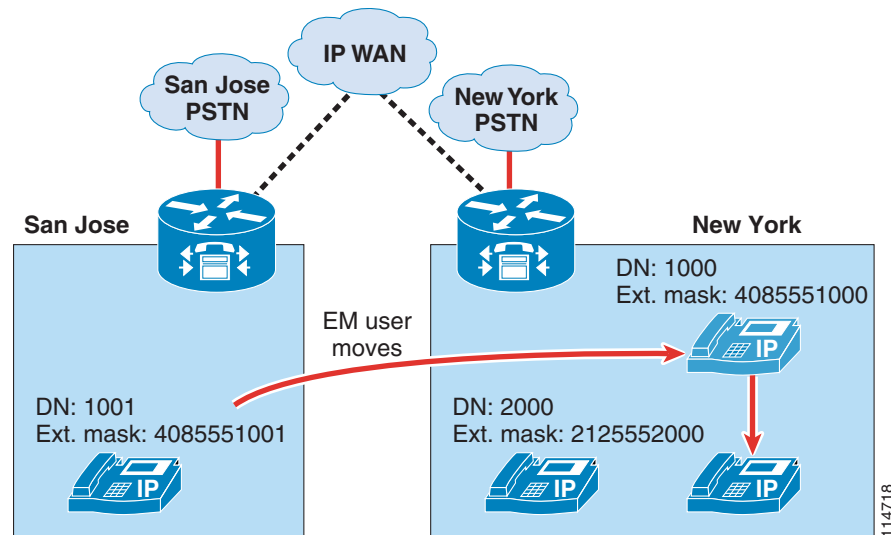
The Extension Mobility feature enables a user to log in to an IP phone and automatically apply his or her profile to that phone, including extension number, speed dials, message waiting indicator (MWI) status, and calling privileges. This mechanism relies on the creation of a device profile associated 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 Unified CM 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 Unified CM cluster. When Extension Mobility is applied to multisite 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 10-27, some limitations exist. Consider the example shown in [Figure 10-13](#), where Extension Mobility and AAR are deployed in a centralized call processing Unified CM cluster with one site in San Jose and one in New York.

Figure 10-13 *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 route patterns pointing to route lists or route groups containing that same gateway.

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.

Immediate Divert (iDivert)

The Immediate Divert (iDivert) function is used to send calls directly to voicemail. It can be invoked when the call is ringing (incoming), when a call is on hold, or when a call is connected. For Cisco Unified CM releases prior to 5.1, the call is always sent to the invoking phone's voicemail box. Specifically, the iDivert action results in a call to the voicemail pilot of the invoking phone's voicemail profile, and it uses the calling search space configured on the voicemail pilot.

iDivert Enhancements in Cisco Unified CM 5.1

The iDivert function has been augmented in Cisco Unified CM 5.1 to allow incoming calls to be diverted to either the invoking phone's voicemail box (legacy behavior) or the voicemail box of the originally called party (enhanced behavior). The enhanced functionality is applicable only to diverted calls such as forwarded calls or calls redirected by an application.

For example:

Assume that phone A calls phone B, whose calls are forwarded to phone C. As phone C is ringing, the user at phone C activates the iDivert softkey, which offers two choices. The first choice results in the call being sent to the original called party's voicemail (in this case, phone B's voicemail box), while the second choice results in the call being sent to the iDivert invoker's voicemail (in this case, phone C's voicemail box). The same choices are available whether phone C invokes the feature while the call is ringing, connected, or placed on hold.

If a call is handled by Auto Call Pickup, Call Transfer, Call Park, Call Park Reversion, Conference, or a MeetMe Conference prior to the invocation of the iDivert function, the call is no longer considered to be a "diverted" call, and the only iDivert functionality available in this case is the legacy iDivert behavior (that is, sending the call to the invoker's voicemail). For example, assume phone A calls phone B, whose calls are forwarded to phone C, and then phone C transfers the call to phone D. This is *not* a diverted call because the last action applied to the call was the transfer to phone D. If phone D invokes the iDivert function, the call will be sent to phone D's voicemail box.

To enable the enhanced iDivert functionality, set the Unified CM service parameter **Use Legacy Immediate Divert** to **False**. When enabled, enhanced iDivert automatically allows the use of the feature over QSIG trunks, thus allowing an invoker's voicemail box to be hosted in a telephony system connected via QSIG.

In cases where iDivert is used in a cluster connected to other telephone systems using QSIG, there might be situations where only the legacy iDivert functionality (sending the call to the invoker's voicemail) is available to a phone when receiving a call. For instance, assume phones A and B are in cluster 1, and phone X is another QSIG-connected telephony system. Phone A calls phone X, which is call-forwarded to phone C. After the call is connected to phone C, iDivert will offer both the legacy (invoker's voicemail) and enhanced (original called party's voicemail) destinations only if QSIG path replacement has not occurred. If phone C invokes iDivert after QSIG path replacement, the only destination available is phone C's voicemail box.

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.

Unified CM 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 10-14](#) depicts the three-tiered architecture of the hunt construct in Cisco Unified CM Release 4.1 or later.



Note

In Cisco Unified CM 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 Unified CM 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. Starting with Cisco Unified CM Release 4.1, these structures are independent. [Table 10-3](#) summarizes a feature comparison of the hunt lists and line groups in Cisco Unified CM Releases 4.0 and 4.1 and hunt groups using the Attendant Console in Cisco Unified CM Releases 3.3 and earlier.

Table 10-3 Comparison of Features Supported by Route/Hunt List, Hunt Pilot, and Hunt Groups

Feature	Hunt Groups in Cisco Unified CM Release 3.3 and Earlier	Route/Hunt List in Cisco Unified CM Release 4.0	Hunt Pilot in Cisco Unified CM Releases 4.1 and 5.x
Skinny Client Control Protocol (SCCP) endpoints	Yes	Yes (line groups)	Yes
SIP endpoints	N/A	N/A	Yes, in releases 5.x
Gateways and trunks (off-net destinations)	No	Yes (route groups)	No
Top-down algorithm	Yes	Yes	Yes
Circular algorithm	Yes	Yes	Yes
Longest-idle algorithm	Yes	Yes	Yes
Broadcast algorithm	Yes	Yes	Yes
Hunt options	No	Yes	Yes
Ring-no-answer reversion	No	Yes	Yes
Performance monitoring (PerfMon)	No	Yes	Yes (except in releases 5.x)
SCCP voicemail port (Cisco Unity)	No	Yes (line groups)	Yes
Simplified Message Desk Interface (SMDI) voicemail systems	Yes	Yes	Yes
Queuing	Yes	No	No
Linked hunt groups and hunt pilot	Yes	No	Yes
Immediate Divert (iDivert)	N/A	No	Only in Release 5.1

For comparison purposes, Figure 10-14 illustrates the hunt pilot architecture in Cisco Unified CM 4.1 and above, while Figure 10-15 shows the hunt pilot architecture in Unified CM 4.0.

Figure 10-14 Three-Tiered Architecture for the Hunt Construct in Unified CM Release 4.1

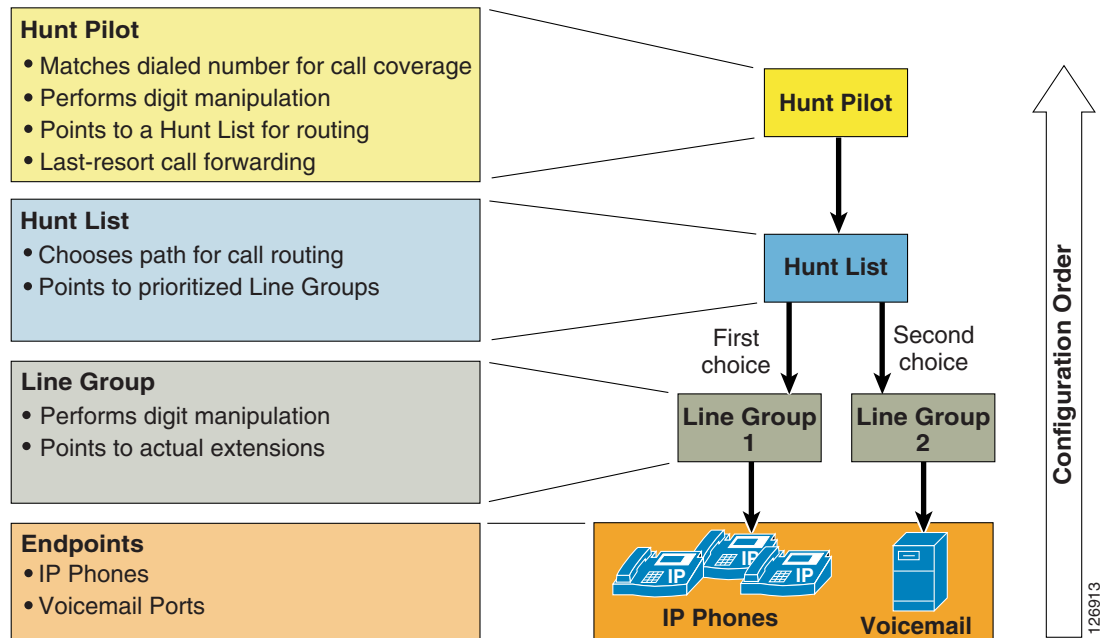
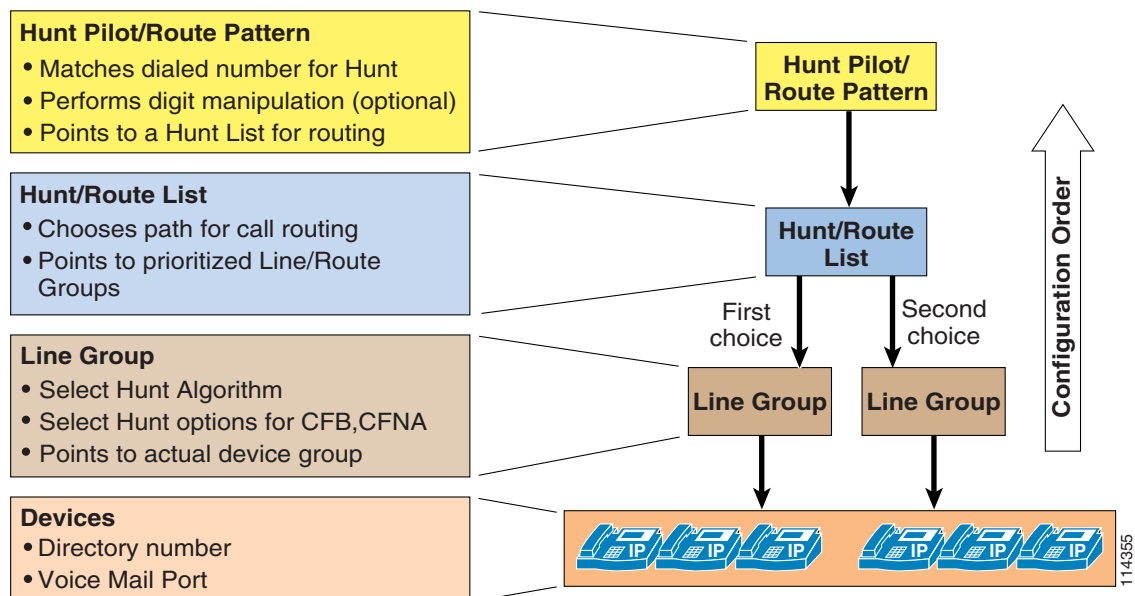


Figure 10-15 Hunt Construct in Unified CM Release 4.0



Hunt Pilot

Hunt pilots are strings of digits and wildcards similar to route patterns, such as 9.[2-9]XXXXXX, configured in Unified CM 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 Unified CM 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 Unified CM
- 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.

**Note**

Cisco Unified CM Release 4.0 does not support call redirection.

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.

**Note**

In Cisco Unified CM Release 5.1, the enhanced iDivert function is available for calls distributed by a hunt pilot.

- The hunt pilot can distribute calls to any of its line 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 sets of phones which are offered a call placed to the hunt pilot number. For example, you can use a hunt list to attempt to find a taker for the call within a set of phones at a particular site. If the call is not taken, then the hunt list attempts to offer the call through a second line group pointing to phones at a second 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 Unified CM. Thus, when the call is being distributed through the line group members, Unified CM 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.
- Unified CM distributes calls to the devices according to the distribution algorithm assigned. Unified CM 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. Unified CM 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 Unified Communications Manager 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 Unified IP Phones, VG248, or ATA 188
- SIP endpoints (requires Cisco Unified CM 5.x)
- Voicemail ports (for Cisco Unity)
- H.323 clients
- FXS extensions attached to an MGCP gateway

Time-of-Day Routing

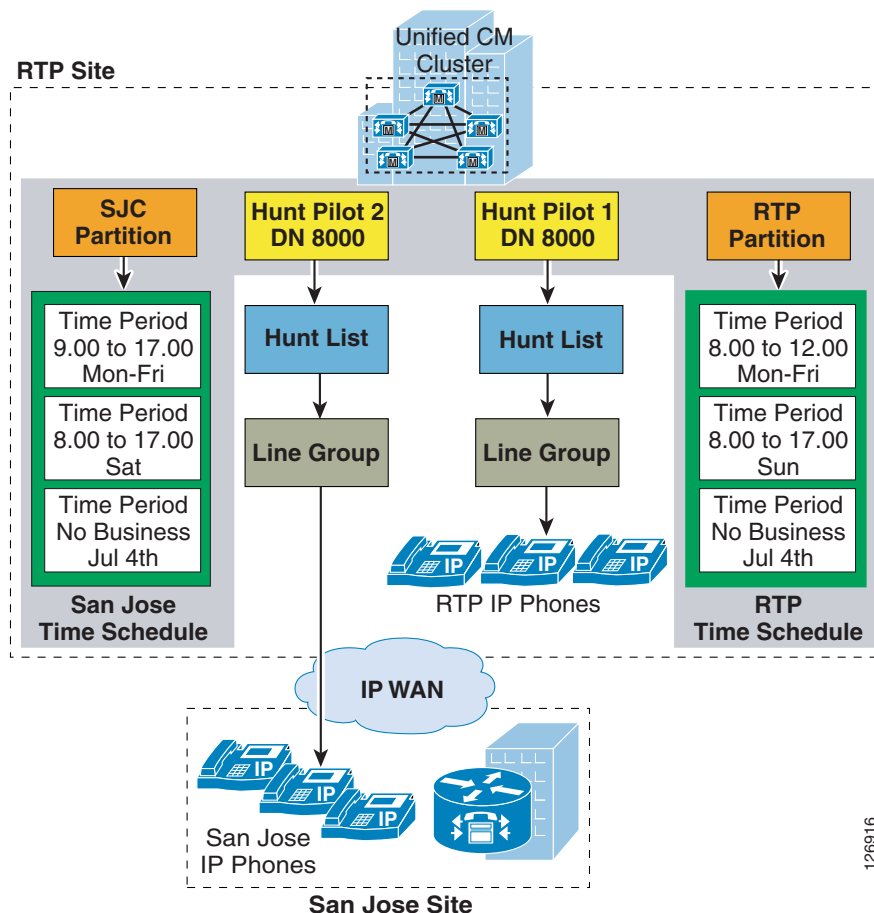
Cisco Unified CM Release 4.1 introduces the time-of-day (ToD) routing feature. To use this feature, configure the following elements:

- 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 10-16](#), 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 10-16 Time-of-Day Routing

For the example in [Figure 10-16](#), 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 Unified CMs

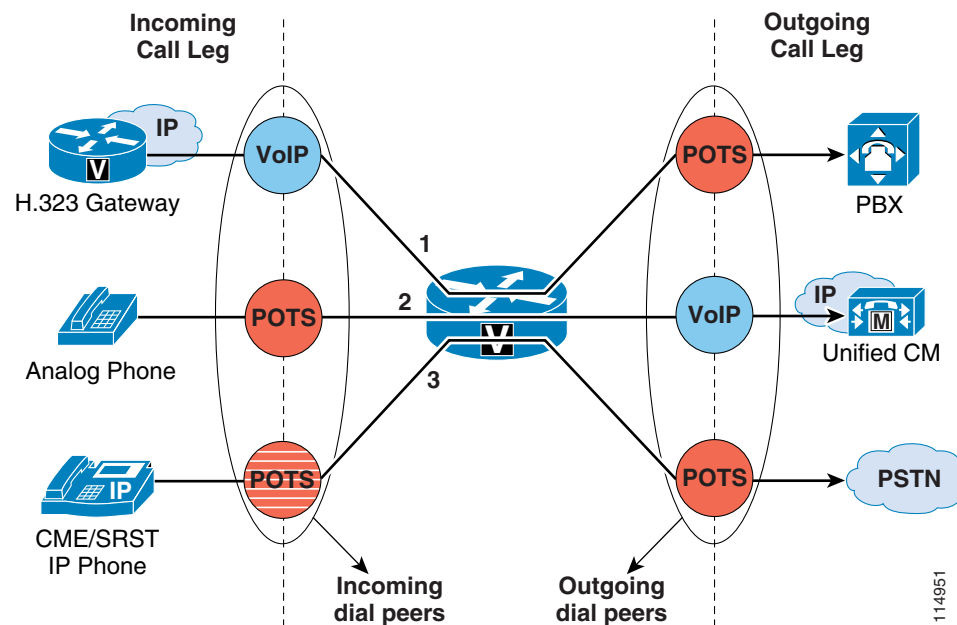
For all calls going through the router, Cisco IOS associates one dial peer to each call leg. Dial peers are also of two main types, according to the type of call leg with which they are associated:

- POTS dial peers, associated with traditional TDM telephony call legs
- VoIP dial peers, associated with IP call legs

Figure 10-17 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 Unified CM 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 Unified Communications Manager Express (Unified CME) 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 10-17 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 criterion is *least recent use*, which selects the destination pattern that has waited the longest since being selected (equivalent to *longest idle* for Unified CM 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 Unified CM cluster, provide redundancy by configuring at least two VoIP dial peers with the same destination pattern pointing to two different Unified CM servers. Use the **preference** attribute to select the priority order between primary and secondary Unified CM servers. The following example shows the use of the **preference** attribute:

```
dial-peer voice 100 voip
  preference 1

!--- Make this the first choice dial peer.

  ip precedence 5
  destination-pattern 1...
  session target ipv4:10.10.10.2

!--- This is the address of the primary Unified CM.

  dtmf-relay h245-alpha

dial-peer voice 101 voip
  preference 2

!--- This is the second choice.

  ip precedence 5
  destination-pattern 1...
  session target ipv4:10.10.10.3

!--- This is the address of the secondary Unified CM.

  dtmf-relay h245-alpha
```

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 Unified Communications Manager Express (Unified CME) and Unified CM 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 8-21](#); for call admission control considerations, refer to [Cisco IOS Gatekeeper Zones, page 9-14](#).

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 10-18](#) 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 Unified CM, 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 10-45](#)). 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.

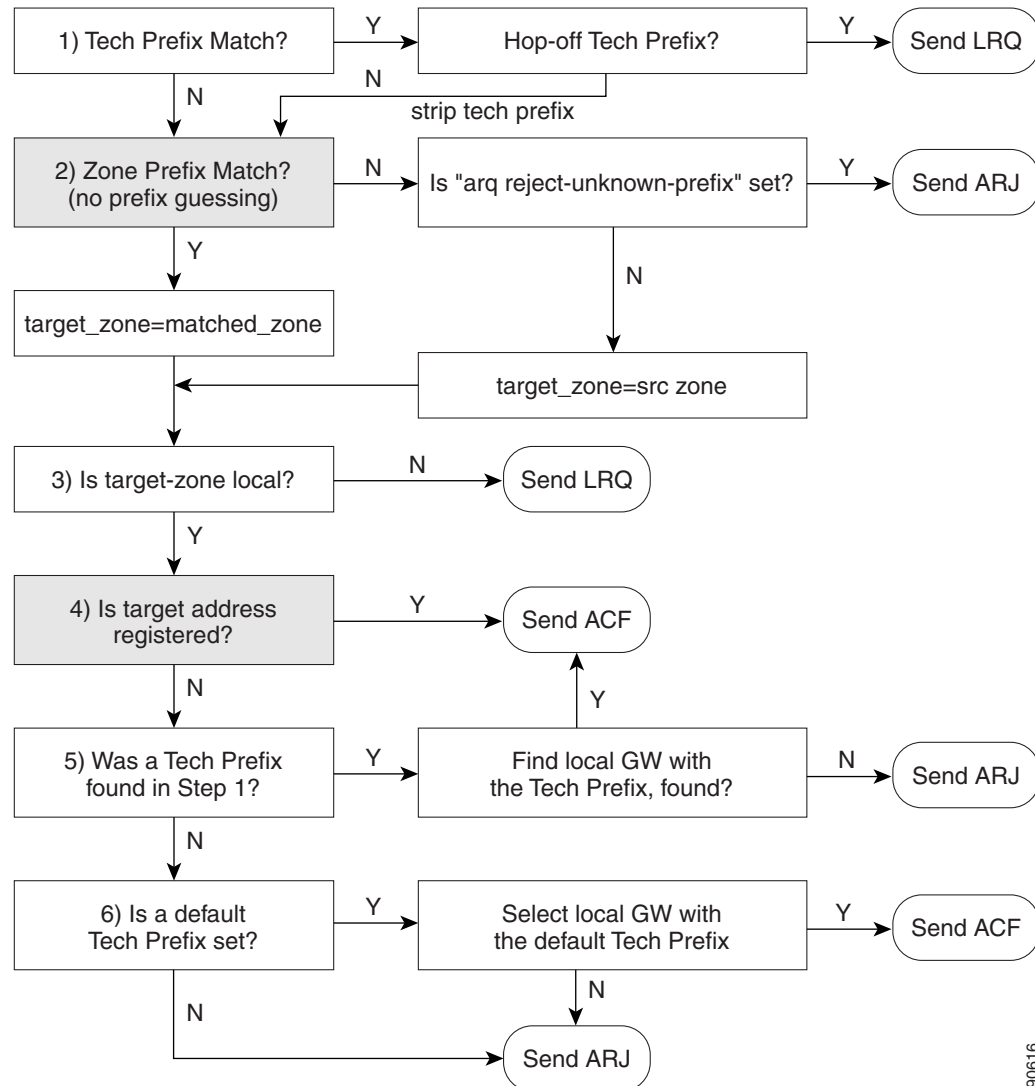
Longest-match is used if multiple potential matches exist. If no zone prefix can be matched, and if the gatekeeper is configured to accept calls with an unknown prefix, the gatekeeper then assumes that the destination zone is equal to the source zone.

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.

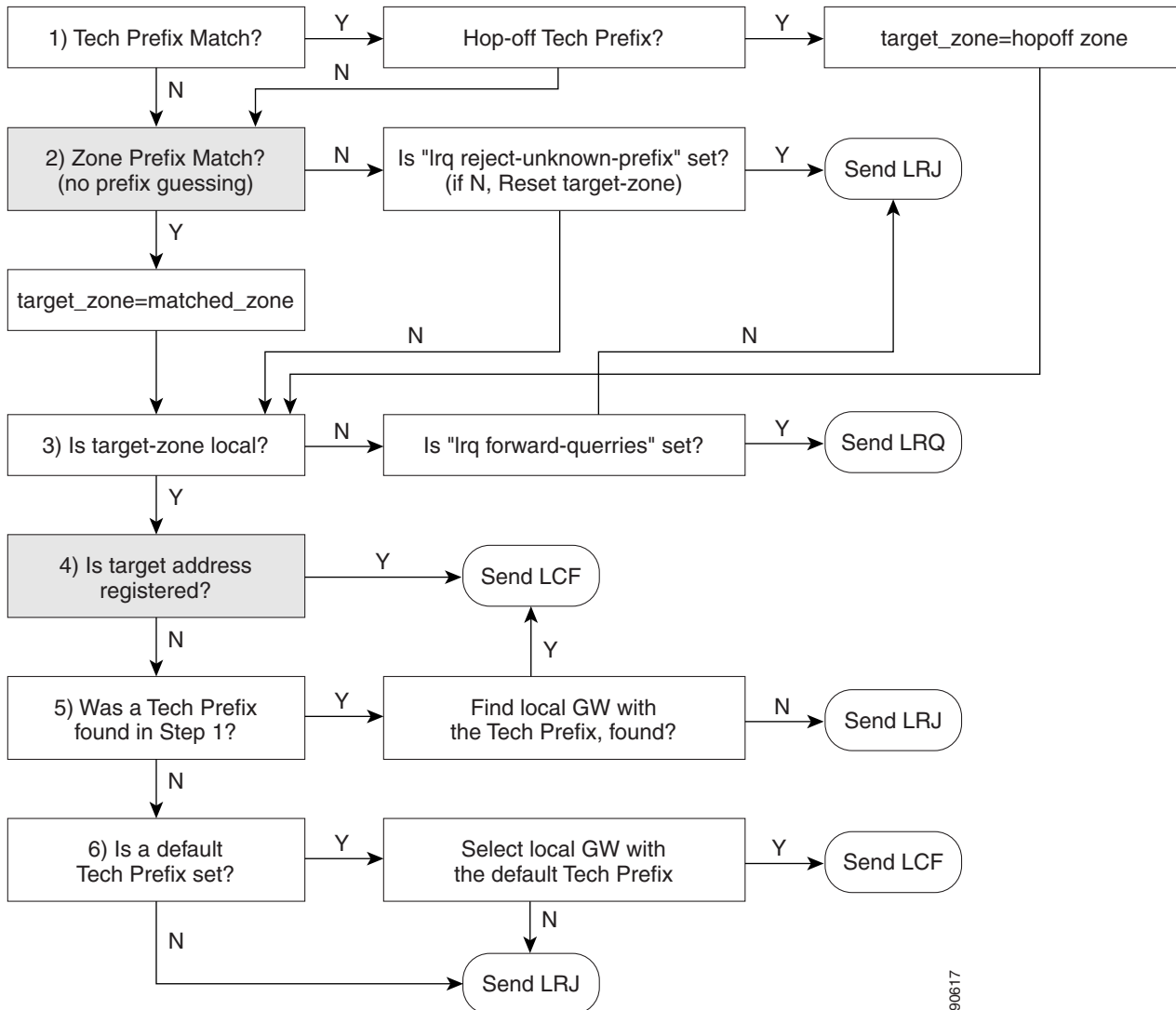
Upon receipt of an ACF from the gatekeeper, the source endpoint can send a setup message directly to the destination endpoint by using the IP address returned in the ACF.

Figure 10-18 Gatekeeper Address Resolution for an ARQ



90616

Figure 10-19 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 10-19 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 [Cisco IOS Gatekeeper and IP-to-IP Gateway with RSVP, page 9-28](#), 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.

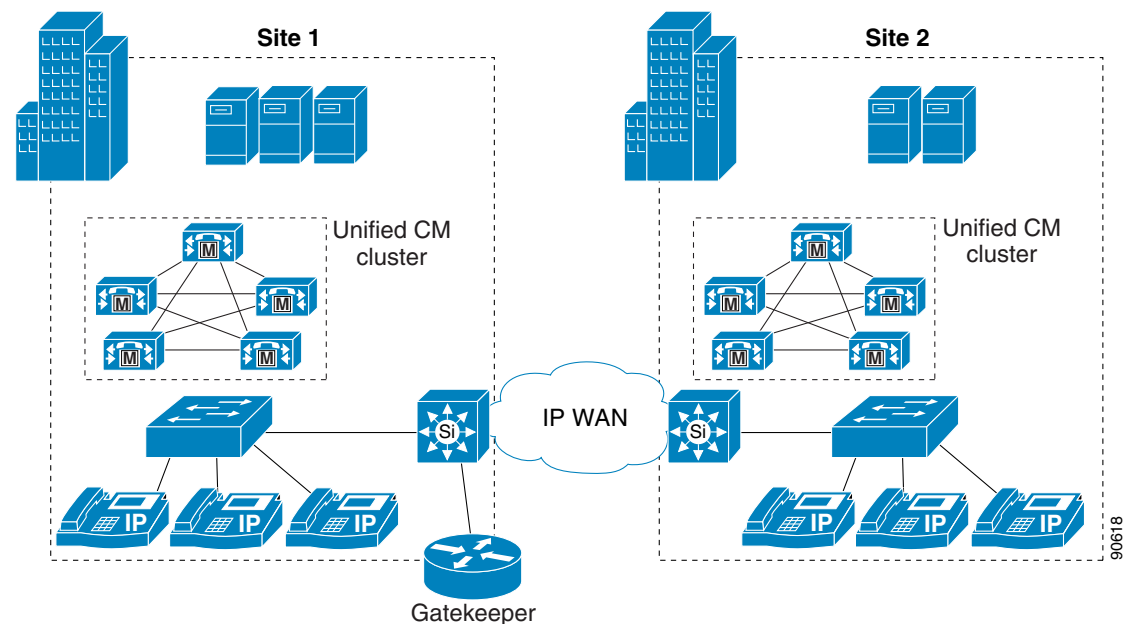
For ARQ messages, the gatekeeper determines if an **outvia** keyword is configured on the destination zone. If the **outvia** keyword is configured and the zone named with the **outvia** keyword is local to the gatekeeper, the call is directed to an IP-IP gateway in that zone by returning an ACF pointing 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 Unified CM clusters. [Figure 10-20](#) illustrates a distributed call processing environment with two Unified CM clusters and a single, centralized gatekeeper.

Figure 10-20 Centralized Gatekeeper Supporting Two Clusters



[Example 10-1](#) shows the gatekeeper configuration for the example in [Figure 10-20](#).

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 10-20](#):

- Each Unified CM cluster has a local zone configured to support Unified CM 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 Unified CM 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.

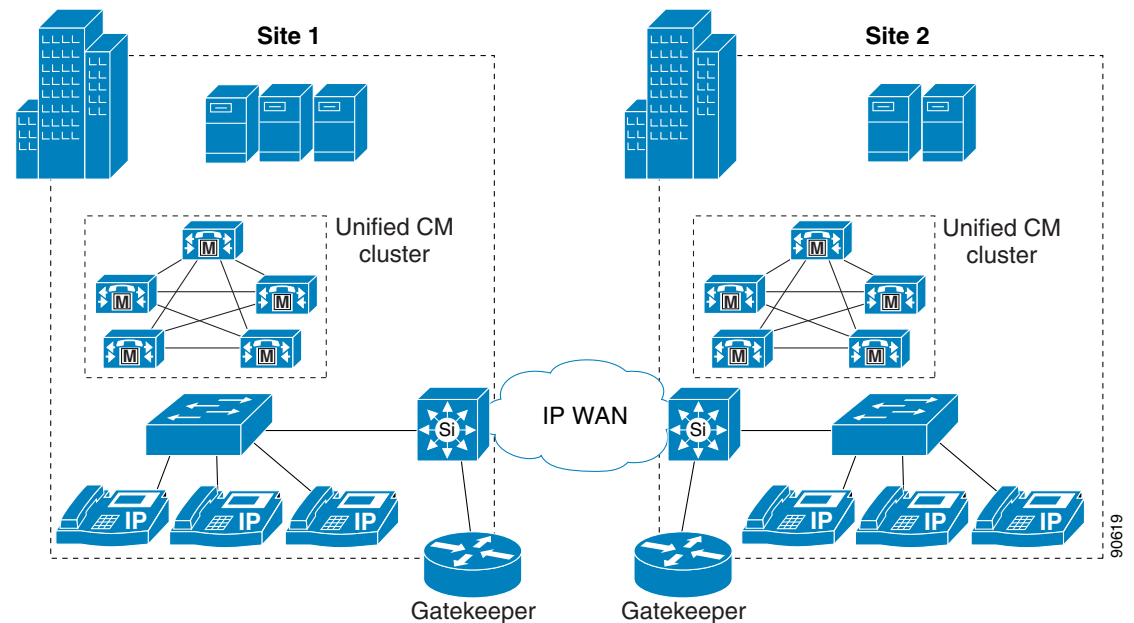
Cisco IOS gateways automatically add a technology prefix to outbound calls if the gateway has a prefix configured. The gateway also automatically strips the prefix from incoming H.323 calls. Unified CM can register with the gatekeeper using a technology prefix, as specified on the configuration page 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 Unified CM. 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 Unified CM 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 10-21](#) illustrates a distributed call processing environment with two clusters and two gatekeepers.

Figure 10-21 Distributed Gatekeepers Supporting Two Clusters



[Example 10-2](#) shows the gatekeeper configuration for Site 1 in [Figure 10-21](#).

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
arg reject-unknown-prefix
no shutdown
```

The following notes apply to [Example 10-2](#):

- A local zone is configured for registration of local Unified CM 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 Unified CM trunks have been configured to register with a 1# prefix.

- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Unified CM trunks.

Example 10-3 shows the gatekeeper configuration for Site 2 in Figure 10-21.

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:

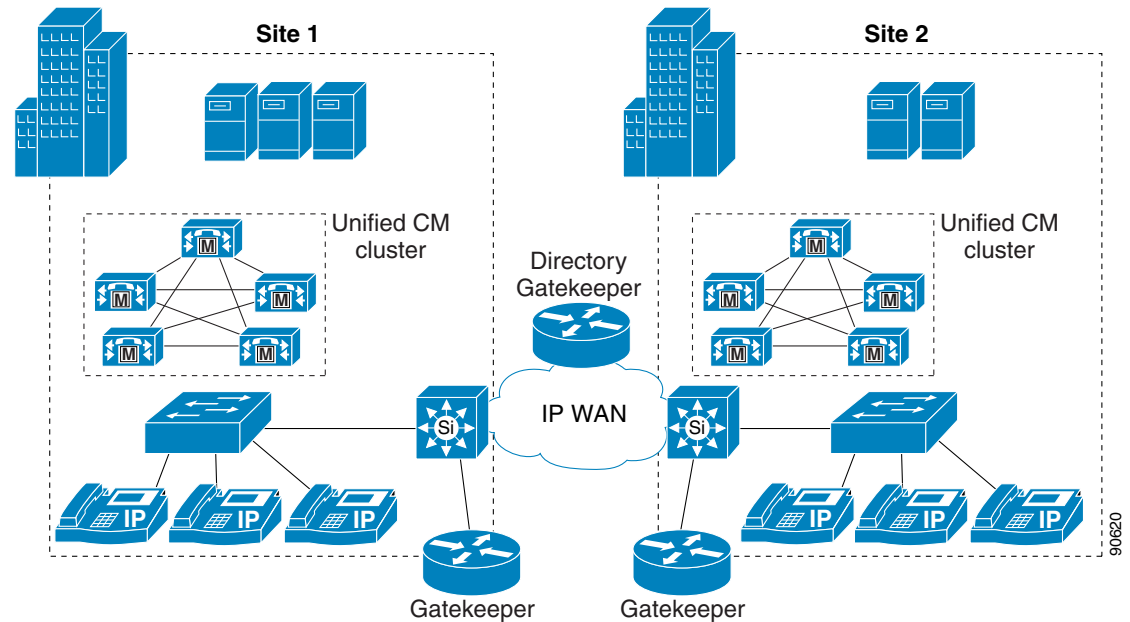
- A local zone is configured for registration of local Unified CM cluster trunks.
- A remote zone is configured for routing calls to the Site 1 gatekeeper.
- Zone prefixes are configured for both zones for inter-zone call routing.
- The **bandwidth remote** command is used to limit bandwidth between 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 Unified CM trunks have been configured to register with a 1# prefix.
- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Unified CM 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 10-22 illustrates a Unified CM distributed call processing environment with distributed gatekeepers for local call routing and a directory gatekeeper to provide call routing between gatekeepers.

Figure 10-22 Distributed Gatekeepers with a Directory Gatekeeper

Example 10-4 shows the gatekeeper configuration for Site 1 in [Figure 10-22](#). 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
arq reject-unknown-prefix
no shutdown
```

The following notes also apply to [Example 10-4](#):

- A local zone is configured for registration of local Unified CM 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 Unified CM trunks have been configured to register with a 1# prefix.

- The **arq reject-unknown-prefix** command guards against potential call routing loops across redundant Unified CM trunks.

Example 10-5 shows the directory gatekeeper configuration for the example in Figure 10-22.

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 Unified Communications Manager Express (Unified CME). 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:

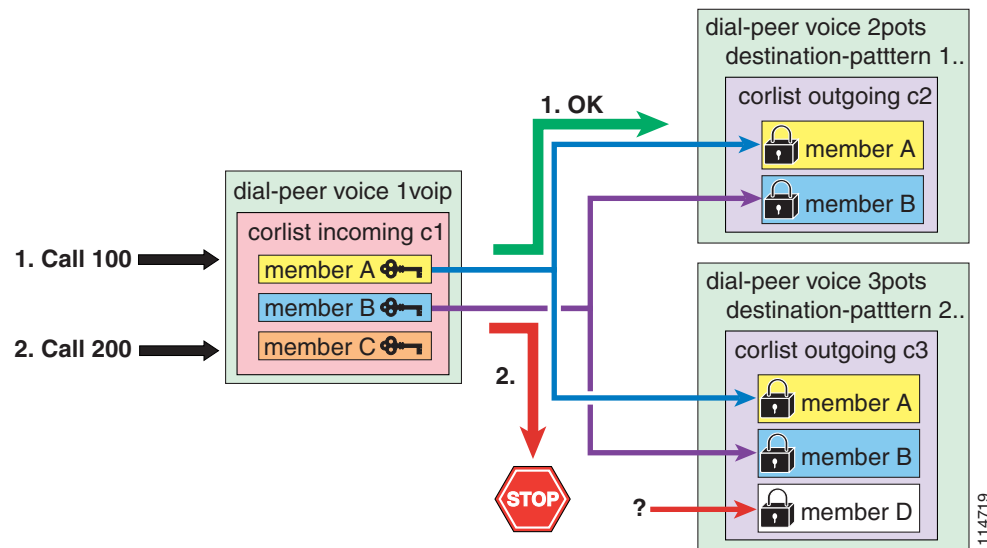
- If the members of the outgoing corlist associated with the outgoing dial peer are a subset of the members of the incoming corlist associated with the incoming dial peer, the call is permitted.
- If the members of the outgoing corlist associated with the outgoing dial peer are *not* a subset of the 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 10-23](#), where one VoIP dial-peer and two POTS dial-peer are defined.

Figure 10-23 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 10-23](#), 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** *corlist-name*.
- Step 3** Associate corlists with existing VoIP or POTS dial-peers by using the command **corlist {incoming | outgoing} corlist-name** 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 Unified CM 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 Unified Communications Manager Express (Unified CME), using Cisco IOS Release 12.2(8)T and later. Because the individual IP phones are specifically configured within Unified CME, 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 10-88, for an example of how to apply all these concepts in practice.

For more details on configuration of Cisco SRST and Unified CME, refer to the *Cisco SRST System Administrator Guide* and the *Cisco Unified Communications Manager Express System Administrator Guide*, both 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 10-4](#) shows the DID ranges and the internal site codes for this example.

Table 10-4 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/docs/ios/termserv/configuration/guide/tsv_reg_express_ps6441_TSD_Products_Configuration_Guide_Chapter.html

Design Considerations

This section presents the following dial plan design considerations for multisite deployments:

- [Design Guidelines for Multisite Deployments, page 10-55](#), covers guidelines and best practices that apply to all multisite deployment models.
- [Choosing a Dial Plan Approach, page 10-57](#), 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.
- [Deploying Dialed Pattern Recognition in the Phones with Unified CM 5.x, page 10-74](#), explains how SIP dial rules can be employed to enable SIP phones to recognize certain dialing patterns.
- The following sections analyze in detail three dial plan approaches and provide configuration guidelines for each:
 - [Deploying Uniform On-Net Dial Plans, page 10-58](#)
 - [Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing, page 10-61](#)
 - [Deploying Variable-Length On-Net Dial Plans with Flat Addressing, page 10-66](#)
- The following sections present two alternative ways of configuring classes of service within Unified CM:
 - [Building Classes of Service for Unified CM with the Traditional Approach, page 10-76](#)
 - [Building Classes of Service for Unified CM with the Line/Device Approach, page 10-80](#)
- [Building Classes of Service in Cisco IOS with H.323, page 10-88](#), explains how to implement classes of service within a Cisco IOS router running the H.323 protocol.
- [Deploying Call Coverage, page 10-91](#), provides guidelines and best practices for implementing call coverage functionality using hunt lists and line groups with Unified CM.

Design Guidelines for Multisite Deployments

The following guidelines and best practices apply in common to all multisite IP Telephony deployments. For deployments that involve more than one Unified CM cluster, also refer to the section on [Additional Considerations for Distributed Call Processing Deployments](#), page 10-56.

- 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 Unified CM.
- 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 automated PSTN failover is possible when the IP WAN is out of bandwidth:
 - 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.! and *9.!! (or *0.! and *0.!!) in the regular partition.

The second approach is preferable because it does not require additional calling search spaces to be defined for AAR. The additional * route patterns also 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, or to force a call through the PSTN when the destination number is located in a branch under SRST mode.

- 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 10-29).
- 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. For example, an emergency situation might prompt someone to place a call to a regular ten-digit number to reach a medical professional. Preemption of this call would abort the ongoing emergency communication and could delay handling of the emergency. Also, incoming calls from emergency service personnel would be at risk of preemption by MLPP.

**Note**

A Unified CM 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 Unified CM 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 multisite deployment involving multiple Unified CM clusters):

- Avoid splitting DID ranges across multiple Unified CM clusters. This practice would make intercluster routing very difficult because summarization would not be possible. Each DID range should belong to a single Unified CM cluster.
- Avoid splitting devices within a remote site between multiple Unified CM clusters using call admission control based on static locations. Static 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 split the available bandwidth between the clusters. Each remote site should belong to a single Unified CM cluster. With Cisco Unified CM 5.x, locations can be configured to use RSVP as the call admission control mechanism, which allows the efficient sharing of a single site's total WAN bandwidth between phones belonging to different clusters. To take full advantage of the efficiency of RSVP-based call admission control, all phones within the remote site must be configured on Unified CM clusters running releases 5.x.
- Use gatekeeper-controlled intercluster trunks to route calls between Unified CM 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 Unified CM and the gatekeeper by using a gatekeeper cluster and by assigning the intercluster trunk to a device pool that uses a Unified CM 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) Unified CM with dial length information for each remote site.
- Within the gatekeeper, configure one zone per Unified CM 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 Unified CM clusters by following these guidelines:
 - Add specific route patterns for the relevant E.164 ranges to the source (originating) Unified CM 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 Unified CM cluster
 - Ensure that the intercluster trunk calling search space in the destination Unified CM 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 Design Considerations, page 8-20](#).

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 10-58](#).

Variable-length on-net dial plans are more scalable but also more complex to design and configure. [Figure 10-24](#) shows the typical requirements for a large-scale deployment using the variable-length on-net dial plan approach.

The diagram illustrates a Unified CM Cluster architecture for E.164 dialing. At the top, a 'Unified CM Cluster' is shown as a group of four Unified Communications Manager (CM) servers connected in a mesh topology. This cluster is connected to a 'Voice mail' server. Below the cluster, a network switch is connected to two IP phones. A central router, labeled with a 'V' (Voice), is connected to the switch and the IP WAN cloud. The IP WAN cloud is connected to three sites: Site 1, Site 2, and Site N. Each site has a local router connected to the IP WAN cloud and two IP phones. The IP phones at each site are labeled 'DN 1XXX', indicating they share the same local directory number. A double-headed arrow between Site 1 and Site N is labeled '4-digit dialing within site', indicating that calls within a site are made using a 4-digit extension. A long double-headed arrow at the bottom is labeled 'E.164 or Site Code dialing between sites', indicating that calls between sites are made using a full E.164 number or a site code followed by a 4-digit extension.

- 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 10-61](#).
- Flat addressing

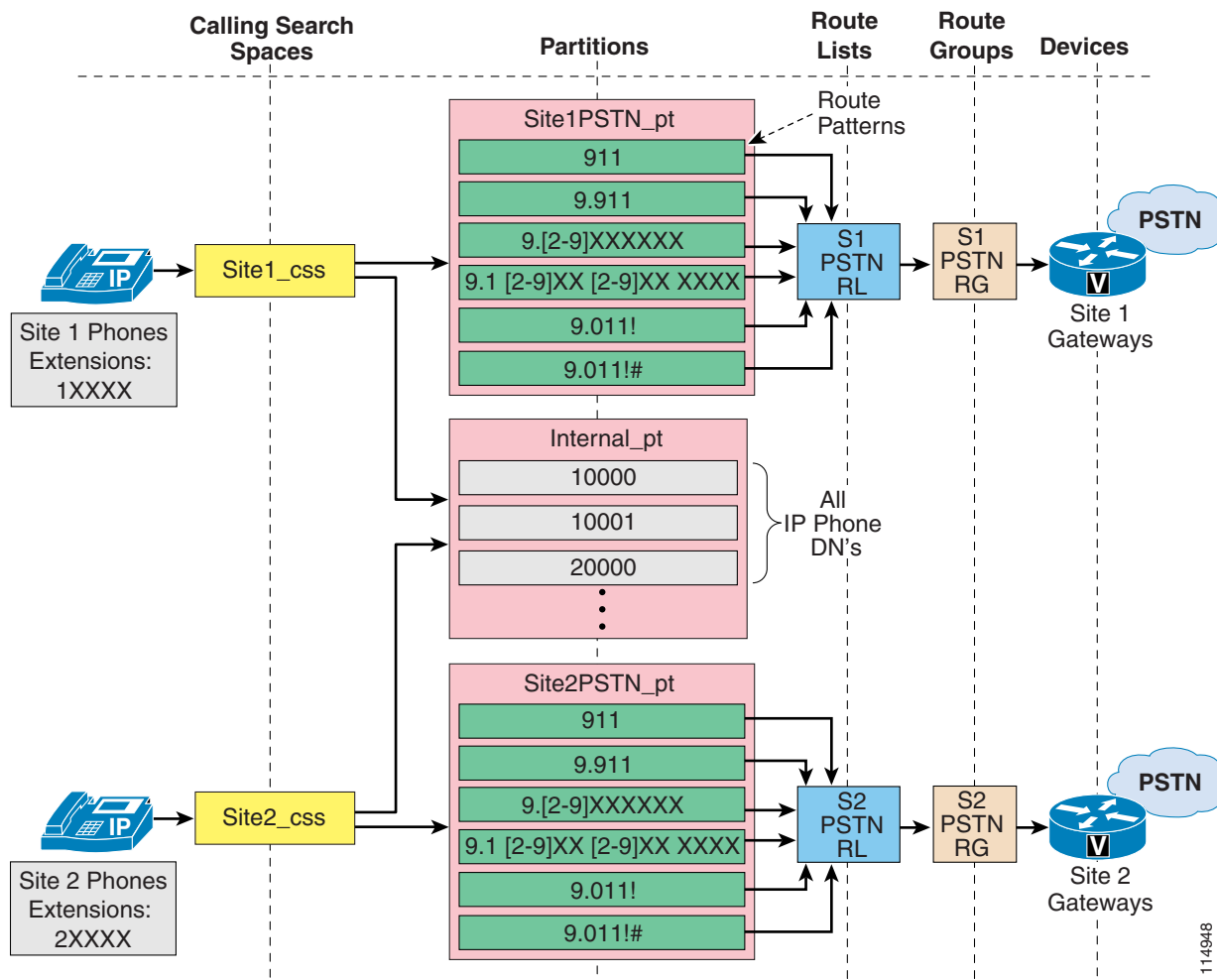
All internal extensions are placed in the same partition. This method is typically based on on-net site codes for inter-site calls and is analyzed in detail in the section on [Deploying Variable-Length On-Net Dial Plans with Flat Addressing, page 10-66](#). In some cases it is possible to use this approach 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 10-72](#).

- Uniquely identify all phones with an abbreviated extension.
- Place all the phone DNs 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 10-25 shows an example implementation for a single Unified CM cluster deployment.

Figure 10-25 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 10-60](#)
- [Outgoing PSTN and IP WAN Calls, page 10-60](#)

- [Incoming Calls, page 10-60](#)
- [Voicemail Calls, page 10-60](#)

Inter-Site Calls Within a Cluster

Because all internal DN's 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 Unified CM.

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 10-25](#)) 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 Unified CM cluster or Cisco Unified Communications Manager Express (Unified CME) 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.

Emergency Calls

If Cisco Emergency Responder is used for managing emergency calls, the partition containing the CTI route point used to send calls to Cisco Emergency Responder should be part of the calling search space of all phones in all branches instead of the site-specific 911 patterns as illustrated in [Figure 10-25](#). Cisco Emergency Responder will be able to identify the calling phone because there is no duplicity of DN's allowed in the internal partition. For more information on Cisco Emergency Responder considerations, refer to the chapter on [Emergency Services, page 11-1](#), and to the Cisco Emergency Responder product documentation available at

<http://www.cisco.com>

Incoming Calls

Incoming PSTN calls simply require stripping the excess digits in order to match the extension length configured in Unified CM. 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 Unified CM.

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 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 DNs defined at the various sites usually overlap with each other.

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 Partitioned Addressing**

Calling Search Space	Partition	Partition Contents
Site1_css	Site1Phones_pt	Site 1 phone DNs (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 DNs (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 10-66](#), because it enables you to greatly simplify the dial plan structure and to better manage and troubleshoot the system.

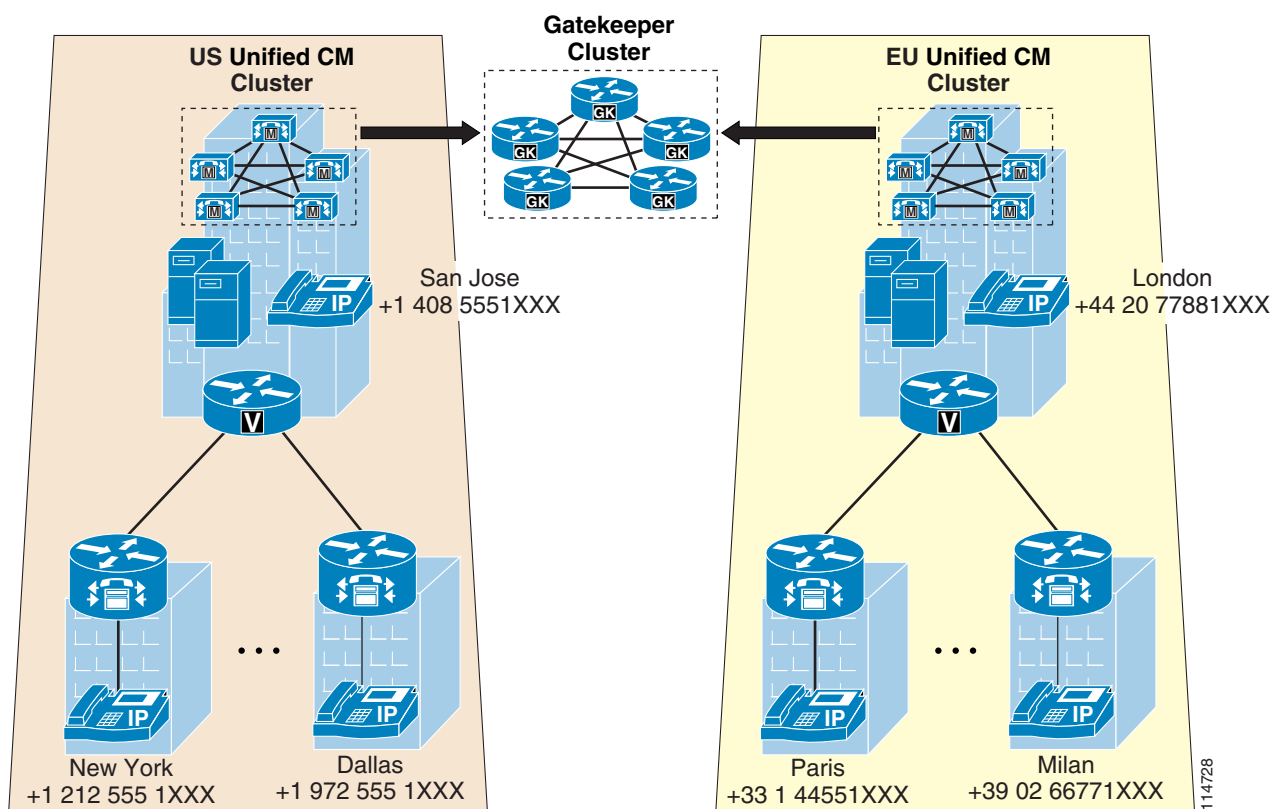
- Policy restrictions must be applied to on-net inter-site calls (that is, some or all users are not allowed to dial other sites on-net).
- Inter-site calls are always routed over the PSTN (as described in the section on [Voice Over the PSTN as a Variant of Centralized Call Processing, page 2-10](#)).
- CTI-based applications are not used across sites.

**Note**

Some CTI-based applications (such as Cisco Emergency Responder), 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 Unified CM Assistant, 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 10-26](#). 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 Unified CM 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 10-26 Example of a Large Multisite Deployment



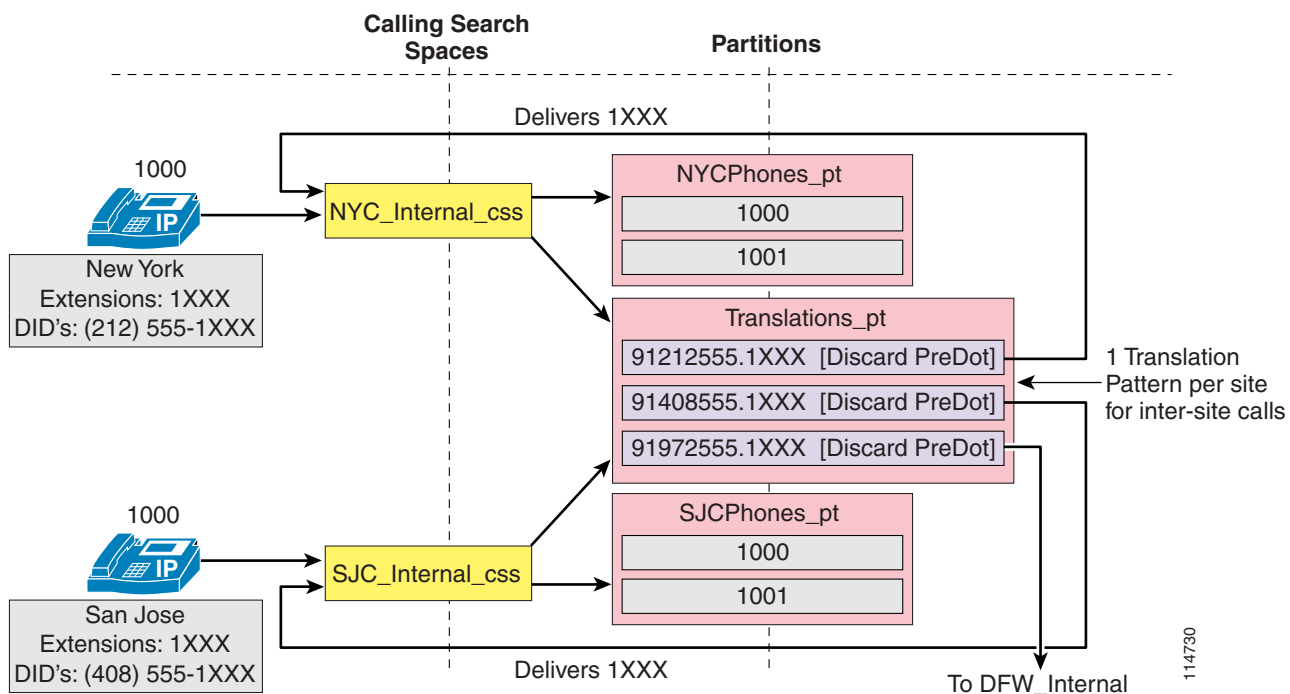
Using the United States (US) cluster from [Figure 10-26](#) as an example, the following sections analyze implementation details and best practices for different types of calls within the framework of the partitioned addressing approach:

- [Inter-Site Calls Within a Cluster, page 10-63](#)
- [Outgoing PSTN and IP WAN Calls, page 10-64](#)
- [Incoming Calls, page 10-65](#)
- [Voicemail Calls, page 10-65](#)

Inter-Site Calls Within a Cluster

[Figure 10-27](#) shows an example configuration for inter-site calls within the US cluster.

Figure 10-27 Inter-Site Calls Within a Cluster for the Partitioned Addressing Method



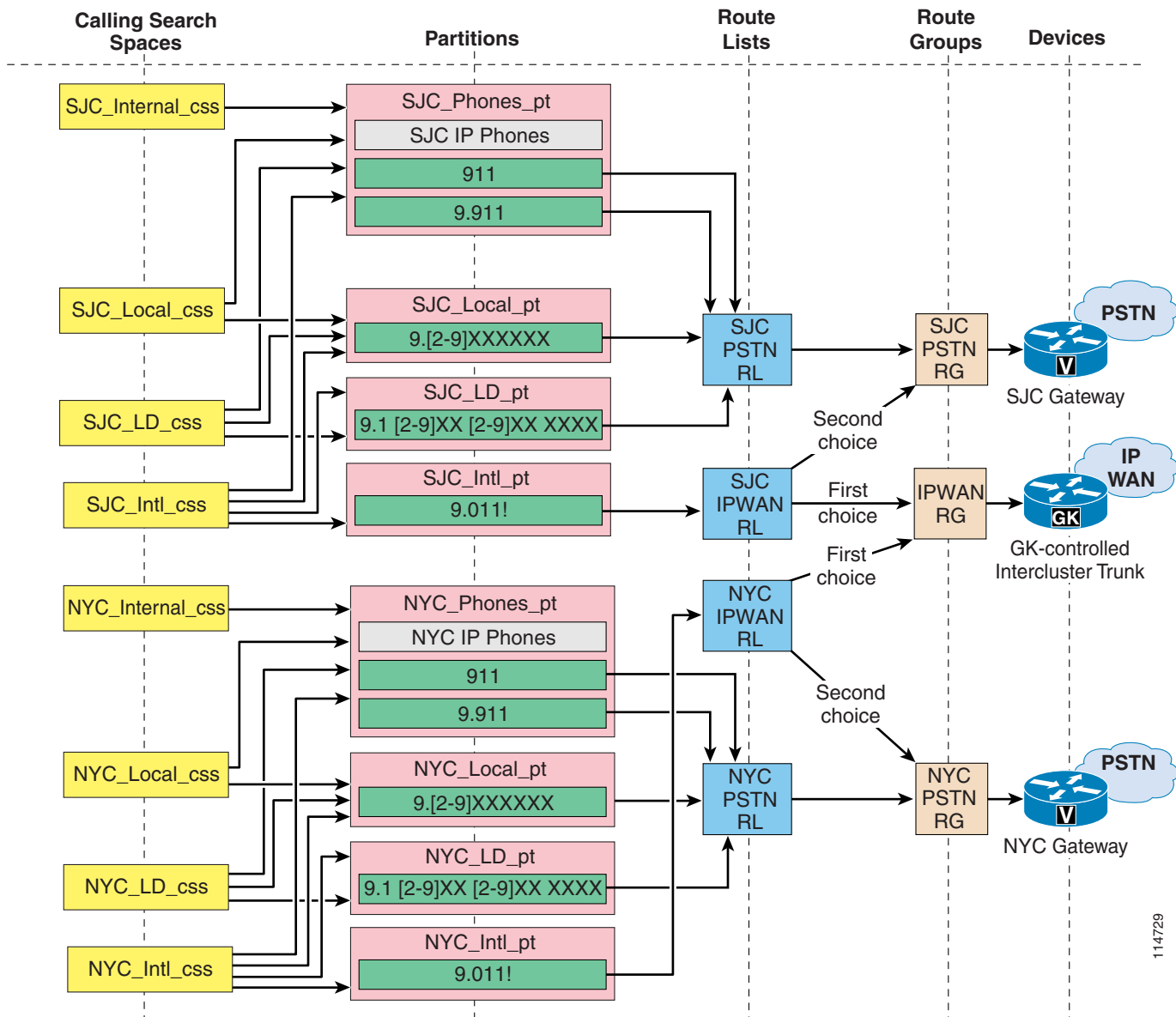
To provide connectivity between sites and partitions, use translation patterns according to the following guidelines:

- 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 10-28](#), 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 10-28 Outgoing PSTN and IP WAN Calls for the Partitioned Addressing Method.



114729

As [Figure 10-28](#) illustrates, for outgoing PSTN and IP WAN calls, there are no special considerations due to the partitioned nature of the internal addressing.

**Note**

Figure 10-28 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 incoming dialed number to match the already defined translation patterns. This approach assumes that the PSTN delivers the fully qualified number of each phone to the gateways, so that if an incoming dialed number is prefixed with 91, Unified CM will be presented with a number that matches the global translation patterns located in the `Translations_pt` partition.

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 Unified CM 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 in the `Translations_pt` partition 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 Unified CM. 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 10-6](#) 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 10-6 *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 10-72](#), for a variant of this approach that can accommodate these needs.

The following considerations apply to this approach:

- The destination numbers of intra-site four-digit calls get expanded to the unique internal DN on the IP phone display.
- The Placed Calls directory will display the original four-digit string as dialed by the user.

- 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 flat addressing approach, consider again the hypothetical customer network shown in [Figure 10-26](#). 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 10-7](#) summarizes the site codes chosen.

Table 10-7 **Site Codes for the Customer Network in Figure 10-26**

	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 10-68](#)
- [Outgoing PSTN and IP WAN Calls, page 10-69](#)
- [Incoming Calls, page 10-72](#)
- [Voicemail Calls, page 10-72](#)
- [Special Considerations for Deployments Without Site Codes, page 10-72](#)

Figure 10-29 shows an example configuration for inter-site calls within the US cluster.

The diagram illustrates the routing of calls from Calling Search Spaces to Partitions. It is divided into two main sections: **Calling Search Spaces** and **Partitions**.

Calling Search Spaces:

- New York:** Site code: 212, Extensions: 1XXX. A phone icon with "IP" is shown.
- San Jose:** Site code: 408, Extensions: 1XXX. A phone icon with "IP" is shown.

Partitions:

- NYC_Internal_css:** A yellow box representing the internal CSS for New York.
- SJC_Internal_css:** A yellow box representing the internal CSS for San Jose.
- NYC_Translations_pt:** A pink box representing the partition for New York. It contains a sub-box labeled "1XXX [Prefix 8212]".
- Internal_pt:** A pink box representing a central partition. It contains a list of phone numbers: 82121000, 82121000, 84081000, 84081000, and vertical dots.
- SJC_Translations_pt:** A pink box representing the partition for San Jose. It contains a sub-box labeled "1XXX [Prefix 8408]".

Routing and Delivery:

- Arrows show that calls from the New York Calling Search Space are routed to NYC_Internal_css, which then routes them to NYC_Translations_pt. A label "Delivers 82121XXX" is placed above this path.
- Arrows show that calls from the San Jose Calling Search Space are routed to SJC_Internal_css, which then routes them to SJC_Translations_pt. A label "Delivers 84081XXX" is placed below this path.
- Arrows also show that NYC_Internal_css routes calls to the Internal_pt partition.
- Annotations on the right side of the diagram state: "1 Translation Pattern per site for 'local' 4-digit dialing" (pointing to the 1XXX boxes) and "Phone DN's are directly reachable" (pointing to the list of phone numbers in the Internal_pt partition).

114733

- 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.


Note

You must ensure that the on-net access code and site code combinations do not overlap with the local abbreviated dialing range at any site.

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

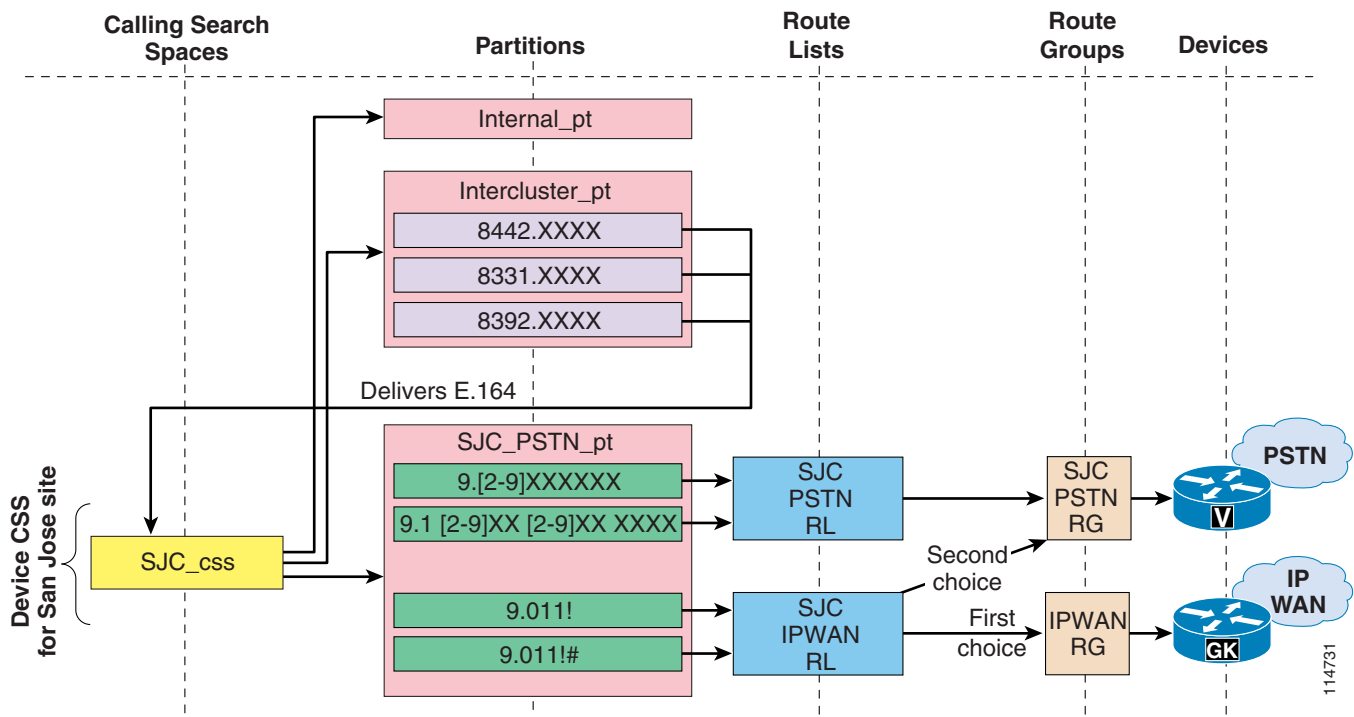
This option relies on a single route pattern that matches all eight-digit ranges (8XXXXXXX) and points to a route list or route group that contains only a gatekeeper-controlled intercluster trunk. The gatekeeper is configured to use the site codes as zone prefixes.

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 10-30](#), 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 10-30 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 10-30](#) assumes that the line/device approach to building classes of service is being used, but the same considerations apply when using the traditional approach.

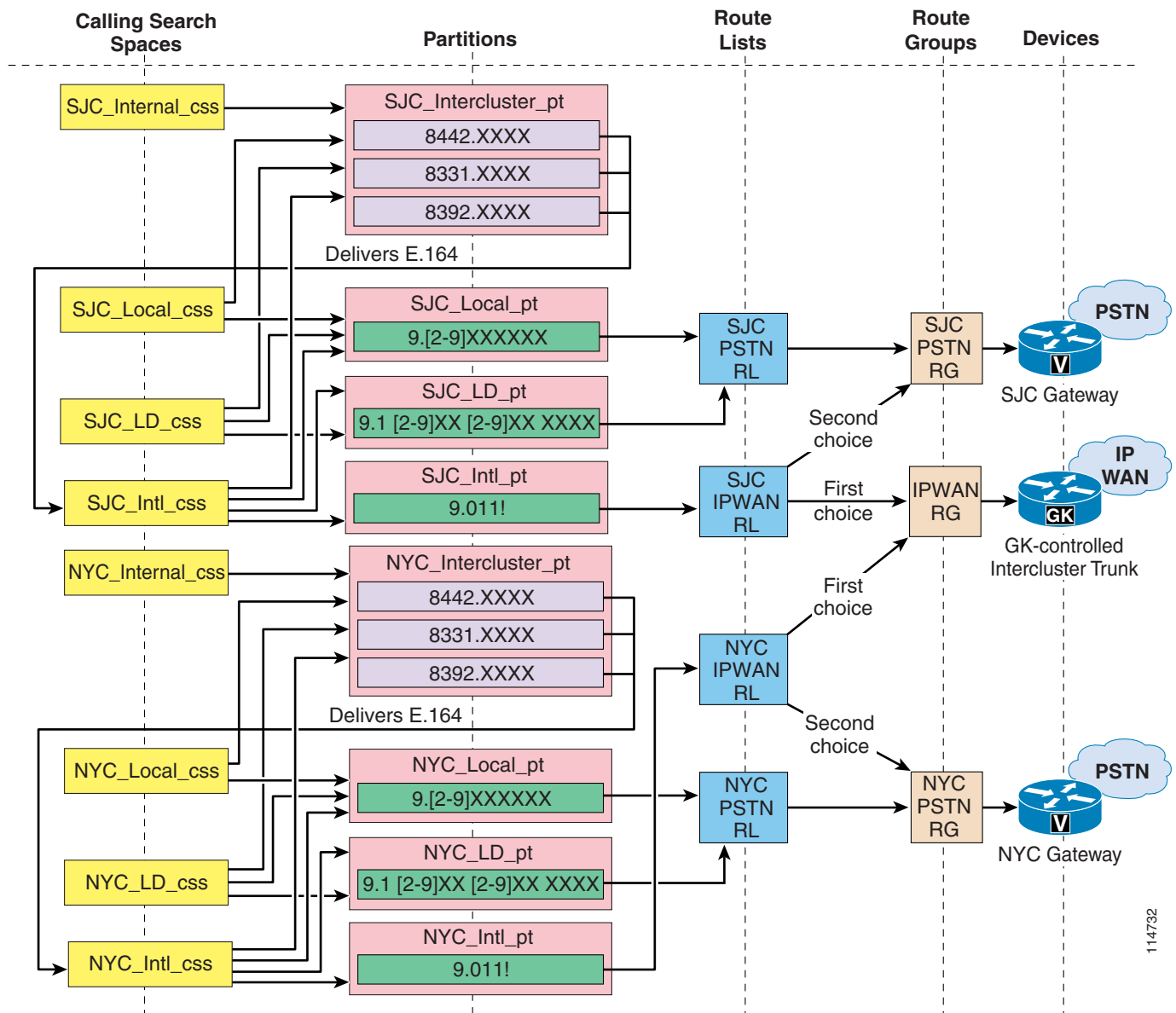
This solution requires a little more configuration and maintenance than that outlined in Option 1 because it requires that you configure and maintain information about other clusters' site codes and E.164 ranges. On the other hand, it provides automatic PSTN failover when the IP WAN is unavailable. PSTN 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 10-31](#), 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 10-31 *Outgoing PSTN and IP WAN Calls for the Flat Addressing Method with Distributed PSTN Failover for IP WAN Calls*



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 Unified CM 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 Unified CM.

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

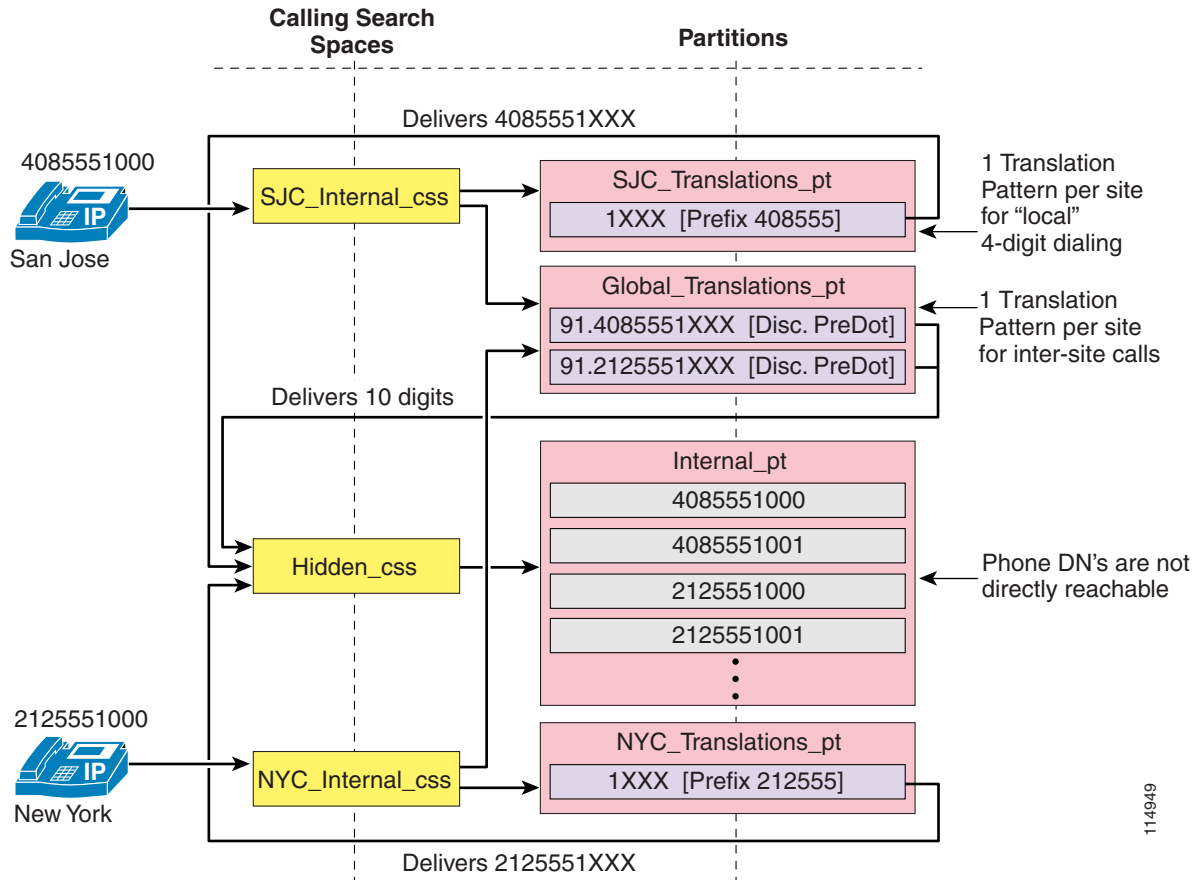
Because every eight-digit extension is unique within the system, the extension itself can be used to configure voicemail boxes within the voicemail system. No translations are necessary to send calls to the voicemail system or to enable Message Waiting Indicators (MWIs) within Unified CM. Note that users are required to use their eight-digit on-net number when prompted for the mailbox number.

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 Unified CM.

To implement this mechanism, follow these guidelines, illustrated in [Figure 10-32](#):

- 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 and country code (for example, 91) 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 10-32 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 Multisite Deployments, page 10-55](#), for details.
- The placed-calls directory displays digit strings as they were dialed by the user. For example, if the user dialed 1000 and a call was placed to phone 4085551000, the placed-calls directory would display 1000, thus allowing for direct redialing of the number without having to edit the dial string.
- The missed-calls and received-calls directories display phone numbers as they appeared when the call was offered to the phone. The calling number thus offered depends on the Calling Number configuration parameters of the translation patterns and the line configuration of the calling phone.
- The configuration shown in [Figure 10-32](#) 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).

Deploying Dialed Pattern Recognition in the Phones with Unified CM 5.x

The dialed pattern recognition capabilities of SIP phones need to take into account the typical dialing habits to be expected from users within the enterprise. Typically any combination of the following patterns may be in use in most enterprises:

- An abbreviated dialing pattern for calls within the same site (In the case of uniform on-net dial plans, the abbreviated dialing pattern could be used for inter-site calls.)
- An inter-site dialing pattern, typically used in variable on-net dial plans when using site codes and an on-net access code such as 8
- An off-net dialing pattern for local calls
- An off-net dialing pattern for long-distance calls
- Emergency call patterns, with and without the off-net access code
- An off-net dialing pattern for international calls

[Table 10-8](#) and [Table 10-9](#) show an example of the SIP dial rules that could be employed in an enterprise with the following dial plan characteristics:

- Abbreviated dialing is four digits (irrespective of whether abbreviated dialing is used for inter-site calls or not)
- Inter-site calls use 8 as an on-net access code, followed by seven digits representing the site code and DN
- Emergency dialing is allowed as 911 and as 9911
- Local seven-digit calls use 9 as an off-net access code, followed by the seven digits
- Local ten-digit calls use 9 as an off-net access code, followed by the ten digits
- Long-distance calls are dialed as 91 and ten digits
- International calls are dialed as 9011 followed by a variable quantity of digits, and dialing can be terminated by #.

Pattern recognition is concerned only with automating the collection of user digit input, to be forwarded automatically to Unified CM without requiring inter-digit timeout or pressing the Dial key. All enforcement of class of service is handled by the various calling search spaces chosen from within Unified CM. That is why all phones are configured with SIP dial rules allowing the recognition of international dialing, for example, even if not all phones will be assigned to an unrestricted class of service.

The dial plan characteristics listed above are representative of the variable-length on-net dial plan with flat addressing (see [Deploying Variable-Length On-Net Dial Plans with Flat Addressing, page 10-66](#)). From the standpoint of pattern recognition, this dial plan is compatible with the uniform on-net dial plan and the variable-length on-net dial plan with partitioned addressing (see [Deploying Uniform On-Net Dial Plans, page 10-58](#), and [Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing, page 10-61](#)).

For each pattern in [Table 10-8](#) and [Table 10-9](#), the description provides the pattern in equivalent Unified CM notation. The tables provide the SIP dial rules for both the 7905_7912 and 7940_7960_OTHER cases.

**Note**

The 7905_7912 SIP dial rules are limited to 128 characters, and the 7940_7060_OTHER SIP dial rules are limited to 8K (8,192) characters.

Table 10-8 **7940_7960_OTHER Dial Rule**

Description	Pattern	Timeout	Effect
Abbreviated 2XXX	2...	0	The combination of these six ranges represents the four-digit abbreviated dialing patterns that could be used at any site. As any string matching [2-7]XXX is dialed, it is sent to Unified CM immediately (timeout = 0).
Abbreviated 3XXX	3...	0	
Abbreviated 4XXX	4...	0	
Abbreviated 5XXX	5...	0	
Abbreviated 6XXX	6...	0	
Abbreviated 7XXX	7...	0	
Inter-site dialing 8.XXXXXXX	8,.....	0	Upon recognition of 8, secondary dial tone is played and seven more digits are collected, followed by immediate forwarding to Unified CM (timeout = 0).
Emergency 911	9,11	0	Upon recognition of 9, secondary dial tone is played and the digits 11 are collected, with immediate forwarding to Unified CM (timeout = 0).
Emergency 9.911	9,911	0	Upon recognition of 9, secondary dial tone is played and the digits 911 are collected, with immediate forwarding to Unified CM (timeout = 0).
Local PSTN 7-digits	9,.....	3	Upon recognition of 9, secondary dial tone is played and seven more digits are collected. Timeout of 3 seconds allows the user to continue dialing when local PSTN ten-digits dialing is configured.
Local PSTN 10-digits	9,.....	0	Upon recognition of 9, secondary dial tone is played and ten more digits are collected, with immediate forwarding to Unified CM (timeout = 0).
Long Distance	9,1.....	0	Upon recognition of 9, secondary dial tone is played and ten more digits are collected, with immediate forwarding to Unified CM (timeout = 0).
International with 6 seconds inter-digit timeout	9,011*	6	Upon recognition of 9, secondary dial tone is played, then 011 and a variable quantity of digits are collected. Timeout of 6 seconds allows for user to pause dialing without triggering a call to an incomplete string.
International with # as end of dialing	9,011*#	0	Upon recognition of 9, secondary dial tone is played, then 011 and a variable quantity of digits are collected, terminated by #. Immediate forwarding to Unified CM (timeout = 0).
Operator	0	0	As soon as 0 is detected, immediate forwarding to Unified CM (timeout = 0).

Table 10-9 7905_7912 Dial Rule

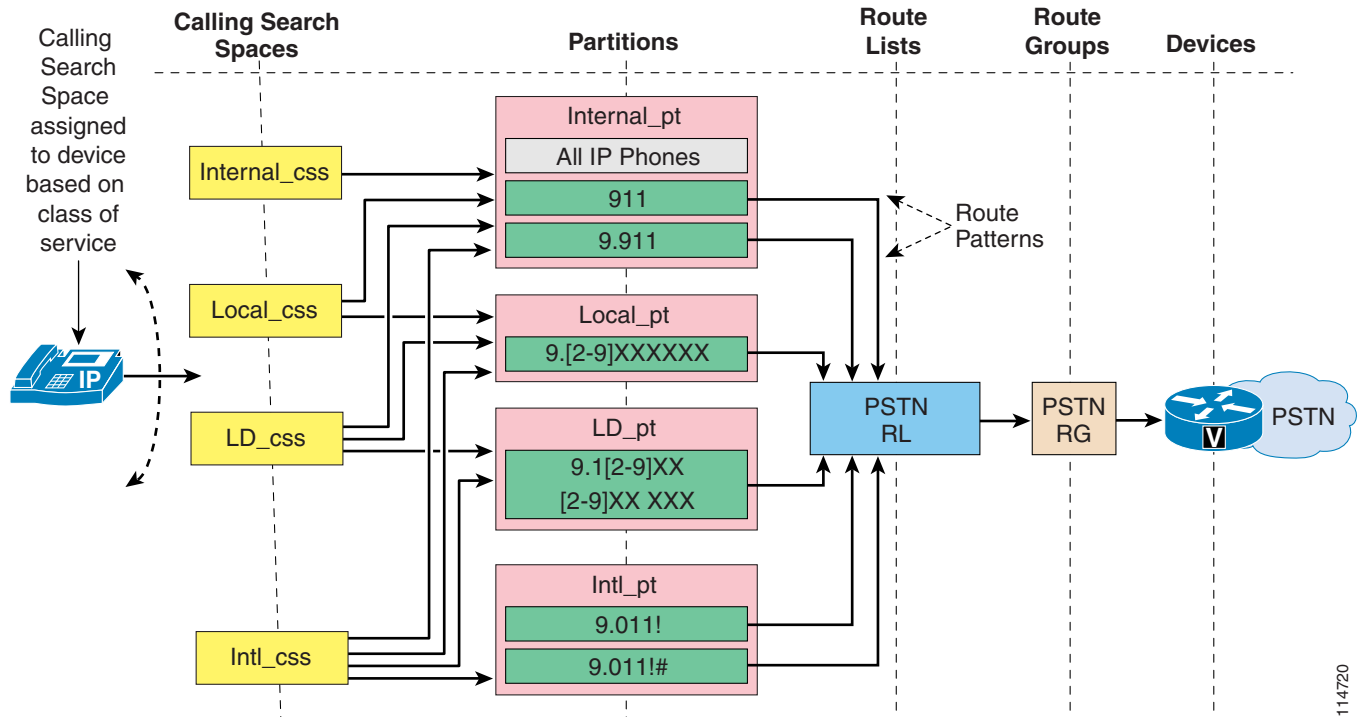
Description	Pattern	Effect
Abbreviated 2XXX	2...t0	The combination of these six ranges represents the four-digit abbreviated dialing patterns that could be used at any site. As any string matching [2-7]XXX is dialed, it is sent to Unified CM immediately (t 0).
Abbreviated 3XXX	3...t0	
Abbreviated 4XXX	4...t0	
Abbreviated 5XXX	5...t0	
Abbreviated 6XXX	6...t0	
Abbreviated 7XXX	7...t0	
Inter-site dialing 8.XXXXXXX	8.....t0	The digit 8 and seven more digits are collected, followed by immediate forwarding to Unified CM (t0).
Emergency 911	911t0	The digits 911 are collected, with immediate forwarding to Unified CM (t0).
Emergency 9.911	9911t0	The digits 9911 are collected, with immediate forwarding to Unified CM (t0).
Local 7-digits and LD	9.....t4>#....t1	The digit 9 and seven more digits are collected, with 4 seconds allowed for up to four other digits to be dialed. If another four digits are entered, they are sent to Unified CM after 1 second. The # would be recognized as the terminating character after 9 and seven digits are entered.
International	9011>#t6-	The digits 9 011 and a variable quantity of other digits are collected. Timeout of 6 seconds allows for user to pause dialing without triggering a call to an incomplete string. The # is allowed as terminating character.
Operator	0	As soon as 0 is detected, immediate forwarding to Unified CM (timeout = 0).

Building Classes of Service for Unified CM with the Traditional Approach

With Unified CM, 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 Unified CM device pages. In this way, all lines configured on the device automatically receive the same class of service.

Figure 10-33 shows a simple example for a single-site deployment.

Figure 10-33 Basic Example of Classes of Service Using the Traditional Approach

114720

With this approach, the device calling search space performs two distinct logical functions:

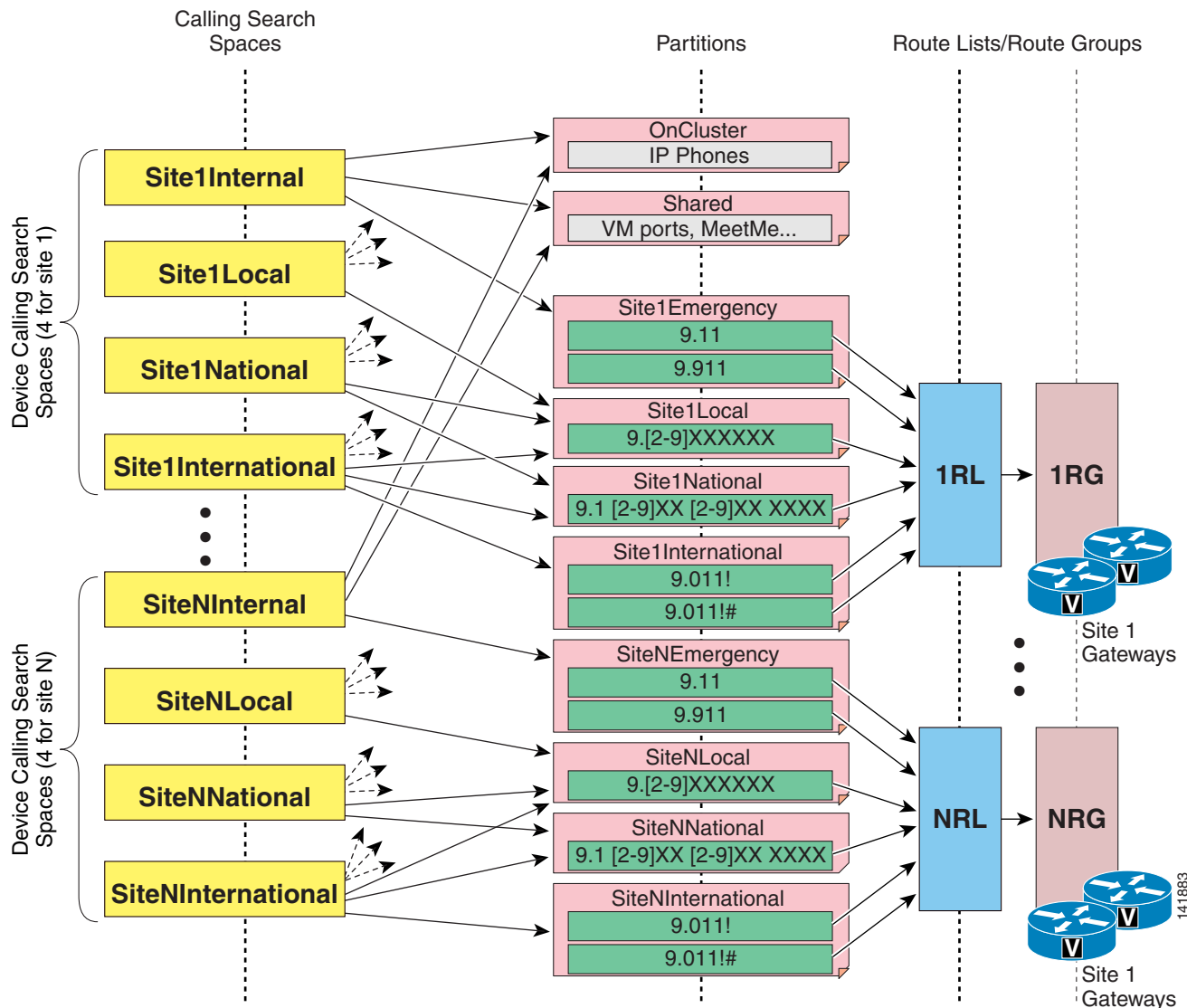
- Path selection

The calling search space contains specific partitions, which in turn contain specific route patterns that point to specific PSTN gateways through route lists and their associated route groups.

- 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 multisite 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 10-34](#).

Figure 10-34 Calling Search Spaces and Partitions Needed with the Traditional Approach

Follow these additional guidelines when applying this dial plan approach to a multisite deployment with centralized call processing:

- To facilitate on-net dialing between sites, place all IP phone DN's 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 DN's overlap. For more details on dial plans that use overlapping extensions, refer to [Deploying Variable-Length On-Net Dial Plans with Partitioned Addressing](#), page 10-61.
- 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 Unified CM with the Line/Device Approach](#), page 10-80.

**Note**

When Cisco Emergency Responder is used, the site-specific calling search space configured on the device should include the partition containing the 911 CTI route point pointing to Cisco Emergency Responder. This same partition can also contain a translation pattern 9.911 pointing to the same 911 CTI route point, to allow users to dial 9911 when trying to reach emergency services.

Building Classes of Service for Unified CM 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 multisite 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.

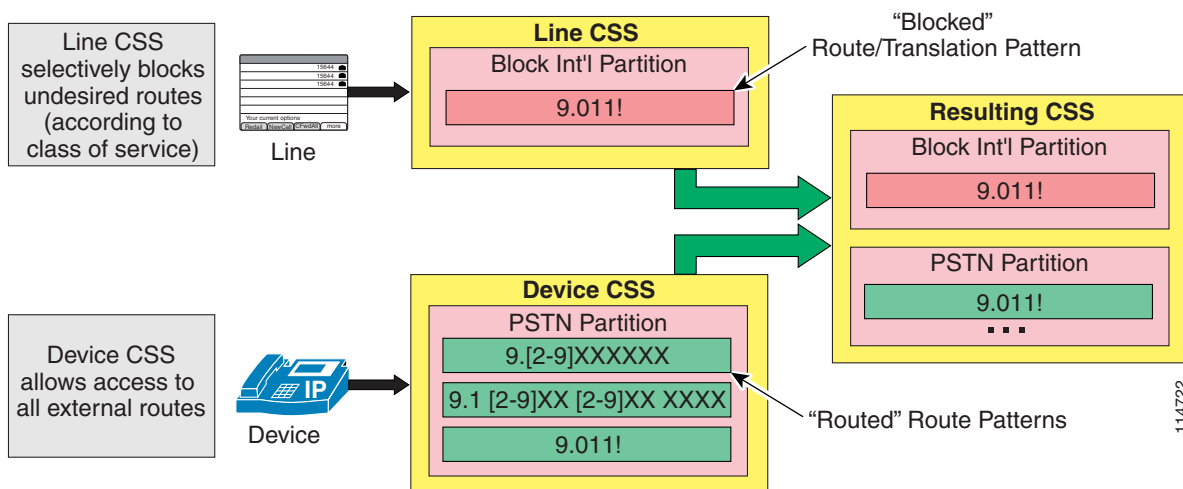
It is possible to significantly decrease the total number of partitions and calling search spaces needed by dividing these two functions between the line calling search space and the device calling search space, in what is called the *line/device approach*.

Keeping in mind how the line calling search space and the device calling search space for each given IP phone are combined within Unified CM, and how the line calling search space partitions appear first in the resulting calling search space (see [Calling Privileges in Unified CM](#), page 10-21), 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 10-35](#), 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 10-35 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:

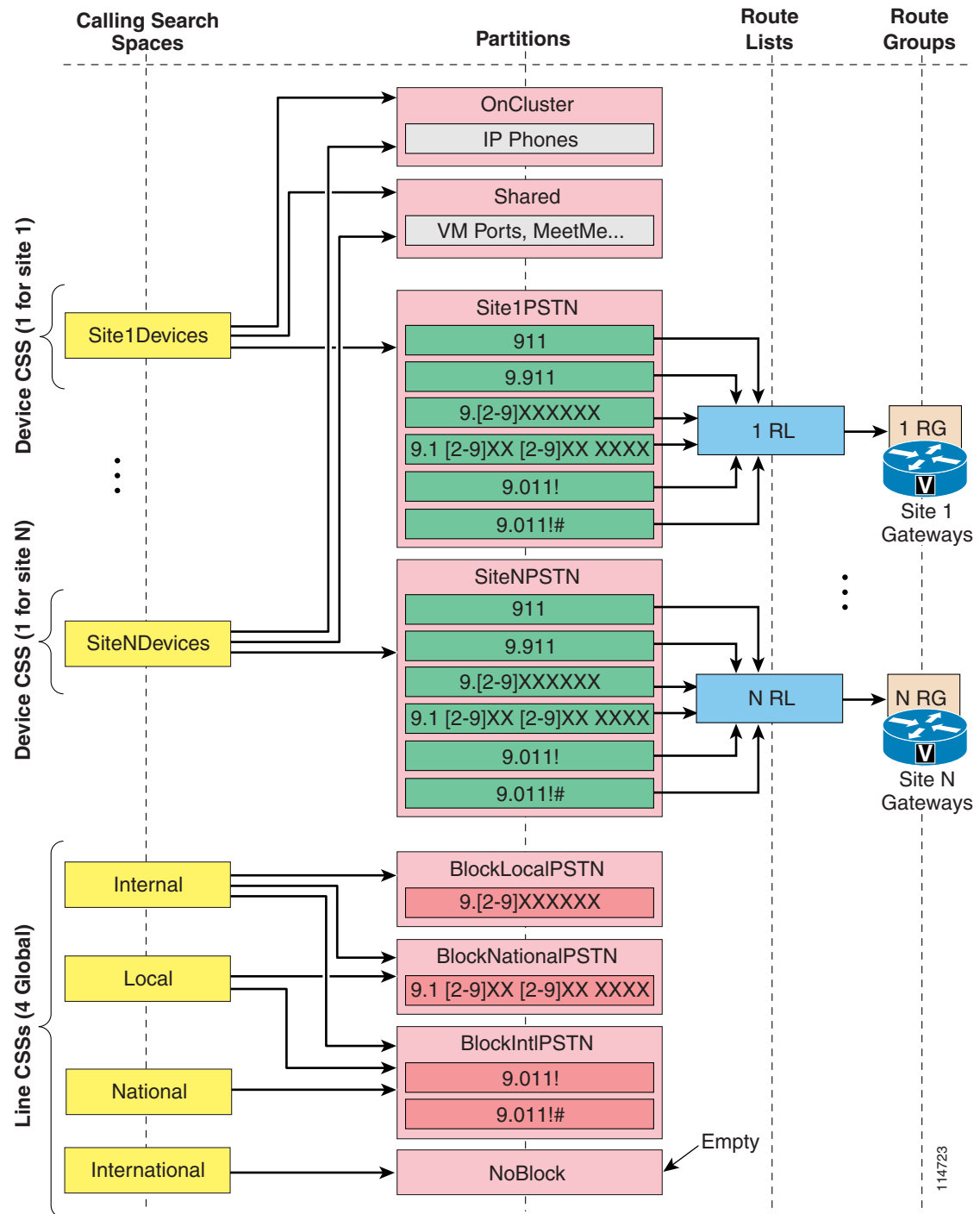
- When using a translation pattern, the end users will be able to dial the entire number and only then will they hear a fast-busy tone.
- When using a route pattern, the end users will hear a fast-busy tone as soon as the number they are dialing can no longer match any allowed pattern. This behavior assumes an IP phone running SCCP, or an Type-B IP phone running SIP with no SIP dial rules configured in the phone.

Follow these guidelines to implement the line/device approach in a multisite 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 10-36 shows an example of how these guidelines can apply to a multisite deployment with N sites.

Figure 10-36 Calling Search Spaces and Partitions Needed with the Line/Device Approach



This approach has the significant advantage that only a single site-specific, 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.

**Note**

If Cisco Emergency Responder is used, the 911 CTI route pattern and 9.911 translation pattern can be placed in the global On-Cluster partition.

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 10-10](#):

Table 10-10 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!#

Table 10-10 The Line/Device Approach Applied to the French National Numbering Plan

Calling Search Space	Partition	Blocked Route Patterns
Local03_css	BlockLD03_pt	0.0[124-6]XXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
Local04_css	BlockLD04_pt	0.0[1-356]XXXXXXXX
	BlockIntl_pt	0.00!, 0.00!#
Local05_css	BlockLD05_pt	0.0[1-46]XXXXXXXX
	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

In Cisco Unified CM 5.x, a new Secondary Calling Search Space for Forward All has been introduced on the line configuration page. This calling search space is concatenated with the line's Forward All calling search space to allow for the line/device dial plan approach to offer the same class of service to Forward All actions as to regular calls made from the same line. The concatenation order is first the Forward All calling search space followed by the Secondary Calling Search Space for Forward All.

Consider the following guidelines when using the line/device approach:

- The Forward Busy and Forward No Answer calling search spaces 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(s) 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 10-85](#).)
 - 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 the same class of service as regular calls, configure Forward All with the same calling search space as configured on the line, and configure the Secondary Calling Search Space for Forward All with the site-specific calling search space configured on the device.



Note

Type-A IP phones running SIP use the Rerouting Calling Search Space for Forward All actions initiated from the phone. Forward All action initiated from the user page or the administration pages in Unified CM use the concatenated calling search spaces as indicated above.

**Note**

In versions prior to Cisco Unified CM 5.0, 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 because 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 DNs are dialed. Access to these DNs 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 unless you are careful not to rely solely on the concatenation order for pattern selection but instead ensure that the desired blocked pattern's precision is greater in all cases than that of the permitted pattern(s).

Extension Mobility Considerations for the Line/Device Approach

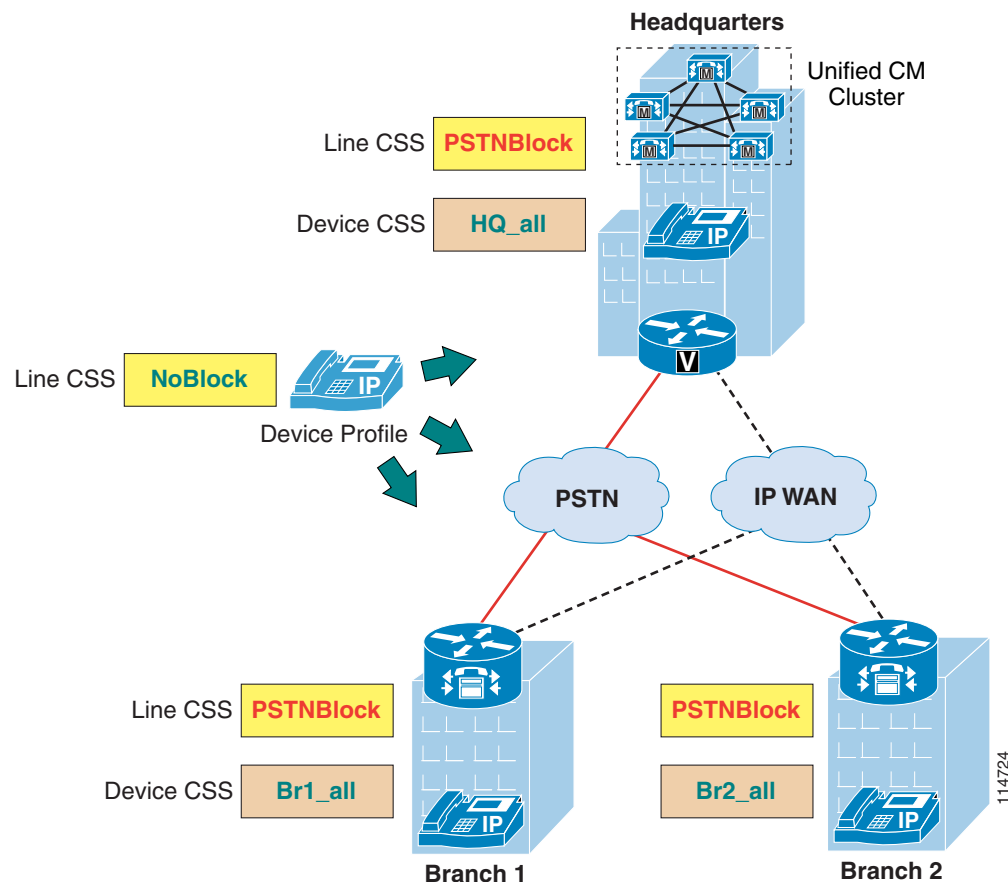
When using the Extension Mobility feature, the line/device approach provides a natural way to implement the dialing restrictions of a phone as a function of the logged-in (or logged-out) status of the phone. Typically, a logged-out phone should be restricted to calling other phones and services (such 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 multisite 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 10-37.

Figure 10-37 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.

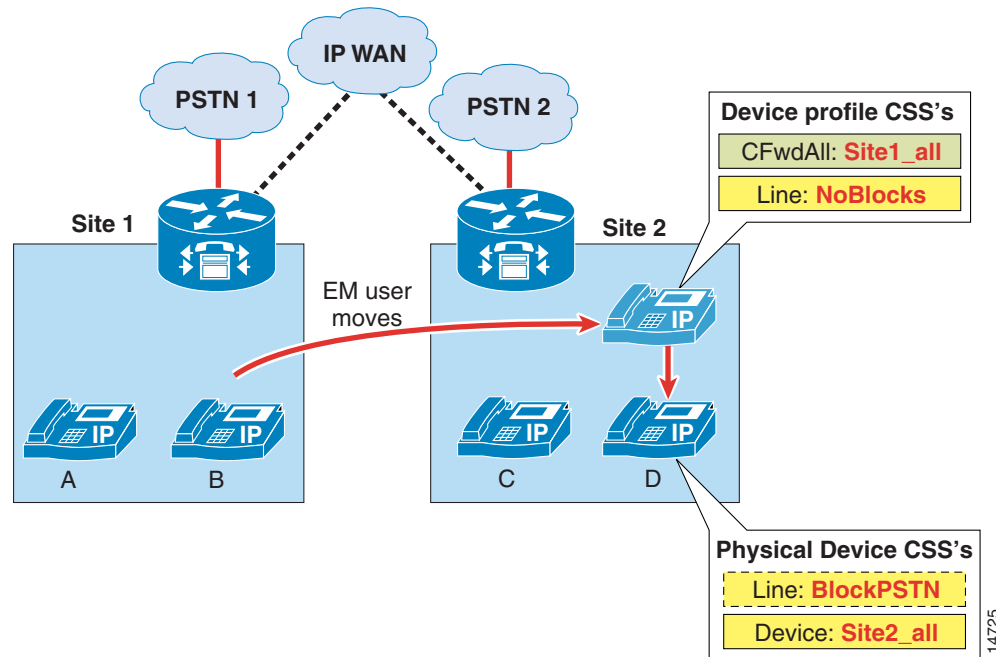
In summary, when you use the line/device approach for Extension Mobility, end-users have to adopt the dialing behavior of the site where they logged in.

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 10-38](#), 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 10-38 *Call Forwarding Considerations for Extension Mobility with the Line/Device Approach*



As described in the section on [Call-Forward Calling Search Spaces](#), [page 10-25](#), 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(s) 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 Unified CM cluster.

Keep this behavior in mind when designing the network and training the users.

**Note**

In Cisco Unified CM 5.x, a new Secondary Calling Search Space for Forward All has been introduced on the line configuration page of the User Device Profile. This calling search space is concatenated with the User Device Profile's line calling search space. Both of these calling search spaces are independent of the calling search spaces of the phone where the device profile is eventually used, therefore Forward All actions are not based on the call routing afforded by the site-specific calling search space of the device where the profile is used. The concatenation order is first the Forward All calling search space followed by the Secondary Calling Search Space for Forward All.

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:

- Unified CM multisite deployments with centralized call processing
- Cisco Unified Communications Manager Express (Unified CME) deployments

Under normal conditions in Unified CM multisite deployments with centralized call processing, classes of service are implemented using partitions and calling search spaces within Unified CM. 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.

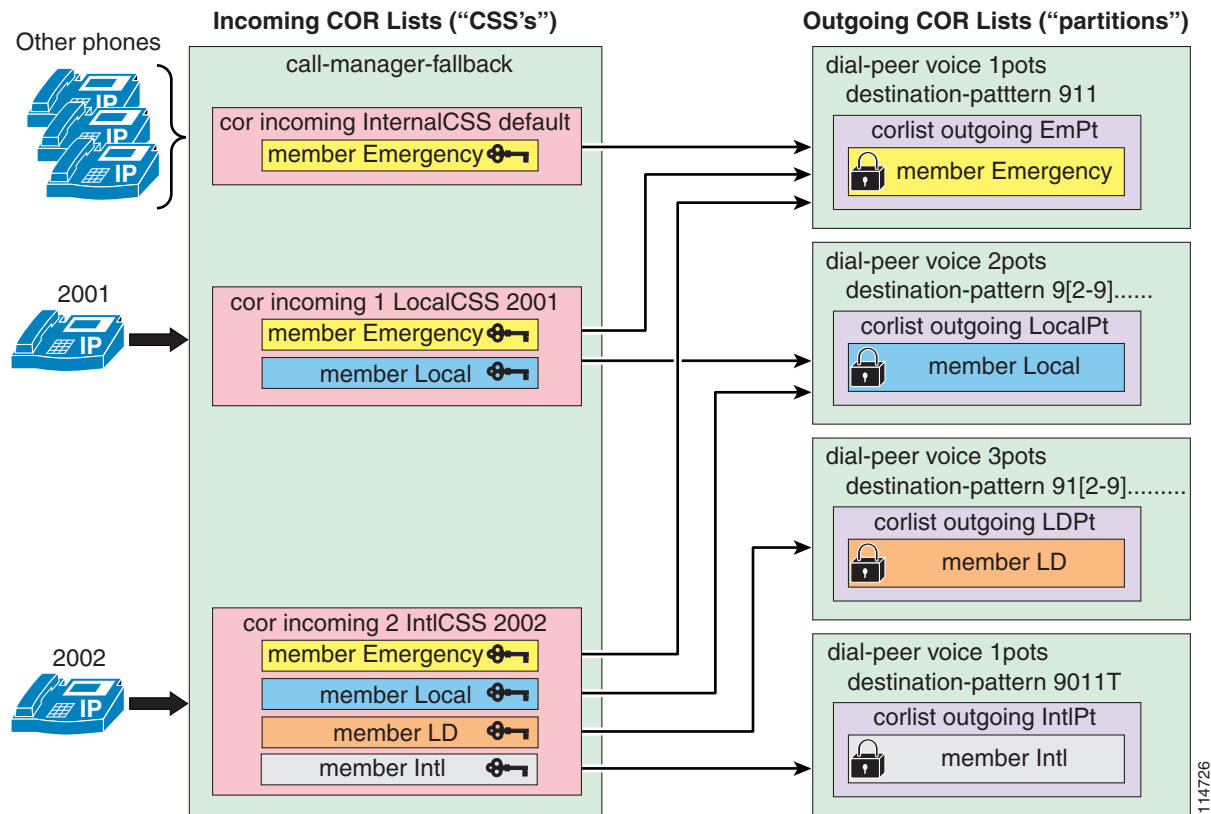
Similarly, in Unified CME deployments, the router needs a mechanism to implement classes of service for the IP phones.

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 10-50](#), for details on COR).

You can adapt the COR functionality to replicate the Unified CM 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 10-39](#) 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 10-39 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 10-39.

- 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 Unified Communications Manager Express (Unified CME) deployment, except for the following considerations:

- With Unified CME, 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. Unified CME 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 Unified CM Release 4.1 to manage call distribution, and it covers the following main topics:

- [Deploying Call Coverage in a Multisite Centralized Call Processing Model, page 10-92](#)
- [Deploying Call Coverage in a Multisite Distributed Call Processing Model, page 10-93](#)



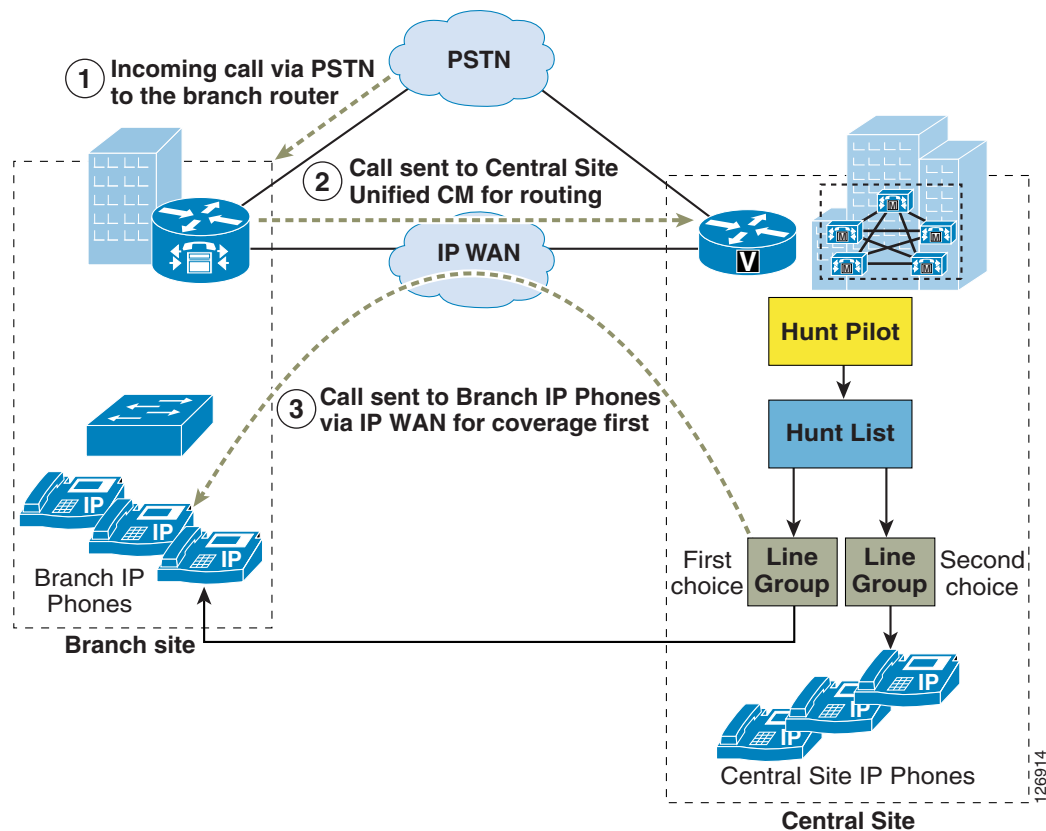
Note

Call coverage functionality does not offer call queues per se, and the caller is presented with ringback tone until a destination is found for the call. To provide prompting, music on hold, and so forth, Cisco offers many contact center technologies such as the Cisco Unified Customer Voice Portal (CVP). For more information on the contact center technologies available from Cisco, refer to the documentation at <http://www.cisco.com/go/designzone>.

Deploying Call Coverage in a Multisite Centralized Call Processing Model

Figure 10-40 shows an example of configuring hunt lists for a multisite 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 10-40 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 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 Unified CM blocks the call to one of its line group members due to insufficient WAN bandwidth. Instead, Unified CM distributes the call to the next available member or next available line group.



Note

Calls distributed by a hunt pilot can use AAR in Cisco Unified CM Release 4.1(3). However, Cisco strongly recommends that you use AAR only when using voicemail ports within line groups.

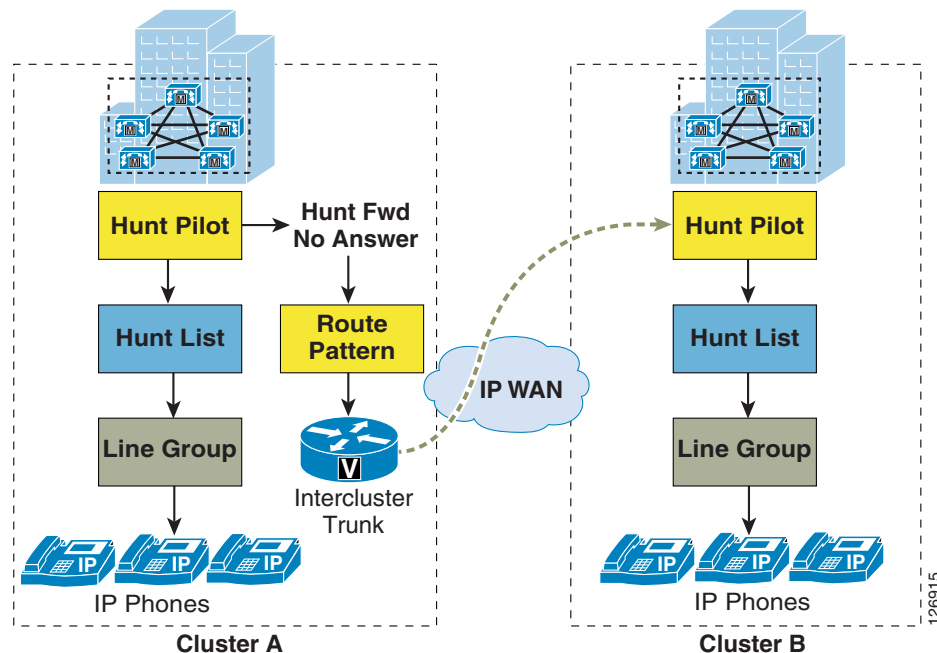
- Survivable Remote Site Telephony (SRST)
 - When Unified CM receives a call for the hunt pilot, and if some of its line group members are at the remote sites operating in SRST mode, then Unified CM skips those members and distributes the call to the next available line group member. From the perspective of Unified CM, 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 Multisite Distributed Call Processing Model

Beginning with Cisco Unified CM 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 Unified CM Release 4.1, can be used to match a route pattern that in turn points to gateways or trunks.

Figure 10-41 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 10-41 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.

Guidelines

Consider the following guidelines when deploying call coverage functionality in a distributed call processing model:

- When calls are distributed across multiple clusters, the route patterns should be properly configured to account for any digit transformations that are done on the outbound or inbound route group devices. If digit transformations are not done, then the configured route patterns and hunt pilot should be the same on all clusters, otherwise the calls will not be distributed appropriately.
- In distributed call processing deployments, load sharing of incoming hunt pilot calls can be managed using Cisco VoIP gateways. In the event that an incoming call is not answered within one cluster, it can overflow to another cluster for service.

**Tip**

Calls can 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 Unified CM 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 Unified CM 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 DN's. Additionally, the number of broadcast line groups depends on the BHCC. If there are multiple broadcast line groups in a Unified CM 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



CHAPTER 11

Emergency Services

Last revised on: September 27, 2007

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 11-1](#)
- [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.nena.org/media/files/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.
- Multisite 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 Police Department and San Jose Police Department are the appropriate PSAPs
- San Francisco Police Department and San Jose Police Department are served by the same 911 network service provider
- However, San Francisco Police Department and San Jose Police Department 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.att.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 Unified Communications Manager 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 Unified Communications Manager, 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 Unified Communications Manager (Unified CM) 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 Unified CM 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, web page, or phone call) to inform them that there is an emergency call in progress. The pager and web page notifications include the calling party name and number, the ERL, and the date and time details associated with the call. Phone notification provides the information about the calling number from which the emergency call was placed.

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/en/US/products/sw/voicesw/ps842/tsd_products_support_series_home.html

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 Unified CM 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 11-11](#)
- [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.

Emergency Responder Version Compatibility with Unified CM

Emergency Responder 1.3(1) is required for compatibility with Cisco Unified CM 5.0. For a full description of the compatibility between Emergency Responder and Unified CM software versions, refer to the *Cisco Emergency Responder Administration Guide 1.3(1)*, available at

<http://www.cisco.com>

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 Unified CM 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 Unified CM 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

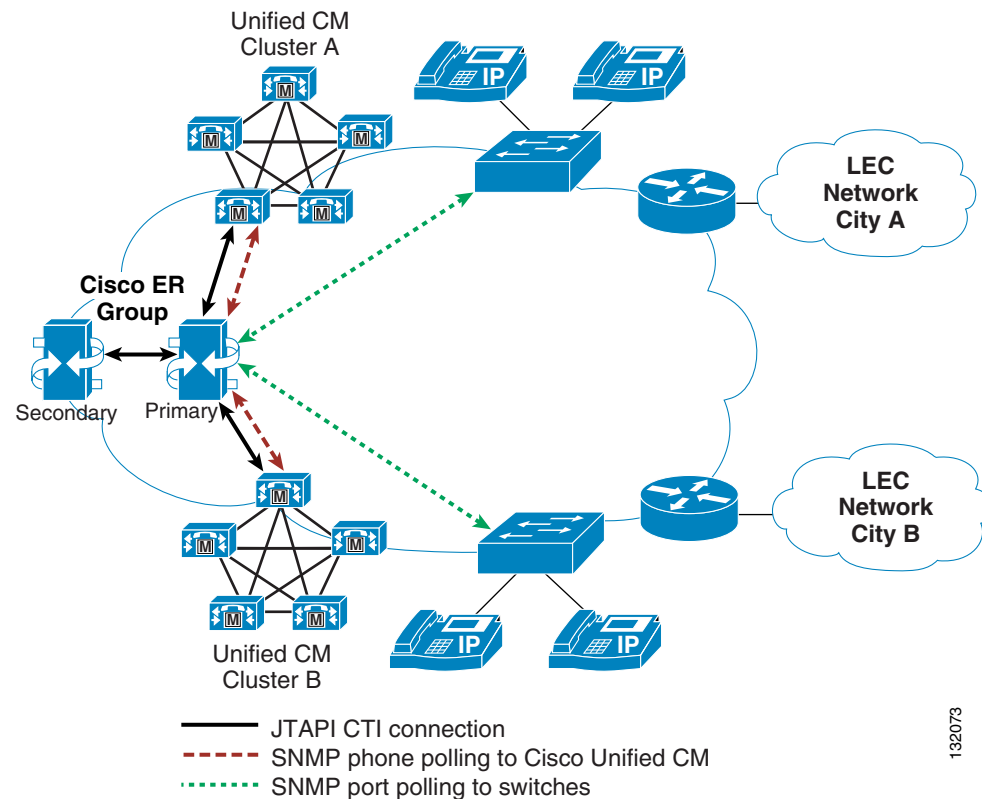
<http://www.cisco.com>

Single Cisco ER Group

A single Emergency Responder group can be deployed to handle emergency calls from two or more Unified CM 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 Unified CM clusters.

Figure 11-1 A Single Cisco ER Group Connected to Two Unified CM Clusters



132073

The single Cisco ER group in [Figure 11-1](#) interfaces with the following components:

- Each Unified CM 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 Unified CM 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 Unified CM software to which it is connected. At system initialization, Cisco ER interrogates the Unified CM 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 Unified CM clusters to which a single Cisco ER group is interfaced *must* run the same version of Unified CM software.

For some deployments, this software version requirement might present some difficulties. For instance, during a Unified CM 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 Unified CM clusters, apply the following guidelines:

- Make Unified CM 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.

**Note**

Emergency Responder version 1.3(1) is required if any of the Unified CM clusters are running Cisco Unified CM Release 4.2 or 5.0.

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 Unified CM 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 Unified CM cluster

- The access switches (via SNMP) to which most of the phones associated with the Unified CM of the Cisco ER group are most likely to be connected

This approach allows Unified CM 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

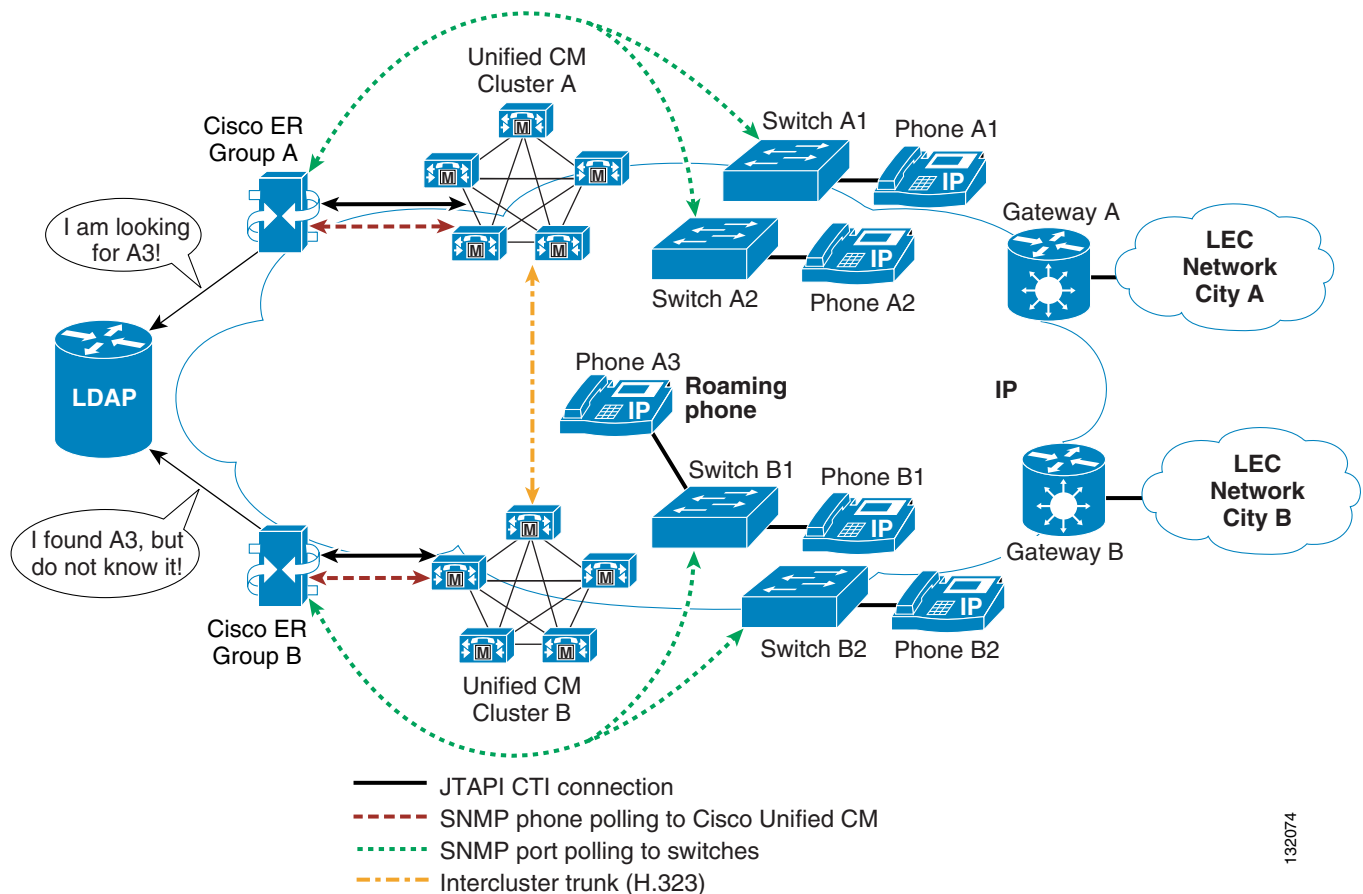


Figure 11-2 illustrates the following topology:

- Cisco ER group A is interfaced to Unified CM cluster A to access switches A1 and A2, and it is deemed to be the home Cisco ER group of all phones registered to Unified CM cluster A.
- Likewise, Cisco ER group B is interfaced to Unified CM cluster B to access switches B1 and B2, and it is deemed to be the home Cisco ER group of all phones registered to Unified CM cluster B.

**Note**

Emergency Responder 1.3(1) requires that all ER groups in an ER cluster run the same version of software. Therefore, if any of the Unified CM clusters are using Cisco Unified CM Release 4.2 or 5.0, then all of the ER groups must use Emergency Responder 1.3(1).

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 Unified CM cluster. For example, a phone moving between access switches A1 and A2 remains registered with Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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 Unified CM cluster A for processing.
2. Unified CM 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 Unified CM cluster B.
4. Unified CM cluster A sends the call to Unified CM cluster B.
5. Unified CM 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 Unified CM 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. Unified CM cluster B routes the call according to the new called number information obtained from Cisco ER group B.
8. Unified CM 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 Unified CM 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 Unified CM 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.

For more information about the ALI formatting tools, refer to the online documentation available at

http://www.cisco.com/en/US/products/sw/voicesw/ps842/prod_maintenance_guides_list.html

For LECs not listed at this URL, the output from Unified CM can be formatted using standard text file editing tools, such as spreadsheet programs and standard text editors.



CHAPTER 12

Third-Party Voicemail Design

Last revised on: September 27, 2007

This chapter discusses various options for deploying third-party voicemail systems with Cisco Unified Communications Manager (Unified CM).



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.

There are many voicemail vendors, and it is not uncommon for customers to want to continue to use an existing voicemail system when deploying Unified CM. 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 Unified CM 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 Unified CM:

- [SMDI, page 12-1](#)
- [Digital Set Emulation, page 12-4](#)
- [Dual PBX Integration, page 12-5](#)
- [Centralized Voicemail, page 12-6](#)
- [Positive Disconnect Supervision, page 12-11](#)
- [Summary of Third-Party Voicemail Integration, page 12-11](#)

SMDI

Unified CM supports use of SMDI through either of the following methods:

- [Cisco Messaging Interface, page 12-2](#)
- [Cisco VG248, page 12-3](#)

Cisco Messaging Interface

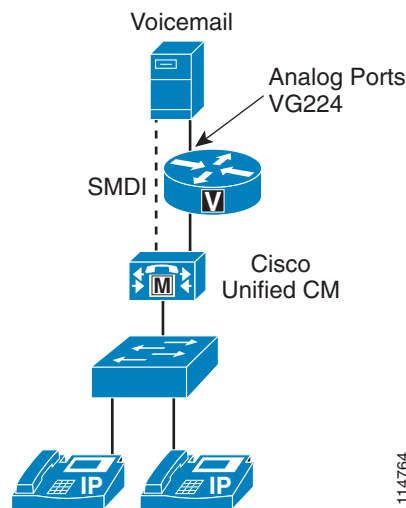
The Cisco Messaging Interface (CMI) is a Unified CM service that should be run only on the publisher server. This service intercepts calls destined for voicemail and generates appropriate SMDI messages, which are then delivered to one of the server's Component Object Model (COM) ports. The CMI service is compatible with any MGCP gateway that supports analog FXS ports or T1 CAS E&M; however, the WS-X6624 and VG224 modules are two of only three gateways that support positive disconnect supervision (see [Positive Disconnect Supervision, page 12-11](#)) and are therefore the only gateways currently recommended for use with the CMI service.


Note

The Cisco WS-X6624 is at End of Sale (EoS).

[Figure 12-1](#) illustrates the use of SMDI through the CMI service in Unified CM.

Figure 12-1 SMDI via Unified CM



Through the CMI, Unified CM supports integration with virtually any voicemail system that can provide SMDI with analog FXS ports, including (but not limited to) the following:

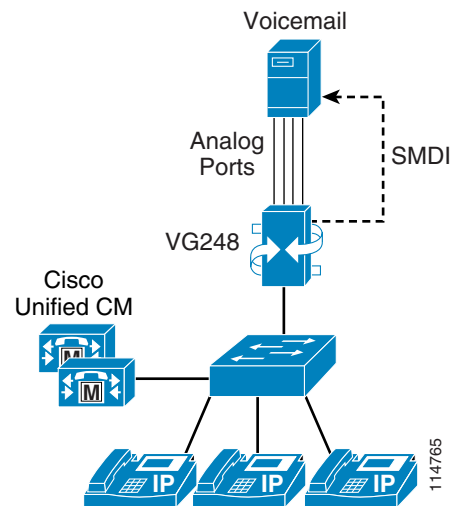
- Octel 100, 200/300, and 250/350
- Intuity Audix
- Siemens PhoneMail
- Centigram/BayPoint (OnePoint and NuPoint Messenger)
- Lyrix ECS
- IBM Message Center

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 and VG224 modules, the VG248 also supports positive disconnect supervision.

Figure 12-2 illustrates the use of SMDI with the VG248.

Figure 12-2 SMDI via the VG248



Voicemail integration through the VG248 provides the following features and advantages:

- Multiple SMDI links per Unified CM
- 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.

- VG224

Use this gateway when there is no physical Catalyst 6500 chassis slot available and when automatic failover of the serial port is not deemed necessary.

- 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 Unified CM 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 Unified CM.

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.

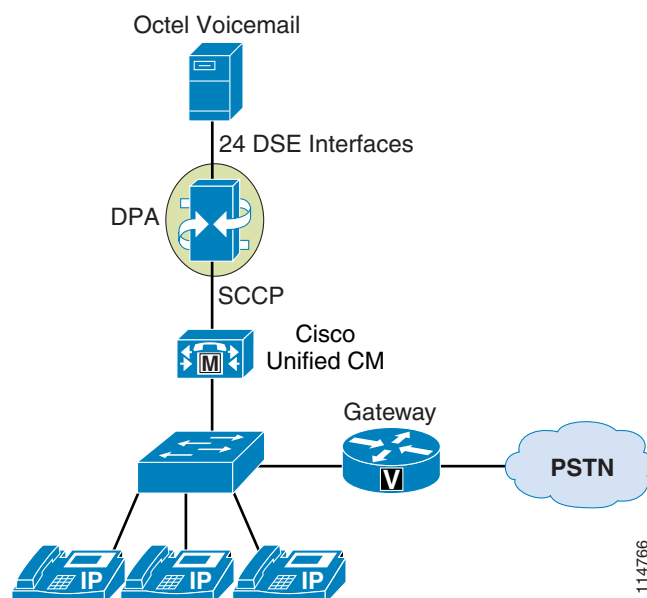


Note

The Cisco DPA 7610 and DPA 7630 are both at End of Sale (EoS).

Figure 12-3 illustrates the DPA integrating an Octel voicemail system with Unified CM.

Figure 12-3 DPA Integrating Octel Voicemail with Unified CM



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.

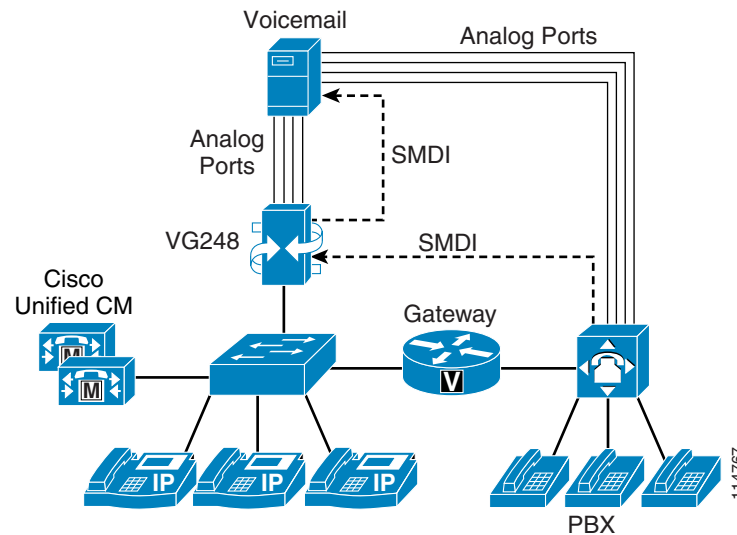


Note

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 12-4](#).)

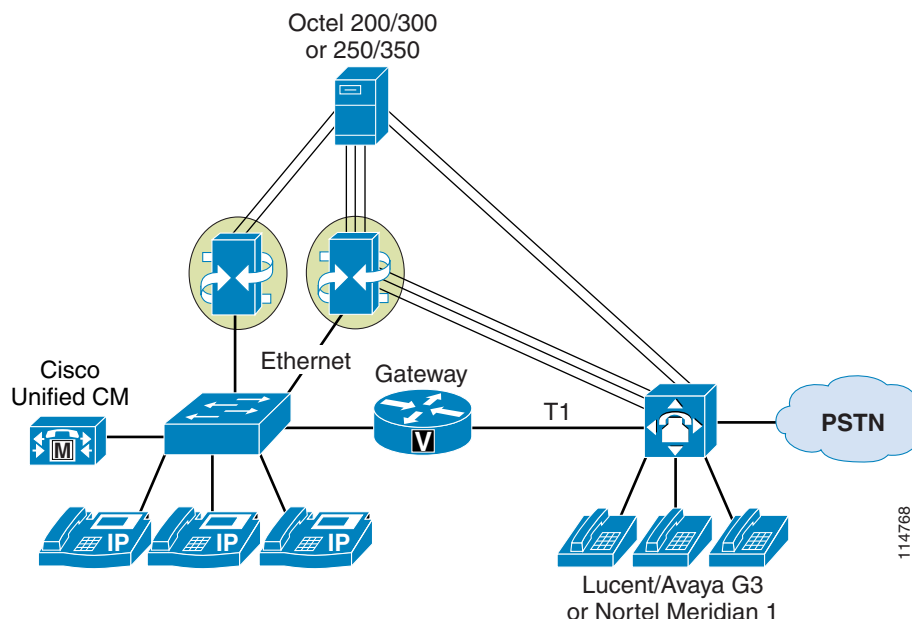
Figure 12-4 Dual Integration via the VG248 and SMDI



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 Unified CM

The Cisco DPA also has the capability to provide dual integration through digital set emulation, as illustrated in [Figure 12-5](#).

Figure 12-5 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:

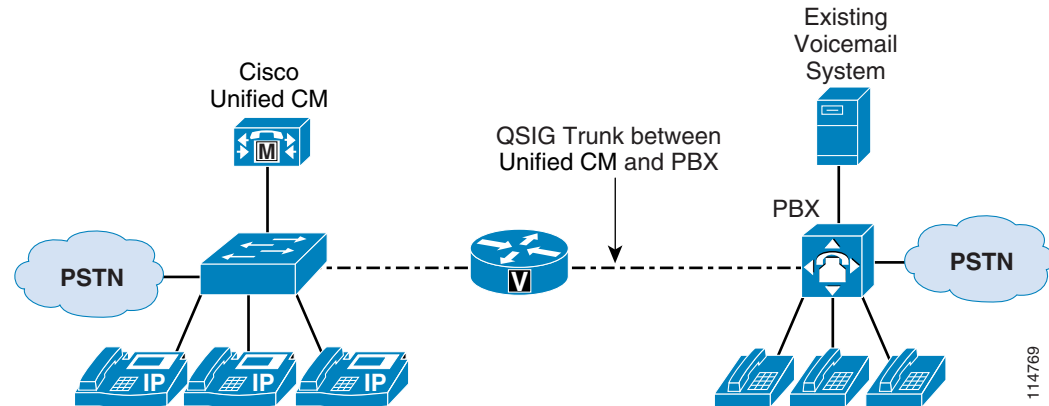
- Uniform dial plan
- Transfer and reconnect sequences
- Connectivity between the PBX and Unified CM

Centralized Voicemail

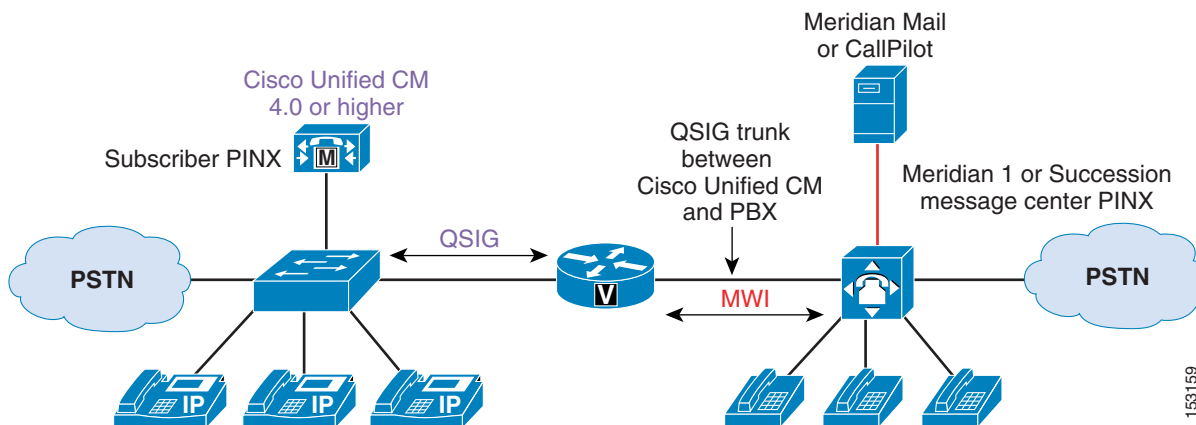
In a centralized voicemail deployment, two or more PBXs share a single voicemail system. The sharing is achieved by integrating the voicemail system to only one PBX and then utilizing an inter-PBX private networking protocol to extend voicemail services to remote subscribers. The networked PBXs look 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 12-6](#).)

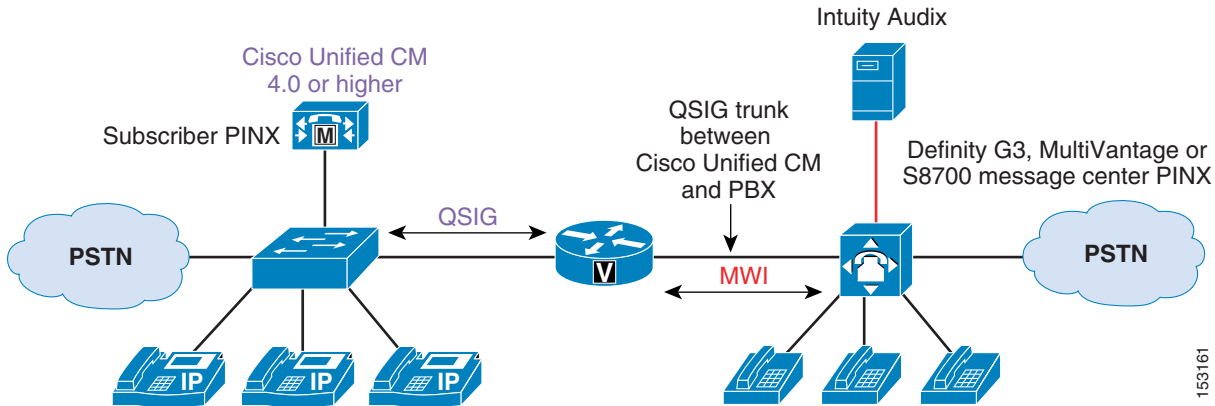
Figure 12-6 Centralized Voicemail with Unified CM and QSIG

If you want to use an existing voicemail system, consider the make and model of that system. If the voicemail system in question is from the same manufacturer as the PBX system, then full voicemail functionality is typically available to Unified CM subscribers. See [Figure 12-7](#) for an example of a Nortel system and [Figure 12-8](#) for an Avaya system.

Figure 12-7 Nortel M1 Centralized Voicemail with Meridian Mail or CallPilot

The system in [Figure 12-7](#) has the following characteristics:

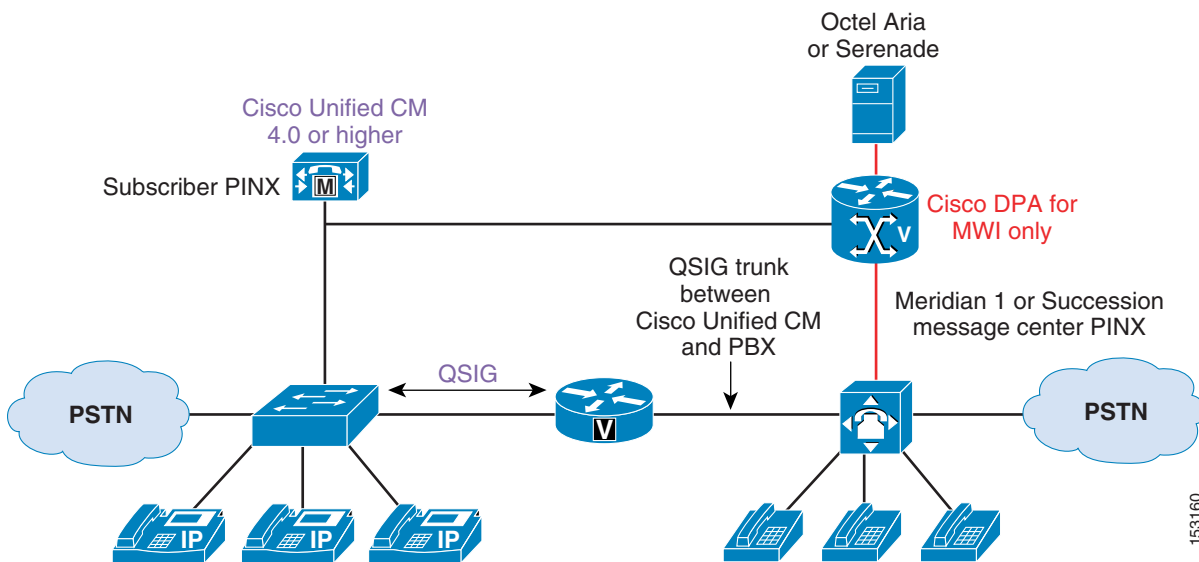
- Voicemail services are available to all subscribers.
- Voicemail is hosted on the message center PINX.
- QSIG MWI works only with Meridian Mail or CallPilot.

Figure 12-8 Avaya G3 Centralized Voicemail with Intuity Audix

The system in [Figure 12-8](#) has the following characteristics:

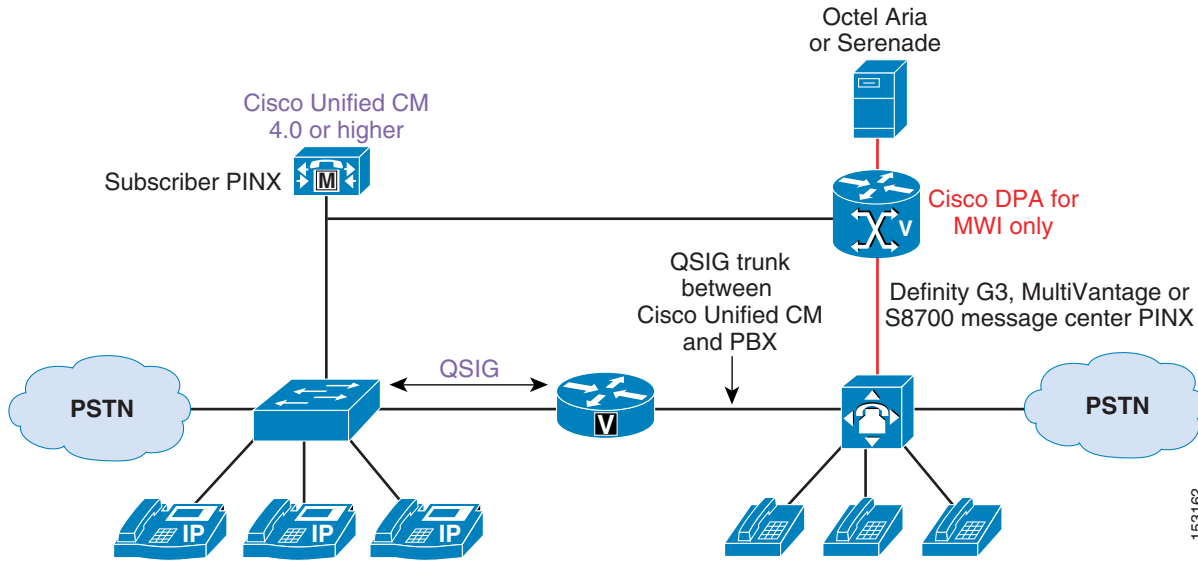
- Voicemail services are available to all subscribers.
- Voicemail is hosted on the message center PINX.
- QSIG MWI works only with Avaya Intuity Audix.

If the voicemail system originates from a different manufacturer than the PBX system, then some functionality might not be passed back through the QSIG trunk to Unified CM. In this instance, the Cisco Digital PBX Adapter (DPA) may be used specifically to deliver MWI directly to Unified CM. See [Figure 12-9](#) for an example of a Nortel system and [Figure 12-10](#) for an Avaya system.

Figure 12-9 Nortel M1 Centralized Voicemail with Octel Aria or Serenade

The system in [Figure 12-9](#) has the following characteristics:

- Voicemail services are available to all subscribers.
- Station-invoked message center functionality is not mapped to QSIG MWI.
- The Cisco DPA is used for MWI to Unified CM only.

Figure 12-10 Avaya G3 Centralized Voicemail with Octel Aria or Serenade

The system in [Figure 12-10](#) has the following characteristics:

- Voicemail services are available to all subscribers.
- Leave Word Calling (LWC) is not mapped to QSIG MWI.
- The Cisco DPA is used for MWI to Unified CM only.

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:

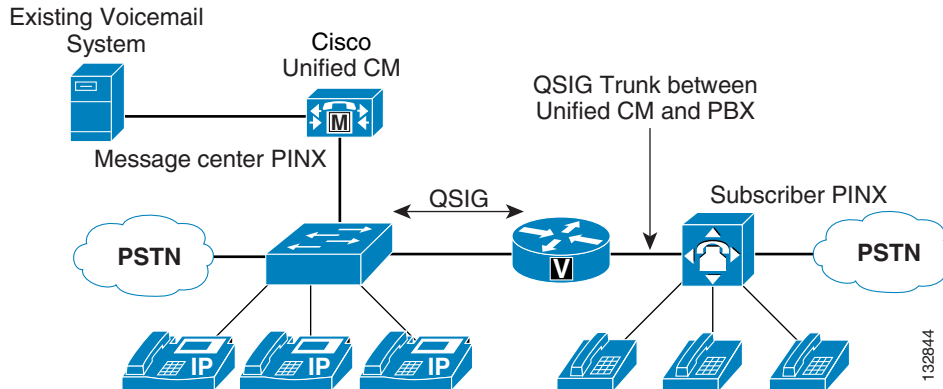
- Message center Private Integrated Services Network Exchange (PINX) — This is the PBX that is "hosting" the voicemail system (the PBX directly connected to the voicemail system).
- Subscriber PINX — This is the PBX that is "remote" from the voicemail system (not directly connected to the voicemail system).

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 Unified CM and have Unified CM act as the message center PINX, with the PBX acting as the subscriber PINX. (See [Figure 12-11](#).)

Figure 12-11 Centralized Voicemail with Unified CM Acting as Message Center PINX**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 Unified CM 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 Unified CM and the PBX. Application Notes for Unified CM have already been written for the leading PBXs, and they cover the scenario of centralized voicemail with Unified CM 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.
- Unified CM 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 Unified CM 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 Unified CM (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 or the Digital PBX Adapter (DPA), therefore they are the recommended solutions.

The method of integrating Unified CM 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 or the VG224 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.



CHAPTER 13

Cisco Unity

Last revised on: September 27, 2007

This chapter focuses on the design aspects of integrating Cisco Unity and Cisco Unity Connection with Cisco Unified Communications Manager (Unified CM). This chapter focuses on Cisco Unified CM Release 5.1, Cisco Unity 4.2, and Cisco Unity Connection 1.2. For information on earlier releases of Cisco Unity, Unity Connection, and Unified CM, refer to the appropriate online documentation available at

<http://www.cisco.com>

The design topics covered in this chapter apply to both voicemail and unified messaging configurations. 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. Cisco Unity Connection has no dependencies on an external message store.

For additional design information about Cisco Unity and Unity Connection, 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 and Unity Connection design topics:

- [Messaging Deployment Models, page 13-2](#)
- [Messaging System Infrastructure Components, page 13-4](#)
- [Port Groups \(Separate Integrations\), page 13-5](#)
- [Managing Bandwidth, page 13-6](#)
- [Cisco Unity and Unity Connection Native Transcoding Operation, page 13-8](#)
- [Configuring a Unified CM SIP Trunk for Voicemail Integration, page 13-9](#)
- [Voice Port Integration with a Unified CM Cluster, page 13-10](#)
- [Voice Port Integration with Dedicated Unified CM Backup Servers, page 13-12](#)
- [Centralized Messaging and Centralized Call Processing, page 13-13](#)
- [Distributed Messaging with Centralized Call Processing, page 13-14](#)
- [Combined Messaging Deployment Models, page 13-16](#)
- [Centralized Messaging with Clustering Over the WAN, page 13-17](#)
- [Distributed Messaging with Clustering Over the WAN, page 13-19](#)
- [Cisco Unity Messaging Failover, page 13-20](#)

- [Cisco Unity Failover and Clustering Over the WAN, page 13-20](#)
- [Centralized Messaging with Multiple Unified CM Servers, page 13-21](#)
- [Considerations for Distributed PIMG Deployments, page 13-22](#)

Cisco Unified CM Release 4.0 introduced *line groups*, *hunt lists*, and *hunt pilots*, which can impact the operation of existing voice ports. Before upgrading from Unified CM versions older than Release 4.0 to Unified CM Release 5.0 or later, refer to the appropriate *Cisco Unified Communications Manager Administration Guide* for your 5.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 Unified CM Release 5.0 and later.

Cisco Unified CM 5.0 also adds new functionality for SIP trunks. SIP trunks now directly support integration to Cisco Unity and Unity Connection without the need of a SIP proxy server.

Messaging Deployment Models

Cisco Unity supports three primary messaging deployment models:

- [Single-Site Messaging, page 13-2](#)
- Multisite WAN deployment with [Centralized Messaging, page 13-3](#)
- Multisite WAN deployment with [Distributed Messaging, page 13-3](#)

Cisco Unity Connection supports single-site messaging and centralized messaging with a single messaging server.

Deployments involving both Cisco Unified CM and Cisco Unity or Unity Connection use one call processing model for Unified CM 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 13-4](#).) Cisco Unity Connection does not currently support failover. Because Cisco Unity Connection uses only a single server without any networking between servers, it supports only the single-site and centralized messaging models.

All messaging deployment models support both voicemail and unified messaging installations. Cisco Unity Connection supports voicemail only.

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.

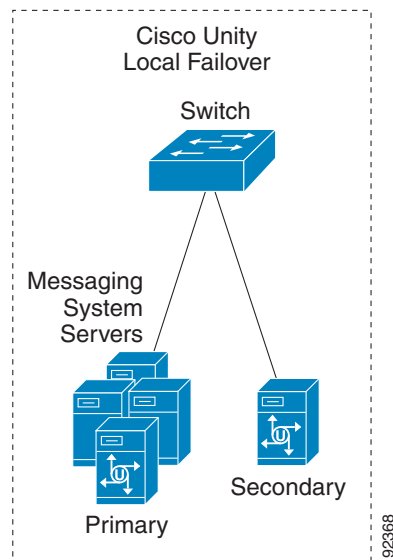
The distributed messaging model has the same design criteria as centralized messaging with regard to local and remote GUI clients, TRaP, and message downloads.

Because Cisco Unity Connection supports only a single server, distribution of messaging servers is not possible with it. Therefore, Cisco Unity Connection does not support the distributed messaging model.

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. This design is only for Cisco Unity; Cisco Unity Connection does not currently support failover.

Figure 13-1 Local Failover of Cisco Unity Messaging



Cisco Unity and Unified CM support the following combinations of messaging and call processing deployment models. Cisco Unity Connection supports all combinations except the distributed messaging 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 and Cisco Unity Connection design guides available at <http://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.) Cisco Unity supports both Microsoft Exchange and IBM Lotus Notes as message stores. Each of these message store types has different infrastructure components (refer to the appropriate deployment type in the *Cisco Unity Design Guide*, available at <http://www.cisco.com>).

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

Figure 13-2 Cisco Unity Messaging System Infrastructure Components (Exchange Specific)

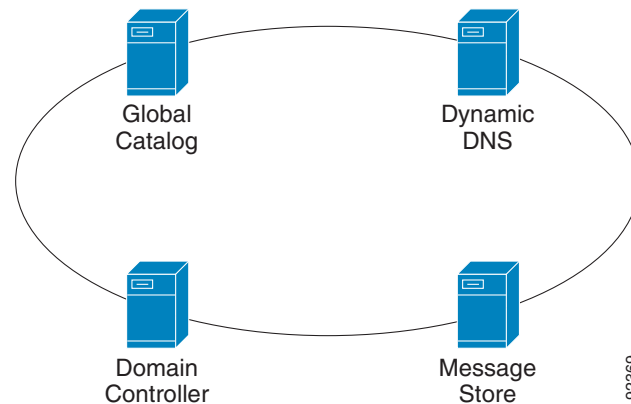


Figure 13-2 is a logical representation of the message system infrastructure components. Some of these components can be located on the same server. In the case of Domino Lotus Notes, the message store and directory (Names.nsf) are located on the same server. Microsoft Windows, the global catalog server, 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 deploying Cisco Unity Connection, it is possible to locate the automatic speech recognition (ASR) service on a separate server. In this deployment environment, the Cisco Unity Connection server and the ASR server should reside in the same location. Cisco Unity Connection does not have the same infrastructure requirements or dependencies as Cisco Unity.

Port Groups (Separate Integrations)

Unity Connection introduces the concept of *port groups*, also known as *separate integrations* in Cisco Unity 4.2. In earlier releases of Unity, under the Cisco Unity Telephony Integration Manager (UTIM) you were able to configure *multiple clusters* of the same type of integration. It is important to understand how port groups and separate integrations are different than multiple clusters under a single integration. With the single-integration multiple-cluster approach, the Unity database associated users to an integration, not a cluster. Because there was only a single integration, all subscribers were associated with that single integration regardless of which cluster they were on. For message waiting indication (MWI), Unity then had to broadcast MWI updates out all ports under the single integration, including to clusters where the subscriber was not present. With the new addition of separate integrations in Unity 4.2

and port groups in Unity Connection, subscribers are associated with the specific integration to which they belong. MWI lights no longer have to be broadcast out all ports but are sent only to the specific ports that are associated with the particular subscriber.

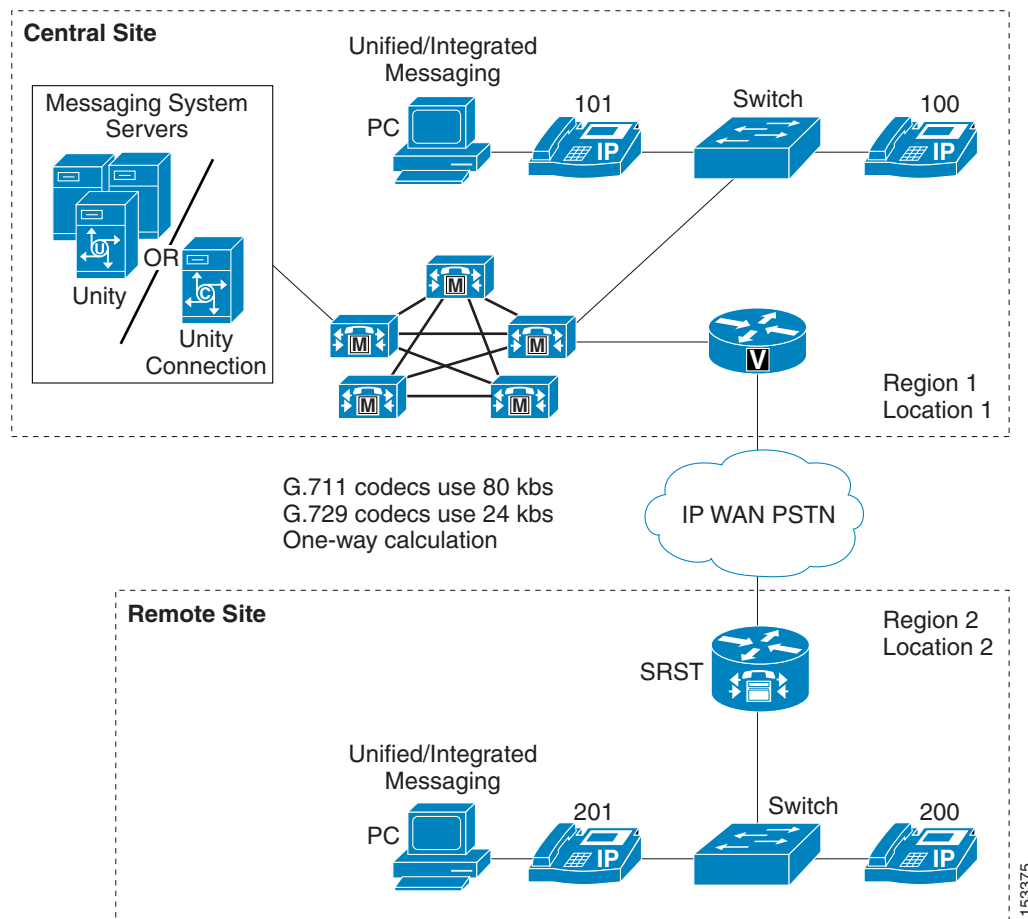
Dual integrations occur when you integrate more than one type of system to a Unity or Unity Connection server, such as when integrating a legacy phone system and a Unified CM phone system or when integrating a Unified CM SIP system and a Unified CM SCCP system. Multiple integrations occur when you integrate Cisco Unity 4.1 or earlier versions to more than one Unified CM cluster via the same type of integration, such as when integrating via SCCP to two or more Unified CM clusters.

With Unity 4.2, separate integrations are configured via the UTIM, while for Unity Connection port groups are configured under the System Administrator.

Managing Bandwidth

Unified CM provides a variety of features for managing bandwidth. Through the use of regions, locations, and even gatekeepers, Unified CM 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 and Unity Connection rely on Unified CM to manage bandwidth and to route calls. If you deploy Cisco Unity or Unity Connection 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 or Unity Connection server is servicing either distributed clients (distributed messaging (Unity only) or distributed call processing) or when Unified CM is remotely located (distributed messaging (Unity only) or centralized call processing). Unified CM provides regions and locations for call admission control.

[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 Unified CM 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 or Unity Connection 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 and Unity Connection Native Transcoding Operation

By default, the Cisco Unity or Unity Connection server performs transcoding automatically. Currently, Unified CM and Cisco Unity support only G.729 and G.711 for the Skinny Client Control Protocol (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. For SCCP integrations only, the native transcoding must be disabled (turned off) on the Cisco Unity server via a registry setting in order for Unified CM to assign hardware transcoders to voice port calls. The registry setting is called "Audio - Enable G.729a codec support," and the tool for setting it is the *Advanced Settings Tool* available at <http://www.CiscoUnityTools.com>. (To disable native transcoding on Cisco Unity Connection, the configuration of the messaging codec is done differently, as described at the end of this section.)

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 Unified CM 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, Unified CM will assign an external transcoding resource for the call. For detailed information on how to configure transcoding resources on Unified CM, 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 System Administration Guide* on <http://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.

For information on disabling native transcoding on Unity when integrated via SIP, refer to the Unity product documentation online at <http://www.cisco.com>.

To change how Cisco Unity Connection advertises which codecs it supports, the configuration is done differently than for Cisco Unity. Instead of using the Cisco Unity Tools Depot, the configuration changes are made from the Cisco Unity Connection Administration page. Under the Telephony Integrations heading, select the Port Groups link. On the Port Groups page, you can change the configuration to advertise either G.711 only, G.729 only, or both, but "prefer" either G.711 or G.729. In the preferred setting, Cisco Unity Connection will advertise that it supports both protocols (native transcoding) but will prefer to use the one specified over the other. With native transcoding disabled, Unified CM will provide a transcoder resource when needed instead of using the CPU on the Unity or Unity Connection server.

Configuring a Unified CM SIP Trunk for Voicemail Integration

Cisco Unified CM 5.0 introduces a new feature for SIP support for line-side devices as well as enhancements to trunk-side SIP. These features enable Unified CM to integrate directly to Cisco Unity and Unity Connection via SIP without the need for a SIP proxy server.

However, while Unified CM allows direct integration to voicemail via SIP trunks, it does not currently have the same Voicemail Port Wizard that Skinny Client Control Protocol (SCCP) ports do. Some manual configuration is required. The following steps are required to configure a SIP trunk for a basic integration scenario.

Step 1 Create a SIP Profile in Unified CM Administration by selecting **Device > Device Setting > SIP Profile**.

There is nothing specific to voicemail integrations here, but for easy and consistency of administration functions you will want to create a SIP profile specific for your voicemail integration(s) and name it accordingly.

Step 2 Create a SIP Trunk Security Profile in Unified CM Administration by selecting **System > Security Profile > SIP Trunk Security Profile**.

The **Incoming Port #** under the SIP Trunk Security Profile is specific to voicemail, and it is the number of the SIP port number that the Cisco Unity or Unity Connection server uses to connect to the Unified CM cluster. Additionally, check (enable) **Accept Unsolicited Notifications** so that Cisco Unity and Unity Connection can notify Unified CM of Message Waiting Indicator (MWI) events.



Note MWI functions are handled via Unsolicited Notifies. There is no need to configure the MWI DN's that are required for SCCP integrations.

Step 3 Create a SIP Trunk in Unified CM Administration by selecting **Device > Trunk [Add New]**. Under the trunk creation page, assign the previously configured SIP Profile and SIP Trunk Security Profile in the appropriate fields. Additionally, configure the destination address of the Unity or Unity Connection server and check (enable) **Unattended Port**. Also check (enable) **Redirecting Number IE Deliver - Outbound**.



Note One exception to the Unattended Port occurs when multiple Cisco Unity servers are attached to a single Unified CM cluster (for example, if more than 72 voicemail ports are deployed, then at least two Cisco Unity servers would be required) and live reply or cross-server logon are configured. In this scenario, you would want to leave Unattended Port unchecked (disabled) to allow calls transferred from voicemail ports on one server to be terminated on the other server. This exception currently applies to Unity only.

Step 4 Create a route pattern and specify the voicemail SIP trunk as the destination.

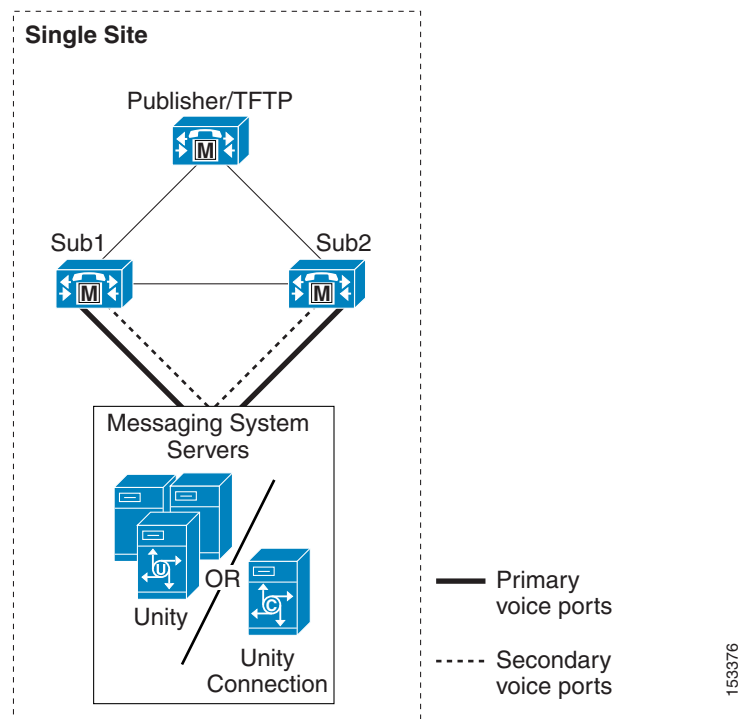
Step 5 Create a Voice Mail Pilot number that is the same number as the route pattern configured in [Step 4](#).

Step 6 Create a Voice Mail Profile using the appropriate Voice Mail Pilot number.

Voice Port Integration with a Unified CM Cluster

When deploying Cisco Unity in a single-site messaging environment, integration with the Unified CM cluster occurs through the SCCP voice ports or SIP ports. Design considerations must include proper deployment of the voice ports among the Unified CM subscribers so that, in the event of a subscriber failure (Unified CM 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 Unified CM Cluster (No Dedicated Backup Servers)



The Unified CM 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 up to 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.

This 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. The second group is configured so that Sub2 is the primary server and Sub1 is the backup.

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 Unity Telephony Integration Manager (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 Unified CM Administration. 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 in this example.

For additional design information on the ratio of inbound to outbound voicemail ports (for MWI, message notification, and TRaP), refer to the *Cisco Unity System Administration Guide* available at <http://www.cisco.com>.

**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 Unified CM Administration.

In Unified CM 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 the unique device prefix. (See [Table 13-1](#).) When the ports register with Unified CM, half will be registered with subscriber server Sub1, and the other half will be registered with Sub2, as shown in [Table 13-1](#).

Table 13-1 Voicemail Port Configuration in Unified CM Administration

Device Name	Description	Device Pool	SCCP Security Profile	Status	IP Address
CiscoUM1-VI1	Unity1	Default	Standard Profile	Registered with sub1	1.1.2.9
CiscoUM1-VI2	Unity1	Default	Standard Profile	Registered with sub1	1.1.2.9
CiscoUM1-VI3	Unity1	Default	Standard Profile	Registered with sub1	1.1.2.9
CiscoUM1-VI4	Unity1	Default	Standard Profile	Registered with sub1	1.1.2.9
CiscoUM2-VI1	Unity1	Default	Standard Profile	Registered with sub2	1.1.2.9
CiscoUM2-VI2	Unity1	Default	Standard Profile	Registered with sub2	1.1.2.9
CiscoUM2-VI3	Unity1	Default	Standard Profile	Registered with sub2	1.1.2.9
CiscoUM2-VI4	Unity1	Default	Standard Profile	Registered with sub2	1.1.2.9

**Note**

The naming convention used for the voicemail ports in Unified CM Administration must match the device name prefix used in Cisco UTIM, otherwise the ports will fail to register.

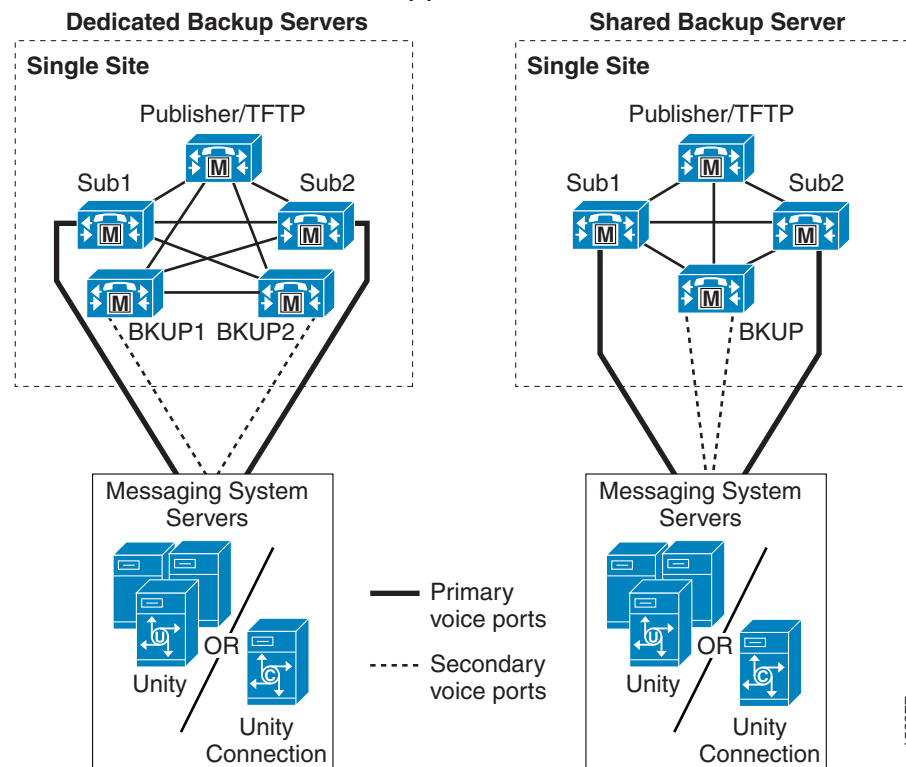
Cisco Unified CM 4.0 introduced numerous changes to the configuration of SCCP voicemail ports with regard to how they hunt and forward calls. For information on how these changes impact the voicemail port operation and configuration, refer to the Cisco Unified CM 4.0 documentation available at

<http://www.cisco.com>

Voice Port Integration with Dedicated Unified CM Backup Servers

This Unified CM 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 13-5](#).) 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 13-5 Cisco Unity Server(s) Integrated with a Single Unified CM 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 Unified CM clustered with a shared backup server, both of the secondary ports for the subscribers servers are configured to use the single backup server.

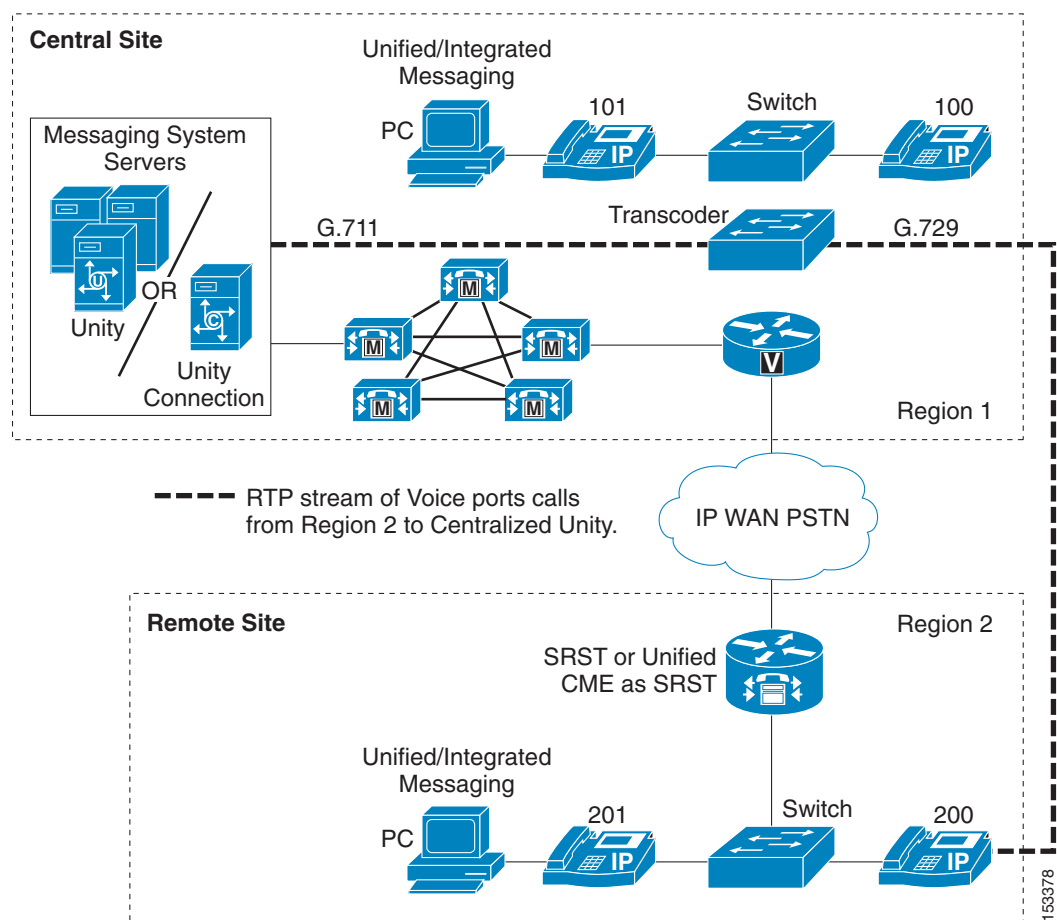
The voice port names (device name prefix) must be unique for each Cisco UTIM group and must be the same as the device names used on the Unified CM server.

To configure the voicemail ports on Cisco Unity, use the UTIM tool. For Cisco Unity Connection, use the Telephony Integration section of the Unity Connection Administration console. For details, refer to the Cisco Unity or Cisco Unity Connection administration guides available at <http://www.cisco.com>.

Centralized Messaging and Centralized Call Processing

In centralized messaging, the Cisco Unity server is co-located with the Unified CM cluster. With centralized call processing, subscribers may be located either remotely and/or locally to the cluster and messaging server(s). (See [Figure 13-6](#).) 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 Unified CM for call admission control. (See [Managing Bandwidth](#), page 13-6.) When making inter-region calls to IP phones or MGCP gateways, IP phones automatically select the inter-region codec that has been configured. Native transcoding should be disabled so that the Cisco Unity voice ports use Unified CM transcoding resources for calls that transverse the WAN (inter-region).

Figure 13-6 Centralized Messaging with Centralized Call Processing



In [Figure 13-6](#), 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 13-6](#) 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. Unified CM transcoding resources must be located at the same site as the voicemail system.

Hairpinning

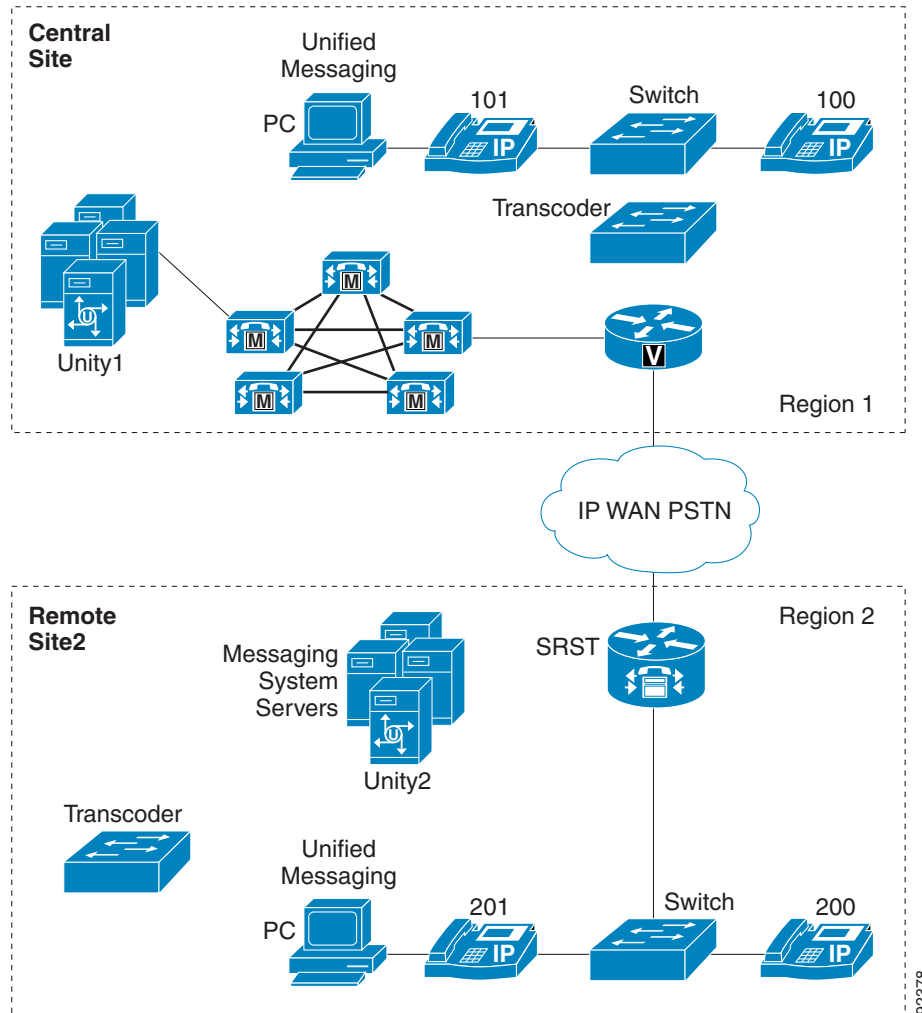
Another issue to consider is hairpinning (or tromboning) of voice calls over multiple Cisco Unity voice ports. Hairpinning is not a concern in environments that use only SCCP TSP voice ports, but it is a concern with dual-integration environments, where hairpinning can occur between the voice ports of the legacy system and the SCCP TSP voice ports.

For more information about dual integration, refer to the *Cisco Unity System 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 13-7](#) illustrates the distributed messaging model with centralized call processing. As with other multisite 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 13-7 Distributed Messaging with Centralized Call Processing

For the configuration in Figure 13-7, 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 Unified CM server. In addition, to associate telephony users with a specific group of voicemail ports, you must configure Unified CM voicemail profiles. For details on configuring calling search spaces, device pools, and voicemail profiles, refer to the applicable version of the *Cisco Unified Communications Manager Administration Guide*, available at

<http://www.cisco.com>

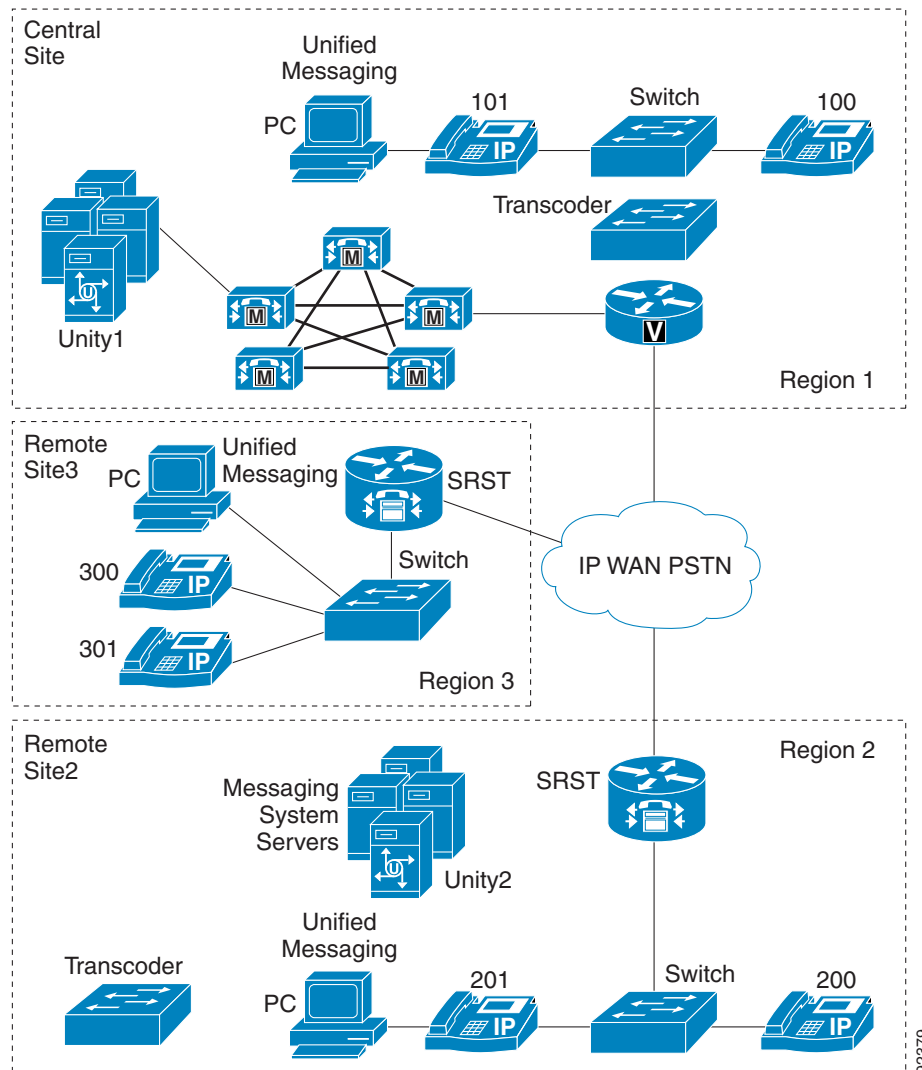
The messaging systems are "networked" together to present a single messaging system to both inside and outside users. For information about Cisco Unity Networking for the distributed Unity servers, Refer to the *Networking in Cisco Unity Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps2237/products_feature_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 13-8](#) shows a user environment in which both centralized and distributed messaging are employed simultaneously.

Figure 13-8 Combined Deployment Models



[Figure 13-8](#) 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 13-8](#), 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

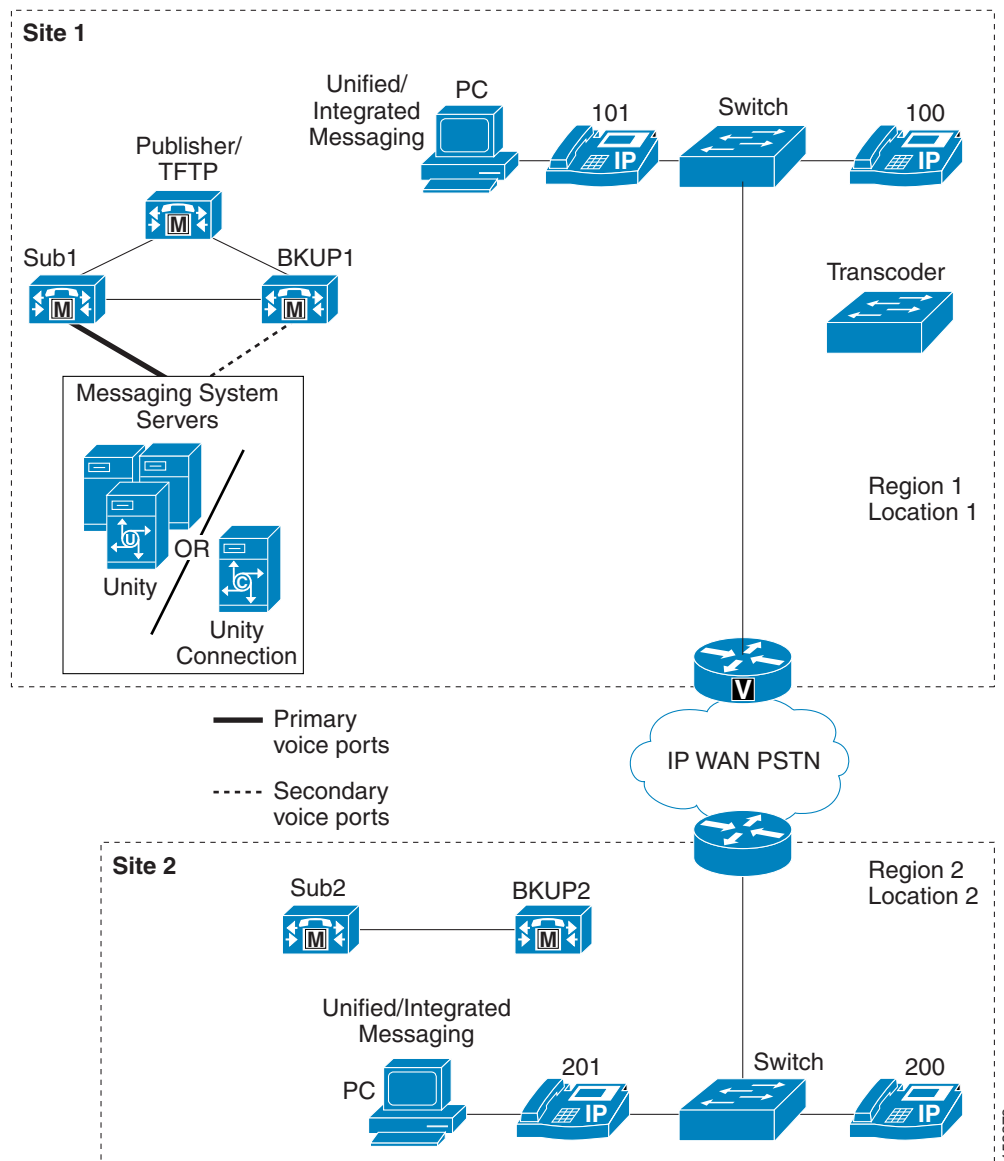
in this chapter. Transcoding resources are located locally to each messaging system site, and they support clients who access messaging services from a remote site (relative to the messaging system), as in the case of a Site2 user leaving a message for a Central Site user.

Centralized Messaging with Clustering Over the WAN

This section addresses Cisco Unity design issues for deploying centralized messaging with Unified CM 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 13-9](#).)

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 2-17.

Figure 13-9 Cisco Unity Centralized Messaging and Clustering Over the WAN with Local Failover

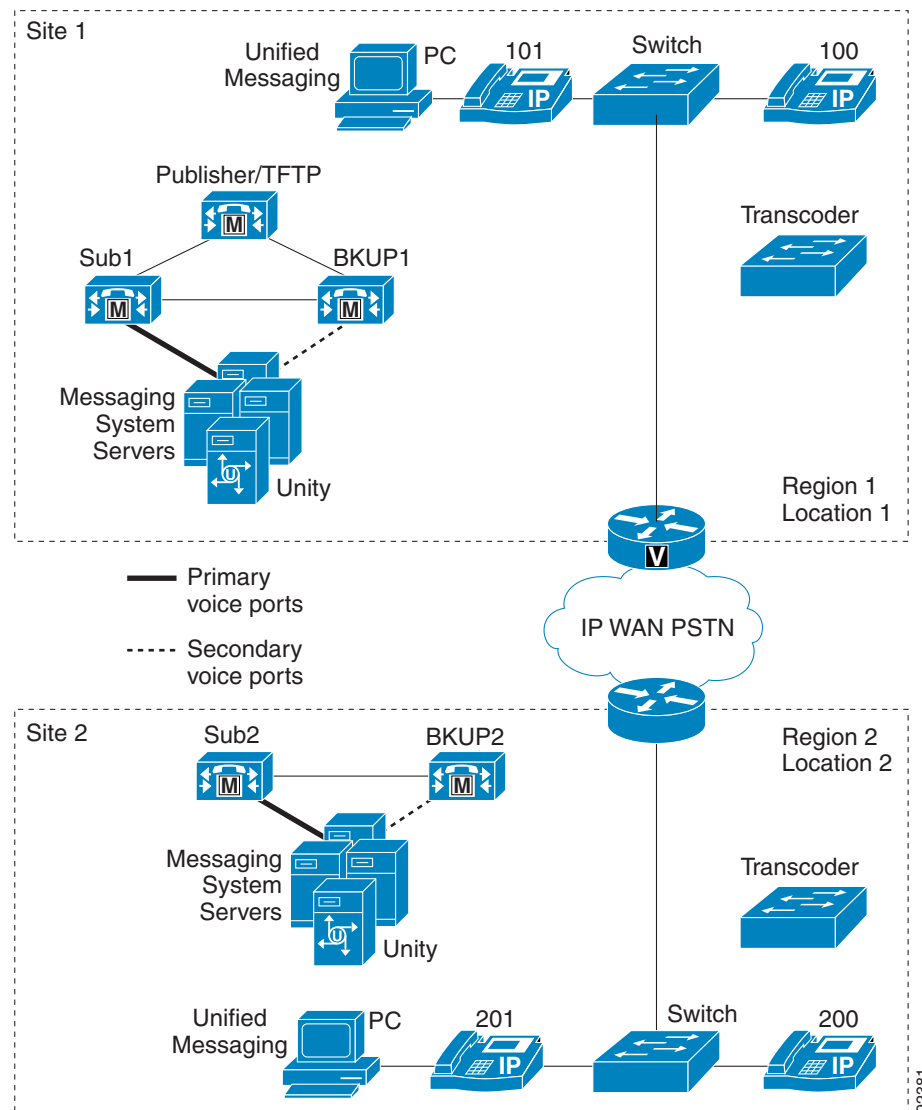
The minimum bandwidth required between cluster sites is a T1 line (1,544 MHz). This amount of bandwidth can support signaling and database traffic for up to 10,000 busy hour call attempts (BHCA), but it does not include the required media bandwidth.

Clustering over the WAN with Cisco Unified CM Release 3.3(3) and earlier supports up to four sites per cluster, but Cisco Unified CM 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 13-9). 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 Unified CM subscriber server(s), similar to the centralized messaging model. For information about configuring the voice ports, see [Voice Port Integration with a Unified CM Cluster](#), page 13-10, and [Voice Port Integration with Dedicated Unified CM Backup Servers](#), page 13-12.

Figure 13-10 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 13-16.) 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 Unified CM to provide call admission control. (See [Managing Bandwidth](#), page 13-6.)

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 <http://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. Cisco Unity Connection currently does not support failover.

For information on configuring Cisco Unity failover, refer to the *Cisco Unity Failover Configuration and Administration Guide*, available at

<http://www.cisco.com>

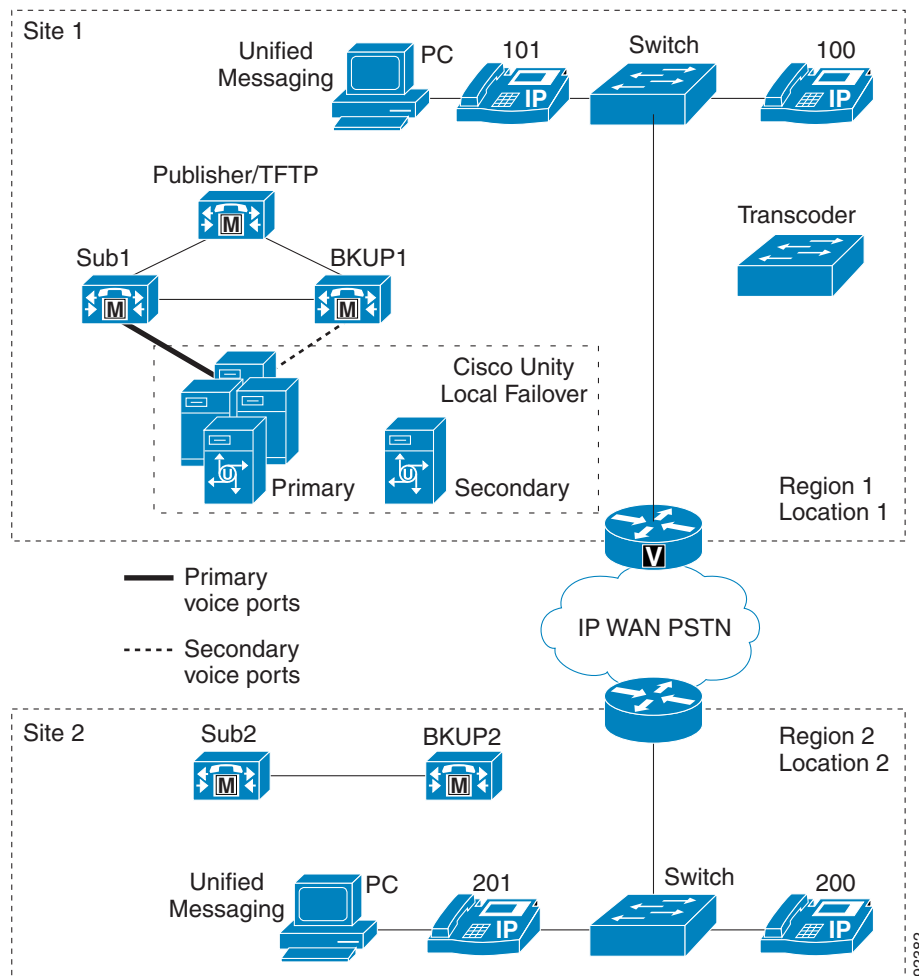
Cisco Unity Failover and Clustering Over the WAN

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 13-17, and [Distributed Messaging with Clustering Over the WAN](#), page 13-19. The voice ports from the primary Cisco Unity server should not cross the WAN during normal operation.

[Figure 13-11](#) 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 Unified CM.

**Note**

Unity Connection does not currently support failover.

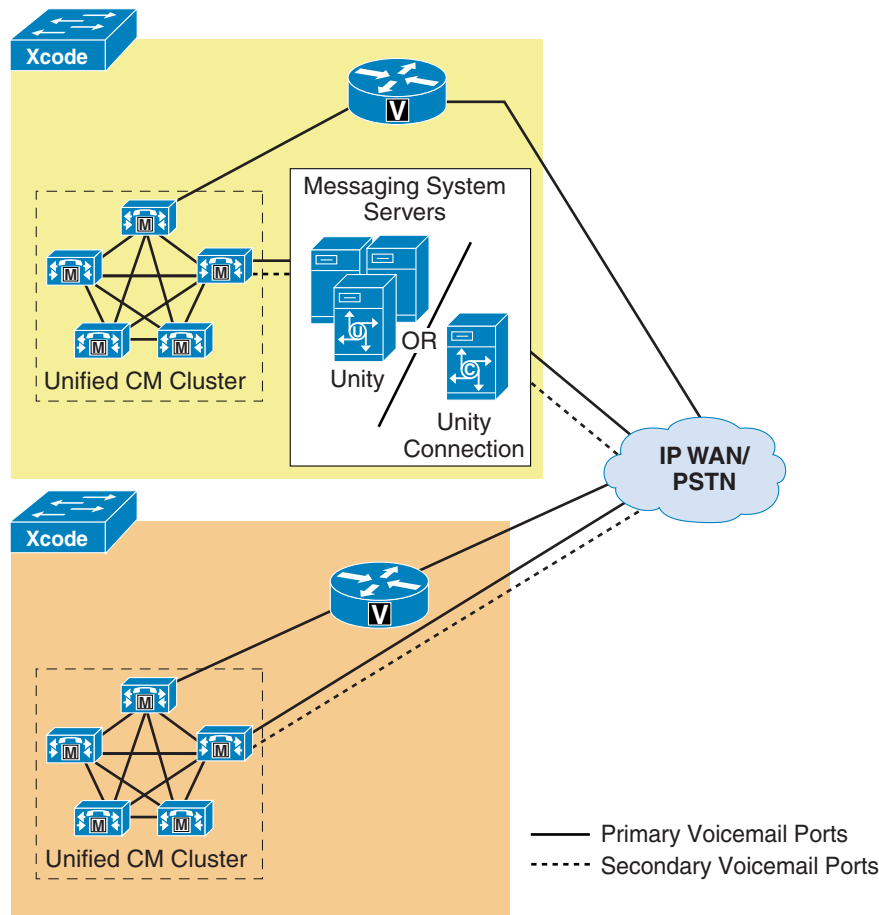
Figure 13-11 Cisco Unity Local Failover and Clustering Over the WAN

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 Unified CM Servers

Cisco Unity and Unity Connection support port groups, which allow subscribers on either of the Unity product servers to be associated with a port group and ultimately an integration. When a subscriber is associated with a particular port group, the MWI function for the user occurs only on the MWI ports belonging to that port group. This feature functions the same way for both SCCP and SIP voicemail ports. Cisco Unity supports up to seven port groups, while Cisco Unity Connection supports up to nine port groups. For details, refer to the appropriate Cisco Unity or Cisco Unity Connection administration guides available at <http://www.cisco.com>.

Figure 13-12 Integrating Cisco Unity or Unity Connection with Multiple Unified CM Clusters

153158

Messaging clients at both Cluster 1 and Cluster 2 sites use the Cisco Unity or Unity Connection messaging infrastructure physically located at Cluster 1. (See [Figure 13-12](#).)

Considerations for Distributed PIMG Deployments

Cisco Unity 4.2 introduced support for Intel NetStructure PBX-IP Media Gateways (PIMGs) distributed at branch locations. From a design perspective, this new feature implies that the SIP signaling between the Cisco Unity server(s) and the PIMGs may now traverse a WAN link. The SIP trunk connects directly between the Cisco Unity server and the PIMG appliance, and there is no locations-based call admission control or gatekeeper call admission control available on Cisco Unity, as is available on Unified CM. Currently the PIMG appliance does not support Quality of Service (QoS) marking for either signaling or media traffic. Therefore, when deploying PIMGs remotely from the Cisco Unity server(s), make sure enough bandwidth exists and that latency/delay, jitter, and dropped packets for best-effort unmarked traffic do not adversely impact voice quality for these calls.



CHAPTER 14

Cisco Unified MeetingPlace Integration

Last revised on: September 27, 2007

This chapter covers system-level design and implementation of Cisco Unified MeetingPlace 5.4 in a Cisco Unified Communications Manager 5.x environment. The following aspects of design and configuration of Cisco Unified MeetingPlace not related to system design with Cisco Unified Communications Manager (Unified CM) are *not* covered in this chapter:

- Hardware requirements
- Software component configuration
- Sizing of voice, web, and video resources
- Audio, web, and video conferencing detailed explanations and configuration outside of system-level design
- Much of the TDM capabilities of Cisco Unified MeetingPlace 5.4

For information on these topics, refer to the Cisco Unified MeetingPlace product documentation available at

http://www.cisco.com/en/US/products/sw/ps5664/ps5669/tsd_products_support_series_home.html

Cisco Unified MeetingPlace 5.4 introduces new and enhanced video capabilities, including:

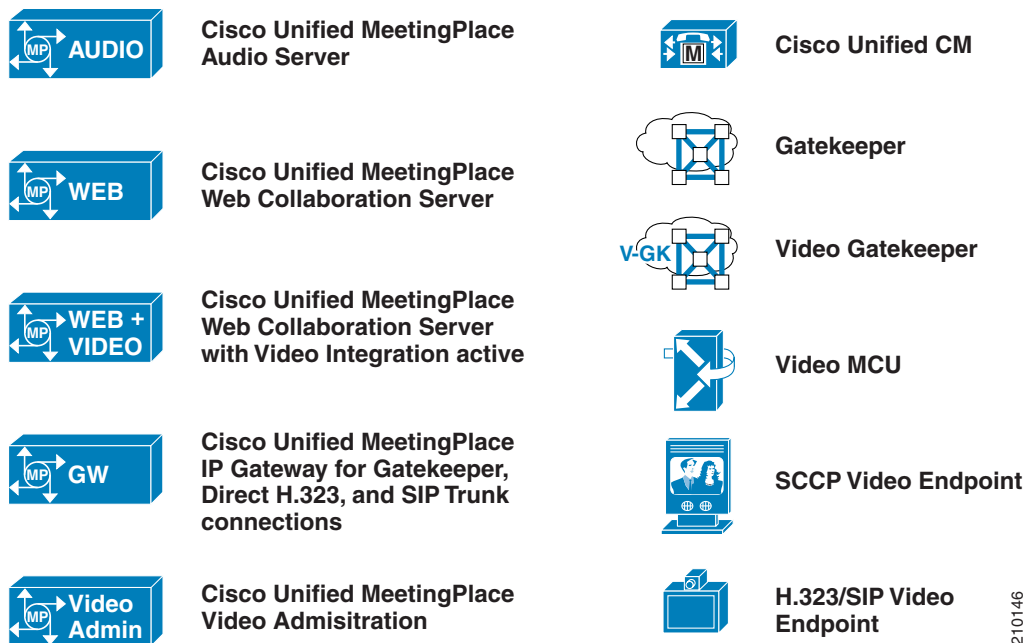
- Video scheduling integrated into meeting scheduling, allowing reservation of video ports and terminals
- Ability to cascade MCU resources, allowing larger video meetings
- Ability to see and control individual video speaker status on participant list

Cisco Unified MeetingPlace Video in many ways acts as a separate entity with full integration into Cisco Unified MeetingPlace 5.4. The first sections of this chapter cover Cisco Unified MeetingPlace implemented as a base deployment with Audio and Web Collaboration. The final section covers the details of adding video to Cisco Unified MeetingPlace 5.4 base deployments. All guidance given in the first sections also applies to deployments with video.

Cisco Unified MeetingPlace Components

Figure 14-1 lists the components of a Cisco Unified MeetingPlace deployment. The icons and symbols in Figure 14-1 are used in diagrams throughout this chapter.

Figure 14-1 Cisco Unified MeetingPlace Components



210146

All components on the left (containing the MP symbol) are specific to Cisco Unified MeetingPlace, but this is not an all-inclusive list. There are other Cisco Unified MeetingPlace components not shown in the diagrams in this chapter. All Cisco Unified MeetingPlace components, with the exception of the Audio Server, are software components that run on Microsoft Windows 2000 or 2003 servers. The Audio Server is a hardware component using a specialized OS. Specifics on server and OS requirements are listed in the product documentation available at

http://www.cisco.com/en/US/products/sw/ps5664/ps5669/tsd_products_support_series_home.html

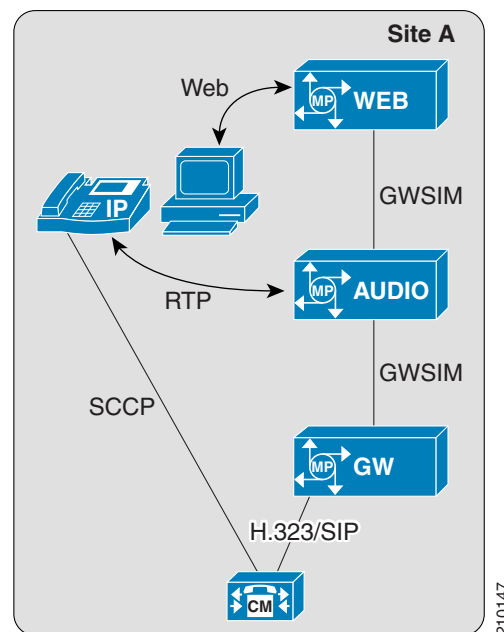
Deployment Models

This section discusses design considerations and recommendations for adding Cisco Unified MeetingPlace to each of the basic Cisco Unified Communications Manager (Unified CM) deployment models. Many variations are possible in the deployments, but the models presented here cover the base deployments without examining every variation.

Single Site

The single-site deployment is the base deployment, with all server components and users located at a single site interconnected by a single LAN (see [Figure 14-2](#)). In this model, Cisco Unified MeetingPlace is co-located with the Unified CM cluster and integrated via H.323 or SIP, as described in sections on [H.323 and SIP Direct Integration](#), page 14-10, and [Gatekeeper Integration](#), page 14-11.

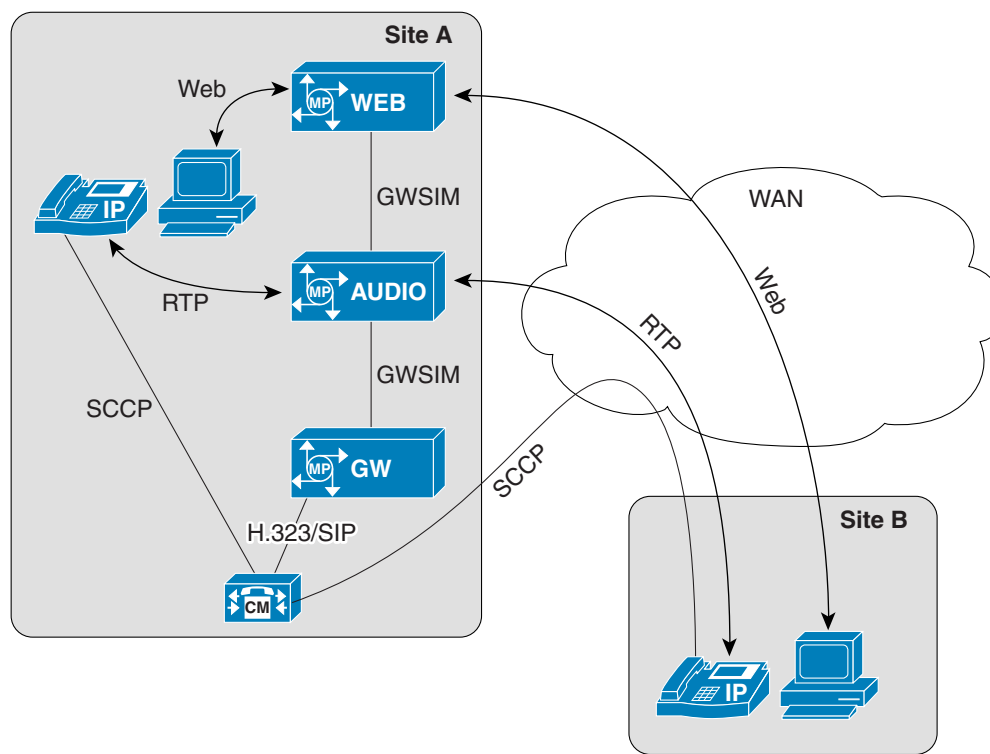
Figure 14-2 **Single Site**



Multisite WAN with Centralized Call Processing

The multisite WAN deployment with centralized call processing builds upon the single-site model by incorporating remote sites (see [Figure 14-3](#)). All Cisco Unified MeetingPlace components are assumed to be centrally located for this model, although distribution of some components is an option. Distributed components are covered in the next model in conjunction with distributed call processing.

Figure 14-3 Multisite Deployment with Centralized Call Processing



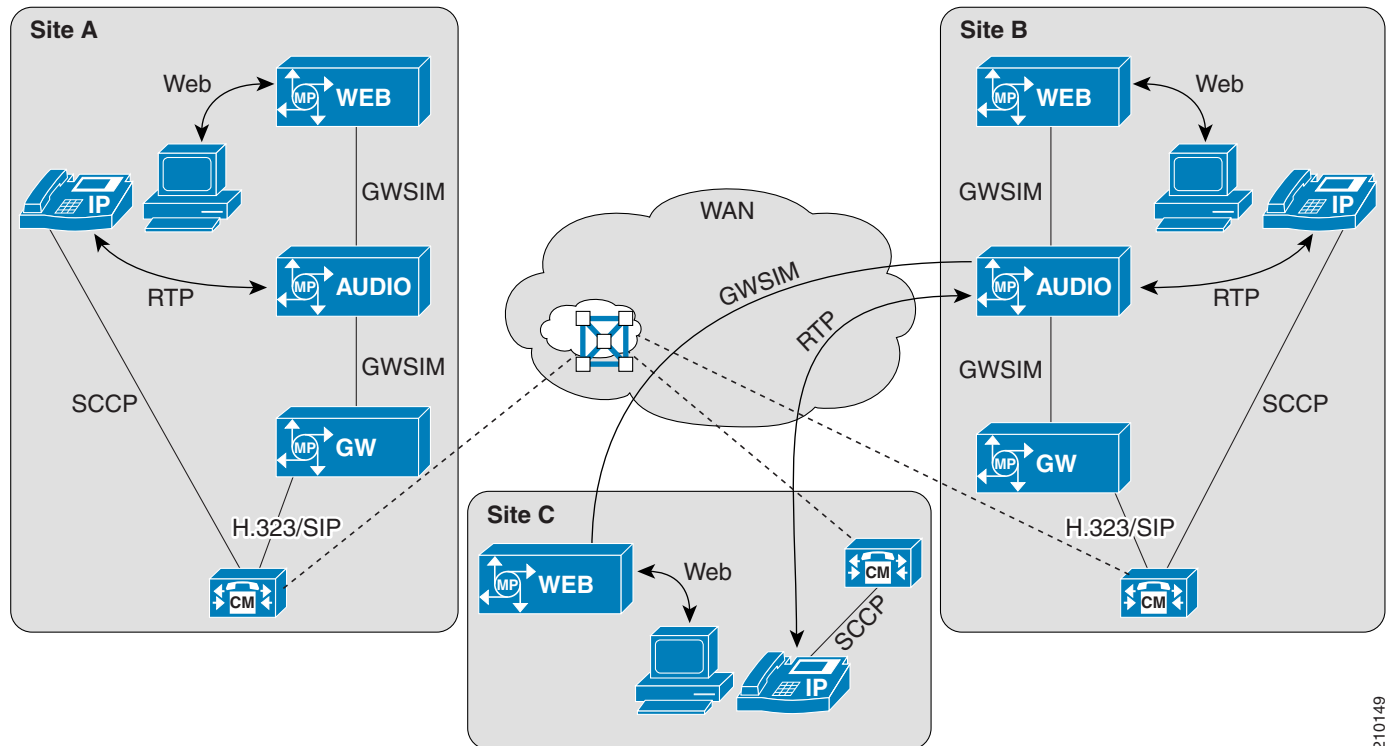
The addition of remote sites, with users separated from Cisco Unified MeetingPlace via WAN links, adds the following design recommendations:

- Remote user web collaboration bandwidth must be taken into account to understand the impact on existing links and to size new links. Bandwidth utilization is covered in the section on [Call Admission Control, QoS, and Bandwidth](#), page 14-8.
- Cisco Unified MeetingPlace supports G.711 and G.729, allowing remote participants to connect via G.729 without the use of transcoders.
- WAN outage to remote sites would require remote participants to connect via the PSTN. Web collaboration connectivity would be unavailable unless a backup data path exists.

Multisite WAN with Distributed Call Processing

The multisite WAN deployment with distributed call processing builds again upon the previous model, moving call processing to multiple locations via multiple Unified CM clusters (see Figure 14-4). In this example environment, Cisco Unified MeetingPlace components are distributed as well.

Figure 14-4 Multisite Deployment with Distributed Call Processing



210149

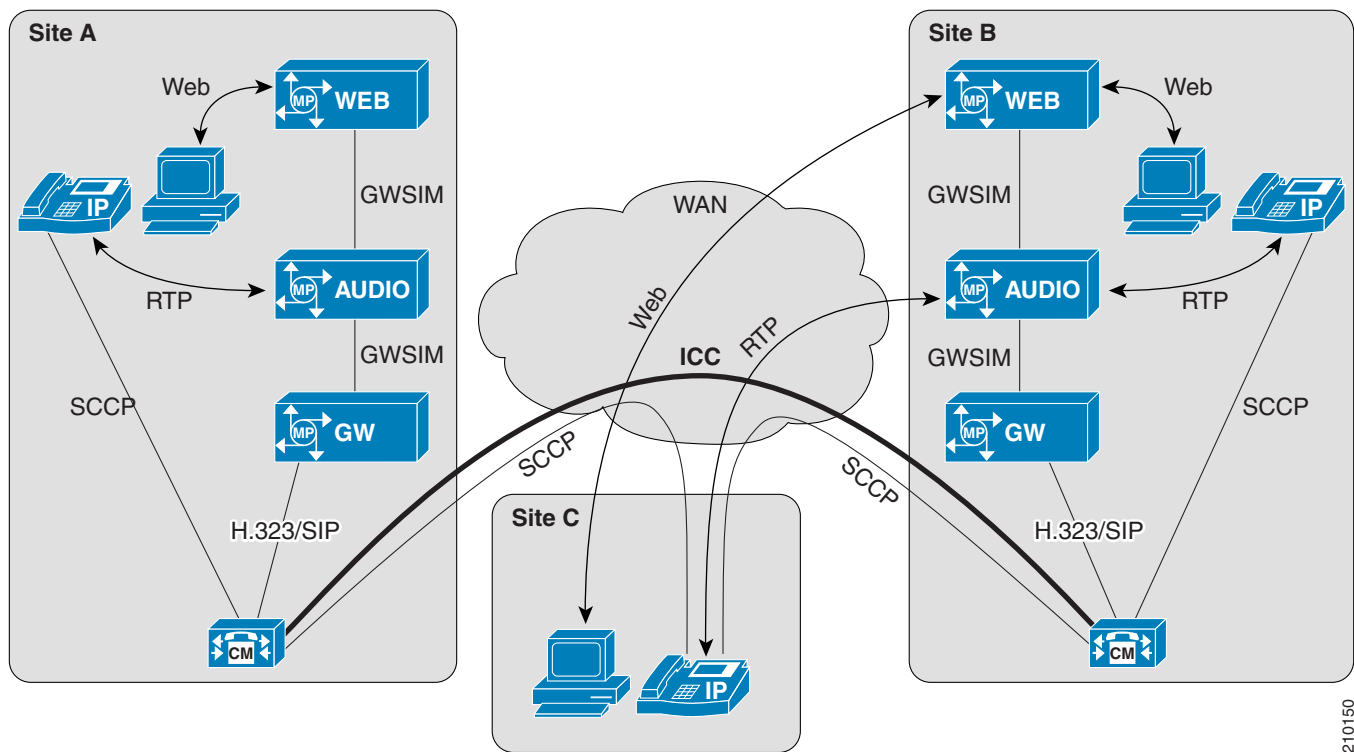
The multisite model with distributed call processing adds the following design considerations:

- Options for distributing Cisco Unified MeetingPlace can be applied to the centralized call processing model.
- Audio servers can be distributed to multiple locations, all with direct connectivity to a single central cluster or to multiple co-located clusters.
- Cisco Unified MeetingPlace Multiserver Meetings allow cascading of audio servers, which can conserve bandwidth by allowing local users to utilize their local audio server.
- Cisco Unified MeetingPlace Web collaboration servers can be distributed across multiple sites, but not geographically cascaded for active conferences. Multiserver or Clustered web conferencing is not geographically aware but does provide load balancing and scalability.

Clustering Over the WAN

The model for clustering over the WAN spreads a single Unified CM clusters across multiple sites, making that single cluster geographically redundant and taking high availability to the next level (see [Figure 14-5](#)). Incorporating Cisco Unified MeetingPlace into this environment implies a focus on redundancy and high availability.

Figure 14-5 Clustering Over the WAN



210150

Multiple Cisco Unified MeetingPlace audio servers and web collaboration servers can be distributed across Unified CM clustered sites. User profile information can be synchronized between servers, keeping user profiles accurate in multiple sites. During a WAN outage, each server will still be able to serve local participants as well as any remote participants having connectivity.

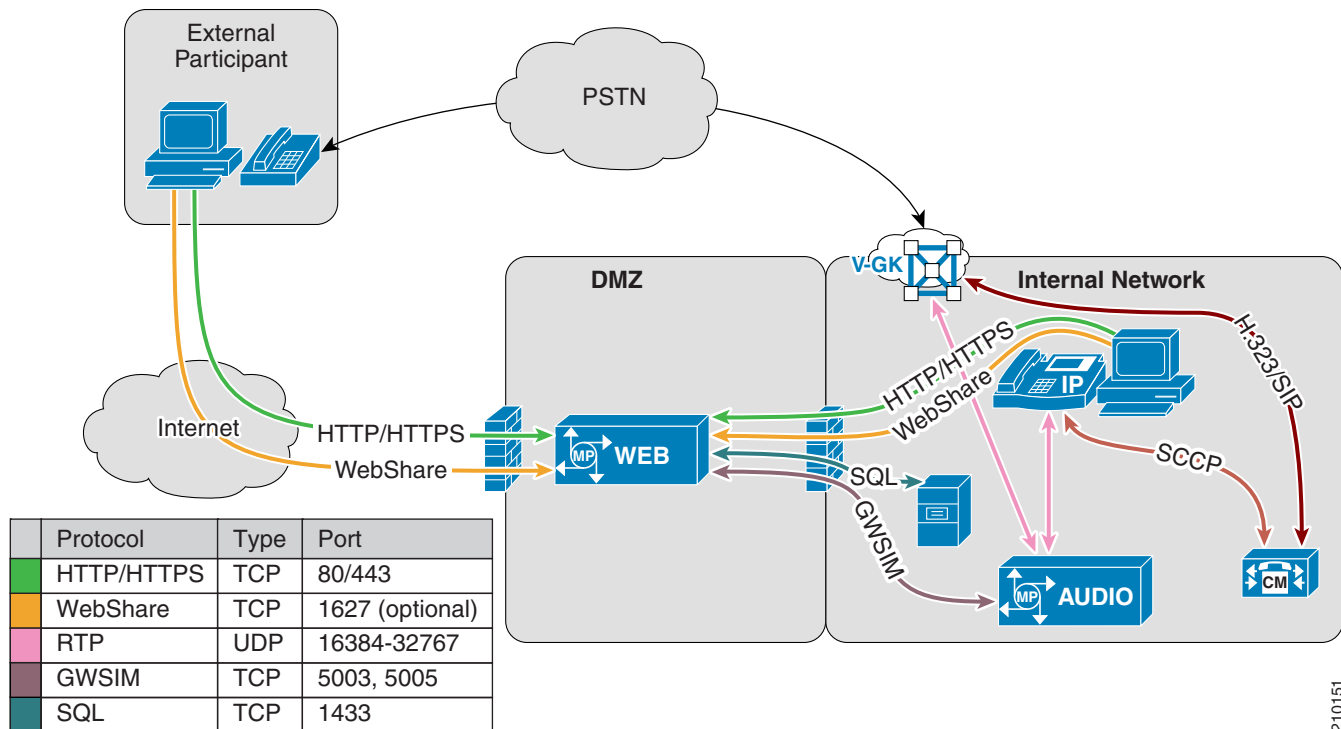
For more information regarding Cisco Unified MeetingPlace redundancy features, refer to the section on [Redundancy and Load Balancing, page 14-12](#), as well as the Cisco Unified MeetingPlace product documentation available at

http://www.cisco.com/en/US/products/sw/ps5664/ps5669/tsd_products_support_series_home.html

DMZ Deployment

Cisco Unified MeetingPlace (Release 5.4 and later) supports the placement of a web collaboration server within the demilitarized zone (DMZ). (See Figure 14-6.) Meetings with external participants automatically use this server for web collaboration. Audio from external participants comes into Cisco Unified MeetingPlace either through a direct TDM trunk to the Audio Server or through a VoIP gateway (controlled by Unified CM) to a VoIP port on the Audio Server.

Figure 14-6 DMZ Deployment



This deployment option, referred to as Segmented Meeting Access (SMA), may be applied to any variation of the deployment models discussed previously in this chapter. SMA implementations require deployment of an internal web server. For more information on SMA, refer to the Cisco Unified MeetingPlace product documentation available at

http://www.cisco.com/en/US/products/sw/ps5664/ps5669/tsd_products_support_series_home.html

Call Admission Control, QoS, and Bandwidth

Call admission control, Quality of Service (QoS), and proper bandwidth allocation are the main mechanisms to ensure voice and video quality. This section describes how these mechanisms apply to Cisco Unified MeetingPlace.

Call Admission Control

Call admission control with Cisco Unified MeetingPlace should be treated the same as call admission control with a Cisco IP Telephony gateway. Cisco Unified MeetingPlace acts like a gateway connected via a gatekeeper, directly via H.323, or via a SIP trunk. For the most effective call admission control implementations, the Cisco Unified MeetingPlace Audio Server and IP Gateway component should be co-located.

For further information regarding call admission control strategies, refer to the chapter on [Call Admission Control, page 9-1](#).

QoS Markings

Cisco Unified MeetingPlace Web Collaboration

Web collaboration traffic is not marked.

Cisco Unified MeetingPlace Audio Server

RTP media from the Audio Server is marked EF (DSCP 0x2E) by default. GWSIM traffic to and from other Cisco Unified MeetingPlace components is not marked.

Cisco Unified MeetingPlace IP Gateway Signaling

H.323 and SIP signaling traffic from the IP Gateway is not marked. Signaling traffic from Unified CM to the IP Gateway is marked CS3 (DSCP 0x18) by Cisco Unified MeetingPlace. If the IP Gateway is co-located with Unified CM, the lack of signal marking will not be an issue unless congestion is present in the locally switched network. For implementations with these components remote from each other, additional configuration may be needed to re-mark signaling traffic on the edge devices.

Bandwidth

Web Collaboration Bandwidth

Web Collaboration is the largest bandwidth consumer, ranging up to several Mbps per web collaboration client. This traffic becomes especially significant for remote users across WAN links. Cisco Unified MeetingPlace Web Collaboration does have the ability to reduce the amount of bandwidth consumed based on congestion encountered, but most available bandwidth will be utilized up to the maximum that Cisco Unified MeetingPlace Web can use.

Audio Bandwidth

Audio traffic consists of Real-Time Transport Protocol (RTP) traffic flowing to and from the Cisco Unified MeetingPlace Audio server. Both G.711 and G.729 codecs are supported with Cisco Unified MeetingPlace, allowing remote users to utilize the lower bandwidth codec without the need for transcoding resources.

Call Control Bandwidth

Call control bandwidth is extremely small but critical. Co-locating the Cisco Unified MeetingPlace IP Gateway component with Unified CM or the gatekeeper helps protect against issues with call control. Remote locations need proper QoS provisioning to ensure reliable operation.

GWSIM

GWSIM traffic flows between the Cisco Unified MeetingPlace Audio server and other Cisco Unified MeetingPlace components including the Cisco Unified MeetingPlace IP Gateway and Cisco Unified MeetingPlace Web.

Traffic between the Cisco Unified MeetingPlace Audio Server and the Cisco Unified MeetingPlace IP Gateway component is minimal but important traffic. Placing the Audio Server and IP Gateway component in the same location helps ensure proper operation. Separation of components by a WAN would require proper QoS provisioning to ensure reliable operation.

Traffic between the Cisco Unified MeetingPlace Audio Server and Cisco Unified MeetingPlace Web includes some database synchronization and might therefore at times be bursty in nature, but this traffic is not real-time.

Directory Integration

The Cisco Unified MeetingPlace Directory Services component provides direct synchronization of directory server information. The Directory Services component supports the following directory servers:

- Microsoft Active Directory 2000 and 2003

Microsoft Active Directory enables you to store, access, and manipulate organizational information about users and resources, and manage all elements of a networked environment, such as computers, groups, users, policies, and other user-defined objects.

- Netscape/SunOne/iPlanet LDAP Directory Server Version 4 and Version 5

Netscape/SunOne/ iPlanet is a general-purpose LDAP directory that stores, publishes, and centrally manages users and network resources.

- Unified CM directory

Synchronizations of user data from the Unified CM directory enable the Cisco Unified MeetingPlace system to support IP telephony users who are configured in Unified CM.

For details about the Directory Services component, refer to the *Administration Guide for Cisco Unified MeetingPlace Directory Services*, available at

http://www.cisco.com/en/US/products/sw/ps5664/ps5669/prod_maintenance_guides_list.html

H.323 and SIP Direct Integration

Cisco Unified MeetingPlace integrates with Unified CM through the Cisco Unified MeetingPlace IP Gateway. This gateway provides multiple methods for integration. The first two methods, discussed in this section, involve direct integration via H.323 or SIP between Unified CM and the Cisco Unified MeetingPlace IP Gateway. The third method involves the use of a gatekeeper for integration and is discussed in the section on [Gatekeeper Integration, page 14-11](#).

Figure 14-7 Direct Integration Through H.323 or SIP



H.323

H.323 integration, by defining Cisco Unified MeetingPlace as an H.323 gateway on Unified CM, is the preferred implementation at this time. This method has the following characteristics:

- All outbound and inbound calls route through a single Unified CM cluster.
- Call admission control and dial plan rules are enforced using the defined H.323 gateway.

SIP

SIP integration, by defining a SIP trunk on Unified CM and Cisco Unified MeetingPlace, is the less preferred implementation at this time. A static media termination point (MTP) must be associated with all calls across a SIP trunk between Unified CM and Cisco Unified MeetingPlace. This requirement is due to the lack of delayed-media support in Cisco Unified MeetingPlace. The SIP trunk has the same capabilities listed above for H.323, but the following limitations apply:

- Static MTP is required
- The audio codec must be G.711 with the static MTP; transcoders must be used for G.729.

Due to these limitations, implementation via H.323 is the preferred method at this time.

Implementation of a SIP trunk also has the following requirements when configuring on Unified CM:

- MTP Required must be checked in the SIP Trunk configuration, and MTP resources must be available.
- A separate SIP Trunk Security profile must be created, with the outbound transport type set to UDP and associated with the SIP Trunk to Cisco Unified MeetingPlace.

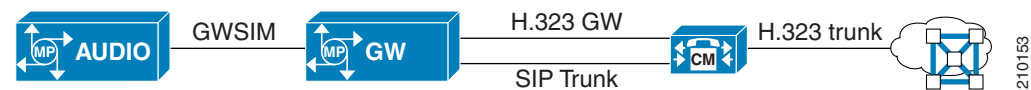
Gatekeeper Integration

Cisco Unified MeetingPlace can be integrated into a gatekeeper environment by either of the two methods detailed in this section. The addition of video support to Cisco Unified MeetingPlace adds more complex gatekeeper requirements, which are outlined in the section on [Video Gatekeeper Integration](#), page 14-20.

Integration Via Unified CM to a Gatekeeper

Integrating via H.323 or SIP directly to Unified CM, as described in the previous section, allows all incoming and outgoing calls to be controlled by Unified CM. Off-cluster incoming and outgoing calls are routed through the Unified CM gatekeeper trunk (see [Figure 14-8](#)). Implementation in this manner takes advantage of Unified CM dial plan restrictions, abbreviated dial, and digit manipulation.

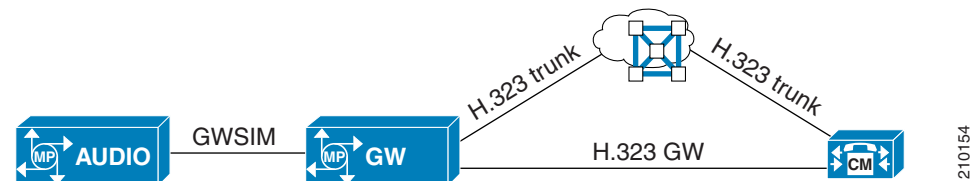
Figure 14-8 Integration Via Unified CM to a Gatekeeper



Integration Directly to a Gatekeeper

Cisco Unified MeetingPlace can register directly with a gatekeeper as a terminal (see [Figure 14-9](#)). The primary E.164 address defined in the Cisco Unified MeetingPlace system is used when registering. Additional E.164 addresses for direct-dial meetings must be assigned statically in the gatekeeper. This model has some additional redundancy because the gatekeeper is able to route outbound calls to multiple nodes.

Figure 14-9 Integration Directly to a Gatekeeper



The destination Unified CM cluster or clusters may be configured with an H.323 gateway pointing to one or more MeetingPlace IP Gateways to match incoming calls from these gateways and enforce dial plan restrictions and call admission control restrictions.

In a Unified CM environment, the following considerations apply to direct registration with a gatekeeper:

- Outdialed calls from Cisco Unified MeetingPlace are resolved in the gatekeeper and are not subject to dial restrictions or call admission control restrictions imposed in Unified CM.
- Outdialed calls from Cisco Unified MeetingPlace do not benefit from abbreviated dialing or digit manipulation defined in Unified CM.

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. The dial plan is responsible for instructing the call processing agent on how to route calls.

Unified CM Dial Plan

As mentioned previously, Cisco Unified MeetingPlace can be integrated with Unified CM by any of the following means:

- H.323 gateway defined on Unified CM
- SIP trunk between Cisco Unified MeetingPlace and Unified CM
- Direct gatekeeper integration

With the first two methods, the dial plan with respect to Cisco Unified MeetingPlace is relatively simple. Cisco Unified MeetingPlace is referenced via a route pattern defined on Unified CM, which points to the H.323 gateway or SIP trunk. All other dial plan configuration is handled through Unified CM and possibly a gatekeeper if Unified CM is connected to a gatekeeper. Use of a route list to point to multiple Cisco Unified MeetingPlace IP gateways adds redundancy and is covered in the section on [Redundancy and Load Balancing, page 14-12](#).

With the third method, direct gatekeeper integration, the dial plan centers more on the gatekeeper. Cisco Unified MeetingPlace registers with the gatekeeper as a terminal with a single E.164 address. You must consider this factor when planning gatekeeper integration so that any abbreviated dialing expansion or digit manipulation can allow for proper call resolution to the single E.164 address.

Reservationless Single Number Access (RSNA)

The Reservationless Single Number Access (RSNA) feature enables multiple Cisco Unified MeetingPlace Audio Servers to appear as one server to the user community. This feature is designed primarily for reservationless meetings. With this feature, users can dial a single directory number to access multiple Audio Servers deployed in the same or different locations. Users enter their reservationless ID, and the system directs the calls to the appropriate server for that meeting. This feature also enables users in a remote location to dial the local server.

RSNA relies on the SIP REFER method. SIP trunks on Unified CM support REFER, enabling the implementation of this feature. Before planning implementation of RSNA, refer to the limitations of integration via a SIP trunk listed in the section on [H.323 and SIP Direct Integration, page 14-10](#).

Redundancy and Load Balancing

This section describes redundancy and load balancing considerations for the following MeetingPlace components:

- [Cisco Unified MeetingPlace Audio Server, page 14-13](#)
- [Cisco Unified MeetingPlace H.323/SIP IP Gateway, page 14-14](#)
- [Cisco Unified MeetingPlace Web, page 14-15](#)
- [Unified CM, page 14-16](#)

Cisco Unified MeetingPlace Audio Server

If the MeetingPlace Audio Server fails, calls in progress will be dropped. If the Audio Server is unreachable by the MeetingPlace H.323/SIP IP Gateway, the MeetingPlace H.323/SIP IP Gateway will immediately reject any incoming calls. If the MeetingPlace H.323/SIP IP Gateway is down, Unified CM will be unable to establish an H.323 connection with the MeetingPlace H.323/SIP IP Gateway. There is no automatic failover mechanism available for the MeetingPlace Audio Server.

Disaster Recovery

A disaster recovery plan provides for orderly restoration of the database, computing, and services. The plan usually includes the following provisions:

- Redundancy (spare parts of hardware)
- Data and software backups
- Alternate emergency locations
- Operations split across multiple sites

A component called the MeetingPlace Network Backup Gateway enables system managers to transfer multiple copies of the MeetingPlace database (audio configuration files and scheduled meeting information) from the Audio Server to a designated storage server on the network. Files are encrypted for security. A maximum of three Network Backup Gateways per system can be implemented, but only one of them can make the database transfer at a time.

Redundancy

If two or more MeetingPlace Audio Servers are deployed, any of the following plans for redundancy can be implemented:

- Shadow Server

A shadow server is a redundant MeetingPlace Audio Server that is not active until the primary Audio Server fails. During the disaster, the shadow server must be switched from shadow mode to active mode using the command line interface (CLI). Once active, the shadow server can store only limited information, such as profile information, meeting information (past, present, and future), and participant information. Recording, attachments, hardware and network configuration, and automatic software upgrades are not available on the shadow server. The shadow server has the following network connectivity requirements:

- Less than 250 ms round trip delay
- Less than 1% packet loss
- Minimum bandwidth of 384 kbps

- Dual Conference Servers

In this plan, two MeetingPlace Audio Servers are active production servers, and each one can overflow to the other. Profiles are synchronized automatically between the servers through the Directory Services. Automatic failover between the two Audio Servers does *not* occur. Meetings have to be uploaded to the remaining server during an emergency, and users must rejoin the meetings manually. On Unified CM, you can configure route groups and route lists so that if one MeetingPlace Audio Servers fails, the calls will be routed to the other active MeetingPlace Audio Server.

- Continuous Meeting Server

In this plan, a MeetingPlace Audio Server is implemented with a set of critical meetings pre-created and always available. This deployment is suitable for any crisis management application, with features such as blast outdial to crisis management teams.

Cisco Unified MeetingPlace H.323/SIP IP Gateway

The following redundancy and load balancing considerations apply to the MeetingPlace H.323/SIP IP Gateway.

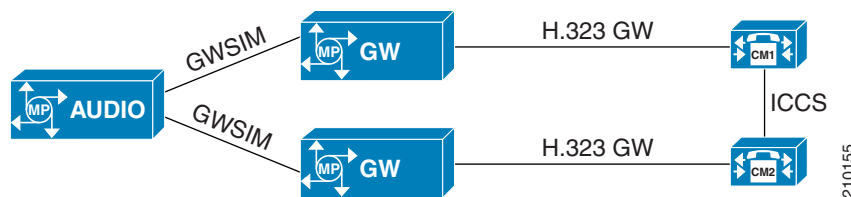
Redundancy

If the MeetingPlace H.323 IP Gateway fails, calls in progress will be dropped.

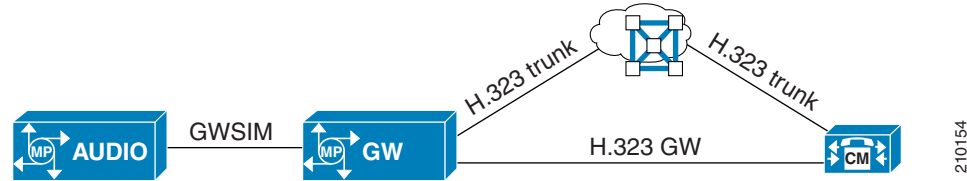
Two or more MeetingPlace H.323 IP Gateways can connect to the same MeetingPlace Audio Server. There are two methods for implementing redundancy, one with and one without a gatekeeper.

- Redundancy without a gatekeeper (see [Figure 14-10](#))
 - Configure all of the MeetingPlace H.323 IP Gateways on Unified CM and put them in the same route group. If the link between Unified CM and the MeetingPlace H.323/SIP IP Gateway fails, Unified CM chooses the next MeetingPlace H.323/SIP IP Gateway in the current route group. This same method also works for load balancing.
 - Configure the MeetingPlace H.323 IP Gateways to point to different Unified CM nodes. If one Unified CM node fails, outbound calls are place through one of the other MeetingPlace IP Gateways to its configured Unified CM node.

Figure 14-10 Redundancy Without a Gatekeeper



- Redundancy with a gatekeeper (see [Figure 14-11](#))
 - Calls outbound from MeetingPlace that are sent to the gatekeeper may be directed to alternate Unified CM nodes in the event of a Unified CM failure. This is an advantage over direct outbound configuration on a single MeetingPlace IP Gateway, which allows only one destination node to be configured.
 - The use of a defined H.323 gateway on Unified CM is still recommended for this model to provide call admission control and dial plan restrictions as well as redundancy for inbound calls if there are multiple MeetingPlace IP Gateways.

Figure 14-11 Redundancy With a Gatekeeper**Load Balancing**

For dial-in load balancing, you can use either of the following same approaches for redundancy:

- Use a route group in Unified CM, but choose **Circular** instead of **Top down** for the Distribution Algorithm. Unified CM will distribute the outbound calls between the gateways in round-robin fashion.
- Use a gatekeeper. For purposes of load balancing, do not configure **gw-priority** to specify a preferred gateway. Instead, let the gatekeeper distribute the calls in round-robin fashion to all the registered gateways.

As explained in the previous section, load balancing for outdialing is done only among those MeetingPlace H.323/SIP IP Gateways that never have outdial failures.

Cisco Unified MeetingPlace Web

The following redundancy and load balancing considerations apply to the MeetingPlace Web.

Redundancy

If a web-conferencing server fails during a web conference, all users are disconnected from the web portion of the meeting, and the current web conference cannot continue. Users can rejoin their web conference on another active web server. If the SQL server database is on the failed server, all web conferencing servers in the cluster become nonfunctional and users will be unable to conduct web conferences until the database is restored.

Load Balancing

MeetingPlace Web servers can be grouped into clusters to increase performance for load balancing (to distribute the meeting requests to different web servers) and to increase the capability of handling more meetings. The MeetingPlace Web cluster can contain a maximum of six web servers in either internal or external configurations.

The MeetingPlace WebConnect feature provides integration between multiple internal and external MeetingPlace systems through a single URL. External systems reside in the DMZ for external users. It can schedule across multiple Audio Servers. If there is not enough capacity on the first Audio Server, it will try the second, and so on.

The following redundancy and load balancing considerations apply to Unified CM.

If Unified CM fails, MeetingPlace calls in progress using that Unified CM will be dropped and users must rejoin the meeting.

For dial-in from endpoints to MeetingPlace, load balancing for calls is achieved through Unified CM redundancy groups and route group configurations.

Integrating video with Cisco Unified MeetingPlace 5.4 differs greatly from video integration with previous versions of Cisco Unified MeetingPlace. Cisco Unified MeetingPlace 5.4 introduces a new component called Video Administration, which is the control point for all video calls to and from the MCU. Video Administration interfaces with the Video Integration component on the MeetingPlace Web server as well as with the gatekeeper and Unified CM directly. (See [Figure 14-12](#).) The remainder of this chapter discusses the impact on system design due to Cisco Unified MeetingPlace Video, and primarily due to the Video Administration component.

The diagram illustrates a VoIP network architecture. Key components and their connections are as follows:

- WEB + VIDEO** (top left) connects to **Video Admin** (top center) via **SOAP/XML**.
- Video Admin** connects to a switch (top right) via **XML** and **H.323**.
- Video Admin** connects to **GW** (middle left) via **H.323**.
- Video Admin** connects to **CM** (middle right) via **H.323**.
- Video Admin** connects to **V-GK** (middle right, cloud icon) via **H.323**.
- AUDIO** (bottom left) connects to **GW** via **GWSIM**.
- GW** connects to **CM** via **H.323** and **H.323/SIP**.
- CM** connects to **V-GK** via **H.323** and **RAS**.
- CM** connects to a desktop phone (bottom center) via **SIP**.
- CM** connects to a mobile phone (bottom right) via **SCCP**.
- V-GK** connects to a switch (middle right) via **H.323**.
- V-GK** connects to a switch (bottom right) via **H.323**.
- V-GK** connects to a switch (bottom right) via **H.323**.

**Note**

In this chapter, the name Cisco Unified MeetingPlace Video refers to the entire video system, including Video Administration, the Video Integration component, MCUs, gatekeepers, endpoints, and Unified CM. One component of the system is the Video Integration component, which resides on one or more Cisco Unified MeetingPlace Web servers. Some other Cisco Unified MeetingPlace documents refer to Video Integration as the entire video system, and this difference will be noted where possible when referencing those documents.

Video Administration

Video Administration is a new component introduced with Cisco Unified MeetingPlace 5.4 Video. It controls all aspects of video conferencing, interfacing with MCUs, gatekeeper, Unified CM, and other Cisco Unified MeetingPlace 5.4 components.

Video Administration is the main control point for all incoming and outgoing video calls, and it makes all decisions regarding resource utilization and cascading of resources. In addition to basic MCU cascading to increase conference size, Video Administration can intelligently select which MCU to use based on internally defined locations for each participant. Multisite video meetings can result in cascading MCUs local to user groups, thus creating one link across a WAN between cascaded MCUs.

Video Administration has the following characteristics:

- Video Administration resides on a separate Windows-based server and cannot reside on the same server with other Cisco Unified MeetingPlace components.
- Video Administration sits between the MCUs and all other components. The Cisco Unified MeetingPlace Video Integration component, Unified CM, and the gatekeeper communicate with Video Administration, which in turn communicates with the MCUs.
- Video Administration contains a special integrated gatekeeper to which only the MCUs register. The MCUs must register to Video Administration.
- All routing decisions for MCU selection and cascading are made by Video Administration.
- Video Administration terminates and authenticates all inbound calls, and only media is sent to the MCUs.
- Video Administration originates and maintains all outbound calls, and only media is sent to the MCUs.
- Video Administration requires the use of an external gatekeeper for outbound calls.
- Video Administration does not register to the gatekeeper it uses for outbound calls.
- Video Administration does not support a direct SIP trunk connection from Unified CM.
- Video Administration does not support a direct H.323 gatekeeper-controlled trunk connection from Unified CM but rather is defined as an H.323 gateway in Unified CM.
- All video meetings scheduled in Cisco Unified MeetingPlace 5.4 are replicated in Video Administration, with appropriate resources reserved
- Participant information and real-time status are relayed from the MCUs to Cisco Unified MeetingPlace components by Video Administration.
- Other than media stream termination, only Video Administration communicates with the MCUs.
- Video Terminals can be defined in Video Administration and reserved within Cisco Unified MeetingPlace Web when users create a meeting.
- Any endpoint (H.323, SCCP, or SIP) can be defined in Video Administration as a Video Terminal.

- MCU selection can be impacted by assigning locations to MCUs and Video Terminals defined in Video Administration.

Video Deployment Models

This section discusses design considerations and recommendations for adding Cisco Unified MeetingPlace Video to each of the basic Unified CM deployment models. Many variations are possible in the deployments, but the models presented here cover the base deployments without examining every variation.

Single-Site Deployment

With a single-site deployment, all components are located in one site with no remote participation. [Figure 14-12](#) illustrates a single site.

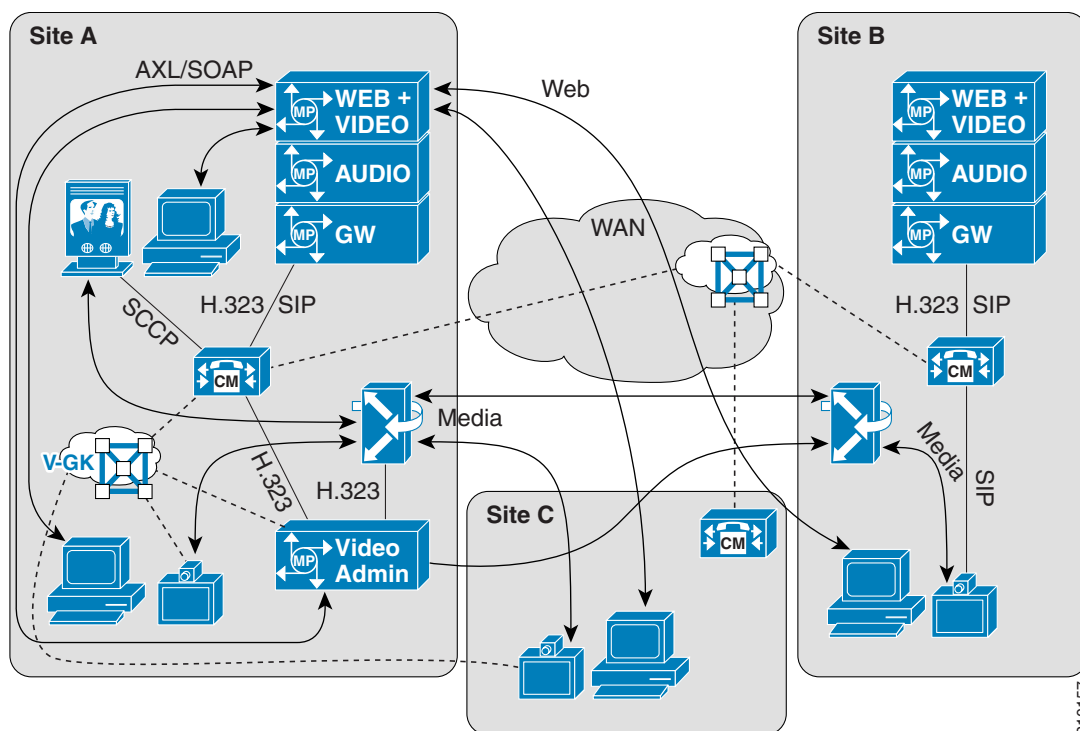
Multisite Centralized Deployment

The multisite model with centralized call processing adds remote users to a centralized deployment. All Cisco Unified MeetingPlace Video components reside in the central data center. Remote participants proper bandwidth allocation and QoS provisioning to ensure proper operation. See [Call Admission Control, QoS, and Bandwidth](#), [page 14-8](#), for more information.

Multisite Distributed Deployment

Multisite distributed deployment adds decentralized call processing and decentralized Cisco Unified MeetingPlace Video components.

Figure 14-13 Multisite Distributed Deployment with Video



210157

The following design notes apply to this model:

- For MCUs that are remote from Video Administration, you must take into account the important call admission control and call routing restrictions that are documented in the section on [Call Admission Control, QoS, and Bandwidth](#), page 14-8.
- Video Administration does have the ability to route calls intelligently to local MCU resources if they are configured properly.
- Remote H.323 video endpoints must be registered to the central video gatekeeper unless multiple video gatekeepers are implemented
- All Cisco Unified MeetingPlace Web participants must use the single active Cisco Unified MeetingPlace Web server with Video Integration enabled.
- The dial plan on the infrastructure gatekeeper should route all calls destined for Cisco Unified MeetingPlace Video to the associated Unified CM cluster.

Clustering Over the WAN

Due to the limitations of having only one Video Administration server, clustering over the WAN does not change from the multisite distributed model with respect to Unified CM Video components.

When implementing clustering over the WAN, it is best to include only local Unified CM nodes when defining the redundancy group for Video Administration connections.

Video Call Admission Control, QoS, and Bandwidth

Call admission control, Quality of Service (QoS), and proper bandwidth allocation are the main mechanisms to ensure voice and video quality. This section describes how these mechanisms apply to Cisco Unified MeetingPlace Video.

Video Call Admission Control

Video Administration is the routing point for all MCUs, and it determines on which MCU each call terminates. As shown in [Figure 14-14](#), all H.225 and H.245 signaling terminates on Video Administration:

- All inbound video calls terminate on Video Administration.
- All outbound video calls originate from Video Administration.
- All signaling to and from the MCU is handled by Video Administration.

Figure 14-14 Video Administration Signaling Termination



Because of this functionality, visibility beyond Video Administration for any current call admission control mechanism is not possible. Video Administration is seen as the terminating endpoint to all call admission control mechanisms.

Implementation of call admission control with MCUs distributed across multiple sites becomes a challenge. Currently the only call admission control capabilities for distributed MCUs are implemented separately within Video Administration and do not tie into any other mechanism in any way. This is an independent, standalone call admission control mechanism for use only with Video Administration MCUs. It is not an optimal way to implement call admission control but is the only available option for this release of Cisco Unified MeetingPlace.

Implementation of call admission control via the Video Administration component is detailed in the *Administration Guide for Cisco Unified MeetingPlace Video Integration Release 5.4*, available at

http://www.cisco.com/en/US/products/sw/ps5664/ps5669/prod_maintenance_guides_list.html

**Note**

The term Video Integration in the above Administration Guide is equivalent to the term Cisco Unified MeetingPlace Video in this chapter.

Video QoS

The following traffic markings apply call control, media, and other flows:

- Video Administration does not mark outgoing H.323 call control traffic.
- Incoming H.323 call control traffic from Unified CM to Video Administration is marked CS3 (DSCP 0x18) by default.
- Traffic between Video Administration and the MCU is not marked.
- The MCU marks outgoing media with AF41 (DSCP 0x22) by default.
- Traffic between Video Administration and the Video Integration component is not marked.

For components separated by WAN links, re-marking may be needed to ensure reliable operation. For further information regarding video QoS, refer to the chapter on [IP Video Telephony, page 16-1](#).

Video Bandwidth

See the chapter on [IP Video Telephony, page 16-1](#), for current information regarding video bandwidth.

Video Gatekeeper Integration

With the introduction of Cisco Unified MeetingPlace 5.4, Gatekeeper is required to outdial video calls from Video Administration to video endpoints, regardless of the type of endpoint. This requirement, among other changes with Cisco Unified MeetingPlace Video, significantly impacts the way Unified CM, Cisco Unified MeetingPlace, and Video Gatekeeper interact in this environment. The following sections detail relevant subjects involving Video Gatekeeper and its critical role in the deployment of video with Cisco Unified MeetingPlace 5.4.

This section assumes Video Administration uses Video Gatekeeper to route all outbound calls through Unified CM. Integration of Video Administration into an infrastructure gatekeeper environment, bypassing Unified CM, is another deployment option not covered here due to the focus on endpoints controlled by Unified CM.

This section covers the following topics:

- [H.323 Video Endpoints and Video Gatekeeper, page 14-21](#)

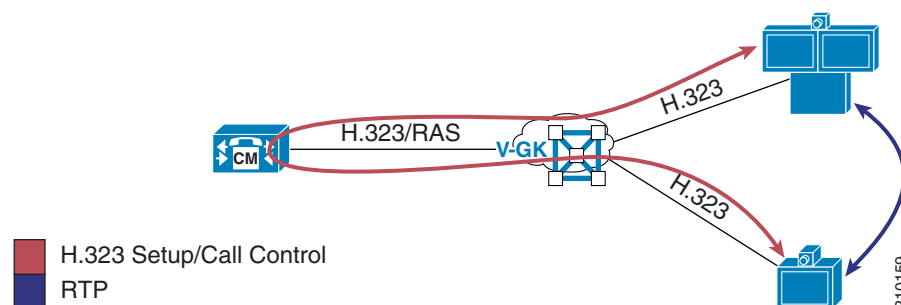
Basic concepts of how H.323 video endpoints interact with Video Gatekeeper.

- [Via-Zone Video Gatekeeper, page 14-21](#)
Description of how a via-zone gatekeeper operates.
- [Video Administration and Video Gatekeeper Overview, page 14-22](#)
Overview of how Video Administration and Video Gatekeeper interact and process inbound and outbound calls.
- [Video Administration and Video Gatekeeper Integration with Unified CM, page 14-23](#)
Specifics around Unified CM configuration and interaction with Video Administration and Video Gatekeeper.
- [Additional Video Gatekeeper Configuration Requirements, page 14-23](#)
Additional steps to ensure proper call resolution within Video Gatekeeper for outbound calls from Video Administration.
- [Zones and RASAggregator Redundancy Groups with H.323 Video Endpoints on Unified CM, page 14-24](#)
Critical information regarding redundancy groups and gatekeeper zone configuration of H.323 video endpoints defined on Unified CM.

H.323 Video Endpoints and Video Gatekeeper

For deployments incorporating H.323 video endpoints, a separate video endpoint gatekeeper is recommended. (See [Figure 14-15](#).) This gatekeeper should be set up as a via-zone gatekeeper, as explained below. Setup of this nature ensures that every call between H.323 video endpoints and any other video endpoint or termination point (SIP, SCCP, Cisco Unified MeetingPlace Video Administration, or other H.323 video endpoints on the same gatekeeper) are handled by Unified CM for digit manipulation, dialing restrictions, and call admission control.

Figure 14-15 Video Endpoints and Video Gatekeeper



Via-Zone Video Gatekeeper

The via-zone gatekeeper can be a complex concept, but the following points present a very simplified way of looking at it and the easiest way to understand the basic operation in this environment:

- Via-zone gatekeepers route calls through an IP-to-IP gateway.
- Video Gatekeeper is a via-zone gatekeeper.
- Unified CM acts as an IP-to-IP gateway.
- Video Gatekeeper routes all calls through Unified CM.

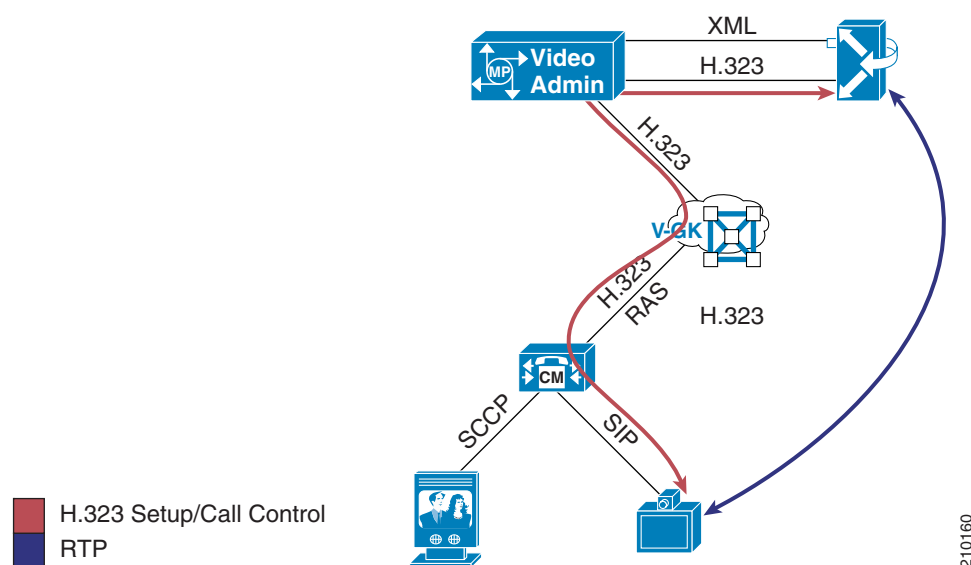
**Note**

You must configure Unified CM to act as an IP-to-IP gateway by setting the H.323 Service Parameter **Send Product ID and Version ID** to **TRUE**.

Video Administration and Video Gatekeeper Overview

Video Administration requires the use of a gatekeeper to resolve outbound calls. Video Administration sends Location Request messages (LRQs) to this gatekeeper while not registering in any way to that gatekeeper. (See [Figure 14-16](#).) This behavior differs greatly from previous Cisco Unified MeetingPlace versions.

Figure 14-16 Video Administration and Video Gatekeeper



In [Figure 14-16](#), an outbound call is placed to a SIP video endpoint controlled by Unified CM. This example is a simplified illustration of basic call flow but is not a detailed or complete example of call flow for outbound calls. For a detailed explanation of call flows, See the section on [Video Call Flows](#), page 14-27.

In a Unified CM environment that contains a Video Gatekeeper for H.323 video endpoints, the logical choice is to use Video Gatekeeper to route all calls through Unified CM.

Sending outbound calls to the Video Gatekeeper would route calls through Unified CM and would apply dial plan restrictions, abbreviated dial, and digit manipulation.

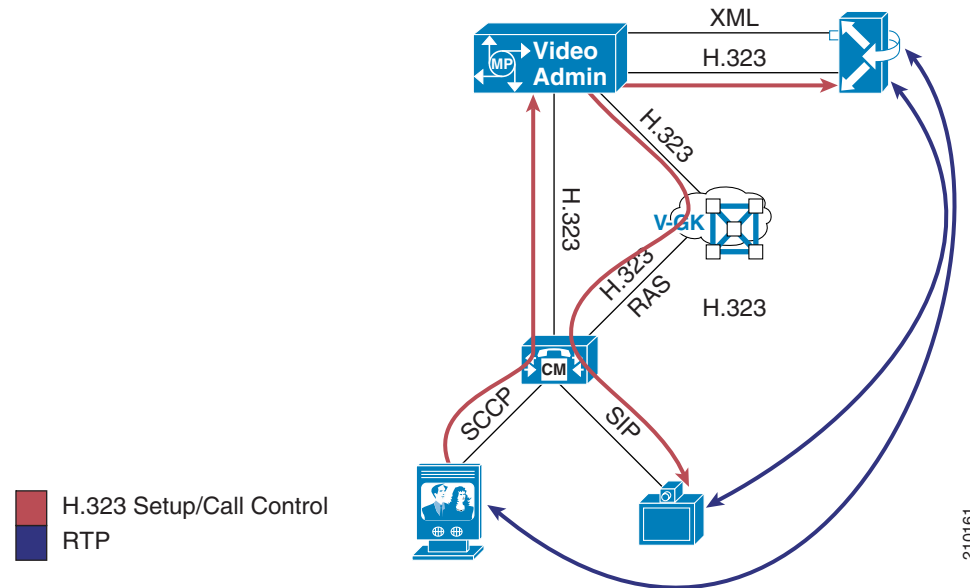
Sending outbound video calls to the primary gatekeeper, if one exists, would have the following effects:

- Outdialed calls from Video Administration are resolved in the gatekeeper and are not subject to dialing restrictions imposed in Unified CM.
- Outdialed calls from Video Administration do not benefit from abbreviated dialing or digit manipulation defined in Unified CM.

Video Administration and Video Gatekeeper Integration with Unified CM

Building upon the previous section, this section takes a closer look at the role of Unified CM configuration in the environment illustrated in [Figure 14-17](#).

Figure 14-17 Video Administration, Video Gatekeeper, and Unified CM



You must configure an H.323 gateway on Unified CM to point to the Video Administration server. The role of this H.323 gateway is critical for calls both inbound to and outbound from Video Administration.

Calls inbound to Unified CM (outbound from Video Administration) resolve via Video Gatekeeper to a particular Unified CM node. H.225 setup is then sent from Video Administration to that Unified CM node. Unified CM matches the incoming call to the defined H.323 gateway and applies any restrictions or manipulation associated with that gateway to the call. The H.323 trunk and Registration Admission Status (RAS) Aggregator to Video Gatekeeper are critical as well to provide call resolution, but the H.323 gateway defined on Unified CM provides control of the ingress call from Video Administration.

Calls outbound to Video Administration (inbound from Unified CM) are resolved via dial pattern matches associated with the defined H.323 gateway on Unified CM. H.225 setup is sent directly to Video Administration, and the call is completed without the use of Video Gatekeeper.

Additional Video Gatekeeper Configuration Requirements

With the use of a via-zone gatekeeper to resolve outbound calls from Video Administration, some additional configuration may be necessary to ensure proper call resolution.

When Video Administration attempts to resolve an outbound call, it begins by sending an LRQ to Video Gatekeeper. This LRQ is resolved to Unified CM for all outgoing calls, based on Video Gatekeeper being a via-zone gatekeeper as discussed previously. For a detailed explanation of the call flow, see the section on [Video Call Flows](#), page 14-27.

One scenario, however, can cause Video Gatekeeper to resolve to something other than Unified CM. If zone prefixes are not used in the Video Gatekeeper and the outdialed number from Video Administration matches exactly with an E.164 address registered to the Video Gatekeeper, the call will bypass Unified CM and complete directly to that endpoint. This behavior occurs regardless of the zone in which

the endpoint is registered. To resolve this issue, you can add a **hopoff** statement to Video Gatekeeper to resolve all LRQs to Unified CM via an intercluster trunk. This method should be implemented on endpoint via-zone gatekeepers only. The **hopoff** will send all calls to a particular zone regardless of origin or destination. (The example below illustrates use of the **hopoff** statement.) If zone prefixes are used on the Video Gatekeeper, this scenario does not occur.

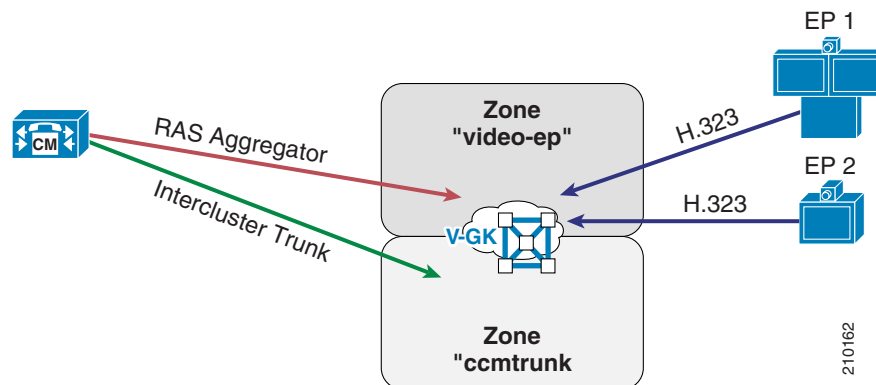
The following notes apply to the example below, which shows a simplified Video Gatekeeper configuration:

- An intercluster trunk (ICT) from Unified CM is registered to the zone ccmtrunk.
- H.323 video endpoints are registered to the zone video-ep.
- Zone prefixes are not used, forcing the use of the **hopoff** statement.
- The H.323 endpoints are defined on Unified CM in the zone video-ep, causing a RASAggregator trunk from Unified CM to register in that same zone.

Gatekeeper

```
zone remote video-admin cisco.com 10.2.2.2 invia ccmtrunk outvia ccmtrunk
zone local video-ep cisco.com 10.1.1.1 invia video-ep outvia video-ep enable-intrazone
zone local ccmtrunk cisco.com invia ccmtrunk outvia ccmtrunk enable-intrazone
send-cisco-circuit-info
gw-type-prefix * hopoff ccmtrunk default-technology
lrq forward-queries
no use-proxy video-ep default inbound-to terminal
no use-proxy video-ep default outbound-from terminal
no use-proxy ccmtrunk default inbound-to terminal
no use-proxy ccmtrunk default outbound-from terminal
```

Figure 14-18 Gatekeeper Configuration Example



Zones and RASAggregator Redundancy Groups with H.323 Video Endpoints on Unified CM

Unified CM redundancy groups (Unified CM groups and device pools) can be defined for H.323 video endpoints to allow failover to another node if the primary node fails. Improper configuration of H.323 video endpoints on Unified CM with respect to these redundancy groups can easily cause call failure.

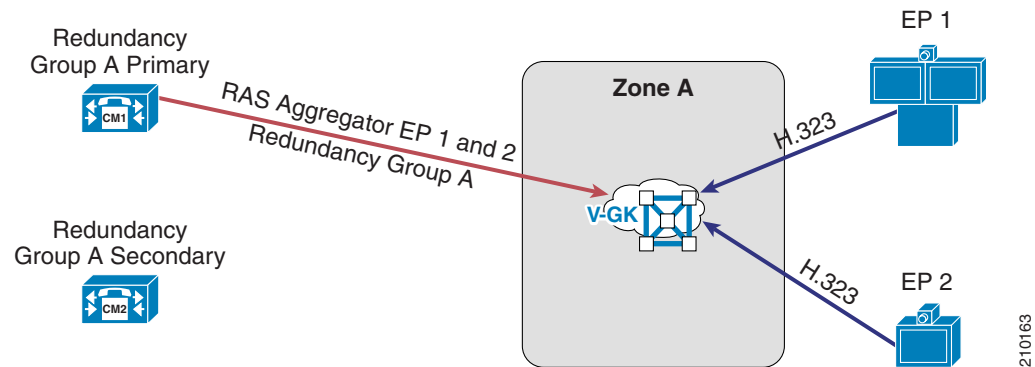
When you configure H.323 video endpoints on Unified CM, a gatekeeper and a zone are defined in addition to a redundancy group. Unified CM registers a single RASAggregator Trunk from the highest priority node listed in the chosen redundancy group. Improper configuration that allows two nodes to register RASAggregator Trunks to the same gatekeeper zone will cause call routing issues and call failure.

The following sections list the possible configuration scenarios.

One Zone and One Redundancy Group

In the most straightforward scenario, all the H.323 video endpoints have the same redundancy group and use the same gatekeeper zone (see Figure 14-19). One RASAggregator Trunk is registered in the single endpoint zone. This configuration ensures proper operation, with no call loss due to mis-routing.

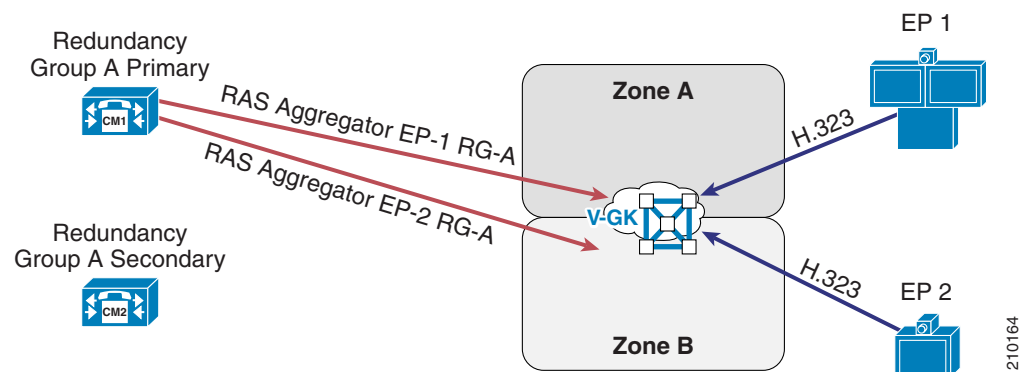
Figure 14-19 One Zone and One Redundancy Group



Multiple Zones and One Redundancy Group

In this scenario, H.323 video endpoints can be defined in Unified CM to be in different zones on the same video gatekeeper (see Figure 14-20). Two RASAggregator Trunks are registered in the two separate endpoint zones. Again this ensures proper operation, with no call loss due to mis-routing.

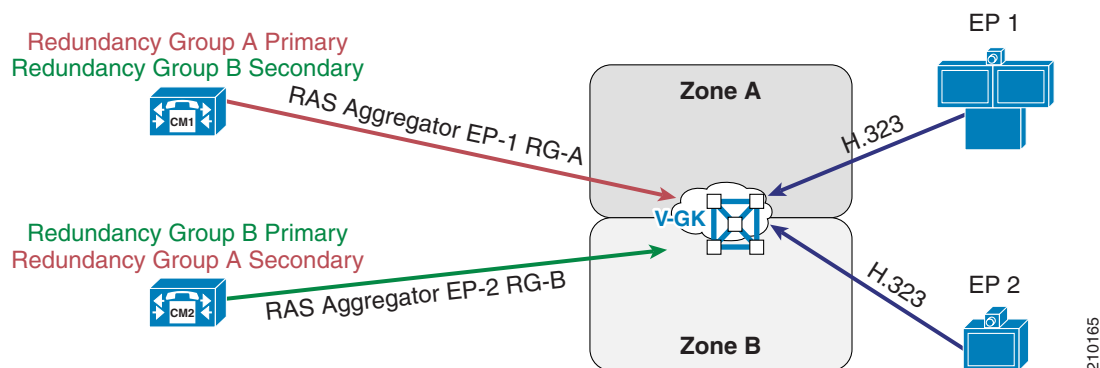
Figure 14-20 Two Zones and One Redundancy Group



Multiple Zones and Multiple Redundancy Groups

This scenario should be avoided if possible due to the complexity that it adds (see Figure 14-21). Improper configuration can easily lead to failed calls.

Figure 14-21 Two Zones and Two Redundancy Groups



Incorrectly Configured Redundancy Groups Cause Failures

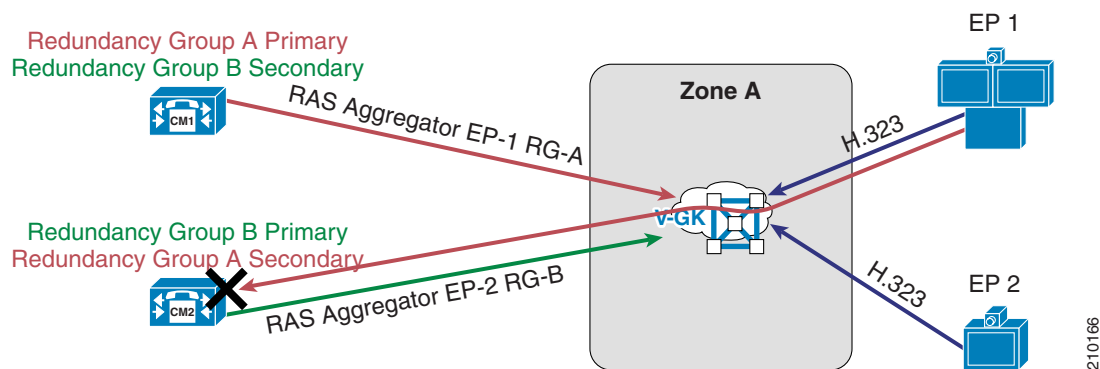


Note

All endpoints in a particular zone *must* use the same redundancy group.

Figure 14-22 shows an *incorrect* configuration of two different redundancy groups registered to the same endpoint zone. This configuration will cause call failures.

Figure 14-22 Incorrect Configuration of One Zone and Two Redundancy Groups



As shown in Figure 14-22, if two endpoints in the same zone use different redundancy groups with different primary Unified CMs, two RASAggregator Trunks will be registered to one zone from different nodes, thus causing the gatekeeper to load-balance between these two trunks. In this case, some calls will be sent across incorrect RASAggregator Trunks and will fail.

Video Call Flows

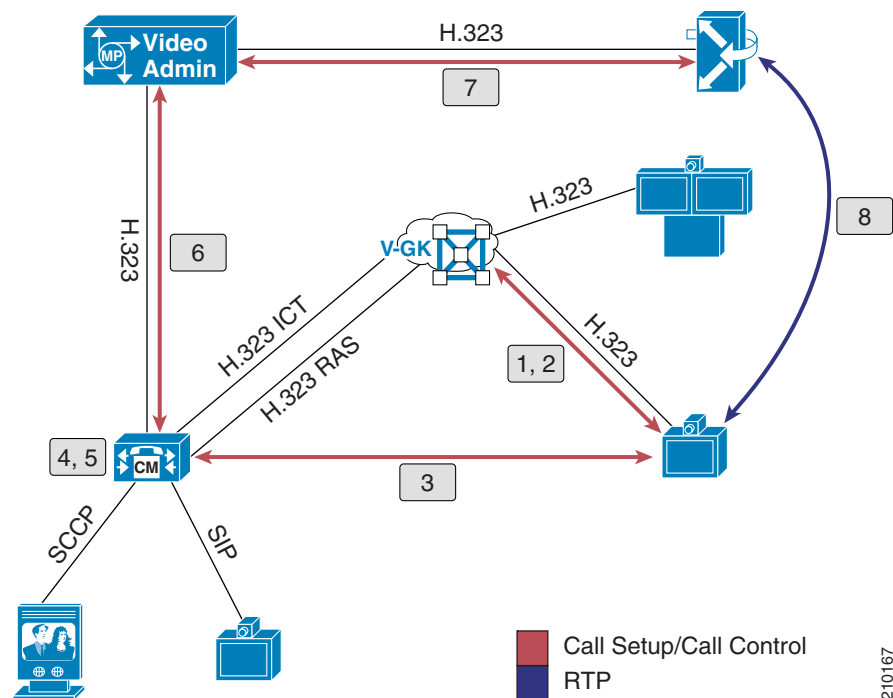
This section shows examples of call flows for inbound and outbound video calls with Cisco Unified MeetingPlace configured.

Inbound Call from an H.323 Video Endpoint

Figure 14-23 illustrates the flow for this call scenario, which involves the following sequence of events:

1. Admission Request (ARQ) sent from the H.323 video endpoint to the video gatekeeper.
2. Admission Confirm (ACF) sent from the video gatekeeper to the H.323 video endpoint, specifying Unified CM.
3. H.225 setup sent from the H.323 video endpoint to Unified CM.
4. Dial plan restrictions and digit manipulation performed by Unified CM.
5. Dialed number matches route pattern for H.323 gateway defined for Video Administration.
6. H.225 setup sent from Unified CM to Video Administration.
7. H.225 setup sent from Video Administration to the MCU.
8. Media is established between the MCU and the H.323 video endpoint.

Figure 14-23 Video Call Flow Into Video Administration

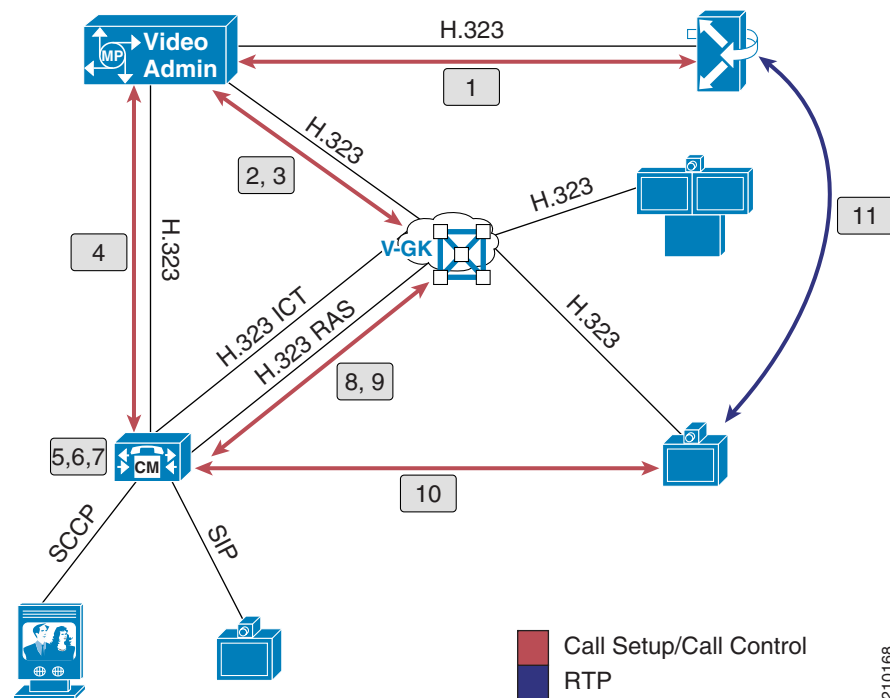


Outbound Call to an H.323 Video Endpoint

Figure 14-24 illustrates the flow for this call scenario, which involves the following sequence of events:

1. H.225 setup sent from Video Administration to the MCU.
2. Location Request (LRQ) sent from Video Administration to the video gatekeeper.
3. Matches to the intercluster trunk (ICT) and Location Confirm (LCF) sent from the video gatekeeper, specifying Unified CM.
4. H.225 setup sent from Video Administration to Unified CM.
5. Incoming H.225 connection matches H.323 gateway defined for Video Administration.
6. Dial plan restrictions and digit manipulation performed by Unified CM.
7. Dialed number matches H.323 endpoint controlled by Unified CM.
8. Admission Request (ARQ) sent from Unified CM to the video gatekeeper across associated RASAggregator Trunk.
9. Admission Confirm (ACF) sent from the video gatekeeper to Unified CM, specifying the H.323 video endpoint.
10. H.225 setup sent from Unified CM to the H.323 video endpoint.
11. Media is established between the MCU and the H.323 video endpoint.

Figure 14-24 Video Call Flow Out of Video Administration



Video Dial Plan

With the video component, video participants dial a prefix and meeting ID as a single dialed number. Because of the incorporated meeting ID, this dialed number is unique per conference and can be of fixed length with fixed-length meeting IDs or of variable length with variable-length meeting IDs.

Example:

*Video prefix*72

*Meeting ID*43221

*Dial-in number*7243221

The video dial-in number for the specific meeting is incorporated into the notifications sent to users, the scheduled meeting information in Cisco Unified MeetingPlace Web, and the Connect box within the web collaboration session. Dial-out from the web collaboration session is also available to connect video endpoints.

Fixed-length meeting IDs ensure a fixed-length dialed number. They are very easy to implement from the standpoint of Unified CM, and they have the least impact to users. The use of vanity IDs, however, is not available with this method.

Variable-length meeting IDs become a challenge due to the unknown length of the dialed number. This factor can cause the need for multiple route patterns in Unified CM to accommodate the various possibilities. Some scenarios (detailed in [Variable-Length Meeting IDs, page 14-29](#)) might require the user to wait for inter-digit timeout to occur before the call is placed.

Fixed-Length Meeting IDs

Restricting meeting IDs to one specific length eliminates issues associated with variable-length meeting IDs and allows for a simple, clean video dial plan implementation. Only one specific route pattern is needed in Unified CM, which will match and process all dialed numbers from video participants immediately.

Example:

*Video prefix*72

*Meeting ID*43221

*Route pattern*72XXXXX

Variable-Length Meeting IDs

The following example shows meeting IDs of differing lengths and how they impact route pattern implementation and operation.

Video prefix 72

Meeting ID #1 43221

Meeting ID #2 5256

Multiple exact route patterns:

*Pattern 1*72XXXX

*Pattern 2*72XXXXX

With this example, dialing into Meeting 1 (7243221) will connect immediately due to an exact match of Pattern 2, with no other possibilities. Dialing into Meeting 2 (725256) will cause the call to be delayed by inter-digit timeout due to the possibility of another, longer, match. Upon timeout, Pattern 1 is matched and the call is connected.

Universal route pattern:

Pattern 172!

With a single route pattern for all dialed numbers of any length, the inter-digit timeout must expire for every call placed. This is a simpler implementation because it does not depend on meeting ID length.

In addition, a second pattern can be added with a terminating # to give users the option of terminating the dialed number:

Pattern 172!

Pattern 272!#

If the dialed number is terminated by #, the inter-digit timeout is not a factor and the call is processed immediately. Calls not terminated by # will be subject to the inter-digit timeout.



Note

Inter-digit timeout is set to 15 seconds by default but can be changed if desired. Changing this value, however, impacts every call processed by that Unified CM cluster. To change the inter-digit timeout, adjust the Cluster Wide CCM Service Parameter, **T302 Timer**.

Video Redundancy

This section explains some of the limitations of video redundancy and provides best practices for achieving the maximum redundancy possible for Cisco Unified MeetingPlace Video. This section focuses specifically on the video component and does not repeat any of the redundancy information provided previously in this chapter.

Video Integration Component

The Video Integration component resides on all Cisco Unified MeetingPlace Web servers, but it can be active on only a single server, enabling that server as the web collaboration server for all conferences involving video. No automatic failover mechanisms exist for the Video Integration component. Failure of the supporting server would require manual intervention to configure and enable Video Integration on another server.

Video Administration

Video Administration is limited to server component redundancy, and it does not have any software-level redundancy. Only one Video Administration server may be implemented per Cisco Unified MeetingPlace deployment. The server on which Video Administration is deployed should contain redundant components to minimize risk of downtime.

Video Gatekeeper

Video Gatekeeper is no different than a standard gatekeeper with respect to redundancy abilities. In this environment, however, some special considerations must be taken into account. Of the following techniques for gatekeeper redundancy, only HSRP is fully compatible with this environment:

Gatekeeper HSRP

Redundancy via Hot Standby Router Protocol (HSRP) provides for outbound calling redundancy from Video Administration. If the primary gatekeeper fails, call resolution requests are automatically sent to the backup gatekeeper. Video Administration does not need to know about the backup gatekeeper or primary gatekeeper state to accomplish this failover.

Gatekeeper Clustering and Alternate Gatekeeper

With the gatekeeper clustering and alternate gatekeeper methods, an issue arises with outbound calls from Video Administration. Outbound calls from Video Administration are sent to the Video Gatekeeper without any registration to that gatekeeper. This prevents Video Administration from being aware of the gatekeeper state and alternate gatekeeper existence. Video Administration will continue blindly to send calls to the failed gatekeeper in this environment.

Unified CM

Unified CM has three vital connections in the Cisco Unified MeetingPlace Video environment:

- H.323 gateway defined for Video Administration
- Intercluster trunk (ICT) registered to Video Gatekeeper
- RASAggregator Trunk registered to Video Gatekeeper

For the first two, H.323 gateway and ICT, the redundancy technique is fairly simple and straightforward. Only one Video Administration may exist, and Video Gatekeeper should be implemented with HSRP for reasons stated in the redundancy section on [Video Gatekeeper, page 14-31](#).

Placement in a redundancy group (Unified CM group and device pool) for the H.323 gateway and ICT allows node-to-node failover within a cluster and is the extent of redundancy available.

With RASAggregator Trunks, you would configure a redundancy group also, but be careful to ensure proper configuration. Improper configuration can easily lead to call failure. For details, see [Zones and RASAggregator Redundancy Groups with H.323 Video Endpoints on Unified CM, page 14-24](#).



CHAPTER 15

Cisco Unified MeetingPlace Express

Last revised on: September 27, 2007

Cisco Unified MeetingPlace Express is a rich-media appliance for voice and web collaboration. This chapter addresses system-level design considerations. For detailed product information, refer to the product documentation available on

<http://www.cisco.com>

Deployment Models

This section describes design recommendations for integrating Cisco Unified MeetingPlace Express with the following major IP Telephony deployment models:

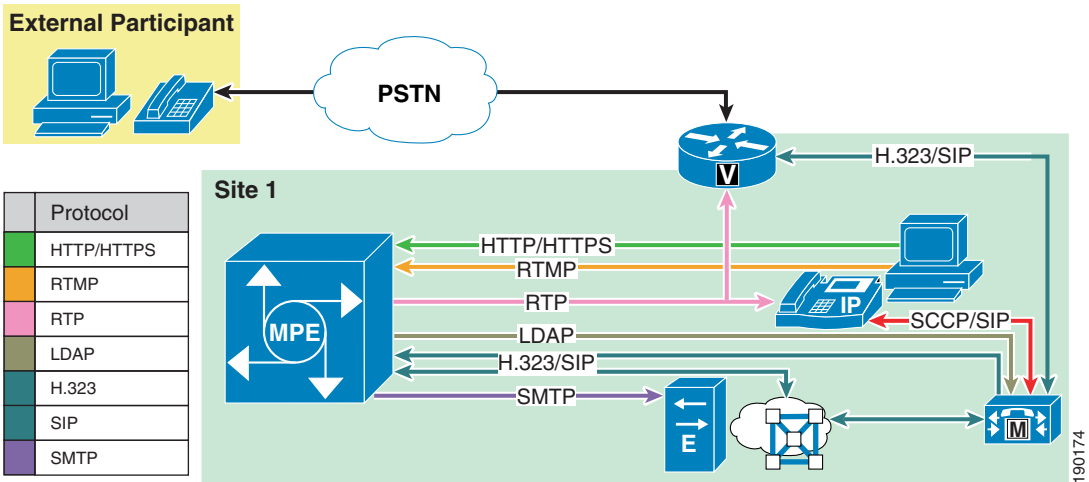
- [Single Site, page 15-2](#)
- [Multisite WAN with Centralized Call Processing, page 15-3](#)
- [Multisite WAN with Distributed Call Processing, page 15-4](#)
- [Clustering Over the WAN, page 15-5](#)

In addition to these deployment models, this section covers deployments based on a demilitarized zone (DMZ). The DMZ deployment is treated separately and is applicable to each of the IP Telephony deployment models listed above. (See [Security and DMZ Deployment, page 15-5](#).)

Single Site

In a single-site deployment, all call processing is local and all participants are local as well. Bandwidth is usually not a concern, as it would be in multisite configurations. In this model, the Cisco Unified MeetingPlace Express server connects to Cisco Unified Communications Manager (Unified CM) through a gatekeeper, H.323 trunk, or SIP trunk. (See [Figure 15-1](#).)

Figure 15-1 Single-Site Deployment

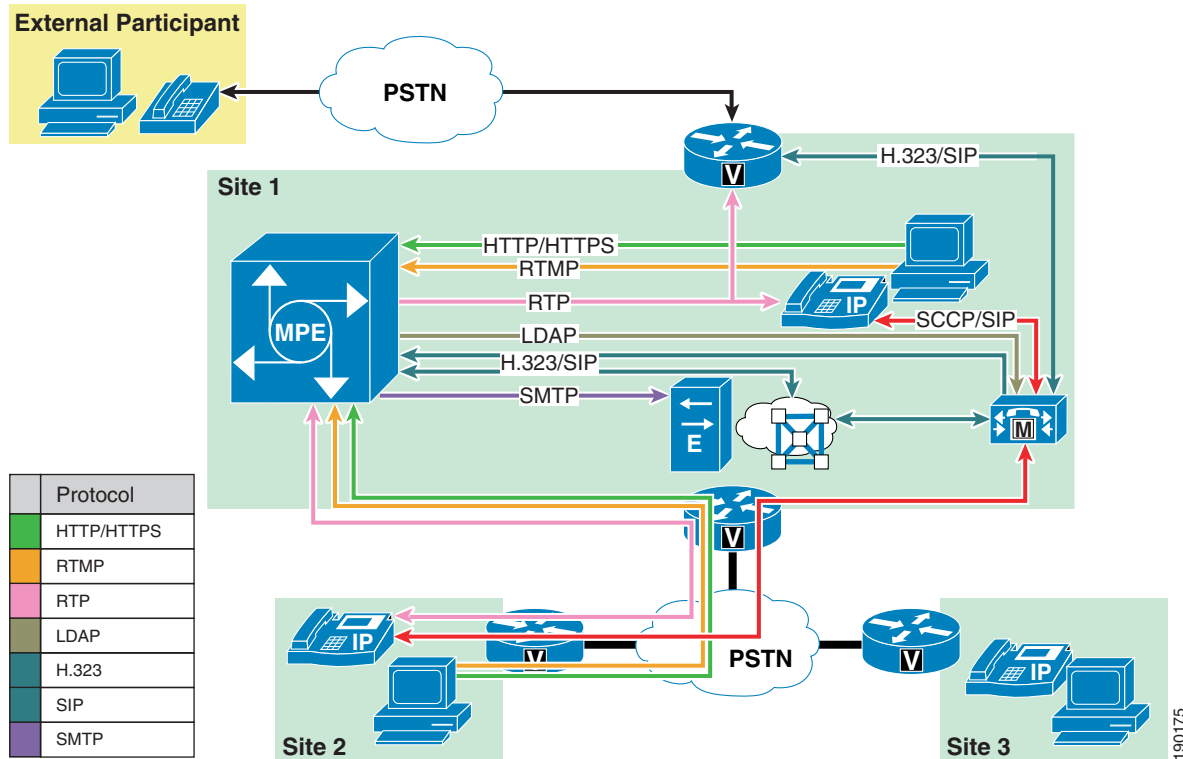


The general design considerations presented in the remaining sections of this chapter apply to this basic deployment model.

Multisite WAN with Centralized Call Processing

In a multisite WAN deployment with centralized call processing, the Cisco Unified MeetingPlace Express server is usually located in the central site, where all call processing occurs. (See [Figure 15-2](#).)

Figure 15-2 Multisite WAN Deployment with Centralized Call Processing



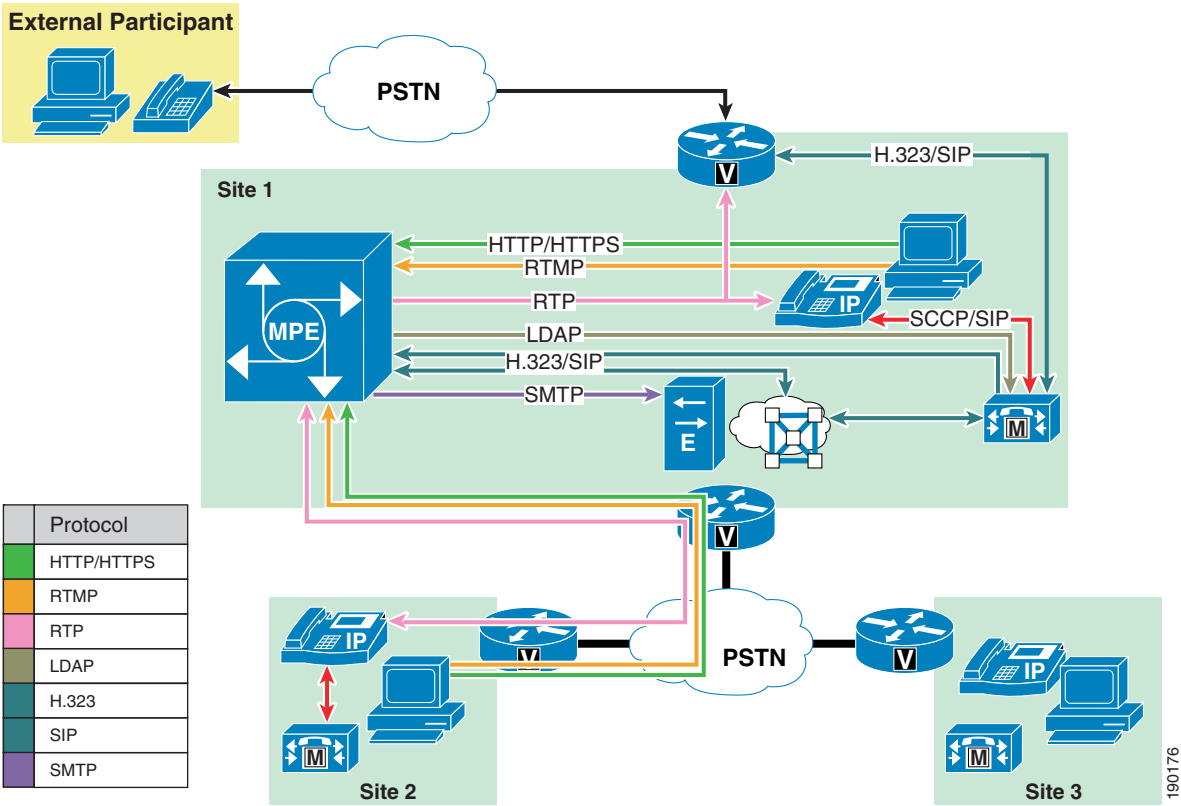
In this deployment, participants in the central site have no special considerations, but participants at remote sites require the following additional considerations:

- Cisco Unified MeetingPlace Express supports only G.711. Transcoders at the central site must be used for all remote calls other than G.711.
- Web collaboration requires significant bandwidth that must be provisioned. (See [Bandwidth Considerations for Web Applications and Screen Sharing](#), page 15-7.)
- The same QoS and call admission control mechanisms used for the existing IP Telephony deployment are required. For purposes of call admission control, the Cisco Unified MeetingPlace Express server should be treated as a telephony endpoint in the central site.

Multisite WAN with Distributed Call Processing

In a multisite WAN deployment with distributed call processing, the Cisco Unified MeetingPlace Express server is usually located in a single site local to Unified CM, similar to a multisite WAN deployment with centralized call processing. [Figure 15-3](#) depicts an example of a multisite WAN deployment with distributed call processing.

Figure 15-3 Multisite WAN Deployment with Distributed Call Processing



This deployment differs from a multisite WAN deployment with centralized call processing because the call processing is distributed to multiple sites. These sites may use Unified CM, Cisco Unified Communications Manager Express (Unified CME), or other methods of call processing. Because Cisco Unified MeetingPlace Express is a single server with no cascading, mirroring, or redundancy capabilities, a single active server at a single site would be typical. An additional server as a secondary active server may be deployed at a separate site but would not allow for a true failover environment. See [Redundancy, page 15-12](#), for more information.

To have a multisite redundant and distributed collaboration system, consider Cisco Unified MeetingPlace as an option. See the chapter on [Cisco Unified MeetingPlace Integration, page 14-1](#), for more information.

190176

Design Considerations

In this deployment, participants local to the Cisco Unified MeetingPlace Express server have no special considerations, but participants remote from the Cisco Unified MeetingPlace Express server require the following additional considerations:

- Cisco Unified MeetingPlace Express supports only G.711. Transcoders at the site containing the Cisco Unified MeetingPlace Express server must be used for all remote calls other than G.711.
- Web collaboration requires significant bandwidth that must be provisioned. (See [Bandwidth Considerations for Web Applications and Screen Sharing, page 15-7](#).)
- The same QoS and call admission control mechanisms used for the existing IP Telephony deployment are required. For purposes of call admission control, the Cisco Unified MeetingPlace Express server should be treated as a telephony endpoint in its local site.
- The Unified CM cluster at the remote site (Site 2 in [Figure 15-3](#)) can directly route calls to Cisco Unified MeetingPlace Express located at the central site (Site 1) by defining an H.323 gateway or SIP trunk directly on the remote cluster. This practice is not recommended due to its potential impact on call admission control. Routing calls directly to MeetingPlace and not through the central cluster via an intercluster trunk or through gatekeeper can cause calls not to be accounted for in call admission control measurements.

Clustering Over the WAN

In a deployment with clustering over the WAN, the Unified CM cluster is split across one or more locations separated by high-capacity and high-speed WAN or MAN links. Cisco Unified MeetingPlace Express does not have any server-to-server redundancy and cannot fully utilize the redundant nature of this model. An additional server as a secondary active server may be deployed at the alternate site but would not allow for a true failover environment. See [Redundancy, page 15-12](#), for more information.

Placement and design considerations would be identical to the previous deployment model, [Multisite WAN with Distributed Call Processing, page 15-4](#).

To have a multisite redundant and distributed collaboration system, consider Cisco Unified MeetingPlace as an option. See the chapter on [Cisco Unified MeetingPlace Integration, page 14-1](#), for more information.

Security and DMZ Deployment

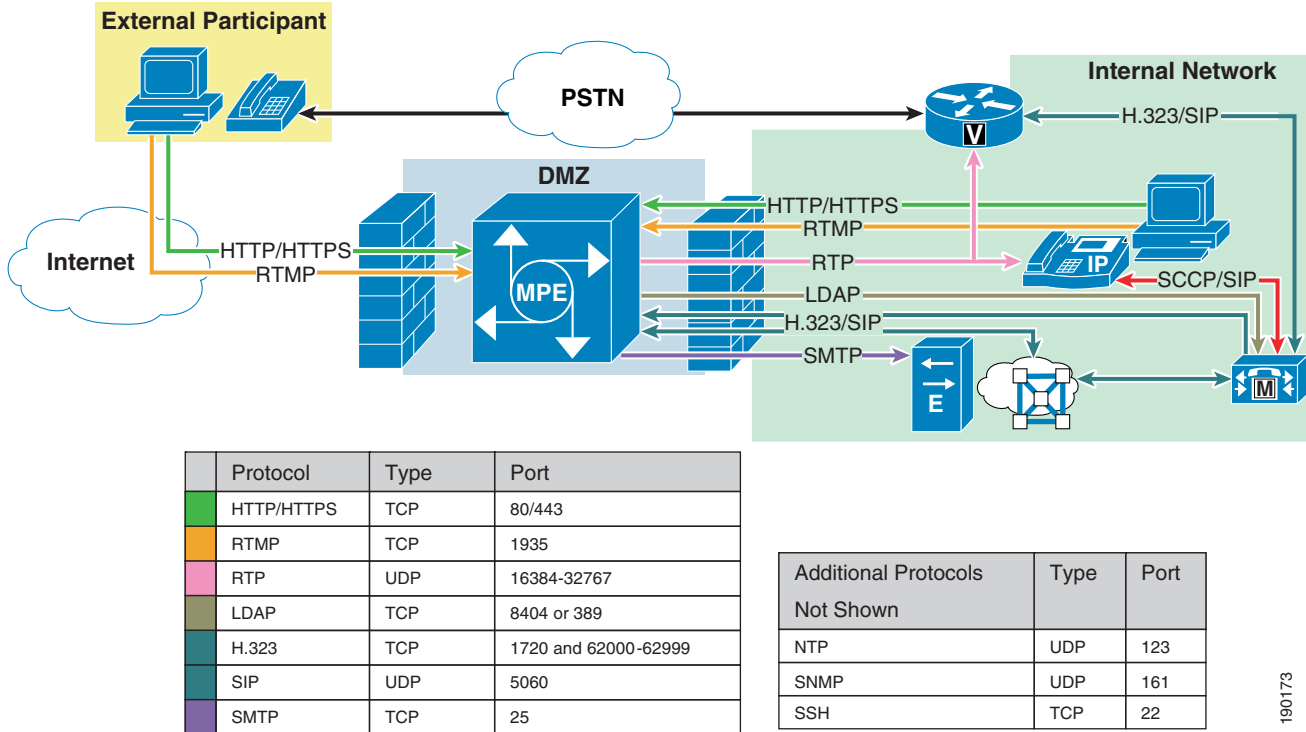
Cisco Unified MeetingPlace Express may be placed in a demilitarized zone (DMZ), although there is no mechanism to allow internal and external applications because it is a single server. All internal and external web collaboration and telephony calls must reach the Cisco Unified MeetingPlace Express server in the DMZ.

External telephony calls would come into a gateway inside the firewall and be treated the same as internal calls being sent into the DMZ to connect to the Cisco Unified MeetingPlace Express server.

External web collaboration would come into the DMZ from outside the firewall, and internal from inside.

[Figure 15-4](#) illustrates communication between the Cisco Unified MeetingPlace Express server in the DMZ and the internal and external networks.

Figure 15-4 DMZ Deployment



Design Considerations for DMZ Deployment

- If port 1935 is not configured in the firewall access control list (ACL), Real Time Messaging Protocol (RTMP) traffic will tunnel over port 80 with a small performance hit.
- If Network Time Protocol (NTP) is to be used, connectivity internally or externally to an NTP source over port 123 is needed.
- Secure Shell (SSH) access over port 22 to the server from the internal network might be needed. Cisco Technical Assistance Center (TAC) might require SSH access to support Cisco Unified MeetingPlace Express properly.
- Real-Time Transport Protocol (RTP) and Real-Time Transport Control Protocol (RTCP) ports are required for voice termination on the Cisco Unified MeetingPlace Express server. With Cisco firewall devices, RTP ports will be opened dynamically to allow the correct traffic flows through the firewall based on the H.323 or SIP call setup. Therefore, it is not necessary to open a range of ports on the firewall.

Call Admission Control, Bandwidth, and QoS

This section describes important call admission control, bandwidth, and QoS considerations for Cisco Unified MeetingPlace Express deployments.

Call Admission Control

Cisco Unified MeetingPlace Express should be treated as a local voice gateway for purposes of call admission control. For additional guidance on call admission control, see the chapter on [Call Admission Control, page 9-1](#).

Bandwidth Considerations for Web Applications and Screen Sharing

Screen sharing consumes large amounts of bandwidth, on the order of 2 Mbps per user in the worst case. Because of the excessive bandwidth and bursty nature of screen sharing, the following design considerations must be observed.

Design Considerations

Use of 100 Mbps for LAN connectivity will limit the number of concurrent web collaboration sessions. When possible, use 1 Gbps connections.

Users at remote sites that cause web collaboration to traverse a WAN will require special consideration. The client flash session bandwidth or room bandwidth setting for these users should be lowered, reducing the load across the WAN. Because web collaboration data is delivered unicast, the largest burst of data should be multiplied by the number of clients at a remote site. For example, assume a remote site has 100 users, 10 of which are on a web collaboration session at any one time. If bursts of 2 Mbps occur in the data from the remote server to each user, 20 Mbps bursts can be experienced across the WAN connection.

When WAN links become congested from excessive web collaboration data or other sources, the degradation of all traffic is compounded by packet loss, retransmissions, and increased latency. Sustained congestion will have a sustained degrading impact on all remote collaboration sessions.

The following settings on the client web collaboration sessions control the rate at which the participant receives data as well as the rate at which the presenter sends data:

- **Modem** — Bandwidth limited to 64 kbps
- **DSL** — Bandwidth limited to 640 kbps
- **LAN** — Bandwidth up to 3,000 kbps (Bandwidth bursts above 3,000 kbps are possible if high-resolution images or photos are shared. Sharing normal to complex presentations or document should not generate bursts above 3,000 kbps unless large complex images are embedded.)

Bandwidth settings are not automatically adjusted when congestion occurs; they must be adjusted manually. Bandwidth settings default to LAN and must be set at the initialization of each collaboration session. A new session is set to LAN regardless of previous settings.

Bandwidth settings can be adjusted in the following locations in the web collaboration client screen:

- My Connection Speed
 - Limits the speed data is delivered to an individual user from the Cisco Unified MeetingPlace Express server.

- Limits the speed data is sent from the presenter to the Cisco Unified MeetingPlace Express server.
- Presenter speed does not limit the rate at which participants receive data.
- Optimize Room Bandwidth
 - Limits the rate data is delivered to all users.
 - Currently has data burst issues and might impact WAN bandwidth utilization.

**Note**

The setting for My Connection Speed on the client web interface for the presenter and participant are independent of each other. Presenter speed does not limit the rate at which participants receive data. If the presenter is set to send data to Cisco Unified MeetingPlace Express at Modem speed (70 kbps) and a participant is set to receive data at LAN speed (3,000 kbps), the participant will still receive data up to 3,000 kbps.

**Note**

While the Optimize Room Bandwidth setting does reduce the rate data is sent to users, bursts in the delivery of that data still occur and can congest WAN links. Changing My Connection Speed on the participant's client eliminates bursts and delivers data to the client at a steady rate. The burst issues with the Optimize Room Bandwidth setting will be addressed in future releases.

The client resolution setting also impacts the bandwidth used by limiting the bits per image. A meeting room set at 640x480 pixels will typically generate less than one-third of the traffic compared to a setting of 1280x1024 pixels.

Design Recommendations Summary

- Connect both Cisco Unified MeetingPlace Express server interfaces to the LAN with 1 Gbps connections.
- For meetings involving remote users, have the remote users set their My Connection Speed to **DSL**. If congestion issues occur, lower the setting to **Modem**, or lower the screen resolution settings.

QoS

Quality of Service (QoS) considerations encompass the following topics.

Traffic Classification or Markings

Traffic classification or markings from the Cisco Unified MeetingPlace Express server are as follows:

- SIP, H.323, and call control — Not marked
- Voice media (RTP) — Marked EF (DSCP 46)
- Web traffic (HTTP and HTTPS) — Not marked
- Flash web collaboration (RTMP) — Not marked

Voice and Call Control Traffic

Voice and call control traffic should be classified with the standard classifications as described in the chapter on [Network Infrastructure, page 3-1](#). If the connection between the Cisco Unified MeetingPlace Express server and the gatekeeper or Unified CM extends over a WAN connection, call control traffic should be re-marked to CS3 (DSCP 24).

Web Interface Traffic

No prioritization is recommended for the main web interface. Web traffic to the Cisco Unified MeetingPlace Express server should be treated the same as traffic to other internal web application servers.

Flash Web Collaboration Traffic

Because of the large amounts of bandwidth consumed and the bursty nature of this data, as described in the previous section, prioritization of web collaboration traffic across the WAN is difficult. If you attempt to prioritize the web collaboration traffic, classify it separately and lower than other prioritized data.

Directory Integration

Cisco Unified MeetingPlace Express integrates directly with Cisco Unified CM 4.x. For third-party directory integration, a plug-in is required. Cisco Unified MeetingPlace Express supports Microsoft Active Directory, Sun One LDAP, and Netscape/iPlanet LDAP.

Cisco Unified MeetingPlace Express also integrates directly with Cisco Unified CM 5.x via Cisco AVVID XML Layer (AXL) Simple Object Access Protocol (SOAP). Third-party directory integration in a Cisco Unified CM 5.x environment requires Cisco Unified MeetingPlace Express to point to Unified CM only.

For more information, refer to the *Administrator's Configuration and Maintenance Guide for Cisco MeetingPlace Express*, available at

<http://www.cisco.com>

H.323 and SIP Direct Integration

Integration with Unified CM can be accomplished by several methods. This section examines two of those methods, H.323 and SIP direct integration, while the third (H.323 via gatekeeper) is examined in the next section.

H.323 Gateway

The simplest integration option is to define Cisco Unified MeetingPlace Express as an H.323 gateway in Unified CM. You must also configure route patterns in Unified CM that point to the defined gateway. Cisco Unified CM 4.1, 4.2, and 5.0, as well as Cisco Unified Communications Manager Express (Unified CME), all support this option.

SIP Trunk

Using a SIP trunk to connect to Cisco Unified MeetingPlace Express has some conditions and version dependencies. The following current combinations are supported between Unified CM and Cisco Unified MeetingPlace Express.

Cisco Unified MeetingPlace Express 1.1.1 and Cisco Unified CM 5.x

With this combination, a SIP trunk can be defined directly from Unified CM to Cisco Unified MeetingPlace Express. The following design rules apply to this integration:

- A separate SIP Security Profile must be created and associated with the SIP trunk to Cisco Unified MeetingPlace Express. This security profile must have the transport set to UDP. The default setting is TCP, which will not work with Cisco Unified MeetingPlace Express.
- The SIP trunk configuration must have **MTP Required** checked and must have MTP resources available. MTPs will be invoked for each call to Cisco Unified MeetingPlace Express for the entire duration of the call. This requirement does not apply to Cisco Unified MeetingPlace Express 1.1.2 and higher.

Cisco Unified MeetingPlace Express 1.1.2 and Cisco Unified CM 5.x

With this combination, a SIP trunk can be defined directly from Unified CM to Cisco Unified MeetingPlace Express. The following design rules apply to this integration:

- Cisco Unified CM 5.0.4 or higher is required for Cisco Unified MeetingPlace Express 1.1.2. Prior Unified CM 5.0 versions are not supported.
- A separate SIP Security Profile must be created and associated with the SIP trunk to Cisco Unified MeetingPlace Express. This security profile must have the transport set to UDP. The default setting is TCP, which will not work with Cisco Unified MeetingPlace Express.
- The SIP trunk does not require MTPs because Cisco Unified MeetingPlace Express 1.1.2 supports enhanced SIP capabilities. Do not check the MTP Required option on the SIP trunk.
- Ensure that MTPs are available in the associated media resource group list. Some endpoints may cause an MTP to be invoked dynamically when completing a call to Cisco Unified MeetingPlace Express.

Cisco Unified MeetingPlace Express 1.1.x and Cisco Unified CM 4.1 and 4.2

With this combination, a SIP trunk can be defined directly from Unified CM to Cisco Unified MeetingPlace Express. The following design rules apply to this integration:

- Outgoing transport must be set to UDP. The default setting is TCP, which will not work with Cisco Unified MeetingPlace Express.
- The SIP trunk configuration must have **MTP Required** checked and must have MTP resources available. MTPs will be invoked for each call to Cisco Unified MeetingPlace Express for the entire duration of the call. This requirement will not be removed in future releases of Cisco Unified MeetingPlace Express due to the requirements of Cisco Unified CM 4.x.

Gatekeeper Integration

Cisco Unified MeetingPlace Express can be integrated into a gatekeeper environment in accordance with the design considerations presented in this section. Integration as an H.323 gateway or via SIP trunk directly from Unified CM is still an option in a gatekeeper environment. The gatekeeper registration method is as a gateway, using a single E.164 address with the registration. The following design considerations must be taken into account when designing for use in a gatekeeper environment.

Design Considerations



Note

Cisco Unified MeetingPlace Express will not register into a specific zone. The default, or first, local zone is always used when multiple local zones exist.

To force Cisco Unified MeetingPlace Express to use a specific local zone, use the **no zone** command on all zones where you do not want Cisco Unified MeetingPlace Express to register. For example, the following commands force Cisco Unified MeetingPlace Express (with address 10.20.110.50) to register to testzone2:

```
gatekeeper
zone local mp2-gk1 mp2.com 10.20.105.50
zone local testzone2 mp2.com
no zone subnet mp2-gk1 10.20.110.50/32 enable
```



Note

The Cisco Unified MeetingPlace Express direct meeting dial-in feature requires additional gatekeeper configuration.

Cisco Unified MeetingPlace Express registers as a single E.164 address, so extensions for direct meeting dial-in cannot be reached through a gatekeeper without adding manual entries to the gatekeeper by one of the methods shown in the following examples.

Option 1: Use Zone Prefix to Route (Recommended)

This example routes extensions 1000 through 1009 to the Cisco Unified MeetingPlace Express server.

```
gatekeeper
zone local mp2-gk1 mp2.com 10.20.105.50
zone prefix mp2-gk1 100.
gw-type-prefix 1#* default-technology gw ipaddr 10.20.110.50 1720
```

Option 2: Use Static E.164 Addresses (Not Recommended)

```
gatekeeper
alias static 10.20.110.50 1720 gkid mp2-gk1 gateway voip ras 10.20.110.50 62675 e164
1005 e164 1008 e164 1007 e164 1006
```

```
show gatekeeper endpoints
```

CallSignalAddr	Port	RASSignalAddr	Port	Zone Name	Type	Flags
-----	----	-----	----	-----	----	-----
10.20.110.50	1720	10.20.110.50	62675	mp2-gk1	UNKN-GW	S
E164-ID: 1000						
H323-ID: MeetingPlace						

```
E164-ID: 1005 (static)
E164-ID: 1008 (static)
E164-ID: 1007 (static)
E164-ID: 1006 (static)
```

Redundancy

This section describes the following types of redundancy:

- [Cisco Unified MeetingPlace Express Server Redundancy, page 15-12](#)
- [Physical Network Redundancy, page 15-12](#)
- [Redundancy Using a Gatekeeper, page 15-12](#)
- [Redundancy Using H.323 Gateway Integration, page 15-13](#)
- [Redundancy Using SIP Trunk Integration, page 15-13](#)

Cisco Unified MeetingPlace Express Server Redundancy

Cisco Unified MeetingPlace Express is a standalone server with no redundancy built in, except for some server components which are themselves redundant. Mirrored drives, dual power supplies, redundant fans, and so forth, give a high level of internal server redundancy and availability. Other methods outside of the server may be incorporated to enhance redundancy characteristics for calls sent from or delivered to Cisco Unified MeetingPlace Express.

Although Cisco Unified MeetingPlace Express is a standalone server, multiple Cisco Unified MeetingPlace Express servers can be deployed using the same Unified CM directory. This configuration enables users to log into an alternate server, via the common directory, and quickly reschedule meetings if the server they were originally using should fail. Two or more fully licensed Cisco Unified MeetingPlace Express servers are needed for this option.

Physical Network Redundancy

Cisco Unified MeetingPlace Express has two Ethernet ports that are utilized for different purposes, one for call control, web interface, and media, and the other for web collaboration. While traffic rerouting should occur, these ports cannot be relied upon for redundancy. If one port on the server or switch (or the cable between them) fails, the server might not operate properly.

Redundancy Using a Gatekeeper

Inbound Calls

The only other redundancy with respect to inbound calls is gatekeeper redundancy, which would apply to outbound calls as well. You can implement gatekeeper redundancy by either of the following methods:

- Alternate gatekeeper

In this method, gatekeepers are set up in a redundant gatekeeper cluster. Cisco Unified MeetingPlace Express registers with its defined gatekeeper. Upon registration, the gatekeeper informs Cisco Unified MeetingPlace Express of an alternate gatekeeper in the cluster for use in the event of a failure.

- Gatekeeper Hot Standby Router Protocol (HSRP)

In this method, two gatekeepers share a single HSRP IP address. In the case of failure, Cisco Unified MeetingPlace Express registers to the secondary gatekeeper using the same gatekeeper IP address.

**Note**

Because the alternate gatekeeper information is delivered during registration and is not permanently stored in Cisco Unified MeetingPlace Express, booting or rebooting of the server when the primary gatekeeper is inaccessible will prevent Cisco Unified MeetingPlace Express from registering with the alternate gatekeeper.

Detailed information about gatekeeper redundancy can be found in the section on [Gatekeeper Redundancy, page 8-21](#).

Outbound Calls

Multiple Unified CM servers in a single cluster can register to a gatekeeper for redundancy. The gatekeeper will choose an active Unified CM to complete an outdial from Cisco Unified MeetingPlace Express.

Gatekeeper redundancy as described in the previous section ([Inbound Calls, page 15-12](#)) applies to outbound calls as well.

Redundancy Using H.323 Gateway Integration

Inbound Calls

If you integrate Cisco Unified MeetingPlace Express as an H.323 gateway, inbound calls will be sent from an available Unified CM server within a cluster. If the server fails, inbound calls will automatically be sent from another server in the cluster.

Outbound Calls

Cisco Unified MeetingPlace Express allows definition of multiple outbound Unified CMs for outbound call resolution. Should one fail, the next one is used.

Redundancy Using SIP Trunk Integration

Inbound Calls

If you integrate Cisco Unified MeetingPlace Express using a SIP trunk, inbound calls will be sent from an available Unified CM server within a cluster. If the server fails, inbound calls will automatically be sent from another server in the cluster.

Outbound Calls

Cisco Unified MeetingPlace Express allows definition of multiple outbound Unified CMs for outbound call resolution via SIP trunk. Should one fail, the next one is used.

Other Important Design Considerations

This section lists other important design considerations not covered previously in this chapter.

Network Connectivity

- Cisco Unified MeetingPlace Express uses two network interface cards (NICs) for connectivity. At this time, both NICs must be placed in the same subnet with a common default gateway. Connectivity issues will arise if the NICs are placed in separate subnets.
- Changing the IP addresses of the Cisco Unified MeetingPlace Express server must be done through the **net** command by connecting to the server via SSH and using the **mpxadmin** login. Changing the IP address through other means will not properly change the configuration of Cisco Unified MeetingPlace Express and will cause configuration issues.

IP Telephony Gateways

- Gateways used with Cisco Unified MeetingPlace Express must be configured with the standard recommendations set forth in the chapter on [Gateways, page 4-1](#).
- The main gateway issue that may impact Cisco Unified MeetingPlace Express is echo cancellation. Cisco recommends increasing the tail allocation on the gateway to 128 ms if echo issues are encountered.



CHAPTER 16

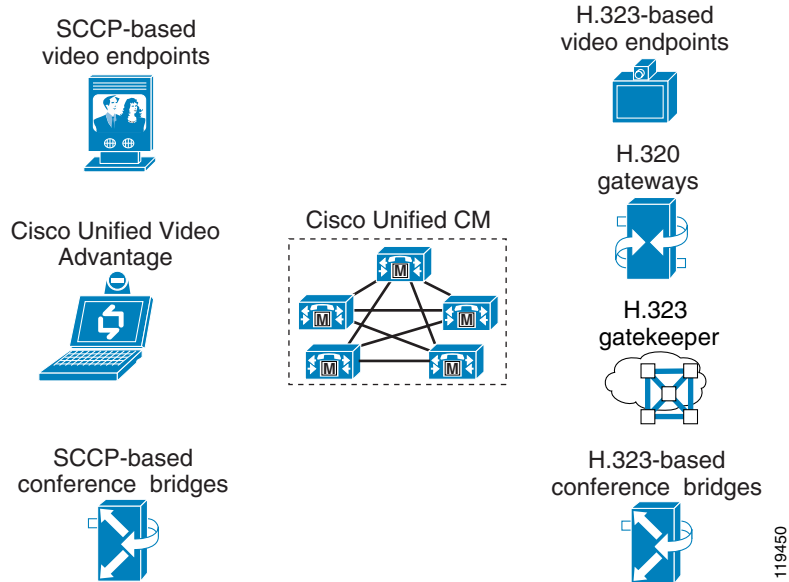
IP Video Telephony

Last revised on: September 27, 2007

Cisco introduced its IP Video Telephony solution in Cisco Unified Communications Manager (Unified CM, formerly Cisco Unified CallManager) Release 4.0. Video is fully integrated into Unified CM, and there are also many video endpoints available from Cisco and its strategic partners. Cisco Unified Video Advantage is just as easy to deploy, manage, and use as a Cisco Unified IP Phone.

IP Video Telephony Solution Components

The Cisco IP Video Telephony solution consists of Cisco Unified Communications Manager (Unified CM); Cisco Unified Videoconferencing 3500 Series Multipoint Control Units (MCUs) for H.323, Session Initiation Protocol (SIP), and Skinny Client Control Protocol (SCCP) conference calls; Cisco Unified Videoconferencing 3500 Series H.320 Gateways; Cisco IOS H.323 Gatekeeper; Cisco Unified Video Advantage; Cisco IP Video Phone 7985; Sony and Tandberg SCCP endpoint solutions; and the existing range of H.323 or SIP-compliant products from partners such as Polycom, Tandberg, Sony, and others. (See [Figure 16-1](#).)

Figure 16-1 IP Video Telephony Components

Video Enhancements in Cisco Unified CM 5.x

Cisco Unified CM Release 4.0 was the first software release to integrate video capability with IP Telephony. Cisco Unified CM Release 5.x (Release 5.0 and higher) provides additional enhancements in the following areas to support IP Video Telephony:

- [Protocols, page 16-2](#)
- [MTP and Transcoder Support for Video Calls, page 16-4](#)
- [Topology-Aware Locations, page 16-4](#)

Protocols

Cisco Unified CM Release 5.0 introduced video support for SIP trunks and SIP endpoints. In previous versions of Unified CM, the SIP support was limited to trunking and did not support video.

Unified CM supports a large number of protocols. Any device can call any other device, but video is supported only on SCCP, H.323, and SIP devices. Specifically, video is not supported in the following protocols in Cisco Unified CM Release 5.x:

- Computer Telephony Integration (CTI) applications (TAPI and JTAPI)
- Media Gateway Control Protocol (MGCP)

Therefore, Unified CM currently supports the types of calls listed in [Table 16-1](#).

Table 16-1 *Types of Calls Supported in Unified CM Release 5.x*

Calling Device Type	Called Device Type				
	SCCP	H.323	MGCP	TAPI/JTAPI	SIP
SCCP	Audio and video	Audio and video	Audio only	Audio only	Audio and video
H.323	Audio and video	Audio and video	Audio only	Audio only	Audio and video
MGCP	Audio only	Audio only	Audio only	Audio only	Audio only
TAPI/JTAPI	Audio only	Audio only	Audio only	Audio only	Audio only
SIP	Audio and video	Audio and video	Audio only	Audio only	Audio and video

Table 16-2 lists the audio and video algorithms and protocols currently supported in Unified CM.

Table 16-2 *Capabilities Supported in Unified CM Release 5.x*

H.323	SCCP	SIP
H.261	H.261	H.261
H.263, H.263+	H.263, H.263+	H.263, H.263+
H.264	H.264	H.264
Cisco VT Camera Wideband Video Codec (H.323 intercluster trunk only)	Cisco VT Camera Wideband Video Codec	
G.711 A-law and mu-law	G.711 A-law and mu-law	G.711 A-law and mu-law
G.723.1	G.723.1	G.723.1
G.728	G.728	G.728
G.729, G.729a, G.729b, and G.729ab	G.729, G.729a, G.729b, and G.729ab	G.729, G.729a, G.729b, and G.729ab
G.722	G.722	G.722
G.722.1		
H.224 far-end camera control (supported by Unified CM but not by all endpoints); No protocol interworking	H.224 far-end camera control (supported by Unified CM but not by all endpoints); No protocol interworking	H.224 far-end camera control (supported by Unified CM but not by all endpoints); No protocol interworking
Out-of-band DTMF (H.245 alphanumeric) RFC2833 AVT Tones (only for H.323 intercluster trunk to SIP calls)	Out-of-band DTMF RFC2833 AVT Tones	RFC2833 AVT Tones Unsolicited SIP Notify KPML
	Cisco Wideband Audio	

MTP and Transcoder Support for Video Calls

Beginning with Cisco Unified CM Release 5.0, the Cisco IOS Enhanced Media Termination Points (MTPs) support video calls through the use of a pass-through codec. The pass-through codec allows an MTP or transcoder that has been dynamically inserted by Unified CM to terminate the video stream of a video call without having to decode the RTP stream.

Topology-Aware Locations

Beginning with Cisco Unified CM Release 5.0, there are now two methods to limit the amount of bandwidth available for calls between locations. Cisco Unified CM 4.0 introduced support for video calls in locations. More specifically, the location option in Cisco Unified CM 4.0 and 4.1 defines the aggregate bandwidth allowed for all calls between that location and all other locations. This aggregate bandwidth value maps well to traditional hub-and-spoke network topologies. Cisco Unified CM Release 5.x includes a new option to use topology-aware locations based on Resource Reservation Protocol (RSVP) to determine if enough bandwidth is available along the path between two sites. RSVP enables per-hop checking that accommodates complex topologies while supporting the separation of audio bandwidth and video bandwidth via the use of RSVP Application IDs.



Note

Static locations and RSVP-based locations use different models to separate voice and video calls. For more details, see [Call Admission Control, page 9-1](#).

RSVP-based locations introduce the concept of an RSVP policy. While there are numerous policy options, there are two main categories:

- The RSVP reservation for the video stream is mandatory for the call to complete. The call will either fail (with busy tone played and the "Bandwidth Unavailable" message displayed to the user) or Automated Alternate Routing (AAR) will try to reroute the call.
- The RSVP reservation for the video stream is desired.

First, the audio codec and the video bandwidth set in the region define the maximum speed (bit rate) of the video call. An RSVP reservation is sent from the Cisco RSVP Agent for the audio and video stream of the call using the maximum bit rate as the reservation requirement. If the RSVP reservation for the video stream fails, Unified CM checks the setting of RSVP policy to determine how to handle that call. If the policy for the audio stream is optional, the call will continue as audio-only. Otherwise, if the RSVP policy for the audio stream is mandatory, the call will continue as audio-only unless the audio stream also fails to get an RSVP Reservation. In that case, the call will either fail (with busy tone played and the "Bandwidth Unavailable" message displayed to the user) or Automated Alternate Routing (AAR) will try to reroute the call. (For more details on topology-aware locations, see [Call Admission Control, page 9-1](#).)



Note

If the video reservation fails while using a video-desired policy, the call will complete as audio-only. However, the user will not have any visual or audible feedback to indicate why the video failed.

Administration Considerations

This section discusses the following configuration elements in Unified CM Administration that pertain to Video Telephony:

- [Regions, page 16-5](#)
- [Locations, page 16-8](#)
- [Retry Video Call as Audio, page 16-8](#)
- [Wait for Far-End to Send TCS, page 16-10](#)

Regions

When configuring a region, you set two fields in Unified CM Administration: the Audio Codec and the Video Bandwidth. The audio setting specifies a codec type, while the video setting specifies the amount of bandwidth you want to allow. However, even though the notation is different, the Audio Codec and Video Bandwidth fields actually perform similar functions. The Audio Codec field defines the maximum bit-rate allowed for audio-only calls as well as for the audio channel in video calls. For instance, if you set the Audio Codec for a region to G.711, Unified CM allocates 64 kbps as the maximum bandwidth allowed for the audio channel for that region. In this case, Unified CM will permit calls using either G.711, G.722, G.728, or G.729. However, if you set the Audio Codec to G.729, Unified CM allocates only 8 kbps as the maximum amount of bandwidth allowed for the audio channel, and it will permit calls using only G.729 because G.728, G.711, and G.722 all take more than 8 kbps.

**Note**

If both endpoints support G.711 and G.722, then G.722 will be negotiated by default.

**Note**

The Audio Codec setting also applies to the audio channel of video calls.

The Video Bandwidth field defines the maximum bit-rate allowed for the video channel of the call. However, for historical continuity with the practices used in traditional videoconferencing products, the value used in this field also includes the bandwidth of the audio channel. For instance, if you want to allow calls at 384 kbps using G.711 audio, you would set the Video Bandwidth field to 384 kbps and not 320 kbps.

In summary, the Audio Codec field defines the maximum bit-rate used for audio-only calls and for the audio channel of video calls, while the Video Bandwidth field defines the maximum bit-rate allowed for video calls and should include the audio portion of the call.

Choosing the correct audio codec bandwidth limit is very important because each device supports only certain audio codecs, as shown in [Table 16-3](#). (For the most recent list of codecs supported by a particular endpoint, refer to the product documentation for that endpoint.)

Table 16-3 *Types of Audio Codecs Supported by Endpoint Devices*

Codec Type	Cisco 7900 Series IP Phones	SCCP Sony and Tandberg Endpoints	Typical H.323 or SIP Endpoints	Cisco Unified Videoconferencing 3500 Series Gateways	Cisco Unified Videoconferencing 3500 Series MCUs
G.729	Yes	Yes, depending on the model	No	No	Yes, depending on the model
G.728	No	Yes, depending on the model	Yes	Yes (with transcoder)	Yes, depending on the model
G.711	Yes	Yes	Yes	Yes	Yes
G.722	Yes, depending on the model	Yes	Yes	Yes (with transcoder)	Yes, depending on the model
Cisco Wideband Audio	Yes, depending on the model	No	No	No	No

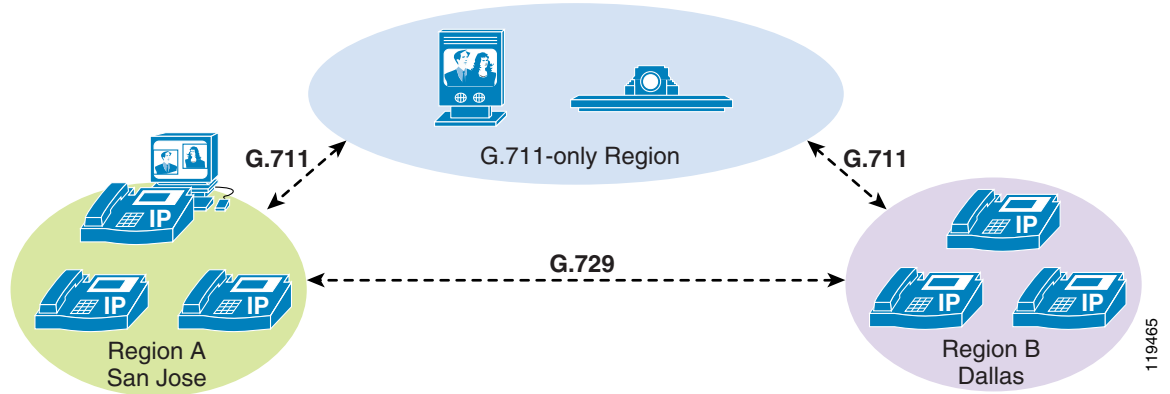
As [Table 16-3](#) indicates, if you set the region to G.729, not all videoconferencing devices are able to support this type of codec. For example, calls between a Cisco Unified Video Advantage endpoint and a Tandberg endpoint would fail, or Unified CM would allocate an audio transcoding resource for the call.

Cisco Unified CM Release 5.0 introduced audio transcoding resources based on Cisco IOS Enhanced Media Termination Point, which are able to support transcoding the audio stream of a video call while still supporting the video stream via a pass-through codec. The pass-through codec is used only for the video stream because the pass-through codec cannot be used for a stream that requires transcoding. The following three conditions must all be true for the pass-through codec to be used:

- The two endpoint devices have a matching CODEC capability.
- **MTP Required** is *not* checked for either endpoint.
- All intermediate resource devices (MTPs and transcoders) support the pass-through codec.

Traditional transcoders do not currently support the pass-through capability, so the call would connect as audio-only and would be transcoded between G.729 and G.711. To avoid this situation without using Cisco IOS Enhanced Transcoders, you would have to set the region to use G.711 instead. However, a region set for G.711 would also use G.711 for audio calls between two IP phones, which you might not want due to the increased consumption of bandwidth over the WAN.

If you want to use G.729 for audio-only calls to conserve bandwidth and to use G.711 for video calls, then you should configure one region to use G.711 for video endpoints that do not support G.729 and a separate region (or regions) to use G.729 for IP phones. (See [Figure 16-2](#).) This method increases the number of regions needed but provides the desired codec and bandwidth allocations.

Figure 16-2 Using G.711 for Video Calls and G.729 for Audio-Only Calls**Note**

It is possible to configure a pair of regions to prohibit video. If two video-capable devices in that region pair try to call each other, they will connect as audio-only unless Retry Video Call as Audio is not checked, in which case AAR rerouting logic will take over.

Table 16-4 lists some example configurations and their outcomes.

Table 16-4 Scenarios for Various Region Settings

Region Setting	Setting of Retry Video as Audio	Result
Region allows video	Enabled	Video calls allowed
Region allows video	Disabled	Video calls allowed
Region does not allow video	Enabled	Video calls will proceed as audio
Region does not allow video	Disabled	If AAR is not configured, video calls fail (with busy tone and "Bandwidth Unavailable" message displayed)

The Video Bandwidth field accepts values in the range of 1 to 8128 kbps. However, to allow for compatibility with H.323 and H.320 videoconferencing devices, Cisco recommends that you always enter values for this field in increments of either 56 or 64 kbps. Therefore, valid values for this field include 112 kbps, 128 kbps, 224 kbps, 256 kbps, 336 kbps, 384 kbps, and so forth.

When the call speed requested by the endpoint exceeds the bandwidth value configured for the region, Unified CM automatically negotiates the call down to match the value allowed in the region setting. For instance, assume that an H.323 endpoint calls another H.323 endpoint at 768 kbps, but the region is set to allow a maximum of 384 kbps. The incoming H.225 setup request from the calling party would indicate that the call speed is 768 kbps, but Unified CM would change that value to 384 kbps in the outgoing H.225 setup message to the called party. Thus, the called endpoint would think that it was a 384-kbps call to begin with, and the call would be negotiated at that rate. The calling endpoint would show the requested bandwidth as 768 kbps, but the negotiated bandwidth would be 384 kbps.

However, if you set the Video Bandwidth to "None" in the region, Unified CM will either terminate the call (and send an H.225 Release Complete message back to the calling party) or will allow the call to pass as an audio-only call instead, depending on whether or not the called device has the Retry Video Call as Audio option enabled. (See [Retry Video Call as Audio](#), page 16-8.)

Locations

When configuring static locations, you also set two fields in Unified CM Administration: the Audio Bandwidth and the Video Bandwidth. Unlike regions however, the Audio Bandwidth for static locations applies only to audio-only calls, while the Video Bandwidth applies to both the audio and video channels of video calls. The audio and video bandwidth are kept separate because, if both types of calls shared a single allocation of bandwidth, then it is very likely that audio calls would take all the available bandwidth and leave no room for any video calls, or vice versa. Also, separate bandwidth pools for audio and video correspond to the way queues are configured in the switches and routers in the network, which typically have a priority queue for voice traffic and a separate priority queue or a Class-Based Weighted Fair Queue for video traffic. (See [WAN Quality of Service \(QoS\)](#), page 3-33, for more details.)

Both the Audio Bandwidth and the Video Bandwidth fields offer three options: None, Unlimited, or a field that accepts numeric values. However, the values entered in these fields use two different calculation models. For the Audio Bandwidth field, the values entered should include the Layers 3 to 7 overhead required for the call. For instance, if you want to permit a single G.729 call to or from a location, you would enter the value of 24 kbps. For a G.711 call, you would enter the value of 80 kbps. The Video Bandwidth field, by contrast, should be entered without the overhead included. For instance, for a 128-kbps call you would enter 128 kbps, and for a 384-kbps call you would enter 384 kbps. As with the values used in the Video Bandwidth field for regions, Cisco recommends that you always use increments of 56 kbps or 64 kbps for the Video Bandwidth field for locations as well.

For example, assume that a company has a three-site network. The San Francisco location has a 1.544-Mbps T1 circuit connecting it to the San Jose main campus. The system administrator wants to allow four G.729 voice calls and one 384-kbps (or two 128-kbps) video calls to/from that location. The Dallas location has two 1.544-Mbps T1 circuits connecting it to the San Jose main campus, and the administrator wants to allow eight G.711 voice calls and two 384-kbps video calls to/from that location. For this example, the administrator would set the San Francisco and Dallas locations to the following values:

Location	Number of Audio Calls Desired	Audio Bandwidth Field Value	Number of Video Calls Desired	Video Bandwidth Field Value
San Francisco	4 using G.729	96 kbps (4 * 24 kbps)	1 at 384 kbps	384 kbps
Dallas	8 using G.711	640 kbps (8 * 80 kbps)	2 at 384 kbps	768 kbps

When the call speed requested by the endpoint exceeds the value configured for the location, Unified CM will *not* automatically negotiate the call speed (as it does for regions) to match the value allowed in the location setting. Instead, the call will be rejected or will be retried as an audio-only call (if the Retry Video as Audio setting is enabled on the called device). Therefore, you should always set the region's video bandwidth to a value lower than the location's video bandwidth value. For example, if you have two regions (region A and region B) and you set the video bandwidth between those two regions to 768 kbps but the devices in region A are in a location that has a video bandwidth set to 384 kbps, then all calls between those two regions will fail or will result in audio-only calls (depending on the Retry Video Call as Audio setting).

Retry Video Call as Audio

This check-box is available on all endpoint types that support video, including Cisco Unified IP Phones 7940, 7941, 7960, 7961, 7970, 7971, and Cisco IP Video Phone 7985, as well as Sony or Tandberg SCCP endpoints, and all H.323 and SIP devices (clients, gateways and all types of H.323 trunks). When this

option is activated (checked), if there is not enough bandwidth to reach the device (for example, if the Unified CM regions or locations do not allow video for that call), then Unified CM will retry the call as an audio-only call. When this option is deactivated (unchecked), Unified CM will not retry the call as audio-only but instead will either fail the call or reroute the call by whatever automated alternate routing (AAR) path is configured. By default, this retry option is enabled (checked).

This feature applies to the following scenarios only:

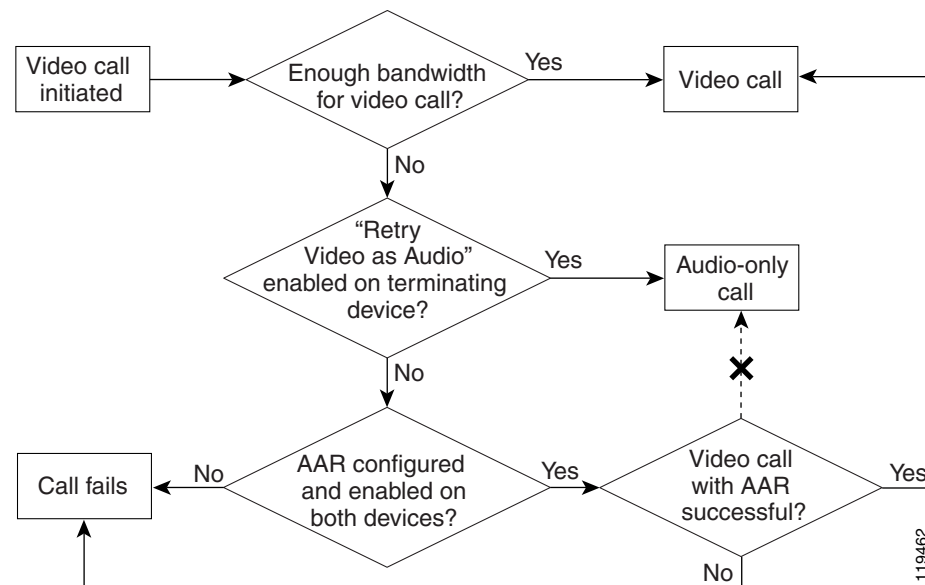
- The region is configured not to allow video.
- The location is configured not to allow video, or the requested video speed exceeds the available video bandwidth for that location.
- For calls between Unified CM clusters, the requested video speed exceeds the gatekeeper's zone bandwidth limits.

The Retry Video Call as Audio option takes effect only on the terminating (called) device, thus allowing the flexibility for the calling device to have different options (retry or AAR) for different destinations.

If the video call fails due to bandwidth limitations but automated alternate routing (AAR) is enabled, Unified CM will attempt to reroute the failed call as a video call to the AAR destination. If AAR is not enabled, the failed call will result in a busy tone and an error message being sent to the caller. (See [Figure 16-3](#).) Depending on the type of device that is calling, the failed call will result in one of the following conditions:

- If the calling device is an SCCP endpoint with an LCD screen (such as a Sony or Tandberg SCCP endpoint or many models of Cisco Unified IP Phones), the caller will hear busy tone and will see the message "Bandwidth Unavailable" displayed on the device.
- If the calling device is an SCCP endpoint without an LCD screen (such as a Cisco Unified IP Phone 7902), the caller will hear busy tone.
- If the calling device is an H.323 or SIP device or a PSTN device connected through a gateway, the caller will hear busy tone and Unified CM will send the appropriate error message (such as a Q.931 Network Congestion cause code) back to the H.323, SIP, or MGCP device.

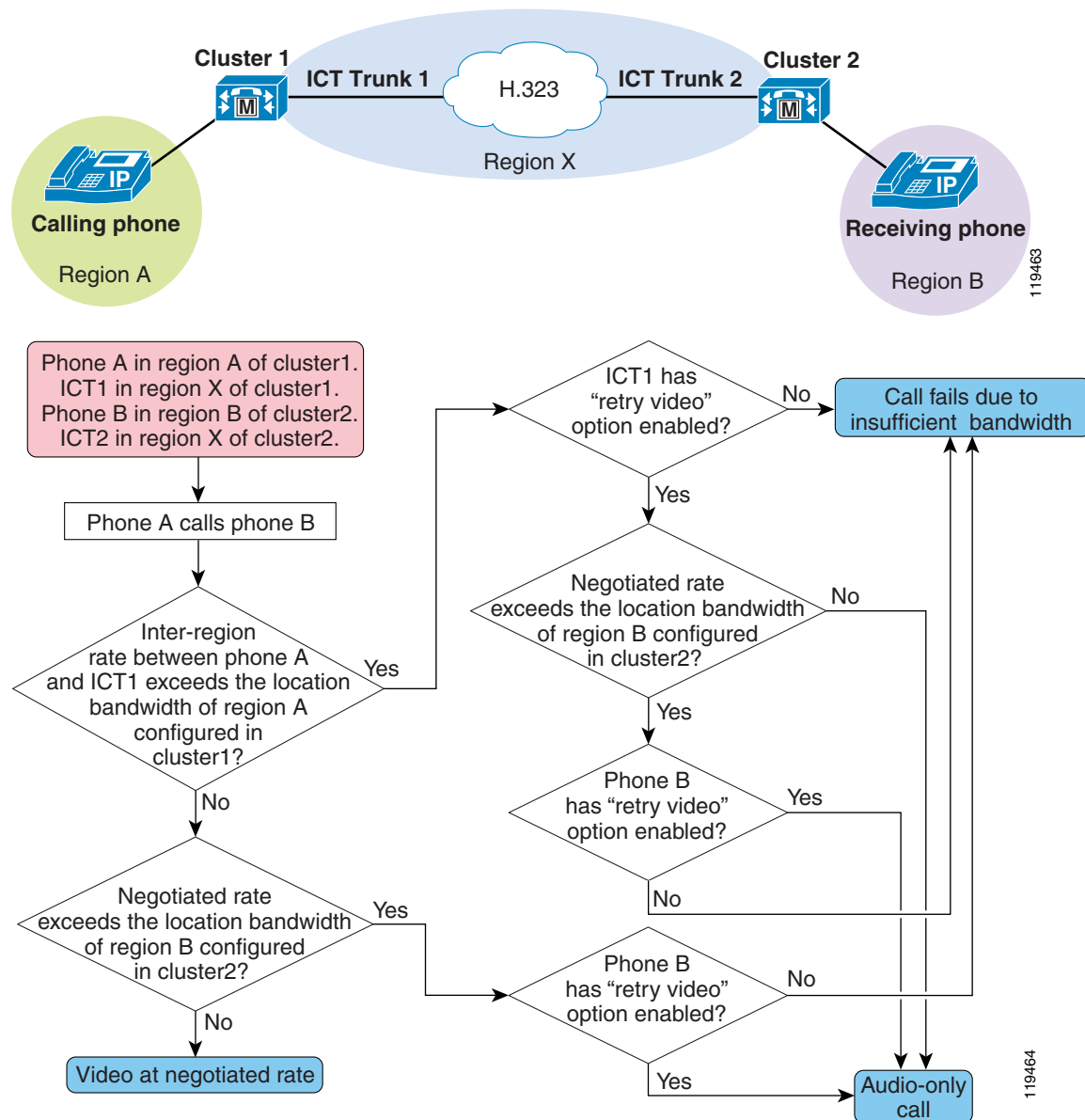
Figure 16-3 Possible Scenarios for a Video Call



See the chapter on [Call Admission Control](#), page 9-1, for further details on the use of AAR.

Figure 16-4 illustrates the steps of a call between two clusters using non-gatekeeper controlled intercluster trunks.

Figure 16-4 Call Flow Between Two Clusters Using Non-Gatekeeper Controlled Intercluster Trunks



Wait for Far-End to Send TCS

This check-box is available on all H.323 devices, including H.323 clients, H.323 gateways, and H.225 gatekeeper-controlled trunks. This feature pertains to the H.245 capabilities-exchange phase of H.323 calls. When this feature is enabled, Unified CM waits for the remote H.323 device to send its Terminal

Capabilities Set (TCS) to Unified CM before Unified CM will send its TCS to the H.323 device. When the option is disabled, Unified CM does not wait but sends its TCS to the remote H.323 device immediately.

By default, the Wait for Far-End to Send TCS option is enabled (checked). However, you must uncheck (disable) it in the following circumstances:

- When the H.323 device communicating with Unified CM is also waiting for the far-end to send its TCS

In this case, a stalemate occurs because neither side sends its TCS, and the H.245 connection times out after a few seconds. Examples of devices that also wait for the far-end to send TCS are some H.323 routed-mode gatekeepers, H.320 gateways, H.323 proxies (or IP-to-IP gateways) and some H.323 multipoint conference bridges. These devices wait for the far-end to send TCS for the same reason Unified CM does: because they are waiting for both sides of the connection to send their TCSs before forwarding them to the other side.

- When communicating with another Unified CM cluster over an intercluster trunk

**Note**

For intercluster trunks and gatekeeper-controlled intercluster trunks, the Wait for Far-End to Send TCS option is always disabled and cannot be enabled.

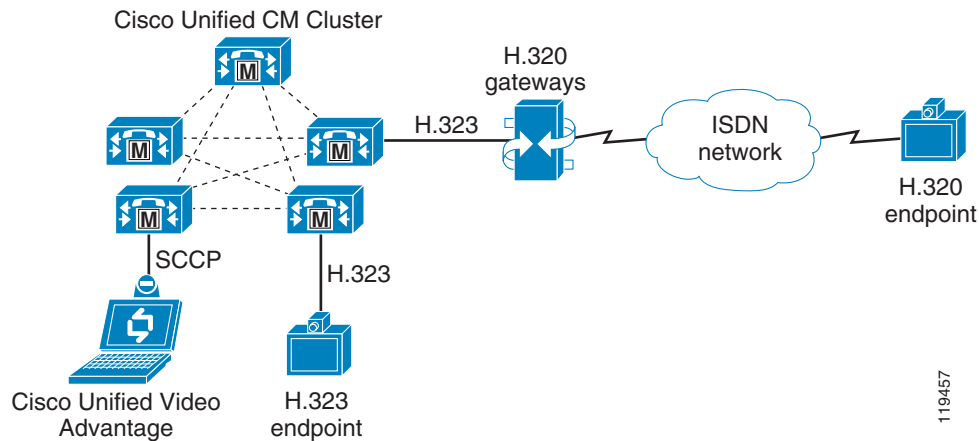
In many scenarios, Unified CM performs the role of a software switch connecting two endpoint devices (such as two H.323 clients that are trying to call each other). In such cases, it is best if Unified CM waits until both devices have sent their TCS messages so that it knows the capabilities of each device and can therefore make the most intelligent decision about what TCS to send back to each party (depending on region and location configurations, among other things). In these cases, the Wait for Far-End to Send TCS feature should be enabled.

However, some other H.323 devices (such as an H.320 gateway, which connects an H.323 device to an H.320 device) also perform the function of connecting two or more parties together. The gateway also prefers to wait until both ends send their TCS messages, so that it can make the most intelligent choice about how to set up the call. A stalemate could result if Unified CM and the gateway both wait for the other side to send their TCSs. To avoid this stalemate situation, disable (uncheck) the Wait for Far-End to Send TCS feature.

For instance, consider the following call scenarios illustrated in [Figure 16-5](#):

- Scenario 1 — Cisco Unified Video Advantage calls an H.320 endpoint
- Scenario 2 — An H.323 client calls an H.320 endpoint

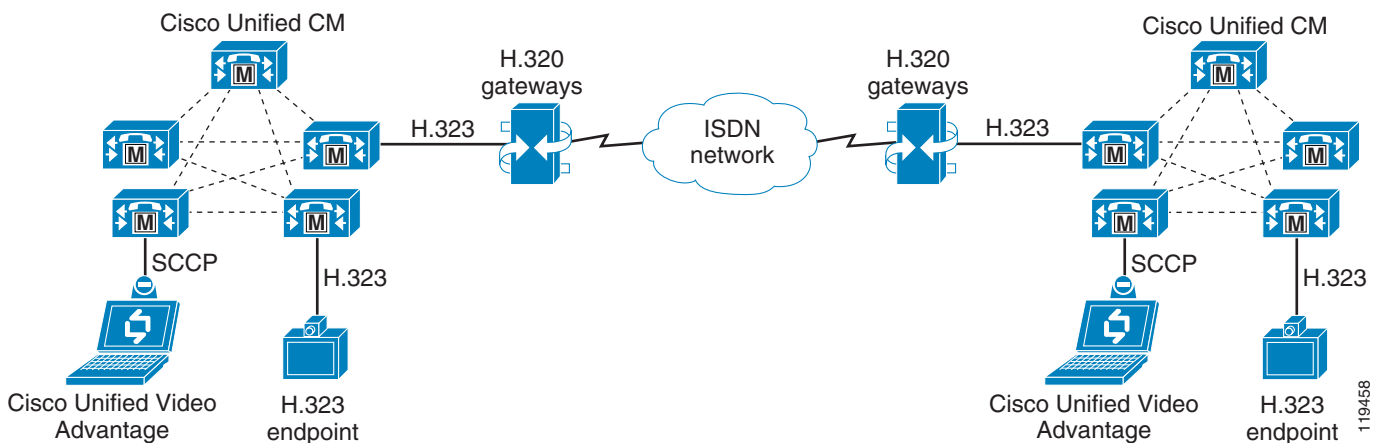
In both of these scenarios, Wait for Far-End to Send TCS feature can be left at its default setting of enabled (checked)

Figure 16-5 Scenarios with the Wait for Far-End to Send TCS Feature Enabled (Checked)

In Scenario 1 from Figure 16-5, Unified CM already knows the capabilities of the Cisco Unified Video Advantage client because SCCP devices provide Unified CM with their media capabilities during registration. But Unified CM does not know the capabilities of the H.320 gateway until the gateway sends its TCS to Unified CM during the H.245 phase of the call. Likewise, the H.320 gateway does not know what TCS to send to Unified CM until the H.320 endpoint sends its TCS to the gateway. In this case it is better to leave the Wait for Far-End to Send TCS feature enabled because the H.320 endpoint will send its TCS to the gateway, the gateway will send its TCS to Unified CM, and Unified CM will then have the TCSs from both endpoints with which to make a decision.

Figure 16-6 shows the following call scenarios, in which the call will fail unless the Wait for Far-End to Send TCS feature is disabled:

- Scenario 1 — Cisco Unified Video Advantage calls another Cisco Unified Video Advantage in a remote cluster via the ISDN network
- Scenario 2 — An H.323 client calls another H.323 client in the remote cluster via the ISDN network

Figure 16-6 Scenarios with the Wait for Far-End to Send TCS Feature Disabled (Unchecked)

In both scenarios from [Figure 16-6](#), there will be a stalemate because both Unified CMs will wait to receive the TCSs from the gateways, and the gateways will both wait to receive the TCS from the ISDN side as well. The call will time-out after a few seconds and fail. From the perspective of the users, the caller will hear ringback tone indicating that the call is progressing and the called party will hear a ring indicating an incoming call. When the called party attempts the answer the call, the H.245 phase will fail due to the stalemate, and the call will then fail by hanging up on both parties.

To work around this issue in scenarios such as these, Cisco recommends that you disable (uncheck) the Wait for Far-End to Send TCS option on the device representing the H.320 gateway in Unified CM. This device could be an H.225 gatekeeper-controlled trunk or an H.323 gateway device, depending on how you have configured Unified CM to reach the H.320 gateway.

However, if the Wait for Far-End to Send TCS option is disabled, there is a possibility that the initial capabilities exchanged will not work for the remote device. For instance, the Unified CM region might be set to 768-kbps video but the H.320 device might only support 384 kbps, or the selected audio codec might not work for the remote party. In such cases, the initially negotiated logical channels might have to be torn down and reopened with the correct speed and codec. Many legacy H.323 and H.320 devices do not handle this situation well and will disconnect the call when Unified CM sends a CloseLogicalChannel message to renegotiate the channel to a different value. Thus, you must be careful where and when you disable the Wait for Far-End to Send TCS option.

Multipoint Conferencing

Whenever three or more parties want to engage in the same video call together, a Multipoint Control Unit (MCU) is required. An MCU consists of the following main components:

- Multipoint Controller (MC)
- Multipoint Processor (MP)

The MC handles all aspects of call setup and teardown for the conference, including media negotiation, call signaling, and choosing which MP to use for the call. The MP processes all the audio and video packets. The MC controls the MP, and one MC can control multiple MPs. The MP can be either software-based or hardware-based. Software-based MPs are typically not capable of advanced transcoding, transrating (multiple speeds), or composition features.

Since 1999, Cisco has offered the Cisco Unified Videoconferencing 3500 Series H.323 Multipoint Conference Units (MCUs). The first product in this family was the Cisco Unified Videoconferencing 3510. This model is no longer available for sale, and is not compatible with Unified CM. In 2002, Cisco introduced the Cisco Unified Videoconferencing 3511 and 3540 models. While these models offer significantly improved features and scalability unavailable on the older 3510 model, the third-generation MCU models (the Cisco Unified Videoconferencing 3515 and 3545) are the recommended platform. The 3515 and 3545 offer an encoder-per-port hardware architecture, which significantly reduces preconfiguration requirements by eliminating the need to limit the bit rates, video formats, and conference features that videoconferencing endpoints and conference participants can use.

In 2003, Cisco introduced software version 3.2+ on the Cisco Unified Videoconferencing 35xx models, which adds support for the Skinny Client Control Protocol (SCCP). SCCP support is not available on the Cisco Unified Videoconferencing 3510. In addition, there are three types of MPs available in the Cisco Unified Videoconferencing MCUs:

- A software-based MP built into the 3510, 3515, and 3540 MCUs
- The Rate Matching (RM) module, a dedicated software-based module for the Cisco Unified Videoconferencing 3540 chassis

- The Enhanced Media Processor (EMP), a hardware-based solution available as a dedicated module for the Cisco Unified Videoconferencing 3540 and 3545 chassis or as an integrated component in the Cisco Unified Videoconferencing 3511 and 3515 models



Note An EMP is required for the 3545 MCU because it does not have a software-based MP.

Cisco Unified CM Release 4.0 (and later) supports the Cisco Unified Videoconferencing 3511, 3515, 3540, and 3545 models in SCCP, H.323 and SIP modes. Each protocol offers different features and is used for different reasons, so each of these MCUs is equipped to run all three protocols. The Cisco Unified Videoconferencing 3511 can be configured to run in either SCCP mode or H.323 and SIP mode, while the Cisco Unified Videoconferencing 3515, 3540, and 3545 models can be configured to run all three protocols simultaneously and divide the total number of available MP resources between the three.

Regardless of signaling protocol, the MCU provides the same basic function of receiving the audio and video streams from each participant and sending those streams back out to all other participants in some sort of combined view. There are two types of views in a multipoint video conference:

- Voice-activated (switched)
- Continuous presence

Voice Activation

Voice-activated conferences take in the audio and video streams of all the participants, decide which participant is the dominant speaker, and send only the dominant speaker's video stream back out to all other participants. The participants then see a full-screen image of the dominant speaker (and the current speaker sees the previous dominant speaker). The audio streams from all participants are mixed together, so everyone hears everyone else, but only the dominant speaker's video is displayed.

You can use any of the following methods to select the dominant speaker:

- Voice activation mode

Using this mode, the MCU automatically selects the dominant speaker by determining which conference participant is speaking the loudest and the longest. To determine loudness, the MCU calculates the strength of the voice signal for each participant. As conditions change throughout the conversation, the MCU automatically selects a new dominant speaker and switches the video to display that participant. A hold timer prevents the video from switching too hastily. To become the dominant speaker, a participant has to speak for a specified number of seconds and be more dominant than all other participants.

- Manual selection of the dominant speaker through the MCU's web-based conference control user interface

The conference controller (or chairperson) can log onto the MCU's web page, highlight a participant, and select that person as the dominant speaker. This action disables voice activity detection, and the dominant speaker remains constant until the chairperson either selects a new dominant speaker or re-enables voice activation mode.

- Configuring the MCU to cycle through the participant list automatically, one participant at a time

With this method, the MCU stays on each participant for a configured period of time and then switches to the next participant in the list. The conference controller (or chairperson) can turn this feature on and off (re-enable voice activation mode) via the web interface.

Continuous Presence

Continuous-presence conferences display some or all of the participants together in a composite view. The view can display from 2 to 16 squares (participants) in a variety of different layouts. Each layout offers the ability to make one of the squares voice-activated, which is useful if there are more participants in the conference than there are squares to display them all in the composite view. For instance, if you are using a four-way view but there are five participants in the call, only four of them will be displayed at any given time. You can make one of the squares in this case voice-activated so that participants 4 and 5 will switch in and out of that square, depending on who is the dominant speaker. The participants displayed in the other three squares would be fixed, and all of the squares can be manipulated via the conference control web-based user interface.

**Note**

Continuous presence requires the use of an Enhanced Media Processor (EMP) in the Cisco Unified Videoconferencing MCU.

MP Resources

For both types of conferences, the type of MP resource determines which video formats, transrating, and transcoding capabilities the MCU can support. If endpoints connect to the conference at different speeds, a transrating-capable MP is required. The RM and EMP modules are both capable of transrating between speeds. If a transrating-capable MP is not available, the MCU will send out flow-control messages to all the endpoints, instructing them to lower their transmit speed to match the maximum receive rate of the slowest endpoint. For example, if three participants are connected in a 384-kbps conference and a fourth participant joins at 128 kbps, the MCU will send flow-control messages to the other three participants instructing them to lower their transmit rates to match the 128-kbps participant. This method causes all participants to suffer degraded quality because one participant is less capable. A transrating-capable MP would, instead, convert the 128-kbps stream to 384 kbps, and vice versa, so that each participant can enjoy the best quality their connection allows.

For continuous-presence conferences, a transrating-capable MP is also very important. The software-based MP built into the MCU combines all the input streams and sends the resulting combination back out to each participant. For instance, if four participants are connected in a continuous-presence conference at 384 kbps using G.711 audio, each participant will transmit 320 kbps of video and 64 kbps of audio into the MCU. The MCU will take these four input video streams and combine them into the four-way composite view. The MCU will then transmit 1280 kbps of video back to each endpoint, along with the mixed 64 kbps of audio, for a total of 1344 kbps per endpoint. This method is known as Asynchronous Continuous Presence, and it can have a negative impact on bandwidth requirements, call admission control mechanisms, and interoperability with certain devices.

**Note**

Cisco strongly advises against the use of Asynchronous Continuous Presence.

With the RM or EMP modules, the MCU is capable of transrating each input stream before combining them, so that the total output bandwidth matches the input bandwidth. For instance, if the MCU is using the four-quadrant layout and each participant transmits 320 kbps of video and 64 kbps of audio into the MCU, the MCU would essentially transrate each input stream down to 80 kbps, combine them so that the resulting four-quadrant view is 320 kbps of video ($4 * 80$ kbps), combine this video with the mixed 64 kbps audio, and transmit the final combination back to each participant. This method is known as Synchronous Continuous Presence. Cisco strongly recommends that all continuous-presence conferences use the Synchronous Continuous Presence mode. However, use of this mode means that each MCU must have a transrating-capable MP (such as an RM or EMP) available, which does increase the cost of the MCU.

**Note**

For H.323 and SIP clients with built-in MCUs, Unified CM does not allow an H.323 client to generate a second call, thereby negating the functionality of the built-in MCU.

SCCP MCU Resources

As stated previously, the Cisco Unified Videoconferencing 3511, 3515, 3540, and 3545 MCUs offer support for SCCP beginning with software version 3.2+ for those models and with Cisco Unified CM Release 4.0. When configured in SCCP mode, Unified CM provides the MC function while the MCU provides the MP function. The SCCP MCU is controlled completely by Unified CM.

Only the following events invoke SCCP MCU resources:

- The user of an SCCP endpoint (such as an IP Phone or a Tandberg endpoint) presses the Conf, Join, or cBarge softkeys to invoke an ad-hoc conference
- The user of an SCCP endpoint (such as an IP Phone or a Tandberg endpoint) presses the MeetMe softkey to invoke a reservationless meet-me conference.

Participants in either of these types of conferences can include any type of endpoint (that is, video and non-video devices using any signaling protocol that Unified CM supports via any supported gateway type); however, only SCCP endpoints can invoke the SCCP MCU resources. In other words, an H.323 video endpoint cannot invoke an SCCP MCU resource, but an SCCP video endpoint can invoke the resource and then join an H.323 video participant to the call. For example, the user at the SCCP endpoint could press the Conf softkey, dial the directory number of an H.323 client, and then press the Conf softkey again to complete the transaction. The H.323 client will be joined as a participant on the SCCP MCU conference.

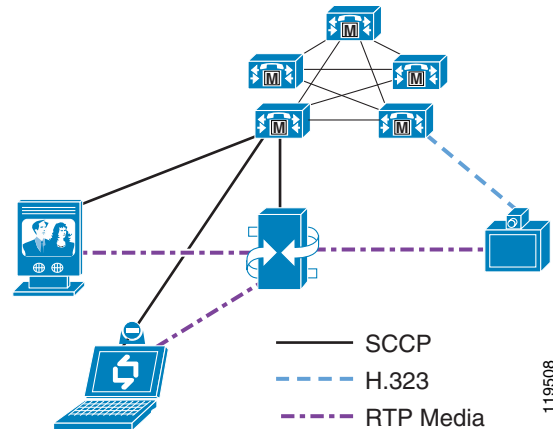
However, for ad-hoc conferences initiated via the Conf, Join, or cBarge softkey, the signaling protocol used by the other participants must support the ability to be placed on hold and then have their audio and video channels redirected to the MCU. For H.323 devices (H.323 clients, H.323 gateways, H.320 gateways, and all types of H.323 trunks) Unified CM uses the Empty Capabilities Set (ECS) method defined in the H.245 specification to achieve this functionality. If the H.323 endpoint does not support receiving an ECS message from Unified CM, it will react by hanging up or possibly even crashing and/or rebooting. To work around this problem, you can enable (check) the "MTP Required" option on the H.323 device but assign it a media resource group list (MRGL) that does not contain any MTP devices, and then set the Unified CM service parameter **Fail Call if MTP Allocation Fails** to **False**. (For details, see [Media Resource Groups and Lists, page 16-17](#).) This configuration will cause the softkeys to be greyed out on the phone, disabling the user from invoking any supplementary services with that endpoint, including placing it on hold, conferencing it into an existing call, joining it with an existing call, or barging into an existing call containing that endpoint.

**Note**

Because the workaround described above requires an MRGL that does not contain any MTP devices, this workaround cannot be used when RSVP-based call admission control is being used.

For reservationless conferences via the MeetMe softkey, the signaling protocol used by the other endpoints does not have to support being placed on hold and transferred. For these types of conferences, each endpoint dials the MeetMe dial-in number arranged by the SCCP client who initiated the conference.

[Figure 16-7](#) illustrates how H.323 and SCCP endpoints can participate in the same SCCP conference. In this example, the conference was initiated by an SCCP endpoint using the Conf softkey to invite the three members.

Figure 16-7 Ad-Hoc Conference Between SCCP and H.323 Endpoints

SCCP conferences support voice-activated mode as well as continuous presence. Furthermore, SCCP conferences support the integrated software-based MP of the MCU and the Rate Matching (RM) module as well as the Enhanced Media Processor (EMP) module.

Media Resource Groups and Lists

Unified CM uses media resource groups (MRGs) and media resource group lists (MRGLs) to determine which specific conference bridge resource to use for a given endpoint. How you group the resources is completely at your discretion, but it is typically done either by geographical placement (so that all endpoints at a given site use the MCU closest to them) or by endpoint type (so that video-capable endpoints use a video-capable MCU while audio-only endpoints use a different conference bridge resource). When a user of an SCCP device activates the Conf, Join, or MeetMe softkey, Unified CM uses the MRGL of the initiating endpoint to determine which conference bridge should be used.

Unified CM applies the following criteria, in the order listed, to select which conference bridge resource to use:

1. The priority order in which the media resource groups (MRGs) are listed in the media resource group list (MRGL)
2. Within the selected MRG, the resource that has been used the least

Because of this selection process, the video-capable MCU must be at the top of the MRGL for the video-capable SCCP endpoint in order for that MCU to be selected when the user activates the Conf, Join, or MeetMe softkey on that endpoint. However, some endpoints are not dedicated video endpoints. For example, the IP Phone used in conjunction with Cisco Unified Video Advantage might be used for audio-only calls most of the time and for video calls only occasionally. Thus, if you place the MCU at the top of the MRGL for that phone, the MCU will always be chosen even for audio-only conferences that do not involve any video-capable participants. In this scenario, the MCU resources might be wasted on audio-only conferences and not be available to satisfy the request for a video conference when it occurs.

Therefore, Cisco recommends that you carefully select which users you want to give access to the video-capable MCU resources through MRGLs. The following considerations can help you make that selection:

- Is the endpoint a dedicated video device (such as a Sony or Tandberg SCCP endpoint), or is it an IP Phone that will only occasionally have a Cisco Unified Video Advantage client associated to it?
- Does the user require video when invoking SCCP conferences, or will audio-only suffice?
- How much are you willing to spend on MCU resources to outfit your network with enough resources to satisfy the SCCP conferencing requirements?

The answers to these selection criteria will be different for every company. One company might deem multipoint video to be a critical feature and be willing to spend the money necessary to outfit their network with enough MCU resources so that, even if a few ports get wasted on audio-only conferences, there will still be resources available to satisfy all of their video conferences. Another company might choose to enable video resources only for their Sony and Tandberg endpoints and assign audio-only conference bridge resources to Cisco Unified Video Advantage users. A third company might choose to enable video resources for all Sony and Tandberg SCCP endpoints and for select Cisco Unified Video Advantage users (perhaps managers and above), but allocate audio-only resources for the rest of the user population.

H.323 and SIP MCU Resources

When configured in H.323 or SIP mode, the MCU provides the MC function and behaves like an H.323 or SIP peer to Unified CM. H.323 and SIP MCU conferences can be invoked in a number of different ways, but they all fall into two major categories:

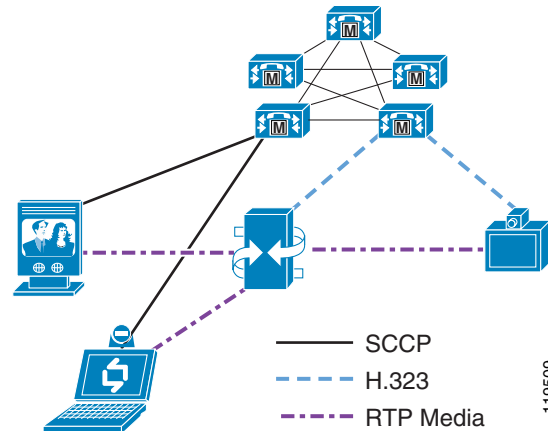
- Scheduled
- Reservationless

A scheduled conference uses a scheduling application to reserve the MCU resources in advance of the call. The scheduling function typically is provided through a web-based user interface such as Cisco Unified MeetingPlace, MagicSoft VCWizard, RADVision VCS, Tandberg TMS, or FVC.COM Click to Meet. The scheduling application usually generates an invitation that provides the user with the date and time of the conference, the number of ports reserved for the conference, and the dial-in information. Alternatively, the scheduling system can be configured to dial out to some or all of the participants at the beginning of the conference.

For a reservationless conference, the MCU has a certain number of resources that are available on demand. To create a conference, a user simply dials into the MCU at any time. If that user is the first participant to dial in, the MCU dynamically creates a new conference using the settings defined in the service template. (For more information on service templates, see [Service Templates and Prefixes, page 16-19](#).) Subsequent users who dial into the same conference number are joined to that conference.

Any type of endpoint can create and participate in scheduled and reservationless H.323 or SIP conferences. For instance, an SCCP endpoint can dial into the H.323 MCU to create a reservationless conference just as well as an H.323 endpoint can.

Figure 16-8 illustrates how H.323 and SCCP endpoints can participate in the same H.323 conference. In this example, the conference was initiated by an SCCP endpoint that dialed into the H.323 MCU to create a new reservationless conference, and the other two parties subsequently dialed into that conference.

Figure 16-8 *SCCP and H.323 Endpoints in a Reservationless Conference*

H.323 and SIP conferences support both voice-activated and continuous-presence modes. Furthermore, H.323 conferences support all of the MP types, including the integrated software-based MP of the MCU, the Rate Matching (RM) module, and the Enhanced Media Processor (EMP) module.

Service Templates and Prefixes

A service on the MCU defines the settings that pertain to each conference. You can have different services defined for the different types of conferences. Each service defines, at a minimum, the following settings:

- Speed of the conference (the video bit-rate)
This setting might include multiple speeds if a transrating-capable MP is used.
- Minimum and/or maximum number of participants
The minimum number defines how many ports will be reserved at the beginning of the conference. The maximum number defines the maximum number of participants that the MCU will allow to join this conference.
- Video codec type (H.261, H.263, or H.264)
- Frame rate (15 or 30 fps)
- Resolution (QCIF or CIF)
- MP resource (Auto, MP, RM, or EMP)
- Video layout to be displayed (voice activated or continuous presence)
A conference can include multiple layouts or even dynamic layouts that change as the number of participants in the conference increases or decreases.
- H.323 and SIP, or SCCP
When the "SCCP service" check-box is enabled (checked), the service is used for SCCP conferences. When this box is disabled (unchecked), the service is used for H.323 and SIP conferences.

For H.323 and SIP services, each service is assigned a service prefix that is dialed by the endpoints to reach that particular service. The service prefix forms the leading digits of the conference number, and the trailing digits define the conference ID. This format allows multiple conferences to run simultaneously using the same service prefix. For instance, you can have a service prefix of 555, while the full dial string of the conference is seven digits. This scheme would allow for conference numbers in the range of 5550000 through 5559999, with four digits for the conference ID. The user must dial the

full string to access the conference. When the call is received by the MCU, the MCU parses the dialed digits to try to match a service prefix. Once it determines which service prefix is being dialed, the MCU then uses the remaining digits as the conference ID. If the conference ID does not exist yet, the MCU creates a new reservationless conference with that ID. If the conference already exists, the user is added to that conference.

It is important to note that, if both H.323 and SIP are enabled on the MCU at the same time, the dial plan must be the same for both protocols. There is no separation between H.323 and SIP as there is with SCCP. If a conference is created using SIP, the MCU will register that conference via H.323. If the gatekeeper or SIP proxy rejects the registration, the conference will fail.

SCCP services must also have a service prefix defined, but the digits themselves do not mean anything because users do not "dial" into an SCCP service. The prefix is used only in the SCCP registration messages between Unified CM and the SCCP MCU resource. Users either use the Conf, Join, or cBarge softkeys to access the SCCP MCU conference (ad hoc), or they dial a MeetMe number assigned by Unified CM to join the conference (reservationless). Therefore, the digits you specify for the SCCP service prefix do not matter and can be any digits you wish, such as 999999 for instance. This prefix is not exposed outside of the SCCP signaling between the MCU and Unified CM (that is, it cannot be dialed, it is not included in the MCU's registration with the gatekeeper, and so forth).

Sizing the MCU

There are several factors involved in determining the types and number of conferences that an MCU can support. These sizing factors are different for different models of MCUs, as described in the following sections.

Cisco Unified Videoconferencing 3515 and 3545

The current MCU models are the Cisco Unified Videoconferencing 3515 and 3545. The Cisco Unified Videoconferencing 3515 MCU supports a fixed number of ports, while the Cisco Unified Videoconferencing 3545 MCU is a modular system that can accept up to three 24-port EMP modules. Unlike previous versions of hardware, all ports of the 3515 and 3545 are fully processed, allowing them to offer any supported connection speed, any supported video and audio codecs, and any supported video resolutions, without any decrease in port capacities or in the number of simultaneous conferences supported. Both MCUs require EMPs to operate; the 3515 has the EMP on-board while the 3545 requires at least one separate module.

Therefore, calculating the size of these MCU is straightforward and depends on the following factors:

- The total number of ports that the MCU can support.
- The number of ports that the MCU can dedicate to each protocol.



Note

A single SCCP conference cannot span multiple EMPs. Each SCCP conference can support up to 24 participants.

Cisco Unified Videoconferencing 3511 and 3540

The previous MCU models are the Cisco Unified Videoconferencing 3511 and 3540. The Cisco Unified Videoconferencing 3511 MCU supports a fixed number of ports, while the Cisco Unified Videoconferencing 3540 MCU is a modular system that can accept various sizes of modules. When calculating the total number of ports available, you must also consider the number of sessions that the

Audio Transcoder Daughter Card and the Rate Matching (RM) module or Enhanced Media Processor (EMP) module can support. Therefore, consider the following factors when calculating the size of these MCUs:

- The number of ports that the MCU can support
This value depends on the speed of the conference; as the speed increases, the number of supported ports decreases.
- The number of ports that the Audio Transcoding Daughter Card can support
This value depends on which audio codecs are used in the conference.
- The number of conferences that the RM or EMP module can support
This value depends on how many participants need to be transrated and how many views are in use in the conference.

For specific information about the number of ports supported, refer to the product documentation for your MCU hardware, available on Cisco.com. Due to the almost infinite number of possible variations, it is very difficult to provide any concrete design guidance in this document. Many customers end up with a mixture of SCCP ad-hoc conferences, H.323 and SIP reservationless conferences, and H.323 and SIP scheduled conferences. The MCUs must be sized to accommodate all of those types of conferences at the correct speeds and video layouts. Needless to say, this can become quite complex to determine. Please consult with your Cisco sales representative for assistance on sizing the MCUs for your particular environment.

IVR for Dial-In Conference

Dial-in conferences typically use an interactive voice response (IVR) system to prompt users to enter the conference ID and the password (if one is configured) of the conference they want to join. You can use either of the following types of IVRs with the Cisco Unified Videoconferencing 3500 Series MCUs:

- The IVR built into the MCU
- Cisco Unified IP IVR

The built-in IVR of the MCU has the following characteristics:

- Can prompt only for the password of the conference
It cannot prompt the user for the conference ID first. In other words, the users must dial the specific conference number they wish to join, then they are prompted for the password of that conference.
- Supports both in-band and out-of-band (H.245 alpha-numeric) DTMF
- Cannot be customized to provide more flexible menus or functionality

The only thing that can be customized is the recorded audio file that is played to the user.

If you want to have a single dial-in number and then prompt the user for the conference ID, you can use Cisco Unified IP IVR in conjunction with the MCU.

Cisco Unified IP IVR has the following characteristics:

- Can prompt for the conference ID and the password (among other things)
- Supports only out-of-band DTMF

That is, the calling device must support an out-of-band DTMF method, such as H.245 alpha-numeric on H.323 devices. These out-of-band DTMF messages are then relayed by Unified CM to the Cisco IP IVR server. If the calling device supports only in-band DTMF tones, the Cisco IP IVR server will not recognize them and the calling device will be unable to enter the conference.

- Can be highly customized to provide more flexible menus and other advanced functionality
- Customizations can include such things as verifying the user's account against a back-end database before permitting that user to enter into the conference, or queuing the participants until the chairperson joins.

**Note**

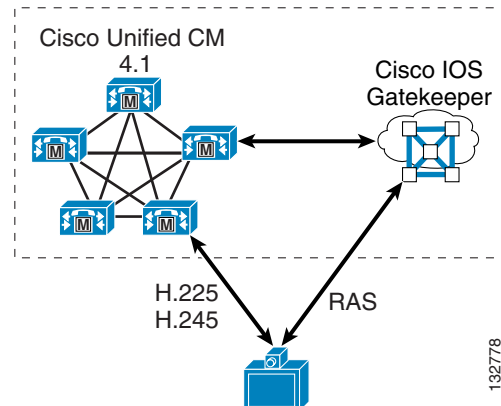
Because Cisco Unified IP IVR supports only out-of-band signaling, it will not work with H.323 endpoints that use in-band DTMF tones.

With Cisco Unified IP IVR, users dial a CTI route point that routes the call to the Cisco Unified IP IVR server instead of dialing a route pattern that routes directly to the MCU. After collecting the DTMF digits of the conference ID, the Cisco Unified IP IVR then transfers the call to the route pattern that routes the call to the MCU. This transfer operation requires that the calling device supports having its media channels closed and reopened to a new destination. For example, an H.323 video device that calls the Cisco Unified IP IVR will initially negotiate an audio channel to the Cisco Unified IP IVR server and then, after entering the appropriate DTMF digits, it will be transferred to the MCU, at which point Unified CM will invoke the Empty Capabilities Set (ECS) procedure described earlier in this chapter to close the audio channel between the endpoint and the Cisco Unified IP IVR server and open new logical channels between the endpoint and the MCU. If the H.323 video endpoint does not support receiving an ECS from Unified CM, it will react by disconnecting the call or, worse, crashing and/or rebooting.

Gatekeepers

Prior to the introduction of video support in Unified CM, H.323 videoconferencing networks relied on gatekeepers to perform device registration management, call routing, and bandwidth control. The Cisco IOS Gatekeeper, formerly known as the Multimedia Conference Manager (MCM), provides these functions. However, most gatekeepers, including Cisco's, offer only rudimentary call routing capabilities compared to what might be expected in a typical enterprise-class PBX. When used to route H.323 video calls, Unified CM supplements the basic gatekeeper features and provides full enterprise-class PBX functionality to H.323 video calls.

Unified CM and the gatekeeper work as a team to manage H.323 video endpoints. The gatekeeper handles all Registration, Admission, and Status (RAS) signaling, while Unified CM handles all of the H.225 call signaling and H.245 media negotiations. Therefore, you have to deploy gatekeepers along with the Unified CM servers if RAS signaling procedures are required for the H.323 endpoints in your network, as illustrated in [Figure 16-9](#).

Figure 16-9 Unified CM and Cisco IOS Gatekeeper Provide RAS Signaling for H.323 Endpoints

RAS signaling is required any time either of the following conditions exists:

- The endpoint does not use a fixed IP address.

If the endpoint uses a static IP address, Unified CM does not require RAS procedures to locate the endpoint. Instead, the endpoint is provisioned in Unified CM Administration with its static IP address, and calls to that H.323 client's directory number are routed directly to that static IP address. If the endpoint does not use a static IP address, then Unified CM must query the gatekeeper to obtain the endpoint's current IP address each time Unified CM extends a call to the endpoint.

- The endpoint requires RAS procedures to place calls to E.164 addresses.

Most H.323 videoconferencing endpoints are capable of dialing another endpoint directly only when dialing by IP address (that is, the user enters the IP address of the destination endpoint in dotted-decimal format and then pushes the call button). However, if the user dials an E.164-formatted number (a numeric value not in the dotted-decimal format of an IP address) or an H.323-ID (in the format of *username* or *username@domain*), most endpoints today provide only one way to resolve these types of destinations – by a RAS query to their gatekeeper. A growing number of endpoints, however, can be configured so that, for any call to an E.164 address, they skip any RAS procedures and instead send an H.225 SETUP message directly to a specified IP address. This method of operation is known as peer-to-peer mode. Tandberg H.323 endpoints are one example that use this mode, in which you can either configure a gatekeeper address for them to register with, or configure the IP address of the Unified CM server they should use. In the latter case, the endpoint sends all calls directly to the specified IP address, bypassing the need for RAS procedures with any gatekeeper.

In addition to managing RAS procedures for H.323 video endpoints, gatekeepers also continue to play an important role in managing dial plan resolution and bandwidth restrictions between Unified CM clusters in large multisite distributed call processing environments. A gatekeeper can also integrate with large numbers of H.323 VoIP gateways within the organization, or it can act as a session border controller between an enterprise IP Telephony network and a service provider VoIP transport network.

Therefore, as it pertains to Cisco IP Video Telephony deployments, the Cisco IOS Gatekeeper can perform one or both of the following roles:

- Endpoint gatekeeper

An endpoint gatekeeper is configured to manage all RAS procedures for calls to, from, and between H.323 clients, MCUs, and H.320 video gateways. The endpoint gatekeeper directs all such calls to the appropriate Unified CM cluster so that Unified CM can perform all of the H.225 call routing and H.245 media negotiations.

- Infrastructure gatekeeper

An infrastructure gatekeeper is configured to manage all dial plan resolution and bandwidth restrictions (call admission control) between Unified CM clusters, between a Unified CM cluster and a network of H.323 VoIP gateways, or between a Unified CM cluster and a service provider's H.323 VoIP transport network.

In Cisco Unified CM Release 4.0, the endpoint gatekeeper and the infrastructure gatekeeper had to run on separate routers, and each endpoint gatekeeper could service only a single Unified CM cluster. If multiple Unified CM clusters existed within the enterprise, a separate endpoint gatekeeper had to be deployed for each Unified CM cluster. With Cisco Unified CM Release 4.1 or above, it is possible to combine these roles on a single gatekeeper, using it as an endpoint gatekeeper for one or more Unified CM clusters and as the infrastructure gatekeeper for managing calls between clusters or between a cluster and other H.323 VoIP networks. However, for the following reasons (among others), Cisco recommends that you still separate these roles onto two or more gatekeepers:

- Scalability

Depending on the Cisco IOS router platform you choose to deploy and your estimated busy hour call volume, you might need several gatekeepers to handle the load.

- Geographical resiliency

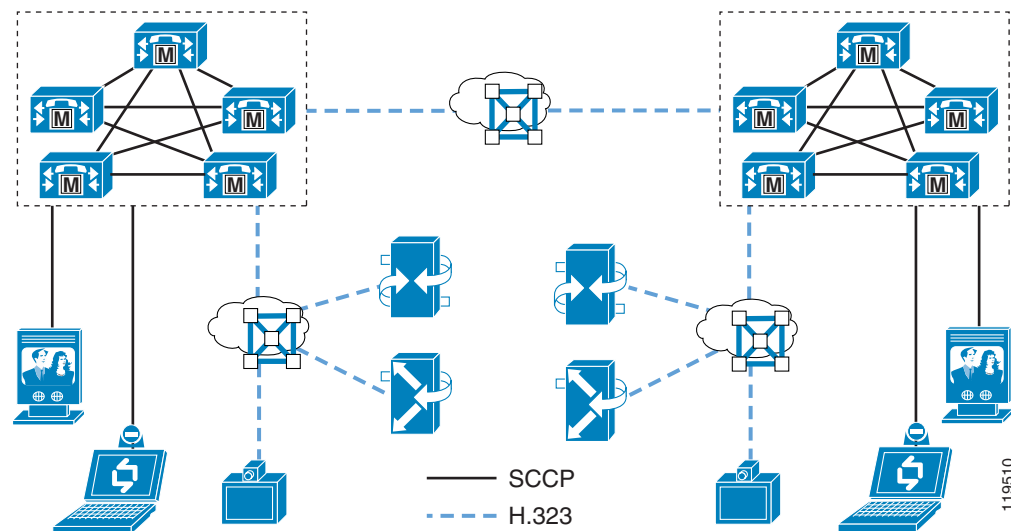
Putting all of your eggs into one basket may not be wise in a large, multi-national VoIP network. Having gatekeepers placed throughout your network (typically by geography) can provide better fault isolation in the event of a gatekeeper failure.

- Incompatibilities

Some configuration aspects of the gatekeeper are global in nature (they pertain to all endpoints registered with that gatekeeper). For example, the command **arq reject-unknown-prefix**, which may be useful in some H.323 VoIP transport environments, conflicts with the use of the **gw-type-prefix <prefix> default-technology** command, which is used in endpoint gatekeepers to route calls to Unified CM. While Cisco IOS does not stop you from configuring both commands on the same gatekeeper, the **arq reject-unknown-prefix** command takes precedence and, therefore, calls to unknown numbers will be rejected instead of being routed to Unified CM. In this case, you would have to use one gatekeeper for the H.323 VoIP transport network and another gatekeeper for the Unified CM cluster(s).

Another example of incompatibility can occur in the way you configure the gatekeeper for redundancy. Most Cisco H.323 voice devices, including Cisco Voice Gateways and Unified CM, support the H.323v3 Alternate Gatekeeper feature, which would allow you to configure the gatekeepers as a gatekeeper cluster using the Gatekeeper Update Protocol (GUP) to keep in sync with each other. However, many H.323 video endpoints do not support Alternate Gatekeeper, so the gatekeepers must be configured to use Hot Standby Routing Protocol (HSRP) for redundancy. You cannot mix and match these two redundancy methods on the same gatekeeper. In this case, you might decide to use a gatekeeper cluster for those endpoints that support Alternate Gatekeeper and an HSRP pair of gatekeepers for those that do not.

Figure 16-10 illustrates a network scenario with two Unified CM clusters. Each cluster consists of SCCP and H.323 clients, H.323 MCUs, and H.320 gateways. To manage the RAS aspects of the H.323 clients, MCUs, and H.320 gateways, an endpoint gatekeeper is deployed with each cluster. A separate infrastructure gatekeeper manages dial plan resolution and bandwidth between the clusters. Gatekeeper redundancy is not shown in the figure, although each of these gatekeepers may actually be multiple gatekeepers configured for either Alternate Gatekeeper or HSRP-based redundancy.

Figure 16-10 Two Unified CM Clusters with Required Gatekeepers

Supported Gatekeeper Platforms

To act as an endpoint gatekeeper with Cisco Unified CM 4.1 and above, the Cisco IOS Gatekeeper must run Cisco IOS Release 12.3(11)T or greater. For minimum Cisco IOS release requirements on the infrastructure gatekeeper, see [Recommended Hardware and Software Combinations](#), page A-1.

The following router platforms support the Cisco IOS Gatekeeper:

- Cisco 2600XM Series and 2691
- Cisco 2800 Series
- Cisco 3640, 3640A, 3660
- Cisco 3725 and 3745
- Cisco 3825 and 3845
- Cisco 7200 Series, 7301, and 7400 Series

To determine which release and feature set you should use for your router platform, use the *Cisco Feature Navigator* (requires a Cisco.com login account), available at

<http://tools.cisco.com/ITDIT/CFN/jsp/index.jsp>

More information is also available at

<http://www.cisco.com/web/psa/products/index.html?c=268438303>

That document states that Cisco IOS Release 12.3(11)T provides integrated voice and video services, therefore it is the recommended minimum release.

Endpoint Gatekeepers

An endpoint gatekeeper is required any time both of the following conditions are met:

- The cluster contains H.323 clients, H.323 MCUs, or H.320 gateways (collectively referred to as H.323 endpoints). If none of these types of endpoints exists (for example, if all clients are SCCP endpoints and there are no MCUs or H.320 gateways), then an endpoint gatekeeper is not needed.
- And either of the following conditions is true:
 - The H.323 endpoints require RAS procedures to initiate calls to E.164 addresses. As mentioned earlier, a growing number of devices are capable of peer-to-peer call signaling, in which case there is no need for those devices to register with a gatekeeper.
 - The H.323 endpoints do not use static IP addresses.

The role of the endpoint gatekeeper is simply to handle the RAS aspects of communications with the endpoints, providing a place for these H.323 endpoints to register. The endpoint gatekeeper responds to all call requests made to, from, or between these endpoints by directing the call to the appropriate Unified CM server(s) so that Unified CM can perform all of the call routing and bandwidth control functions. To accomplish this call routing and bandwidth control, you configure Unified CM to register H.323 trunk(s) with the gatekeeper and configure the gatekeeper to route calls to those trunks for all calls to, from, or within that zone.

Cisco Unified CM Release 4.1 introduced a new type of H.323 trunk called the RASAggregator trunk. This type of trunk is used for all H.323 client, H.323 MCU, or H.320 gateway zones, while the gatekeeper-controlled intercluster trunk and gatekeeper-controlled H.225 trunk from previous Unified CM releases are used to integrate with infrastructure gatekeepers. If you have deployed H.323 video endpoints based on the recommendations documented in the *IP Video Telephony SRND for Cisco Unified CallManager 4.0*, you should modify your configuration to use the new RASAggregator trunk so that you can take advantage of its flexibility. This configuration change should be carefully planned and done at a time when it is convenient for the administrator so as to minimize disruption to the existing H.323 video endpoints. For a description of the procedure for migrating from a Cisco Unified CM 4.0 deployment, see [Migrating from Cisco Unified CM 4.0, page 16-44](#).

Provisioning H.323 Clients

H.323 clients are provisioned much the same way as other phones are, in that you create a new phone (model type = H.323 Client), assign a directory number to it, and assign it a calling search space, device pool, and so forth. You configure the H.323 clients in Unified CM in one of the following ways. The method you use depends on whether or not the client uses a static IP address and whether or not the client requires RAS procedures to dial E.164 addresses.

- Gatekeeper controlled

This type of configuration is used for clients that do not have a static IP address assigned to them (they use a DHCP-assigned address) and that require RAS procedures to dial E.164 addresses. A RASAggregator trunk is used to communicate to and from these clients. (See [Figure 16-11](#) and [Figure 16-12](#).)

- Non-gatekeeper controlled, asynchronous

This type of configuration is used for clients that have a static IP address assigned to them but that require RAS procedures to dial E.164 addresses. While Unified CM can signal directly to them without the need of a gatekeeper to resolve their IP addresses, they are not able to signal directly to Unified CM but instead must query the gatekeeper to resolve the E.164 address they are trying to dial (thus, asynchronous communications). To support these types of clients, you must have at least one gatekeeper-controlled client defined in Unified CM for each zone on the gatekeeper, even if all

the clients actually use static IP addresses. In this case, the non-gatekeeper controlled client may be a "dummy" client that does not actually exist. Its purpose is merely to create the RASAggregator trunk so that the gatekeeper will be able to route calls from the clients to Unified CM. (See [Figure 16-13](#) and [Figure 16-14](#).)

- Non-gatekeeper controlled, synchronous

This type of configuration is used for clients that have a static IP address and are also capable of peer-to-peer signaling (that is, they do not require RAS procedures to dial E.164 numbers). Unified CM signals directly to them, and they signal directly to Unified CM (thus, synchronous communications). No gatekeeper or RASAggregator trunk is needed for this type of client. (See [Figure 16-15](#) and [Figure 16-16](#).)

[Figure 16-11](#) through [Figure 16-16](#) illustrate the call signaling flows used in these three scenarios.

Figure 16-11 Call to Gatekeeper-Controlled Client from Unified CM

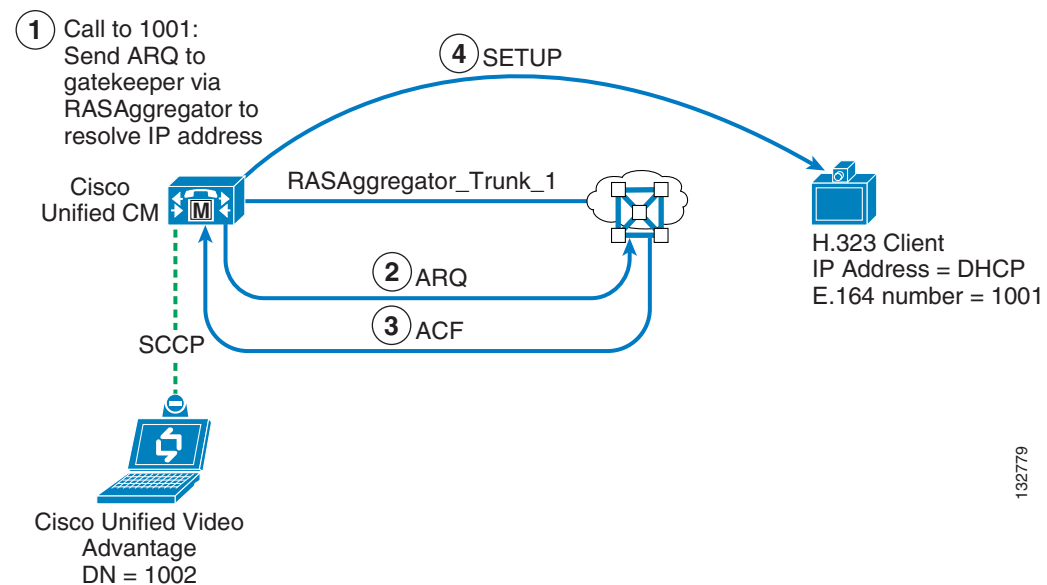


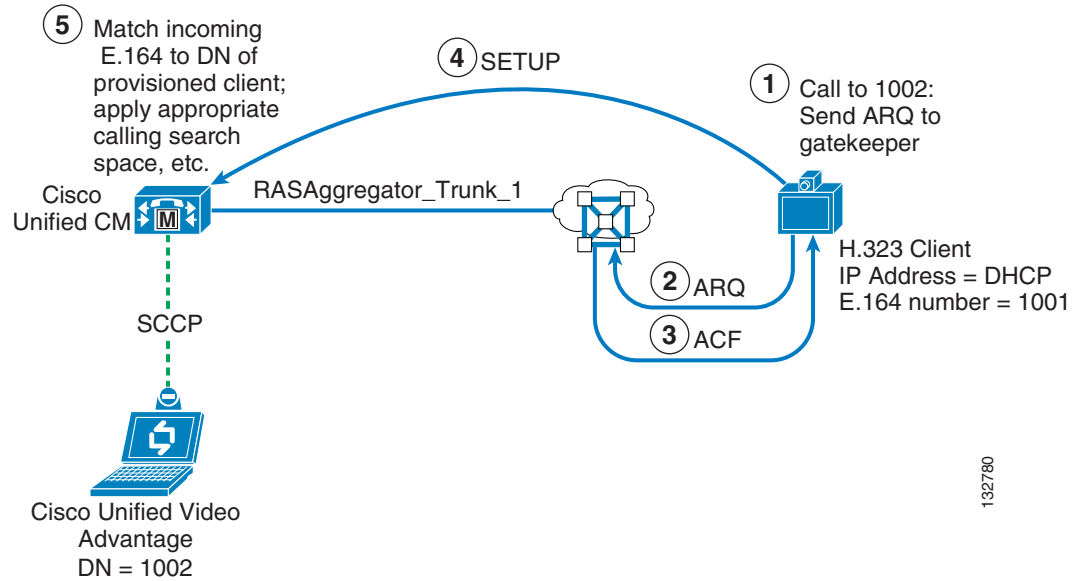
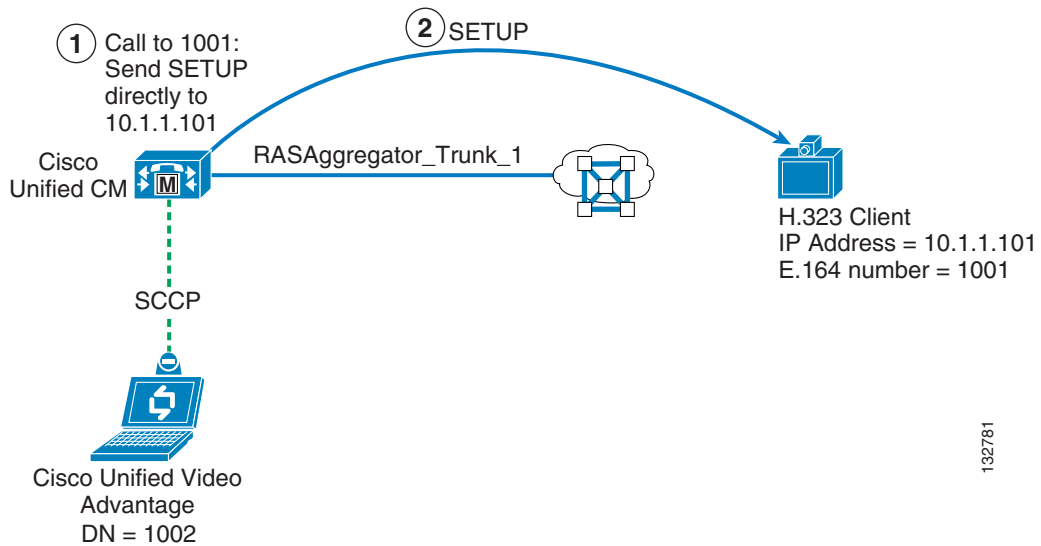
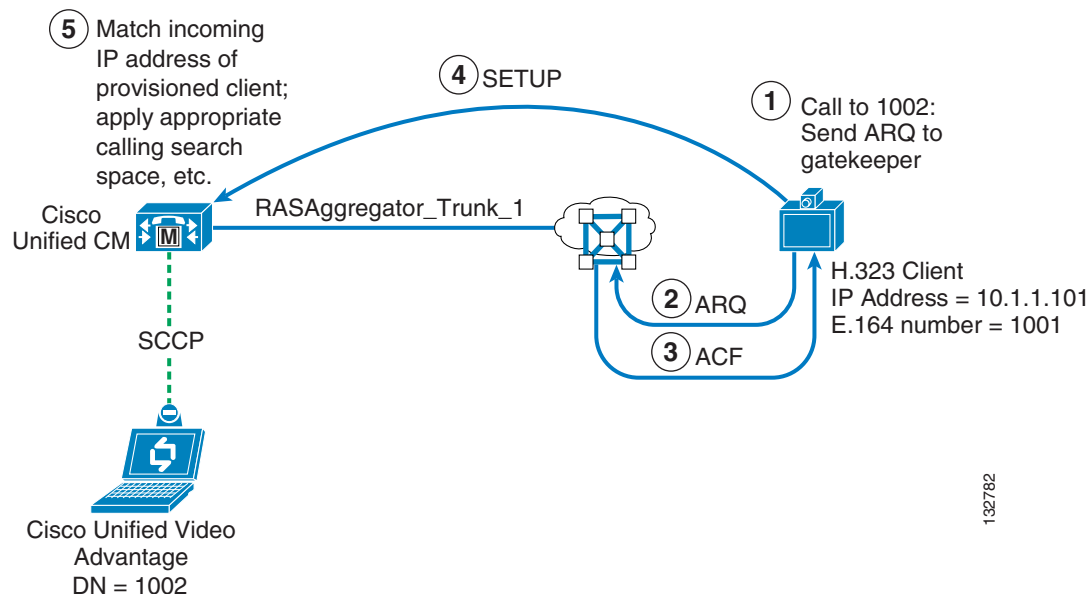
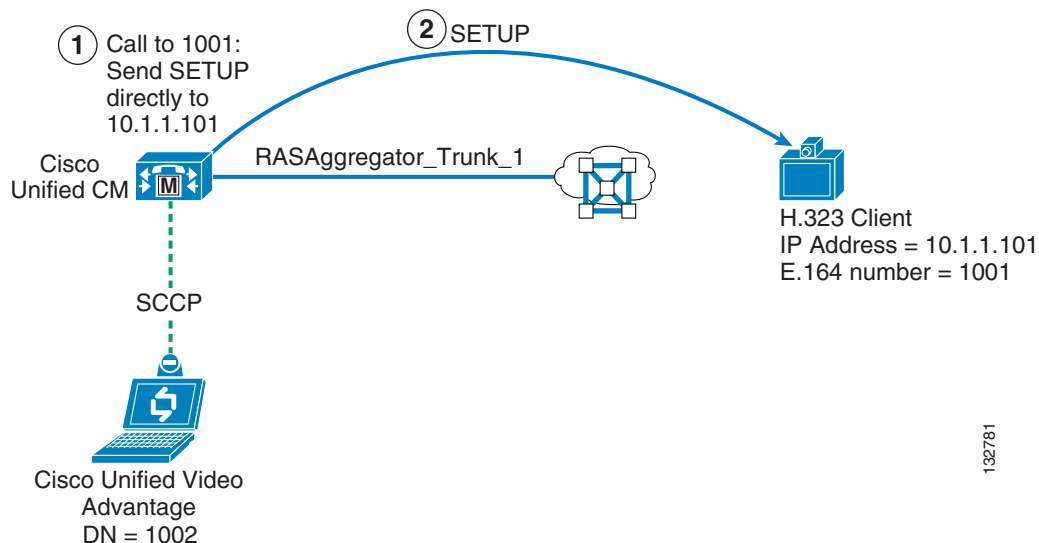
Figure 16-12 Call from Gatekeeper-Controlled Client to Unified CM**Figure 16-13** Call to Non-Gatekeeper Controlled Client from Unified CM (Asynchronous)

Figure 16-14 Call from Non-Gatekeeper Controlled Client to Unified CM (Asynchronous)

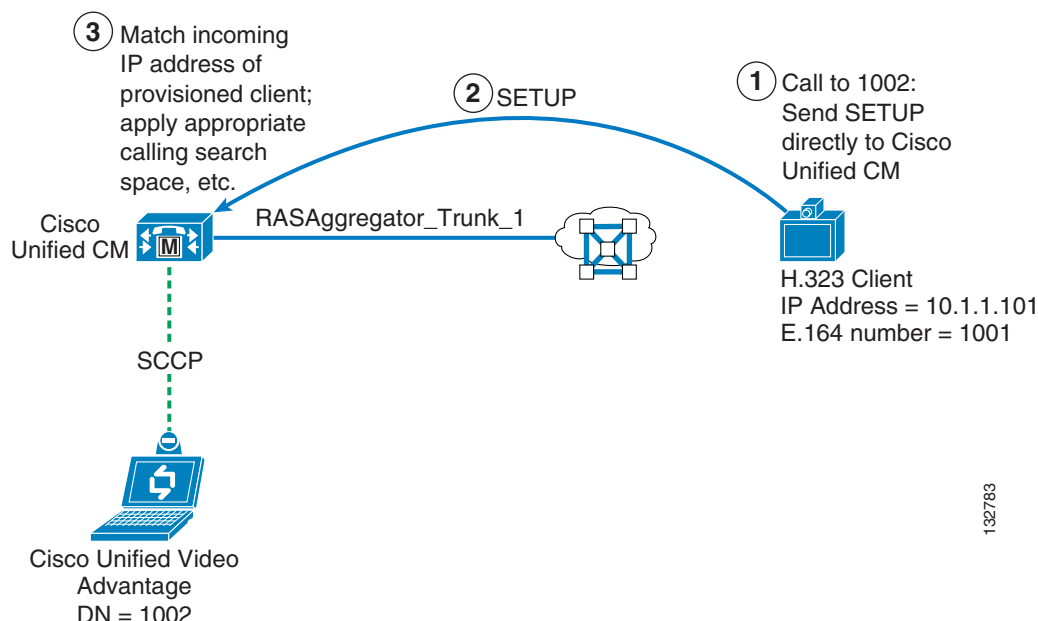


132782

Figure 16-15 Call to Non-Gatekeeper Controlled Client from Unified CM (Synchronous)



132781

Figure 16-16 Call from Non-Gatekeeper Controlled Client to Unified CM (Synchronous)

Gatekeeper-Controlled Clients

When you configure an H.323 client as gatekeeper-controlled, you may enter any alpha-numeric string (such as a descriptive name) in the Device Name field, check the **Gatekeeper-controlled** box, and fill in the following fields:

- **Device Pool**
The device pool you want the client to use. All H.323 clients (whether gatekeeper-controlled or non-gatekeeper controlled) that are registered in the same zone must use the same device pool. If you accidentally assign different device pools across the endpoints, Unified CM will register multiple RASAggregator trunks within the zone, and an inbound call might be rejected by Unified CM if the call is directed to the wrong RASAggregator trunk.
- **Gatekeeper**
A drop-down list of gatekeeper IP addresses. You must define the gatekeeper in Unified CM before configuring any gatekeeper-controlled H.323 clients.
- **Technology Prefix**
The technology prefix used by the RASAggregator trunk to register in the client zone on the gatekeeper. This technology prefix must match what is configured as the default technology prefix on the gatekeeper. All gatekeeper-controlled H.323 clients that are registered in the same zone must use the same technology prefix. If you accidentally assign different technology prefixes across the endpoints, Unified CM will register multiple RASAggregator trunks within the zone, and an inbound call might be rejected by Unified CM if the call is directed to the wrong RASAggregator trunk. Cisco recommends that you use **1#** for this prefix.
- **Zone Name**
The (case-sensitive) name of the client zone as configured in the gatekeeper. All gatekeeper-controlled H.323 clients that are registered in the same zone must use the same zone name. If you accidentally assign different zone names (remember, the field is case sensitive) across

the endpoints, Unified CM will attempt to register multiple RASAggregator trunks with the gatekeeper (but the one with the incorrect zone name will fail to register), and an inbound call might be rejected by Unified CM if the call is directed to the wrong RASAggregator trunk.

Also, you must set the Unified CM service parameter **Send Product ID and Version ID** to **True**. This parameter allows the RASAggregator trunk to register with the gatekeeper as an H323-GW, so that the gatekeeper can direct all H.323 calls to, from, or within the client zone to the RASAggregator trunk.

Non-Gatekeeper Controlled Clients

When provisioning an H.323 client as non-gatekeeper controlled, you must enter the static IP address of the client into the Device Name field and leave all of the settings under the Gatekeeper-controlled section blank (unchecked). Unified CM then uses the static IP address to reach the client any time a call is extended to its directory number.

If the client is configured to use peer-to-peer mode, then no further configuration is required. If the client requires RAS procedures to place calls to E.164 addresses, then you must also configure a dummy gatekeeper-controlled H.323 client in order to create the RASAggregator trunk, by filling in the following fields:

- Device Name

A descriptive name that identifies this client as a dummy client used for the purpose of creating the RASAggregator trunk for the client zone.

- Device Pool

The device pool you chose when configuring the non-gatekeeper controlled H.323 client(s). If the device pool assigned to the dummy client is different than that assigned to the real clients, inbound calls from the real clients might be rejected by Unified CM.

- Gatekeeper

A drop-down list of gatekeeper IP addresses. You must define the gatekeeper in Unified CM before configuring the dummy gatekeeper-controlled H.323 client.

- Technology Prefix

The technology prefix used by the RASAggregator trunk to register in the client zone on the gatekeeper. This technology prefix must match what is configured as the default technology prefix on the gatekeeper. Cisco recommends that you use **1#** for this prefix.

- Zone Name

The (case-sensitive) name of the client zone as configured in the gatekeeper.

Also, you must set the Unified CM service parameter **Send Product ID and Version ID** to **True**. This parameter allows the RASAggregator trunk to register with the gatekeeper as an H323-GW, so that the gatekeeper can direct all H.323 calls to, from, or within the client zone to the RASAggregator trunk.

Provisioning H.323 MCUs

H.323 MCUs are provisioned in Unified CM as H.323 gateways, and then route patterns are configured to extend calls to these devices. When provisioning an H.323 gateway, you must enter the static IP address and TCP signaling port of the MCU into the Device Name field. Unified CM then uses the static IP address and TCP port to reach the MCU any time a call matches the route pattern(s) associated with it.

**Note**

The Cisco Unified Videoconferencing 3500 Series MCUs do not listen on TCP port 1720 by default. (The Cisco Unified Videoconferencing 3500 Series MCUs listen on port 2720 by default.) You must verify which TCP port they are listening on, and either change it to 1720 or provision the correct port in Unified CM.

If the MCU is configured to use peer-to-peer mode, then no further configuration is required. (Cisco Unified Videoconferencing MCUs do not currently support peer-to-peer mode, but some third-party MCUs do.) If the MCU requires RAS procedures to place calls to E.164 addresses, then you must also configure a dummy gatekeeper-controlled H.323 client in order to create the RASAggregator trunk, by filling in the following fields:

- Device Name

A descriptive name that identifies this client as a dummy client used for the purpose of creating the RASAggregator trunk for the MCU zone.

- Device Pool

The device pool you chose when configuring the H.323 gateway representing the MCU. If the device pool assigned to the dummy client is different than that assigned to the H.323 gateway device representing the MCU, inbound calls from the MCU might be rejected by Unified CM.

- Gatekeeper

A drop-down list of gatekeeper IP addresses. You must define the gatekeeper in Unified CM before configuring the dummy gatekeeper-controlled H.323 client.

- Technology Prefix

The technology prefix used by the RASAggregator trunk to register in the MCU zone on the gatekeeper. This technology prefix must match what is configured as the default technology prefix on the gatekeeper. Cisco recommends that you use **1#** for this prefix.

- Zone Name

The (case-sensitive) name of the MCU zone as configured in the gatekeeper.

Also, you must set the Unified CM service parameter **Send Product ID and Version ID** to **True**. This parameter allows the RASAggregator trunk to register with the gatekeeper as an H323-GW, so that the gatekeeper can direct all H.323 calls to, from, or within the MCU zone to the RASAggregator trunk.

MCU Service Prefixes

H.323 MCUs can use either E.164 addresses or technology prefixes (also referred to as service prefixes in the MCU) as the dial-in number(s) to reach reservationless or scheduled H.323 conferences running on them. Cisco recommends that you configure the MCUs to use E.164 addresses by setting the MCU Mode to **MCU** instead of **Gateway** in the MCU administration screens. If the **MCU** setting is not available on the model of MCU you are using, then you must use the following special configuration to properly route calls placed from other H.323 endpoints to the MCU:

If the MCU is configured in **Gateway** mode or is another vendor's MCU that (for whatever reason) requires its conference IDs to register as technology prefixes instead of as E.164 addresses, then the service prefix(s) of the MCU must begin with a # character. For example, if the MCU's service prefix is 8005551212, then you must provision the service prefix on the MCU as #8005551212. Thus, when other H.323 endpoints dial 8005551212, the gatekeeper will not find a matching technology prefix registered and will instead route the call to the RASAggregator trunk that is registered with the default technology prefix in the zone of the endpoint that is placing the call. Unified CM must then prepend the # character to the beginning of the called number before extending the call to the MCU.

This character is prepended on the route pattern(s) associated with the H.323 gateway representing the MCU. Calls to the MCU from SCCP clients will therefore also have this # character prepended to the calling number.

If the MCU is configured in **MCU** mode or is another vendor's MCU that uses E.164 addresses for its conference IDs, then you do not have to prepend the # character. Also note that, if the MCU uses peer-to-peer mode and hence does not need to register its technology prefixes with any gatekeeper, then this situation does not apply and you do not have to prepend a # character.

Provisioning H.320 Gateways

As with H.323 MCUs, H.320 gateways are provisioned in Unified CM as H.323 gateways, and then route patterns are configured to extend calls to these devices. When provisioning an H.323 gateway, you must enter the static IP address and TCP signaling port of the H.320 gateway into the Device Name field. Unified CM then uses the static IP address and TCP port to reach the gateway any time a call matches the route pattern(s) associated with it.



Note

The Cisco Unified Videoconferencing 3500 Series Gateways do not listen on TCP port 1720 by default. (The Cisco Unified Videoconferencing 3500 Series Gateways listen on port 1820 by default.) You must verify which TCP port they are listening on, and either change it to 1720 or provision the correct port in Unified CM.

If the gateway is configured to use peer-to-peer mode, then no further configuration is required. If the gateway requires RAS procedures to place calls to E.164 addresses, then you must also configure a dummy gatekeeper-controlled H.323 client in order to create the RASAggregator trunk, by filling in the following fields:

- **Device Name**
A descriptive name that identifies this client as a dummy client used for the purpose of creating the RASAggregator trunk for the gateway zone.
- **Device Pool**
The device pool you chose when configuring the H.323 gateway representing the H.320 gateway. If the device pool assigned to the dummy client is different than that assigned to the gateway, inbound calls from the gateway might be rejected by Unified CM.
- **Gatekeeper**
A drop-down list of gatekeeper IP addresses. You must define the gatekeeper in Unified CM before configuring the dummy gatekeeper-controlled H.323 client.
- **E.164**
This field requires an entry. Ensure that it is "non-dialable" within Unified CM.
- **Technology Prefix**
The technology prefix used by the RASAggregator trunk to register in the gateway zone on the gatekeeper. This technology prefix must match what is configured as the default technology prefix on the gatekeeper. Cisco recommends that you use **1#** for this prefix.
- **Zone Name**
The (case-sensitive) name of the gateway zone as configured in the gatekeeper.

Also, you must set the Unified CM service parameter **Send Product ID and Version ID** to **True**. This parameter allows the RASAggregator trunk to register with the gatekeeper as an H323-GW, so that the gatekeeper can direct all H.323 calls to, from, or within the gateway zone to the RASAggregator trunk.

Gateway Service Prefixes

H.320 gateways use technology prefixes (also referred to as service prefixes in the gateway) as the prefix that users should dial to reach an ISDN destination. For calls to route correctly, you must configure the service prefix(s) of the gateway to begin with a # character. For example, if the gateway's service prefix that clients dial to reach an ISDN number is 9, then you must provision the service prefix on the gateway as #9. In this way, when H.323 clients dial 9 plus the PSTN number (such as 918005551212), the gatekeeper will not find a matching technology prefix registered and will instead route the call to the Unified CM trunk that is registered with the default technology prefix. Unified CM must then prepend the # character to the beginning of the called number before extending the call to the gateway. Note that, if the gateway uses peer-to-peer mode and hence does not need to register its technology prefixes with any gatekeeper, then this situation does not apply and you do not have to prepend a # character.

Gatekeeper Zone Configuration

The preceding sections discuss how to provision the endpoints in Unified CM Administration. You must also configure the endpoint gatekeeper(s) with the appropriate zone definitions. You must configure a zone for each type of endpoint (client, MCU, or gateway) and, optionally, for each device pool associated with these endpoints in Unified CM.

Each zone is configured to route all calls placed to, from, or within the zone to the RASAggregator trunk registered in that zone. You configure the zones on the endpoint gatekeeper by using the following command syntax:

```
zone local <zone_name> <domain_name> <ip_address> invia <zone_name>
outvia <zone_name> enable-intrazone
```

The command argument **invia** applies to calls placed to the zone from any other zone, **outvia** applies to calls placed from the zone to any other zone, and **enable-intrazone** applies to calls placed within the zone. The following sections illustrate the use of these commands.

Client Zones

The number of client zones you have to configure within each endpoint gatekeeper depends on the following factors:

- The device pools to which the H.323 clients are associated
The device pool determines which Unified CM servers are primary, secondary, and tertiary servers for each H.323 client. If you assign all H.323 clients to the same device pool, then you need to define only a single client zone in the endpoint gatekeeper. In other words, for each device pool used by H.323 clients, you must configure a separate client zone in the gatekeeper.
- Whether the endpoint gatekeeper provides services for a single Unified CM cluster or multiple Unified CM clusters

Each client zone is configured to route calls to a particular RASAggregator trunk. Therefore, if one endpoint gatekeeper is used to service multiple Unified CM clusters, then you must define a separate client zone for each cluster that the gatekeeper services.

To illustrate, the following three examples show how client zones may be configured. [Example 16-1](#) shows a single client zone defined for a single Unified CM cluster in which all H.323 clients are associated with the same device pool. [Example 16-2](#) shows a single Unified CM cluster in which the H.323 clients are divided between two different device pools. [Example 16-3](#) shows two Unified CM clusters in which the H.323 clients are divided between two different device pools in each cluster.

**Note**

Some of the commands shown in the following examples are the default values applied in the Cisco IOS Gatekeeper and, therefore, would not have to be configured explicitly, nor would they appear in the running configuration. They are included here for thoroughness but are marked by a ! at the beginning of the command line.

Example 16-1 Client Zone for a Single Unified CM Cluster and Single Device Pool

```
gatekeeper
zone local clients domain.com invia clients outvia clients enable-intrazone
gw-type-prefix 1# default-technology
no use-proxy clients default inbound-to terminal
no use-proxy clients default outbound-from terminal
! no arg reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Example 16-2 Client Zones for a Single Unified CM Cluster and Two Device Pools

```
gatekeeper
zone local dp1-clients domain.com invia dp1-clients outvia dp1-clients enable-intrazone
zone local dp2-clients domain.com invia dp2-clients outvia dp2-clients enable-intrazone
gw-type-prefix 1# default-technology
no use-proxy dp1-clients default inbound-to terminal
no use-proxy dp1-clients default outbound-from terminal
no use-proxy dp2-clients default inbound-to terminal
no use-proxy dp2-clients default outbound-from terminal
! no arg reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Example 16-3 Client Zones for Two Unified CM Clusters with Two Device Pools per Cluster

```
gatekeeper
zone local clstr1-dp1-clients domain.com invia clstr1-dp1-clients outvia
clstr1-dp1-clients enable-intrazone
zone local clstr1-dp2-clients domain.com invia clstr1-dp2-clients outvia
clstr1-dp2-clients enable-intrazone
zone local clstr2-dp1-clients domain.com invia clstr2-dp1-clients outvia
clstr2-dp1-clients enable-intrazone
zone local clstr2-dp2-clients domain.com invia clstr2-dp2-clients outvia
clstr2-dp2-clients enable-intrazone
gw-type-prefix 1# default-technology
no use-proxy clstr1-dp1-clients default inbound-to terminal
no use-proxy clstr1-dp1-clients default outbound-from terminal
no use-proxy clstr1-dp2-clients default inbound-to terminal
no use-proxy clstr1-dp2-clients default outbound-from terminal
no use-proxy clstr2-dp1-clients default inbound-to terminal
no use-proxy clstr2-dp1-clients default outbound-from terminal
no use-proxy clstr2-dp2-clients default inbound-to terminal
no use-proxy clstr2-dp2-clients default outbound-from terminal
```

```
! no arg reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Disabling The Use of Proxy

The Cisco IOS Gatekeeper, formerly known as the Cisco Multimedia Conference Manager (MCM), previously offered an H.323 proxy function that has been at End of Life (EOL) for some time and is not compatible with Unified CM, but the commands in the gatekeeper to use a proxy for all calls to and from terminals (clients) are still enabled by default. You must disable this function for each client zone by using the following command syntax:

```
gatekeeper
no use-proxy <zone_name> default [inbound-to | outbound-from] terminals
```

The Cisco MCM proxy was replaced by a new solution called the Cisco IOS Multiservice IP-to-IP Gateway and the associated via-zone-enabled Cisco IOS Gatekeeper. This document does not discuss the IP-to-IP Gateway, but Cisco Unified CM Release 4.1 and above leverages the via-zone and IP-to-IP gateway constructs by registering its RASAggregator trunks with the gatekeeper, effectively mimicking an IP-to-IP gateway so that the gatekeeper will route all invia, outvia, and enable-intrazone calls to the RASAggregator trunk as if it were an IP-to-IP gateway.

Client Zone Prefixes

For H.323 client zones, there is no need to configure zone prefixes or technology prefixes of any kind, except for the default technology prefix. Instead, the **invia**, **outvia**, **enable-intrazone**, and **gw-type-prefix <I#> default-technology** commands ensure that all calls placed are routed to the RASAggregator trunk associated with the zone in which the call originated.

MCU Zones

The number of MCU zones you have to configure within each endpoint gatekeeper depends on the following factors:

- The device pools to which the MCUs are associated

The device pool determines which Unified CM servers are primary, secondary, and tertiary servers for each MCU. If you assign all MCUs to the same device pool, then you need to define only a single MCU zone in the endpoint gatekeeper. In other words, for each device pool used by MCUs, you must configure a separate MCU zone in the gatekeeper.

- Whether the endpoint gatekeeper provides services for a single Unified CM cluster or multiple Unified CM clusters

Each MCU zone is configured to route calls to a particular RASAggregator trunk. Therefore, if one endpoint gatekeeper is used to service multiple Unified CM clusters, then you must define a separate MCU zone for each cluster that the gatekeeper services.

To illustrate, the following three examples show how MCU zones may be configured. [Example 16-4](#) shows a single MCU zone defined for a single Unified CM cluster in which all MCUs are associated with the same device pool. [Example 16-5](#) shows a single Unified CM cluster in which the MCUs are divided between two different device pools. [Example 16-6](#) shows two Unified CM clusters in which the MCUs are divided between two different device pools.

**Note**

Some of the commands shown in the following examples are the default values applied in the Cisco IOS Gatekeeper and, therefore, would not have to be configured explicitly, nor would they appear in the running configuration. They are included here for thoroughness but are marked by a ! at the beginning of the command line.

Example 16-4 MCU Zone for a Single Unified CM Cluster and Single Device Pool

```
gatekeeper
zone local MCUs domain.com invia MCUs outvia MCUs enable-intrazone
gw-type-prefix 1# default-technology
! no use-proxy MCUs default inbound-to [MCU | gateway]
! no use-proxy MCUs default outbound-from [MCU | gateway]
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Example 16-5 MCU Zones for a Single Unified CM Cluster and Two Device Pools

```
gatekeeper
zone local dp1-MCUs domain.com invia dp1-MCUs outvia dp1-MCUs enable-intrazone
zone local dp2-MCUs domain.com invia dp2-MCUs outvia dp2-MCUs enable-intrazone
gw-type-prefix 1# default-technology
! no use-proxy dp1-MCUs default inbound-to [MCU | gateway]
! no use-proxy dp1-MCUs default outbound-from [MCU | gateway]
! no use-proxy dp2-MCUs default inbound-to [MCU | gateway]
! no use-proxy dp2-MCUs default outbound-from [MCU | gateway]
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Example 16-6 MCU Zones for Two Unified CM Clusters with Two Device Pools per Cluster

```
gatekeeper
zone local clstr1-dp1-MCUs domain.com invia clstr1-dp1-MCUs outvia clstr1-dp1-MCUs
enable-intrazone
zone local clstr1-dp2-MCUs domain.com invia clstr1-dp2-MCUs outvia clstr1-dp2-MCUs
enable-intrazone
zone local clstr2-dp1-MCUs domain.com invia clstr2-dp1-MCUs outvia clstr2-dp1-MCUs
enable-intrazone
zone local clstr2-dp2-MCUs domain.com invia clstr2-dp2-MCUs outvia clstr2-dp2-MCUs
enable-intrazone
gw-type-prefix 1# default-technology
! no use-proxy clstr1-dp1-MCUs default inbound-to [MCU | gateway]
! no use-proxy clstr1-dp1-MCUs default outbound-from [MCU | gateway]
! no use-proxy clstr1-dp2-MCUs default inbound-to [MCU | gateway]
! no use-proxy clstr1-dp2-MCUs default outbound-from [MCU | gateway]
! no use-proxy clstr2-dp1-MCUs default inbound-to [MCU | gateway]
! no use-proxy clstr2-dp1-MCUs default outbound-from [MCU | gateway]
! no use-proxy clstr2-dp2-MCUs default inbound-to [MCU | gateway]
! no use-proxy clstr2-dp2-MCUs default outbound-from [MCU | gateway]
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Disabling The Use of Proxy

By default, the Cisco IOS Gatekeeper is set to not use a proxy for calls to and from MCUs or gateways. However, if you have enabled the use of proxy for those types of endpoints, you must disable it for each MCU zone by using the following command syntax:

```
gatekeeper
no use-proxy <zone_name> default [inbound-to | outbound-from] [MCU | gateway]
```

If your MCU is registering as an MCU, then use the **MCU** argument at the end of the **no use-proxy** command; if your MCU is registering as a gateway, then use the **gateway** argument instead.

MCU Zone Prefixes

For H.323 MCU zones, there is no need to configure zone prefixes or technology prefixes of any kind, except for the default technology prefix. Instead, the **invia**, **outvia**, **enable-intrazone**, and **gw-type-prefix <1#> default-technology** commands ensure that all calls placed are routed to the RASAggregator trunk associated with the zone in which the call originated.

If your MCUs are registering their service prefixes as technology prefixes instead of E.164 addresses, use the special configuration described previously for prepending a # character to the MCU's service prefixes (see [MCU Service Prefixes, page 16-32](#)). Due to the way the Cisco IOS Gatekeeper selects a via-zone for calls to a technology prefix, when the endpoint dials the service prefix of the MCU, the call will fail if the gatekeeper finds a matching technology prefix registered. You must ensure that the client does not dial the # character, so that the gatekeeper will not find a matching technology prefix and will instead route the call to the RASAggregator trunk associated with the zone in which the call originated.

H.320 Gateway Zones

The number of H.320 gateway zones you have to configure within each endpoint gatekeeper depends on the following factors:

- The device pools to which the H.320 gateways are associated

The device pool determines which Unified CM servers are primary, secondary, and tertiary servers for each H.320 gateway. If you assign all gateways to the same device pool, then you need to define only a single gateway zone in the endpoint gatekeeper. In other words, for each device pool used by H.320 gateways, you must configure a separate gateway zone in the gatekeeper.

- Whether the endpoint gatekeeper provides services for a single Unified CM cluster or multiple Unified CM clusters

Each gateway zone is configured to route calls to a particular RASAggregator trunk. Therefore, if one endpoint gatekeeper is used to service multiple Unified CM clusters, then you must define a separate gateway zone for each cluster that the gatekeeper services.

To illustrate, the following three examples show how gateway zones may be configured. [Example 16-7](#) shows a single gateway zone defined for a single Unified CM cluster in which all H.320 gateways are associated with the same device pool. [Example 16-8](#) shows a single Unified CM cluster in which the gateways are divided between two different device pools. [Example 16-9](#) shows two Unified CM clusters in which the gateways are divided between two different device pools.



Note

Some of the commands shown in the following examples are the default values applied in the Cisco IOS Gatekeeper and, therefore, would not have to be configured explicitly, nor would they appear in the running configuration. They are included here for thoroughness but are marked by a ! at the beginning of the command line.

Example 16-7 Gateway Zone for a Single Unified CM Cluster and Single Device Pool

```

gatekeeper
zone local gateways domain.com invia gateways outvia gateways enable-intrazone
gw-type-prefix 1# default-technology
! no use-proxy gateways default inbound-to gateway
! no use-proxy gateways default outbound-from gateway
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown

```

Example 16-8 Gateway Zones for a Single Unified CM Cluster and Two Device Pools

```

gatekeeper
zone local dp1-gateways domain.com invia dp1-gateways outvia dp1-gateways enable-intrazone
zone local dp2-gateways domain.com invia dp2-gateways outvia dp2-gateways enable-intrazone
gw-type-prefix 1# default-technology
! no use-proxy dp1-gateways default inbound-to gateway
! no use-proxy dp1-gateways default outbound-from gateway
! no use-proxy dp2-gateways default inbound-to gateway
! no use-proxy dp2-gateways default outbound-from gateway
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown

```

Example 16-9 Gateway Zones for Two Unified CM Clusters with Two Device Pools per Cluster

```

gatekeeper
zone local clstr1-dp1-gateways domain.com invia clstr1-dp1-gateways outvia
clstr1-dp1-gateways enable-intrazone
zone local clstr1-dp2-gateways domain.com invia clstr1-dp2-gateways outvia
clstr1-dp2-gateways enable-intrazone
zone local clstr2-dp1-gateways domain.com invia clstr2-dp1-gateways outvia
clstr2-dp1-gateways enable-intrazone
zone local clstr2-dp2-gateways domain.com invia clstr2-dp2-gateways outvia
clstr2-dp2-gateways enable-intrazone
gw-type-prefix 1# default-technology
! no use-proxy clstr1-dp1-gateways default inbound-to gateway
! no use-proxy clstr1-dp1-gateways default outbound-from gateway
! no use-proxy clstr1-dp2-gateways default inbound-to gateway
! no use-proxy clstr1-dp2-gateways default outbound-from gateway
! no use-proxy clstr2-dp1-gateways default inbound-to gateway
! no use-proxy clstr2-dp1-gateways default outbound-from gateway
! no use-proxy clstr2-dp2-gateways default inbound-to gateway
! no use-proxy clstr2-dp2-gateways default outbound-from gateway
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown

```

Disabling The Use of Proxy

By default, the Cisco IOS Gatekeeper is set to not use a proxy for calls to and from gateways. However, if you have enabled the use of proxy for those types of endpoints, you must disable it for each H.320 gateway zone by using the following command syntax:

```

gatekeeper
no use-proxy <zone_name> default [inbound-to | outbound-from] gateway

```

Gateway Zone Prefixes

There is no need to configure zone prefixes of any kind for H.320 gateway zones. Instead, the **invia**, **outvia**, **enable-intrazone**, and **gw-type-prefix <I#> default-technology** commands ensure that all calls placed are routed to the RASAggregator trunk associated with the zone in which the call originated.

You must also use the special configuration described previously for prepending a # character to the gateway's service prefixes (see [Gateway Service Prefixes, page 16-34](#)). Due to the way the Cisco IOS Gatekeeper selects a via-zone for calls to a technology prefix, when the endpoint dials the service prefix of the gateway, the call will fail if the gatekeeper finds a matching technology prefix registered. You must ensure that the client does not dial the # character, so that the gatekeeper will not find a matching technology prefix and will instead route the call to the RASAggregator trunk associated with the zone in which the call originated.

Zone Subnets

As mentioned previously, the H.323 specification permits a single gatekeeper to manage multiple zones. However, the gatekeeper needs a way to decide which zone an endpoint should be placed in when it receives a Registration Request (RRQ) from that device. The RRQ message contains a Gatekeeper Identifier field that enables the endpoint to indicate the zone in which it would like to register. However, many H.323 video endpoints do not populate this field, and if the gatekeeper has multiple zones defined, it will not know which zone to place the endpoint into. Therefore, you must use of the **zone subnet** command to tell the gatekeeper which zone to associate with the endpoint. This command defines which IP addresses or IP address ranges are permitted to register in each zone. The command syntax requires that you enter a network mask. Therefore, you can specify either a particular host address by entering a 32-bit (/32) network mask or a range of addresses by specifying a smaller network mask.

Because MCUs, H.320 gateways, and Unified CM servers typically use fixed IP addresses but H.323 clients can use DHCP addresses, Cisco recommends that you define **zone subnet** commands only for the MCU and gateway zones but leave the client zones open so that any IP address is permitted in them. Note that you must also permit the Unified CM servers to register in the MCU and gateway zones, as illustrated in [Example 16-10](#).



Note

Some of the commands shown in the following example are the default values applied in the Cisco IOS Gatekeeper and, therefore, would not have to be configured explicitly, nor would they appear in the running configuration. They are included here for thoroughness but are marked by a ! at the beginning of the command line.

Example 16-10 Defining Zone Subnets

```
gatekeeper
no zone subnet MCUs default enable
zone subnet MCUs [MCUs_IP_addr]/32 enable
zone subnet MCUs [RASAggregators_IP_addr]/32 enable
no zone subnet gateways default enable
zone subnet gateways [gateways_IP_addr]/32 enable
zone subnet gateways [RASAggregators_IP_addr]/32 enable
! zone subnet clients default enable
no zone subnet clients [MCUs_IP_addr]/32 enable
no zone subnet clients [gateways_IP_addr]/32 enable
```

The configuration in [Example 16-10](#) explicitly permits the MCU and the RASAggregator for the MCU zone to register in the MCU zone, and it explicitly permits the gateway and RASAggregator for the gateway zone to register in the gateway zone. It also explicitly denies the MCU and gateway from registering in the client zone, while implicitly permitting all other IP addresses (including the RASAggregator for the client zone) to register in the client zone.

Endpoint Time to Live

Endpoints send lightweight Registration Requests (RRQs) to their gatekeeper periodically to maintain their registration status. The frequency with which they send these RRQs is referred to as the Time to Live (TTL) value. The endpoint may specify the TTL it wishes to use in the body of its RRQs. The gatekeeper may then honor the endpoint's requested TTL value by echoing it in the Registration Confirm (RCF) response or, alternatively, may override the endpoint's request by specifying a different TTL value in the RCF.

If the TTL value is not specified in the RRQ, the gatekeeper should specify one in its RCF response. The endpoint should then honor the TTL specified by the gatekeeper. The Cisco IOS Gatekeeper honors all TTL values specified by the endpoints. However, many H.323 video endpoints do not specify a TTL value in their RRQs. In such cases, the Cisco IOS Gatekeeper defaults to specifying a TTL value of 1800 seconds (30 minutes). The Cisco IOS Gatekeeper will flush the endpoint's registration after three TTL intervals have passed without receiving any messages from the endpoint ($3 * 30 \text{ minutes} = 90 \text{ minutes}$).

A large TTL value can cause problems with H.323 clients that do not use static IP addresses. For example, with the default TTL value of 1800 seconds, if you disconnect the client from the network and move it to another location in which it receives a different DHCP address, it will fail to register with the gatekeeper (Registration Reject (RRJ) cause value "duplicate alias") until three TTL intervals have passed, and the gatekeeper will flush that endpoint's original registration.

Therefore, Cisco recommends that you consider reducing the TTL value to as low a number as possible without causing any negative effect on your network. The Cisco IOS Gatekeeper permits you to set the TTL value anywhere in the range of 60 seconds to 3600 seconds. In most cases, 60 seconds should work well. However, if your gatekeeper is already heavily utilized, adjusting the TTL from the default of 1800 seconds to 60 seconds might cause it to become overwhelmed.

Use the following command syntax to set the TTL value:

```
gatekeeper
endpoint ttl <seconds>
```

Summary of Endpoint Gatekeepers

This section summarizes some key points to remember about endpoint gatekeepers and provides some example configurations that combine techniques used in the previous examples.

- Configure a separate zone in the endpoint gatekeeper for each type of endpoint (clients, MCUs, and H.320 gateways). If the endpoints are associated with multiple device pools, configure multiple zones for each type of endpoint.
- Configure a RASAggregator trunk to register in each zone. This trunk is automatically created when you configure gatekeeper-controlled H.323 clients in Unified CM Administration. However, for non-gatekeeper-controlled H.323 clients, H.323 MCUs, and H.320 gateways, you must configure a dummy gatekeeper-controlled H.323 client in order to create the RASAggregator trunk for that zone.
- Set the service parameter **Send Product ID and Version ID** to **True** in order for the RASAggregator trunk to register with the gatekeeper as an IP-to-IP gateway. This setting enables the RASAggregator to be selected by the gatekeeper for all calls placed to, from, or within each zone due to the use of the **invia**, **outvia**, **enable-intrazone**, and **gw-type-prefix <I#> default-technology** commands applied to each local zone definition.

- You do not have to associate any zone prefixes for any of the endpoint zones. No matter what the endpoint dials, the gatekeeper should not find a matching zone prefix or technology prefix but should instead route the call to the RASAggregator trunk associated with the zone from which the call originated. To avoid having the gatekeeper accidentally match the dialed number to the technology prefix of your MCUs or gateways, mask all MCU and gateway service prefixes with a # character, and then prepend the # character in the route pattern associated with that MCU or gateway.
- Configure zone subnets if any of the H.323 endpoints do not support the ability to specify the Gatekeeper Identifier (name of the zone) with which they wish to register.
- Disable the use of the old MCM Proxy for all zones.
- Set the endpoint registration Time to Live (TTL) to as low of a value as you can without creating undo stress on the gatekeeper. In extreme cases where the gatekeeper is serving hundreds of endpoint registrations, setting the TTL to 60 seconds might cause an unmanageable amount of RAS traffic. In smaller environments, setting it to 60 seconds should work well.

Example 16-11 shows a configuration for an endpoint gatekeeper servicing a single Unified CM cluster in which a single device pool is used to service all H.323 video endpoint types.

**Note**

Some of the commands shown in the following examples are the default values applied in the Cisco IOS Gatekeeper and, therefore, would not have to be configured explicitly, nor would they appear in the running configuration. They are included here for thoroughness but are marked by a ! at the beginning of the command line.

Example 16-11 Endpoint Gatekeeper Configuration for a Single Cluster and a Single Device Pool

```
gatekeeper
zone local clients domain.com invia clients outvia clients enable-intrazone
zone local MCUs domain.com invia MCUs outvia MCUs enable-intrazone
zone local gateways domain.com invia gateways outvia gateways enable-intrazone
! zone subnet clients default enable
no zone subnet clients [MCUs_IP_addr]/32 enable
no zone subnet clients [gateways_IP_addr]/32 enable
no zone subnet MCUs default enable
zone subnet MCUs [MCUs_IP_addr]/32 enable
zone subnet MCUs [RASAggregators_IP_addr]/32 enable
no zone subnet gateways default enable
zone subnet gateways [gateways_IP_addr]/32 enable
zone subnet gateways [RASAggregators_IP_addr]/32 enable
no use-proxy clients inbound-to terminals
no use-proxy clients outbound-from terminals
! no use-proxy MCUs inbound-to [MCU | gateway]
! no use-proxy MCUs outbound-from [MCU | gateway]
! no use-proxy gateways inbound-to gateway
! no use-proxy gateways outbound-from gateway
gw-type-prefix 1# default-technology
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown
```

Example 16-12 shows a configuration for an endpoint gatekeeper servicing two Unified CM clusters. Each cluster has two different device pools for its H.323 video endpoints.

Example 16-12 Endpoint Gatekeeper Configuration for Two Clusters and a Two Device Pools

```
gatekeeper
zone local clstr1-dp1-clients domain.com invia clstr1-dp1-clients outvia
clstr1-dp1-clients enable-intrazone
```

```

zone local clstr1-dp1-MCUs domain.com invia clstr1-dp1-MCUs outvia clstr1-dp1-MCUs
enable-intrazone
zone local clstr1-dp1-gateways domain.com invia clstr1-dp1-gateways outvia
clstr1-dp1-gateways enable-intrazone
zone local clstr1-dp2-clients domain.com invia clstr1-dp2-clients outvia
clstr1-dp2-clients enable-intrazone
zone local clstr1-dp2-MCUs domain.com invia clstr1-dp2-MCUs outvia clstr1-dp2-MCUs
enable-intrazone
zone local clstr1-dp2-gateways domain.com invia clstr1-dp2-gateways outvia
clstr1-dp2-gateways enable-intrazone
zone local clstr2-dp1-clients domain.com invia clstr2-dp1-clients outvia
clstr2-dp1-clients enable-intrazone
zone local clstr2-dp1-MCUs domain.com invia clstr1-dp2-MCUs outvia clstr2-dp1-MCUs
enable-intrazone
zone local clstr2-dp1-gateways domain.com invia clstr2-dp1-gateways outvia
clstr2-dp1-gateways enable-intrazone
zone local clstr2-dp2-clients domain.com invia clstr2-dp2-clients outvia
clstr2-dp2-clients enable-intrazone
zone local clstr2-dp2-MCUs domain.com invia clstr2-dp2-MCUs outvia clstr2-dp2-MCUs
enable-intrazone
zone local clstr2-dp2-gateways domain.com invia clstr2-dp2-gateways outvia
clstr2-dp2-gateways enable-intrazone
! zone subnet clstr1-dp1-clients default enable
no zone subnet clstr1-dp1-clients [clstr1-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr1-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr2-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr2-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr1-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr1-dp2 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr2-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp1-clients [clstr2-dp2 gateways_IP_addr]/32 enable
! zone subnet clstr1-dp2-clients default enable
no zone subnet clstr1-dp2-clients [clstr1-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr1-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr2-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr2-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr1-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr1-dp2 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr2-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp2-clients [clstr2-dp2 gateways_IP_addr]/32 enable
! zone subnet clstr2-dp1-clients default enable
no zone subnet clstr2-dp1-clients [clstr1-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr1-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr2-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr2-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr1-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr1-dp2 gateways_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr2-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr2-dp1-clients [clstr2-dp2 gateways_IP_addr]/32 enable
zone subnet clstr2-dp2-clients default enable
no zone subnet clstr2-dp2-clients [clstr1-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr1-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr2-dp1 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr2-dp2 MCUs_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr1-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr1-dp2 gateways_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr2-dp1 gateways_IP_addr]/32 enable
no zone subnet clstr2-dp2-clients [clstr2-dp2 gateways_IP_addr]/32 enable
no zone subnet clstr1-dp1-MCUs default enable
zone subnet clstr1-dp1-MCUs [clstr1-dp1 MCUs_IP_addr]/32 enable
zone subnet clstr1-dp1-MCUs [clstr1-dp1 RASAggregators_IP_addr]/32 enable
no zone subnet clstr1-dp2-MCUs default enable
zone subnet clstr1-dp2-MCUs [clstr1-dp2 MCUs_IP_addr]/32 enable
zone subnet clstr1-dp2-MCUs [clstr1-dp2 RASAggregators_IP_addr]/32 enable

```

```

no zone subnet clstr2-dp1-MCUs default enable
zone subnet clstr2-dp1-MCUs [clstr2-dp1 MCUs_IP_addr]/32 enable
zone subnet clstr2-dp1-MCUs [clstr2-dp1 RASAggregators_IP_addr]/32 enable
no zone subnet clstr2-dp2-MCUs default enable
zone subnet clstr2-dp2-MCUs [clstr2-dp2 MCUs_IP_addr]/32 enable
zone subnet clstr2-dp2-MCUs [clstr2-dp2 RASAggregators_IP_addr]/32 enable
no zone subnet clstr1-dp1-gateways default enable
zone subnet clstr1-dp1-gateways [clstr1-dp1 gateways_IP_addr]/32 enable
zone subnet clstr1-dp1-gateways [clstr1-dp1 RASAggregators_IP_addr]/32 enable
no zone subnet clstr1-dp2-gateways default enable
zone subnet clstr1-dp2-gateways [clstr1-dp2 gateways_IP_addr]/32 enable
zone subnet clstr1-dp2-gateways [clstr1-dp2 RASAggregators_IP_addr]/32 enable
no zone subnet clstr2-dp1-gateways default enable
zone subnet clstr2-dp1-gateways [clstr2-dp1 gateways_IP_addr]/32 enable
zone subnet clstr2-dp1-gateways [clstr2-dp1 RASAggregators_IP_addr]/32 enable
no zone subnet clstr2-dp2-gateways default enable
zone subnet clstr2-dp2-gateways [clstr2-dp2 gateways_IP_addr]/32 enable
zone subnet clstr2-dp2-gateways [clstr2-dp2 RASAggregators_IP_addr]/32 enable
no use-proxy clstr1-dp1-clients inbound-to terminals
no use-proxy clstr1-dp1-clients outbound-from terminals
no use-proxy clstr1-dp2-clients inbound-to terminals
no use-proxy clstr1-dp2-clients outbound-from terminals
no use-proxy clstr2-dp1-clients inbound-to terminals
no use-proxy clstr2-dp1-clients outbound-from terminals
no use-proxy clstr2-dp2-clients inbound-to terminals
no use-proxy clstr2-dp2-clients outbound-from terminals
! no use-proxy clstr1-dp1-MCUs inbound-to [MCU | gateway]
! no use-proxy clstr1-dp1-MCUs outbound-from [MCU | gateway]
! no use-proxy clstr1-dp2-MCUs inbound-to [MCU | gateway]
! no use-proxy clstr1-dp2-MCUs outbound-from [MCU | gateway]
! no use-proxy clstr2-dp1-MCUs inbound-to [MCU | gateway]
! no use-proxy clstr2-dp1-MCUs outbound-from [MCU | gateway]
! no use-proxy clstr2-dp2-MCUs inbound-to [MCU | gateway]
! no use-proxy clstr2-dp2-MCUs outbound-from [MCU | gateway]
! no use-proxy clstr1-dp1-gateways inbound-to gateway
! no use-proxy clstr1-dp1-gateways outbound-from gateway
! no use-proxy clstr1-dp2-gateways inbound-to gateway
! no use-proxy clstr1-dp2-gateways outbound-from gateway
! no use-proxy clstr2-dp1-gateways inbound-to gateway
! no use-proxy clstr2-dp1-gateways outbound-from gateway
! no use-proxy clstr2-dp2-gateways inbound-to gateway
! no use-proxy clstr2-dp2-gateways outbound-from gateway
gw-type-prefix 1# default-technology
! no arq reject-unknown-prefix
endpoint ttl 60
no shutdown

```

Migrating from Cisco Unified CM 4.0

The process of migrating from the design methodology documented in the *Cisco IP Video Telephony SRND for Cisco Unified CallManager 4.0* to the design methodology presented in this chapter consists of the following fundamental steps:

1. Reconfigure your H.323 clients in Unified CM Administration to take advantage of the new gatekeeper-controlled H.323 client feature.
2. Remove any H.225 gatekeeper-controlled trunks that were used for the H.323 client, MCU, and H.320 gateway zones, and replace them with RASAggregator trunks by creating dummy gatekeeper-controlled H.323 clients.

3. Reconfigure your endpoint gatekeeper(s) to use the new via-zone definition constructs, removing any zone prefixes that might be applied to the H.323 client, MCU, and gateway zones.
4. Modify all MCU and H.320 gateway service prefixes as necessary to include a # character at the beginning of the prefix.

Each of these steps will cause service disruptions to existing H.323 video endpoints, so ensure that these configuration changes are carefully planned and done at a time when the administrator will be able to minimize disruption to the existing H.323 video endpoints.

Applications

Cisco IP Communications provides an expanding portfolio of applications that extend the features of Unified CM and provide advanced capabilities and integration with other communication media. Many of these applications can be used in conjunction with IP Video Telephony devices, even if they do not specifically support video. For instance, Cisco Unified CM Release 4.1 does not support the negotiation of video channels for CTI applications using the TAPI/JTAPI protocols, but that does not necessarily preclude using a CTI application in conjunction with a video call. This section reviews some of the Cisco and third-party applications and discusses whether or not they can be used to provide advanced call treatment for video calls.

CTI Applications

The following applications are based on the Computer Telephony Integration (CTI) interface.

Cisco Emergency Responder

Cisco Emergency Responder (ER) routes emergency (911) calls to the correct Public Safety Answering Point (PSAP). It also provides the PSAP with the correct calling line ID of the originating device so that the PSAP can respond to the correct physical location of the incident and call the party back in the event that the call is disconnected. Cisco ER uses JTAPI to integrate with Unified CM. Emergency calls are routed to Cisco ER via a CTI route point, then Cisco ER decides which PSAP to forward the call to and what calling line ID to display. Cisco ER tracks each endpoint on the network to determine its physical location by using Simple Network Management Protocol (SNMP) and Cisco Discovery Protocol (CDP) to discover the physical port and specific Cisco Catalyst Ethernet switch to which the endpoint is connected. If CDP is not available, Cisco ER can be configured to locate endpoints by their IP subnet instead. Cisco ER then correlates this information with the physical location of the switch and stores the information in its database.

Both Cisco Unified Video Advantage and the Cisco IP Video Phone 7985 support CDP for the purpose of Cisco ER discovery. Cisco Unified Video Advantage does not send CDP messages to the switch directly but relies on its associated Cisco Unified IP Phone for this support. Therefore, if a Video Telephony user dials 911, Cisco ER is able to route the call to the correct PSAP.

Because Sony and Tandberg SCCP endpoints do not support CDP, Cisco ER must track these endpoints by their IP subnet. Cisco ER is therefore able to route the call to the correct PSAP.

Because H.323 videoconferencing clients do not support CDP, Cisco ER must track them by their IP subnet. Cisco ER is therefore able to route the call to the correct PSAP. However, the H.323 device must support the Empty Capabilities Set (ECS) procedure in order to have its call routed by Cisco ER. If the H.323 endpoint does not support receiving an ECS from Unified CM, calls to 911 that are handled by Cisco ER will fail.

Cisco Unified Communications Manager Assistant

Cisco Unified Communications Manager Assistant enables administrative assistants to provide coverage for the managers with whom they are associated. Unified CM Assistant uses JTAPI to integrate with Unified CM. While Unified CM Assistant is not specifically video-capable, there is nothing stopping Unified CM Assistant from being used with phones that are video-enabled. Once Unified CM Assistant has handled the call and the call is transferred to the final destination device, the two devices in the call communicate with each other directly and could establish video channels at that point. For example, if a video-capable endpoint dialed the directory number of a Manager, and the Assistant covered the call using the Unified CM Assistant application, there might not be any video established during the initial handling of the call. But once the Assistant transferred the caller to the Manager, Unified CM could then negotiate video channels. However, H.323 devices must support the Empty Capabilities Set (ECS) procedure in order to interoperate with Unified CM Assistant. If the H.323 endpoint does not support receiving an ECS from Unified CM, calls that are intercepted by Unified CM Assistant will fail when the Assistant attempts to transfer the call to the Manager.

Cisco Unified IP Interactive Voice Response and Cisco Unified Contact Center

Cisco Unified IP Interactive Voice Response (Unified IP IVR) and Cisco Unified Contact Center (Unified CC) use JTAPI to integrate with Unified CM. If a video-capable device calls into an IVR application (such as a help desk), the communication is audio-only while the caller is connected to the application server (that is, while the caller browses the IVR menu or waits in queue for a help-desk member to take the call). However, once the IVR application transfers the call to its final destination, video channels can be negotiated at that time. H.323 devices must support the Empty Capabilities Set (ECS) procedure in order to interoperate with Cisco Unified IP IVR and Unified CC. If the H.323 endpoint does not support receiving an ECS from Unified CM, calls that are intercepted by Cisco Unified IP IVR or Unified CC will fail when the application attempts to transfer the caller to the final destination.

IVR applications often use DTMF tones to select options in the IVR menu. An alternative is speech recognition, which enables the caller to speak commands to the IVR server instead of pressing keys on the phone. Because Cisco Unified IP IVR and Unified CC both use JTAPI to integrate with Unified CM, they pass DTMF tones through out-of-band signaling messages. Many H.323 devices on the market today use in-band DTMF tones, and these H.323 clients would not be able to use DTMF to navigate an IP IVR or Unified CC menu. However, these H.323 clients could use speech recognition if the IVR server is enabled for it. Video-capable devices such as Cisco Unified Video Advantage, Sony or Tandberg SCCP devices, and any H.323 endpoint that uses H.245 alphanumeric out-of-band signaling for DTMF, can navigate the IVR menus using DTMF tones.

Cisco Attendant Console

Cisco Attendant Console uses JTAPI to integrate with Unified CM. The Attendant Console is used as a front-office device to handle incoming calls. Although Attendant Console does not specifically support video, video channels can be negotiated once the call is transferred to its final destination. However, H.323 devices must support the Empty Capabilities Set (ECS) procedure in order to interoperate with the Attendant Console. If the H.323 endpoint does not support receiving an ECS from Unified CM, calls that are intercepted by an Attendant Console will fail when the attendant attempts to transfer the caller to the final destination.

Cisco Personal Assistant

Cisco Personal Assistant (PA) provides two main functions:

- Dialing assistance either through DTMF tones or voice commands
- User-programmed call routing preferences based on the time of day, caller ID information, and other factors

For both of these functions, no video is established while the call is connected to the PA server, but video channels can be negotiated once PA transfers the call to its final destination. H.323 devices must support the Empty Capabilities Set (ECS) procedure in order to interoperate with PA. If the H.323 endpoint does not support receiving an ECS from Unified CM, calls that are intercepted by PA will fail when the application attempts to transfer the caller to the final destination.

Cisco IP SoftPhone and Cisco IP Communicator

Cisco IP SoftPhone uses TAPI to integrate with Unified CM and can be configured either as a standalone softphone or as a software interface to control an associated SCCP hardware phone. Cisco IP SoftPhone does not specifically support video, but it can be used in conjunction with an IP Phone that has a Cisco Unified Video Advantage client associated to it. Cisco IP SoftPhone cannot be used to control a Sony or Tandberg SCCP device.

Cisco IP Communicator differs from IP SoftPhone in that it is an SCCP client and therefore acts like a Cisco 7970 Series IP Phone.

As of version 2.0, Cisco IP Communicator can associate to Cisco Unified Video Advantage 2.0 when both applications reside on the same PC. For more information, refer to the chapter on [IP Telephony Endpoints](#), page 20-1.

Collaboration Solutions

The following technologies are sometimes used to provide video communications between endpoints.

T.120 Application Sharing

Some videoconferencing endpoints use the T.120 protocol to share documents, whiteboards, and text among participants in a conference. Unified CM does not support negotiating a T.120 channel. Instead of T.120, Cisco recommends using web-based collaboration solutions such as Cisco MeetingPlace or other third-party collaboration solutions such as Placeware, Web-Ex, or IBM E-Collaborate.

Cisco Unified MeetingPlace

Cisco Unified MeetingPlace combines a high-end audio and video conferencing solution with a web-based front end for scheduling and participating in conferences. For more information, refer to the chapter on [Cisco Unified MeetingPlace Integration](#), page 14-1.

Wireless Networking Solutions

Because video is so bandwidth intensive, Cisco does *not* recommend using a shared wireless medium such as 802.11b/g for video endpoints.

Cisco Aironet 802.11b Networking Solutions

The Tandberg T-1000 model endpoint offers a PCMCIA interface in which you can install a Cisco Aironet PCMCIA 802.11b Wireless Adapter. However, you should take care to ensure that the video endpoints do not share the wireless bandwidth with any production IP Phones because video will consume much of the bandwidth and make it difficult to support video, audio, and data all on the same wireless medium.

Cisco Unified Video Advantage relies on a physical Ethernet connection to the physical IP Phone with which it is associated. It is not uncommon for a user to have both a physical Ethernet interface and an Aironet 802.11b Wireless Adapter installed in the same PC. This configuration could cause problems for Cisco Unified Video Advantage if the wireless interface happens to be the preferred path to the network because Cisco Unified Video Advantage will not associate over that interface. Cisco recommends that you always make the physical Ethernet interface the preferred path. Also, when users connect to the PC port on the back of the IP Phone, instruct them to disable their Aironet Adapter to keep it from accidentally taking precedence.

Cisco Unified Wireless IP Phone 7920

The Cisco Unified Wireless IP Phone 7920 does not support video. Nothing prevents a video endpoint from calling a Cisco Unified Wireless IP Phone 7920; however, it will negotiate as an audio-only call. The Wireless IP Phone user can then put the call on hold, transfer it, or conference it. If the caller is an H.323 video endpoint, it must support the Empty Capabilities Set (ECS) procedure in order for these supplementary services to work.

XML Services

Currently there are no XML applications specifically written for the Cisco Unified Video Advantage client solution, Cisco IP Video Phone 7985, or the Sony or Tandberg SCCP endpoints. However, with the exception of the Sony endpoints, these endpoints do support XML applications. Cisco Unified Video Advantage uses a Cisco Unified IP Phone, so any XML application supported on those phone models should also work with Unified Video Advantage.

The Tandberg SCCP endpoints support XML, but not all XML applications currently work on the Tandberg endpoints. For example, Cisco Extension Mobility and the Berbee InformaCast product are two popular XML applications that currently do not work on the Sony or Tandberg SCCP endpoints. Support for these applications will require firmware upgrades to the endpoints as well as changes in Unified CM Administration in some cases.



CHAPTER 17

LDAP Directory Integration

Last revised on: October 30, 2008

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.

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 a Cisco IP Communications system based on Cisco Unified Communications Manager (Unified CM) 5.x with a corporate LDAP directory. The main topics include:

- [What is Directory Integration?, page 17-2](#)

This section analyzes the various requirements for integration with a corporate LDAP directory in a typical enterprise IT organization.

- [Directory Access for IP Telephony Endpoints, page 17-4](#)

This section describes the technical solution to enable directory access for Cisco Unified Communications endpoints and provides design best-practices around it.

- [Directory Integration with Unified CM 5.x, page 17-6](#)

This section describes the technical solutions and provides design best-practices for directory integration with Cisco Unified CM 5.x, including the LDAP synchronization and LDAP authentication functions.

The considerations presented in this chapter apply to Cisco Unified CM 5.x as well as the following applications bundled with it: Cisco Extension Mobility, Cisco Unified Communications Manager Assistant, WebDialer, Bulk Administration Tool, and Real-Time Monitoring Tool.

For earlier releases of Unified CM, refer to the appropriate version of the *Cisco Unified Communications SRND*; and for all other Cisco voice applications, refer to the respective product documentation available at:

<http://www.cisco.com>

In particular, for Cisco IP Contact Center refer to the *Cisco Cisco Unified Contact Center Enterprise Edition SRND* and the *Cisco Cisco Unified Contact Center Express SRND*, both available at

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

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*, also available at

<http://www.cisco.com>

What's New in This Chapter

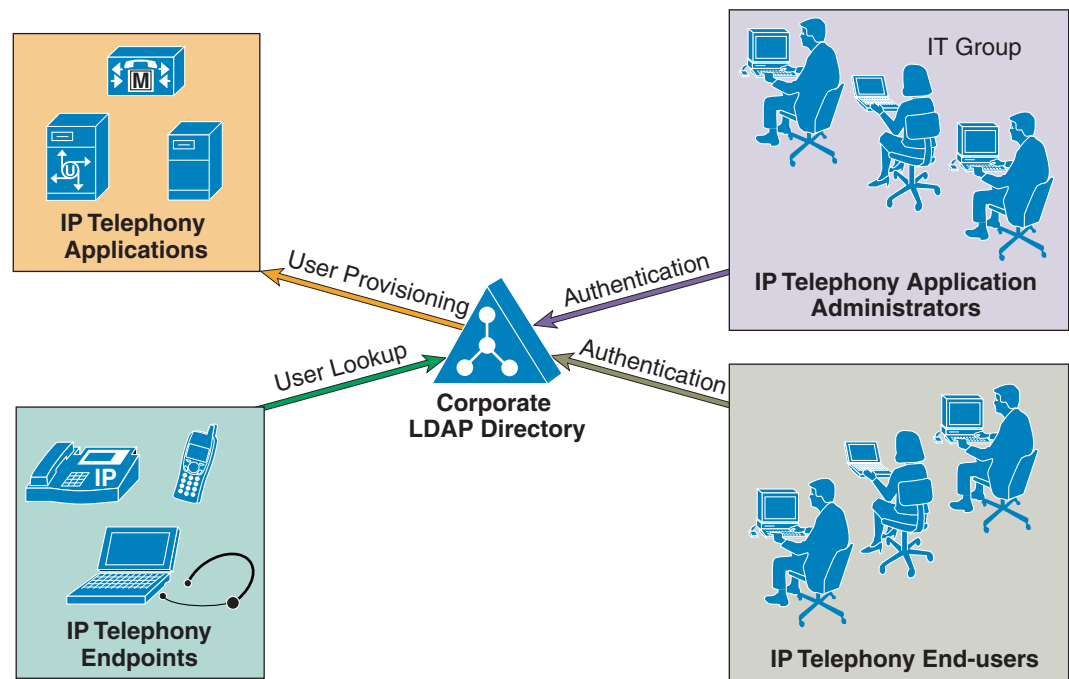
Table 17-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 17-1 ***New or Changed Information Since the Previous Release of This Document***

New or Revised Topic	Described in:
Directory synchronization	Using the LDAP Structure to Control Synchronization, page 17-25
The section on LDAP Query Filter was deleted from this chapter.	

What is Directory Integration?

Integration between voice applications and a corporate LDAP directory is a common task for many enterprise IT organizations. However, the exact scope of the integration varies from company to company, and it can translate to one or more specific and independent requirements, as shown in [Figure 17-1](#).

Figure 17-1 Various Requirements for Directory Integration

153279

For example, one common requirement is to enable user lookups (sometimes called the "white pages" service) from IP phones or other voice and/or video endpoints, so that users can dial a contact directly after looking up their number in the directory.

Another requirement is to provision users automatically from the corporate directory into the user database of voice and/or video applications. This method avoids having to add, remove, or modify core user information manually each time a change occurs in the corporate directory.

Often authentication of end-users and administrators of the voice and/or video applications using the corporate directory credentials is also required. This method enables the IT department to deliver single log-on functionality and reduces the number of passwords that each user needs to maintain across different corporate applications.

Each of these requirements can be satisfied by a Cisco IP Communications system using different mechanisms according to the Unified CM version used, as summarized in [Table 17-2](#).

Table 17-2 Directory Requirements and Cisco Solutions

Requirement	Cisco Solution	Cisco Unified CM 4.x Feature	Cisco Unified CM 5.x Feature
User lookup for endpoints	Directory access	Cisco Unified IP Phone Services SDK	Cisco Unified IP Phone Services SDK
User provisioning	Directory integration	Cisco Customer Directory Configuration Plugin	LDAP Synchronization

Table 17-2 **Directory Requirements and Cisco Solutions (continued)**

Requirement	Cisco Solution	Cisco Unified CM 4.x Feature	Cisco Unified CM 5.x Feature
Authentication for IP Telephony end users	Directory integration	Cisco Customer Directory Configuration Plugin	LDAP Authentication
Authentication for IP Telephony application administrators	Directory integration	Cisco Customer Directory Configuration Plugin + Cisco Multilevel Administration	LDAP Authentication

As shown in [Table 17-2](#), within the context of a Cisco IP Communications system, the term *directory access* refers to mechanisms and solutions that satisfy the requirement of user lookups for IP Telephony endpoints, while the term *directory integration* refers to mechanisms and solutions that satisfy the requirements of user provisioning and authentication (for both end users and administrators).

The remainder of this chapter describes how to address these requirements in a Cisco IP Communications system based on Cisco Unified CM Release 5.x. For a detailed description of directory integration solutions with earlier releases of Unified CM, refer to the appropriate version of the *Cisco Unified Communications SRND*, available at

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

**Note**

Another interpretation of the term *directory integration* revolves around the ability to add application servers to a Microsoft Active Directory domain in order to centralize management and security policies. Cisco Unified CM 5.x is an appliance that runs on a customized embedded operating system, and it cannot currently be added to a Microsoft Active Directory domain. Server management for Unified CM is provided through the Cisco Real Time Monitoring Tool (RTMT). Strong security policies tailored to the application are already implemented within the embedded operating system, and a customized version of Cisco Security Agent can be installed on every server. It is also possible to manage security policies centrally with the CiscoWorks Management Center for Cisco Security Agents.

Directory Access for IP Telephony Endpoints

This section describes how to configure corporate directory access to any LDAP-compliant directory server to perform user lookups from Cisco Unified Communications endpoints (such as Cisco Unified IP Phones). The guidelines contained in this section apply regardless of whether Unified CM or other IP Telephony applications have been integrated with a corporate directory for user provisioning and authentication.

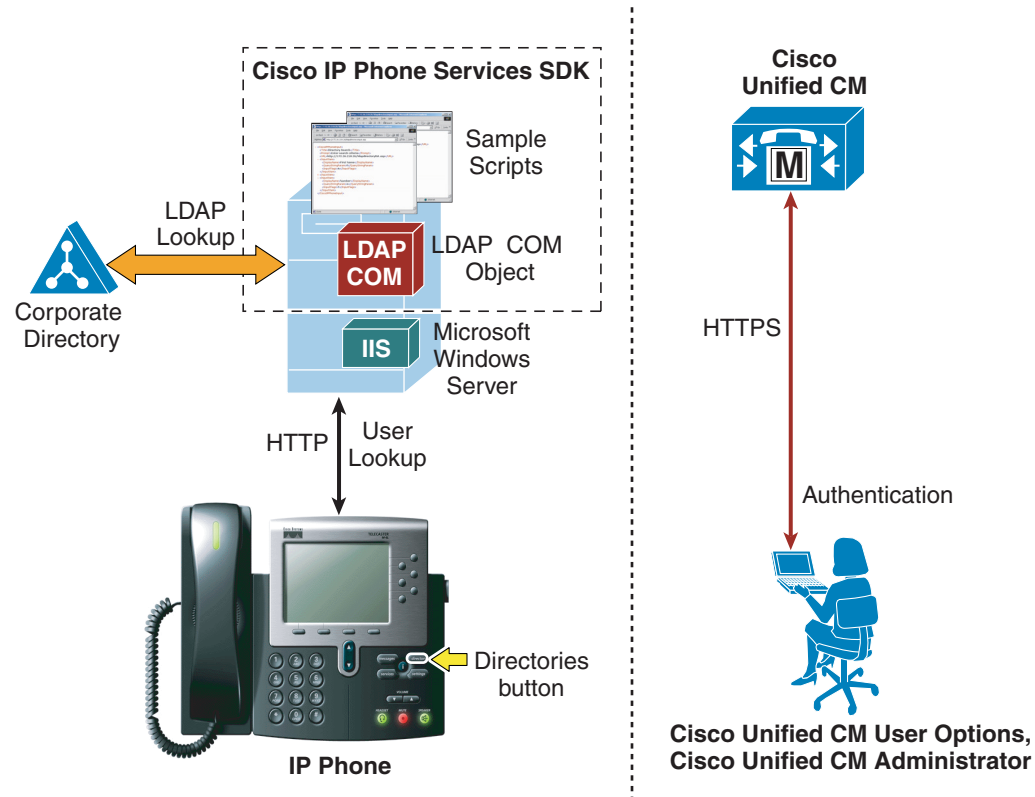
Cisco Unified IP Phones equipped with a display screen can search a user directory when a user presses the Directories button on the phone. The IP Phones use Hyper-Text Transfer Protocol (HTTP) to send requests to a web server. The responses from the web server must contain some specific Extensible Markup Language (XML) objects that the phone can interpret and display.

By default, Cisco Unified IP Phones are configured to perform user lookups against Unified CM's embedded database. However, it is possible to change this configuration so that the lookup is performed on a corporate LDAP directory. In this case, the phones send their HTTP requests to an external web

server that operates as a proxy and translates these requests into LDAP queries against the corporate directory. The LDAP responses are then encapsulated in the appropriate XML objects and sent back to the phones via HTTP.

Figure 17-2 illustrates this mechanism in a deployment where Unified CM has not been integrated with the corporate directory. Note that, in this scenario, Unified CM is not involved in the message exchange related to the user lookup.

Figure 17-2 Directory Access for Cisco Unified IP Phones Using the Cisco Unified IP Phone Services SDK



In the example shown in Figure 17-2, the web server proxy function is provided by the Cisco LDAP Search Component Object Model (COM) server, which is included in the Cisco Unified IP Phone Services Software Development Kit (SDK) version 3.3(4) or later. You can download the latest Cisco Unified IP Phone Services SDK from Cisco Developer Support Central at

http://www.cisco.com/en/US/products/svcs/ps3034/ps5408/ps5418/serv_home.html

The IP Phone Services SDK can be installed on a Microsoft Windows web server running IIS 4.0 or later, but it cannot be installed on a Unified CM server. The SDK includes some sample scripts to provide simple directory lookup functionality.

To set up a corporate directory lookup service using the IP Phone Services SDK, perform the following steps:

-
- Step 1** Modify one of the sample scripts to point to your corporate LDAP directory, or write your own script using the LDAP Search COM Programming Guide provided with the SDK.
- Step 2** In Unified CM, configure the URL Directories parameter (under **System > Enterprise Parameters**) to point to the URL of the script on the external web server.
- Step 3** Reset the phones to make the changes take effect.
-

**Note**

If you want to offer the service only to a subset of users, configure the URL Directories parameter directly within the Phone Configuration page instead of the Enterprise Parameters page.

In conclusion, the following design considerations apply to directory access with the Cisco Unified IP Phone Services SDK:

- User lookups are supported against any LDAP-compliant corporate directory.
- When querying Microsoft Active Directory, you can perform lookups against the Global Catalog by pointing the script to a Global Catalog server and specifying port 3268 in the script configuration. This method typically results in faster lookups.
- There is no impact on Unified CM for this functionality, and only minimal impact on the LDAP directory server.
- The sample scripts provided with the SDK allow only a minimal amount of customization (for example, you can prefix a digit string to all returned numbers). For a higher degree of manipulation, you will have to develop custom scripts, and a programming guide is included with the SDK to aid in writing the scripts.
- This functionality does not entail provisioning or authentication of Unified CM users with the corporate directory.

Directory Integration with Unified CM 5.x

This section describes the mechanisms and best practices for directory integration with Cisco Unified CM 5.x to allow for user provisioning and authentication with a corporate LDAP directory. This section covers the following topics:

- [Comparison with the Unified CM 4.x Approach, page 17-7](#)

The approach to directory integration has changed significantly between Unified CM releases 4.x and 5.x, and this section introduces the new approach by comparing it with the old one.

- [Unified CM 5.x Directory Architecture, page 17-8](#)

This section provides an overview of the user-related architecture in Unified CM 5.x.

- [LDAP Synchronization, page 17-11](#)

This section describes the functionality of LDAP synchronization and provides design guidelines for its deployment, with additional considerations for Microsoft Active Directory.

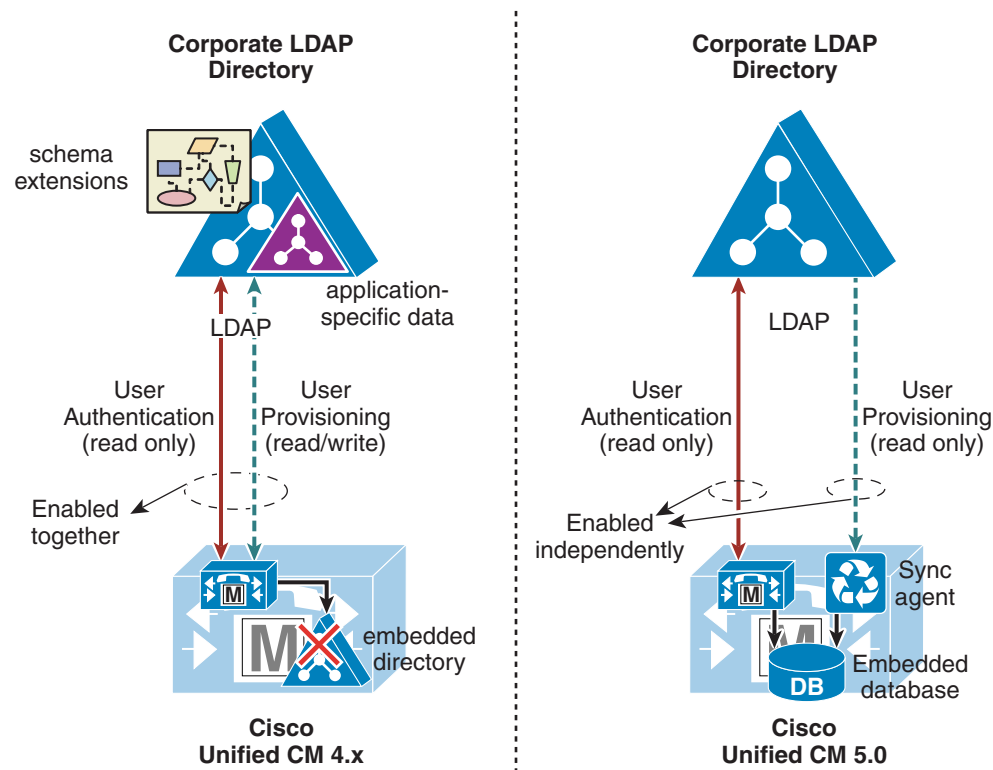
- [LDAP Authentication, page 17-19](#)

This section describes the functionality of LDAP authentication and provides design guidelines for its deployment, with additional considerations for Microsoft Active Directory.

Comparison with the Unified CM 4.x Approach

Figure 17-3 shows high-level functional diagrams of the directory integration approaches for user provisioning and authentication with Unified CM 4.x and 5.x.

Figure 17-3 Directory Integration Approaches in Unified CM 4.x and 5.x



Cisco Unified CM Release 4.x used an embedded LDAP directory to store user-related information. Directory integration was enabled by extending the corporate directory schema, shutting down the embedded directory, and using the corporate directory to store the application-specific data related to the users. Because the corporate directory was effectively used as the back-end storage repository for user information, this method satisfied the requirement for both user provisioning and user authentication. Any changes to user data in the corporate directory were immediately picked up by Unified CM because it accessed the same data store.

However, this approach had an impact on the corporate directory in terms of schema extensions and additional data, and it also introduced dependencies between the real-time functionality of the IP Communications system and the availability of the directory. When connectivity was lost or the directory became unavailable, Unified CM was unable to access all user-related configurations, which impacted applications such as Extension Mobility, Attendant Console, and IP Contact Center Express. In this approach, the user provisioning and user authentication functions had to be enabled at the same

time because they relied on the same integration process. In addition, using the corporate directory as the storage repository for application-specific data also imposed limitations on the day-to-day maintenance operations on the corporate directory itself.

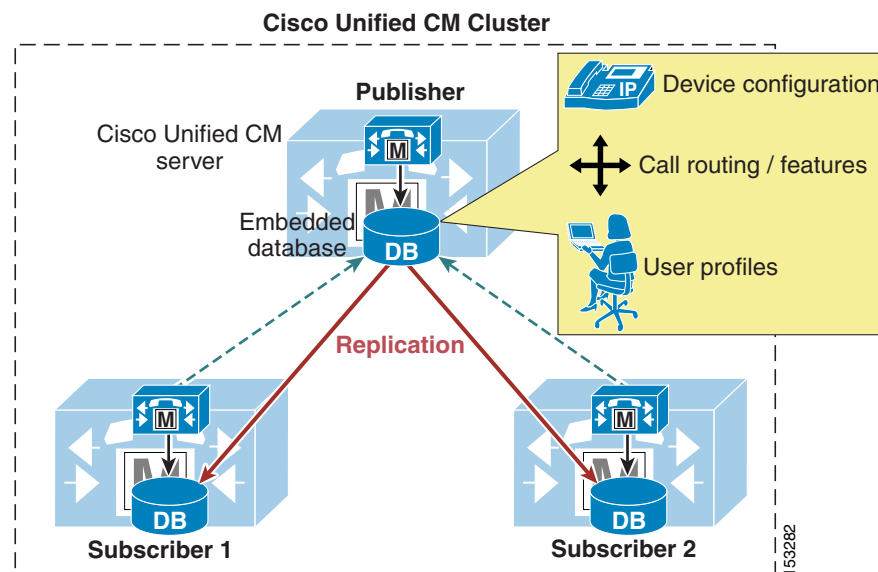
By contrast, the approach to directory integration adopted by Cisco Unified CM 5.x relies on two separate components to satisfy the user provisioning and user authentication requirements independently. User provisioning is performed with a one-way synchronization of user data from the corporate directory to the Unified CM embedded database. The synchronization uses standard LDAPv3 and can be triggered manually or scheduled periodically to ensure that changes are incorporated into the Cisco IP Communications system. This solution avoids the need to write anything to the corporate directory, and it does not require any schema extensions.

User authentication is enabled independently from user provisioning, and it provides authentication of end-user passwords against the corporate directory credentials. With this approach, the Cisco IP Communications system preserves all of its real-time functionality even when the corporate directory is unavailable or unreachable.

Unified CM 5.x Directory Architecture

Figure 17-4 shows the basic architecture of a Unified CM 5.x cluster. The embedded database stores all configuration information, including device-related data, call routing, and other features and user profiles. The database is present on all servers within a Unified CM cluster and is replicated automatically from the publisher server to all subscriber servers.

Figure 17-4 Cisco Unified CM 5.x Architecture



By default, all users are provisioned manually into the database via the Unified CM Administration interface. Cisco Unified CM 5.x introduces an important new concept by dividing the users in its database in two categories:

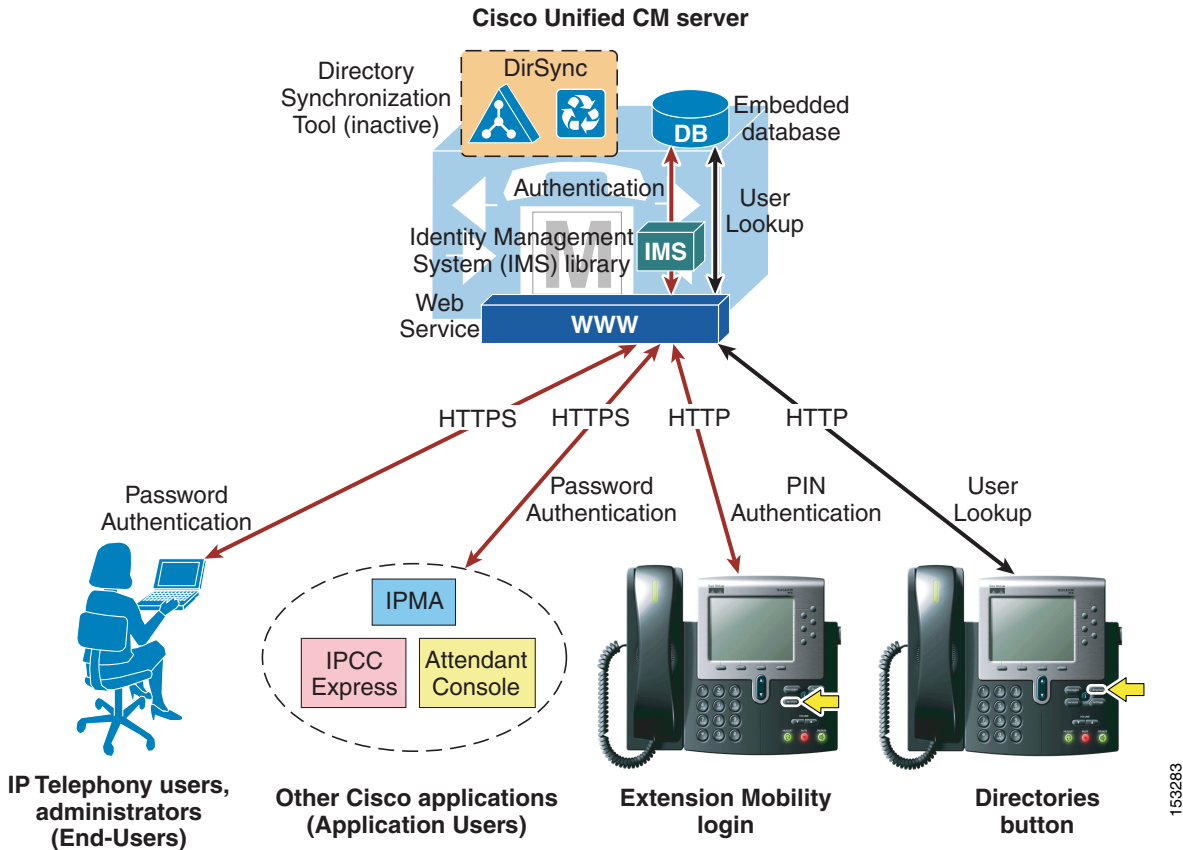
- End users — All users associated with a physical person and an interactive login. This category includes all IP Telephony users as well as Unified CM administrators when using the User Groups and Roles configuration (equivalent to the Cisco Multilevel Administration feature in prior Unified CM versions).
- Application users — All users associated with other Cisco IP Communications features or applications, such as Cisco Attendant Console, Cisco IP Contact Center Express, or Cisco Unified Communications Manager Assistant. These applications need to authenticate with Unified CM, but these internal "users" do not have an interactive login and serve purely for internal communications between applications.

Table 17-3 lists the application users created by default in the Unified CM database, together with the feature or application that uses them. Additional application users can be created manually when integrating other Cisco IP Communications applications (for example, the **ac** application user for Cisco Attendant Console, the **jtapi** application user for Cisco IP Contact Center Express, and so forth).

Table 17-3 **Default Application Users for Unified CM 5.x**

Application User	Used by:
CCMAdministrator	Unified CM Administration (default "super user")
CCMQRTSecureSysUser	Cisco Quality Reporting Tool
CCMQRTSysUser	
CCMSysUser	Cisco Extension Mobility
IPMASecureSysUser	Cisco Unified Communications Manager Assistant
IPMASysUser	
WDSecureSysUser	Cisco WebDialer
WDSysUser	

Based on these considerations, Figure 17-5 illustrates the default behavior in Unified CM 5.x for user-related operations such as lookups, provisioning, and authentication.

Figure 17-5 Default Behavior for User-Related Operations for Unified CM 5.x

159283

End users access the Unified CM User Options page via HTTPS and authenticate with a user name and password. If they have been configured as administrators by means of User Groups and Roles, they also access the Unified CM Administration pages with the same credentials.

Similarly, other Cisco features and applications authenticate to Unified CM via HTTPS with the user name and password associated with their respective application users.

The authentication challenges carried by the HTTPS messages are relayed by the web service on Unified CM to an internal library called Identity Management System (IMS). In its default configuration, the IMS library authenticates both end users and application users against the embedded database. In this way, both "physical" users of the IP Communications system and internal application accounts are authenticated using the credentials configured in Unified CM.

End users may also authenticate with their user name and a numeric password (or PIN) when logging into the Extension Mobility service from an IP phone. In this case, the authentication challenge is carried via HTTP to Unified CM but is still relayed by the web service to the IMS library, which authenticates the credentials against the embedded database.

In addition, user lookups performed by IP Telephony endpoints via the Directories button communicate with the web service on Unified CM via HTTP and access data on the embedded database.

The importance of the distinction between End Users and Application Users becomes apparent when integration with a corporate directory is required. As mentioned in the previous section, this integration is accomplished by means of the following two separate processes:

- LDAP synchronization

This process uses an internal tool called Cisco Directory Synchronization (DirSync) on Unified CM to synchronize a number of user attributes (either manually or periodically) from a corporate LDAP directory. When this feature is enabled, users are automatically provisioned from the corporate directory. This feature applies only to End Users, while Application Users are kept separate and are still provisioned via the Unified CM Administration interface. In summary, End Users are defined in the corporate directory and synchronized into the Unified CM database, while Application Users are stored only in the Unified CM database and do not need to be defined in the corporate directory.

- LDAP authentication

This process enables the IMS library to authenticate user credentials against a corporate LDAP directory. When this feature is enabled, End User passwords are authenticated against the corporate directory, while Application User passwords are still authenticated locally against the Unified CM database. Cisco Extension Mobility PINs are also still authenticated locally.

Maintaining and authenticating the Application Users internally to the Unified CM database provides resilience for all the applications and features that use these accounts to communicate with Unified CM, independently of the availability of the corporate LDAP directory.

Cisco Extension Mobility PINs are also kept within the Unified CM database because they are an integral part of a real-time application, which should not have dependencies on the responsiveness of the corporate directory.

The next two sections describe in more detail LDAP synchronization and LDAP authentication, and they provide design best-practices for both functions.

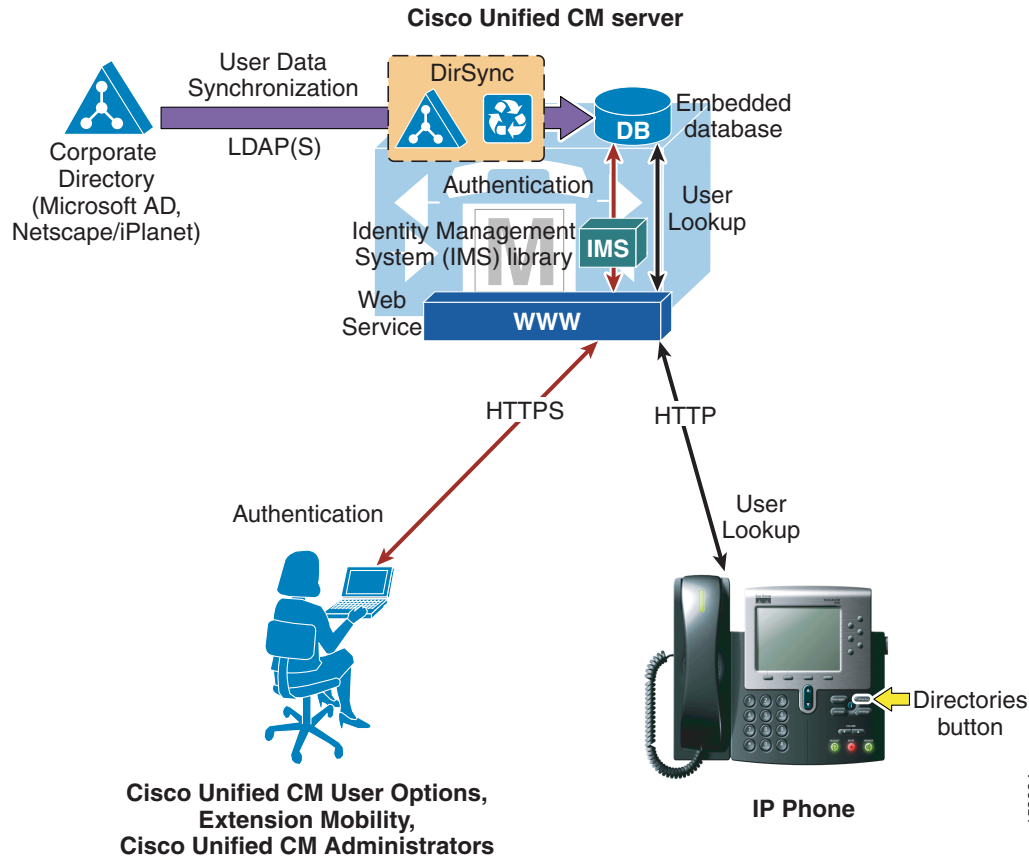
**Note**

As illustrated in the section on [Directory Access for IP Telephony Endpoints](#), page 17-4, user lookups from endpoints can also be performed against a corporate directory by configuring the Cisco Unified IP Phone Services SDK on an external web server.

LDAP Synchronization

Synchronization of Unified CM with a corporate LDAP directory allows reuse of user data stored in the LDAP directory and allows the corporate LDAP directory to serve as the central repository for that information. Unified CM has an integrated database for storing user data and a web interface within Unified CM Administration for creating and managing user data in that database. When synchronization is enabled, that local database is still used, but the Unified CM facility to create user accounts becomes disabled. Management of user accounts is then accomplished through the interface of the LDAP directory. (See [Figure 17-6](#).)

The user account information is imported from the LDAP directory into the database located on the Unified CM publisher server. Information that is imported from the LDAP directory may not be changed by Unified CM. Additional user information specific to the Unified CM implementation is managed by Unified CM and stored only within its local database. For example, device-to-user associations, speed dials, and user PINs are data that are managed by Unified CM, and they do not exist in the corporate LDAP directory. The user data is then propagated from the Unified CM publisher server to the subscribers via the built-in database synchronization.

Figure 17-6 Enabling Synchronization of User Data

The following directories are supported by Unified CM for synchronization:

- Microsoft Active Directory (AD) 2000 and 2003
- Netscape Directory Server 4.x, iPlanet Directory Server 5.1, and Sun ONE Directory Server 5.2

When LDAP synchronization is activated, only one of the above groups of LDAP products may be chosen for the cluster at any one time. Also, one attribute of the directory user is chosen to map into the Unified CM User ID field. Unified CM uses standard LDAPv3 for accessing the data.

The data that Unified CM imports are all from standard attributes. [Table 17-4](#) lists the attributes that are used, and they differ between the two groups of LDAP implementations. The data of the directory attribute that is mapped to the Unified CM User ID must be unique within all entries for that cluster. The **sn** attribute must be populated with data, otherwise that record will not be imported from the corporate directory. If the primary attribute used during import of end user accounts matches any Application User in the Unified CM database, that user is not imported from the LDAP directory.

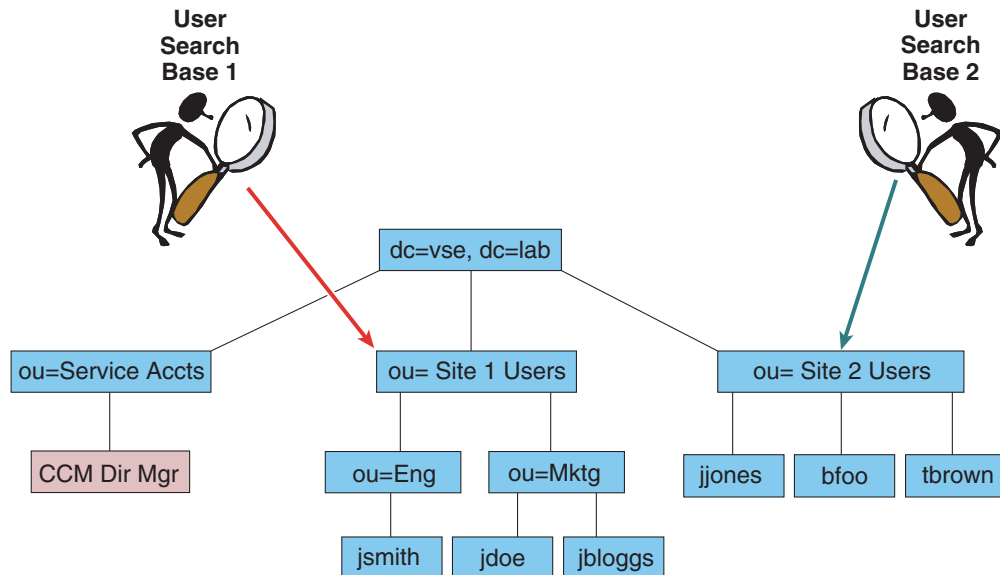
[Table 17-4](#) lists the attributes that are imported from the LDAP directory into corresponding Unified CM user fields, and it describes the mapping between those fields. Some Unified CM user fields might be mapped from one of several LDAP attributes.

Table 17-4 Data Attributes Imported by Unified CM

Unified CM User Field	Microsoft Active Directory (AD) Attribute	Netscape, iPlanet, or Sun ONE Attribute
User ID	<i>One of:</i> sAMAccountName mail employeeNumber telephoneNumber UserPrincipalName	<i>One of:</i> uid mail employeeNumber telephonePhone
First Name	givenName	givenname
Middle Name	<i>One of:</i> middleName initials	initials
Last Name	sn	sn
Manager ID	manager	manager
Department	department	departmentnumber
Phone Number	<i>One of:</i> telephoneNumber ipPhone	telephonenumber
Mail ID	<i>One of:</i> mail sAMAccountName	<i>One of:</i> mail uid

The synchronization is performed by a process called Cisco DirSync, which is enabled through the Serviceability web page. When enabled, it allows one of five synchronization agreements to be configured in the system. An agreement specifies a search base that is a position in the LDAP tree where Unified CM will begin its search for user accounts. Unified CM can import only users that exist in the domain specified by the search base for a particular synchronization agreement.

In [Figure 17-7](#), two synchronization agreements are represented. One synchronization agreement specifies User Search Base 1 and imports users jsmith, jdoe and jblogs. The other synchronization agreement specifies User Search Base 2 and imports users jjones, bfoo, and tbrown. The CCMDirMgr account is not imported because it does not reside below the point specified by a user search base. When users are organized in a structure in the LDAP directory, you can use that structure to control which user groups are imported. In this example, a single synchronization agreement could have been used to specify the root of the domain, but that search base would also have imported the Service Accts. The search base does not have to specify the domain root; it may specify any point in the tree.

Figure 17-7 User Search Bases

153285

To import the data into the Unified CM database, the system performs a bind to the LDAP directory using the account specified in the configuration as the LDAP Manager Distinguished Name, and reading of the database is done with this account. The account must be available in the LDAP directory for Unified CM to log in, and Cisco recommends that you create a specific account with permissions to allow it to read all user objects within the sub-tree that was specified by the user search base. The sync agreement specifies the full Distinguished Name of that account so that the account may reside anywhere within that domain. In the example in [Figure 17-7](#), `CCMDirMgr` is the account used for the synchronization.

It is possible to control the import of accounts through use of permissions of the LDAP Manager Distinguished Name account. In this example, if that account is restricted to have read access to `ou=Eng` but not to `ou=Mktg`, then only the accounts located under `Eng` will be imported.

Synchronization agreements have the ability to specify multiple directory servers to provide redundancy. You can specify an ordered list of up to three directory servers in the configuration that will be used when attempting to synchronize. The servers are tried in order until the list is exhausted. If none of the directory servers responds, then the synchronization fails, but it will be attempted again according to the configured synchronization schedule.

Synchronization Mechanism

The synchronization agreement specifies a time for synchronizing to begin and a period for re-synchronizing that can be specified in hours, days, weeks, or months (with a minimum value of 6 hours). A synchronization agreement can also be set up to run only once at a specific time.

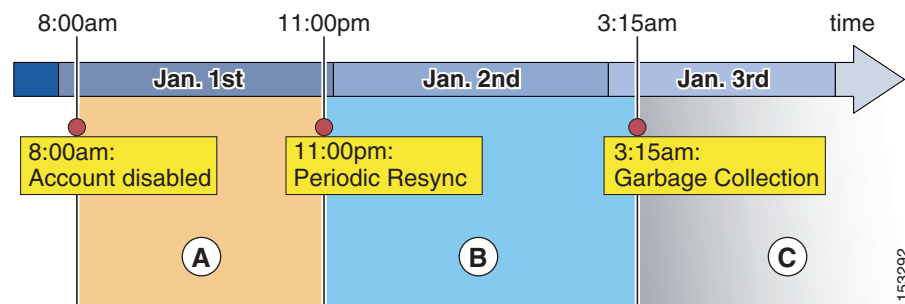
When synchronization is enabled for the first time on a Unified CM publisher server, user accounts that exist in the corporate directory are imported into the Unified CM database. Then either existing Unified CM end-user accounts are activated and data is updated, or a new end-user account is created according to the following process:

1. If end-user accounts already exist in the Unified CM database and a synchronization agreement is configured, all pre-existing accounts are marked inactive in Unified CM. The configuration of the synchronization agreement specifies a mapping of an LDAP database attribute to the Unified CM UserID. During the synchronization, accounts from the LDAP database that match an existing Unified CM account cause that Unified CM account to be marked active again.
2. After the synchronization is completed, any accounts that were not set to active are permanently deleted from Unified CM when the garbage collection process runs. Garbage collection is a process that runs automatically at the fixed time of 3:15 AM, and it is not configurable. The deletion of Unified CM accounts that do not match LDAP directory accounts is necessary because Unified CM cannot manage accounts while synchronization is configured.
3. Subsequently when changes are made in the corporate directory, the synchronization from Microsoft Active Directory occurs as a full re-synchronization at the next scheduled synchronization period. On the other hand, the Netscape, iPlanet, and Sun ONE products perform an incremental synchronization triggered by a change in the directory. The following sections present examples of each of these two scenarios.

Account Synchronization with Active Directory

Figure 17-8 shows an example timeline of events for a Unified CM deployment where LDAP Synchronization and LDAP Authentication have both been enabled. The re-synchronization is set for 11:00 PM daily.

Figure 17-8 Change Propagation with Active Directory



After the initial synchronization, the creation, deletion, or disablement of an account will propagate to Unified CM according to the timeline shown in Figure 17-8 and as described in the following steps:

1. At 8:00 AM on January 1, an account is disabled or deleted in AD. From this time and during the whole period A, password authentication (for example, Unified CM User Options page) will fail for this user because Unified CM redirects authentication to AD. However, PIN authentication (for example, Extension Mobility login) will still succeed because the PIN is stored in the Unified CM database.
2. The periodic re-synchronization is scheduled for 11:00 PM on January 1. During that process, Unified CM will verify all accounts. Any accounts that have been disabled or deleted from AD will at that time be tagged in the Unified CM database as inactive. After 11:00 PM on January 1, when the account is marked inactive, both the PIN and password authentication by Unified CM will fail.

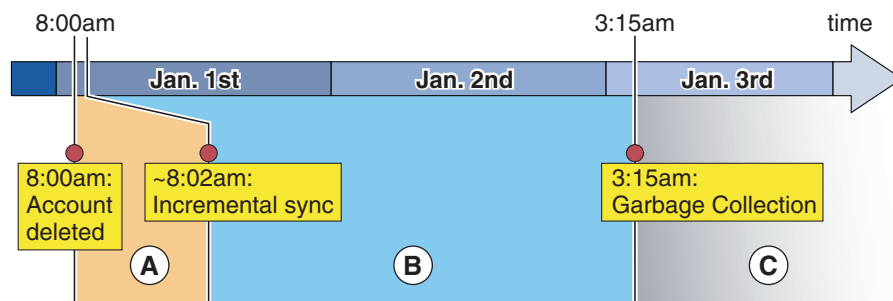
- Garbage collection of accounts occurs daily at the fixed time of 3:15 AM. This process permanently deletes user information from the Unified CM database for any record that has been marked inactive for over 24 hours. In this example, the garbage collection that runs at 3:15 AM on January 2 does not delete the account because it has not been inactive for 24 hours yet, so the account is deleted at 3:15 AM on January 3. At that point, the user data is permanently deleted from Unified CM.

If an account has been created in AD at the beginning of period A, it will be imported to Unified CM at the periodic re-synchronization that occurs at the beginning of period B and will immediately be active on Unified CM.

Account Synchronization with Netscape, iPlanet, or Sun ONE

The Netscape, iPlanet, and Sun ONE products support incremental synchronization agreements and use a different synchronization timeline than Active Directory. Figure 17-9 shows an example of this synchronization timeline for a Unified CM deployment with LDAP Synchronization and LDAP Authentication both enabled.

Figure 17-9 Change Propagation with Netscape, iPlanet, and Sun ONE



The example in Figure 17-9 involves the following steps:

- An account is deleted from the corporate directory at 8:00 AM on January 1, which causes an incremental update to be sent from the LDAP server to Unified CM. Unified CM sets its corresponding copy of the data to inactive. Because LDAP authentication is configured, the user will be unable to log in via password as soon as the LDAP server has deleted the record. Also, the PIN may not be used for login at the moment the Unified CM record is marked inactive.
- During period B, the user's record is still present in Unified CM, albeit inactive.
- When the garbage collection runs at 3:15 AM on January 2, the record has not yet been inactive for 24 hours. The data remains in the Unified CM database until the beginning of period C on January 3, when the garbage collection process runs again at 3:15 AM and determines that the record has been inactive for 24 hours or more. The record is then permanently deleted from the database.

Accounts that are newly created in the directory are synchronized to Unified CM via incremental updates as well, and they may be used as soon as the incremental update is received.

Security Considerations

During the import of accounts, no passwords or PINs are copied from the LDAP directory to the Unified CM database. If LDAP authentication is not enabled on Unified CM, the password and PIN for the end user are managed using Unified CM Administration. By default, the password is set to **ciscocisco** and the PIN is set to **12345** when the account is created. These settings can be changed by the user via

the user web pages or by the administrator via the administrator web pages. The password and PIN are stored in an encrypted format in the Unified CM database. If you want to use the directory password to authenticate an end user, see the section on [LDAP Authentication, page 17-19](#).

The connection between the Unified CM publisher server and the directory server can be secured by enabling Secure LDAP (SLDAP) on Unified CM and the LDAP server. Secure LDAP enables LDAP to be sent over a Secure Socket Layer (SSL) connection and can be enabled by uploading the SSL certificate from within the Unified CM Platform Administration. For detailed procedure steps, refer to the Unified CM product documentation available at <http://www.cisco.com>.

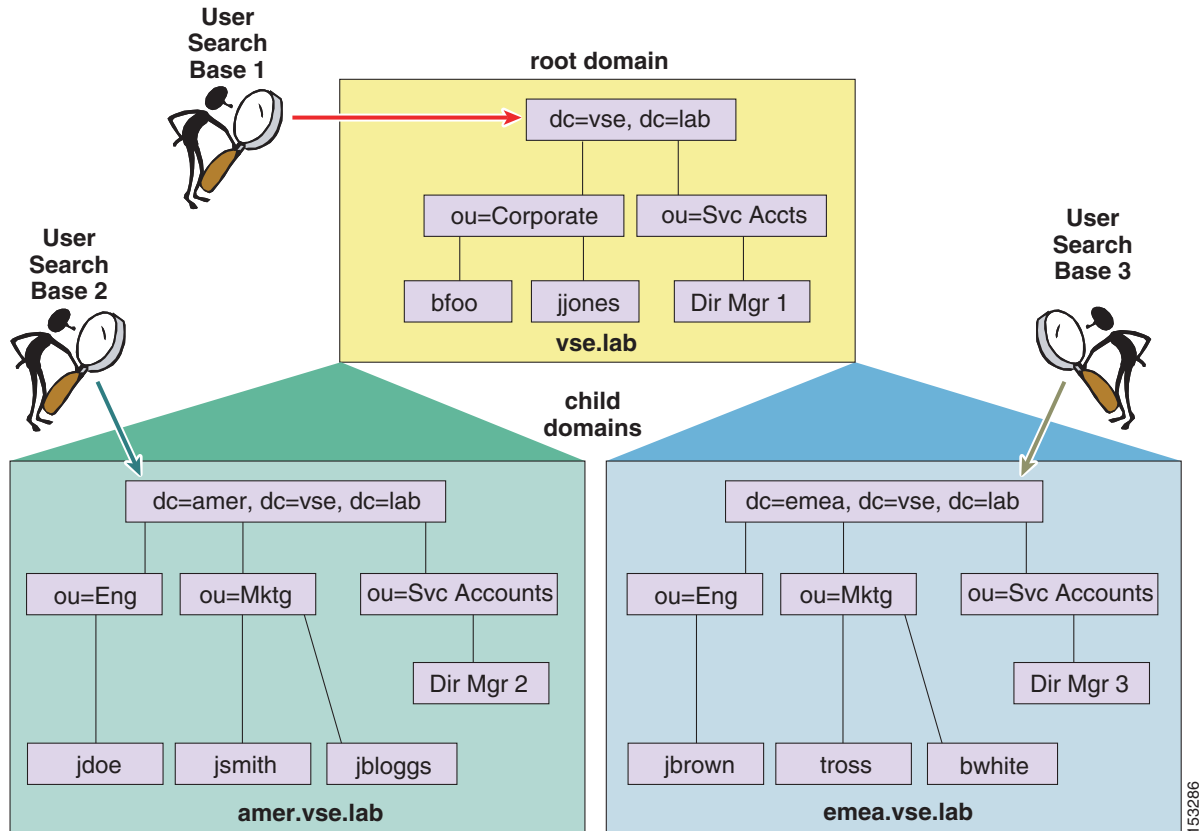
Best Practices for LDAP Synchronization

Observe the following design and implementation best practices when deploying LDAP synchronization with Cisco Unified CM 5.x:

- Use a specific account within the corporate directory to allow the Unified CM synchronization agreement to connect and authenticate to it. Cisco recommends that you use an account dedicated to Unified CM, with minimum permissions set to "read" all user objects within the desired search base and with a password set never to expire. (When the password for this account changes in the directory, Unified CM must be reconfigured to take the change into account.)
- All synchronization agreements on a given cluster must integrate with the same family of LDAP servers (either Microsoft AD or Netscape, iPlanet, and Sun ONE).
- Stagger the periodicity of synchronization agreements so that multiple agreements are not querying the same LDAP servers simultaneously. Choose synchronization times that occur during quiet periods.
- If security of user data is of concern, enable Secure LDAP (SLDAP) by checking the **Use SSL** field on the LDAP Directory configuration page in Unified CM Administration.
- Ensure that the LDAP directory attribute chosen to map into the Unified CM UserID field is unique within all synchronization agreements for that cluster.
- The attribute chosen as UserID must not be the same as that for any of the Application Users defined in Unified CM.
- An existing account in the Unified CM database before synchronization is maintained only if an account imported from the LDAP directory has a matching attribute. The attribute that is matched to the Unified CM UserID is determined by the synchronization agreement.
- Configure at least two LDAP servers for redundancy. You can use IP addresses instead of host names to eliminate dependencies on Domain Name System (DNS) availability.
- Administer end-user accounts through the LDAP directory's management tools, and manage the Cisco-specific data for those accounts through the Unified CM Administration web page.

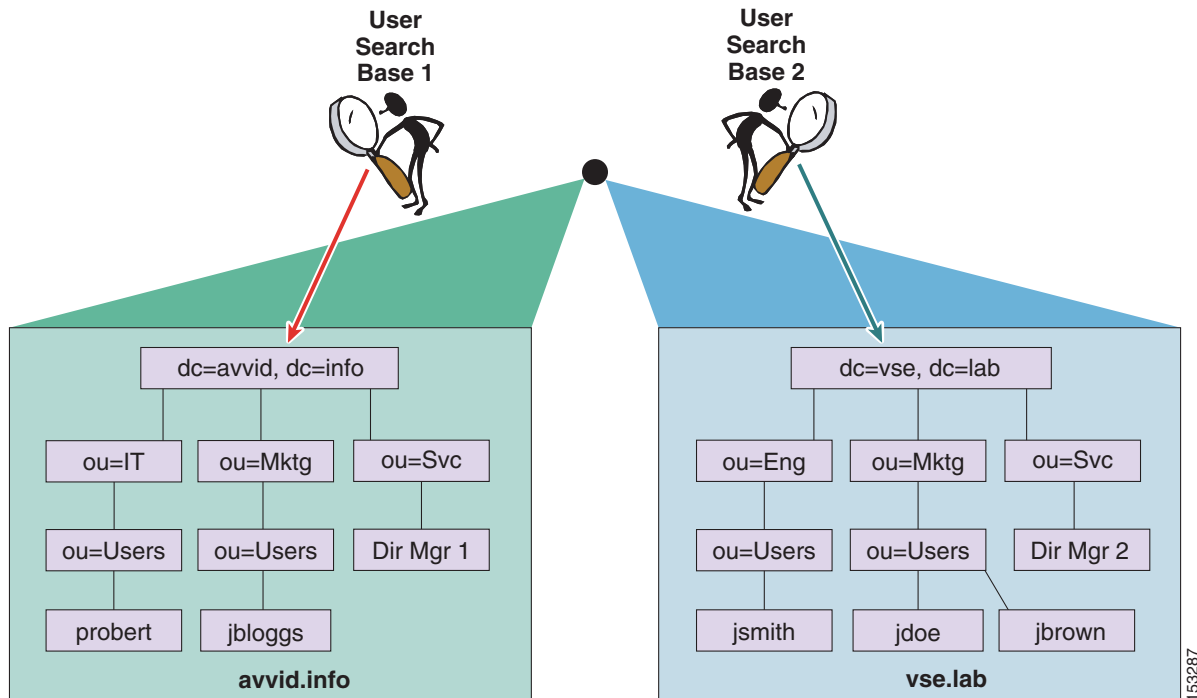
Additional Considerations for Microsoft Active Directory

A synchronization agreement for a domain will not synchronize users outside of that domain nor within a child domain because Unified CM does not follow AD referrals during the synchronization process. The example in [Figure 17-10](#) requires three synchronization agreements to import all of the users. Although Search Base 1 specifies the root of the tree, it will not import users that exist in either of the child domains. Its scope is only VSE.LAB, and separate agreements are configured for the other two domains to import those users.

Figure 17-10 Synchronization with Multiple Active Directory Domains

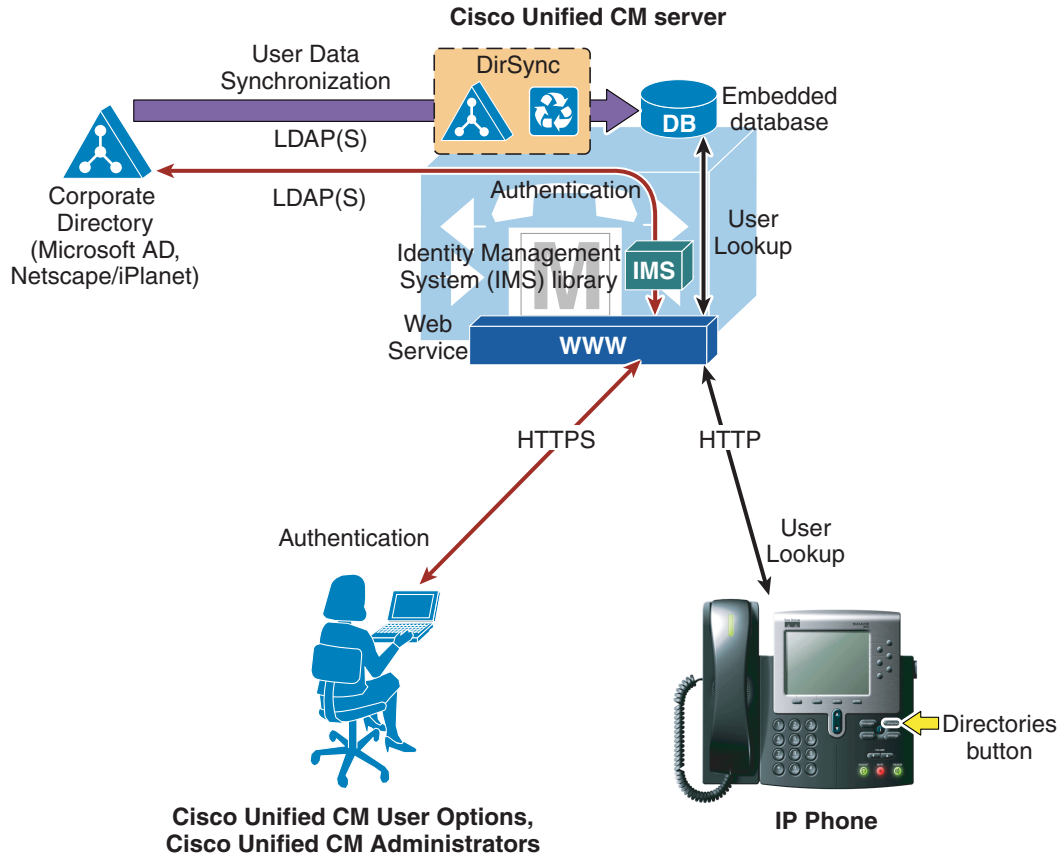
In Figure 17-10, each of the domains and sub-domains contains at least one domain controller (DC) associated to them, and the three synchronization agreements each specify the appropriate domain controller. The DCs have information only on users within the domain where they reside, therefore three synchronization agreements are required to import all of the users.

When synchronization is enabled with an AD forest containing multiple trees, as shown in Figure 17-11, multiple synchronization agreements are still needed for the same reasons listed above. Additionally, the UserPrincipalName (UPN) attribute is guaranteed by Active Directory to be unique across the forest and must be chosen as the attribute that is mapped to the Unified CM UserID. For additional considerations on the use of the UPN attribute in a multi-tree AD scenario, see the section on [Additional Considerations for Microsoft Active Directory](#), page 17-22.

Figure 17-11 Synchronization with Multiple AD Trees (Discontiguous Namespaces)

LDAP Authentication

The LDAP authentication feature enables Unified CM to authenticate end user passwords against a corporate LDAP directory instead of using the embedded database. This authentication is accomplished with an LDAPv3 connection established between the IMS module within Unified CM and a corporate directory server, as shown in [Figure 17-12](#).

Figure 17-12 Enabling LDAP Authentication

153288

As in the case of the LDAP synchronization feature, the supported corporate directory products are:

- Microsoft Active Directory (AD) 2000 and 2003
- Netscape Directory Server 4.x, iPlanet Directory Server 5.1, and Sun ONE Directory Server 5.2

To enable authentication, a single authentication agreement may be defined for the entire cluster. The authentication function allows configuration of up to three servers for redundancy, and it also supports secure connections to the directory server if you optionally enable LDAP over SSL (SLDAP). It is not possible to enable the authentication functionality without also enabling the synchronization functionality.

The following statements describe Unified CM's behavior when authentication is enabled:

- End user passwords are authenticated against the corporate directory.
- Application user passwords are authenticated against the Unified CM database.
- End user PINs are authenticated against the Unified CM database.

This behavior is in line with the guiding principle of providing single logon functionality for end users while making the operation of the real-time IP Communications system independent of the availability of the corporate directory, and is shown graphically in [Figure 17-13](#).

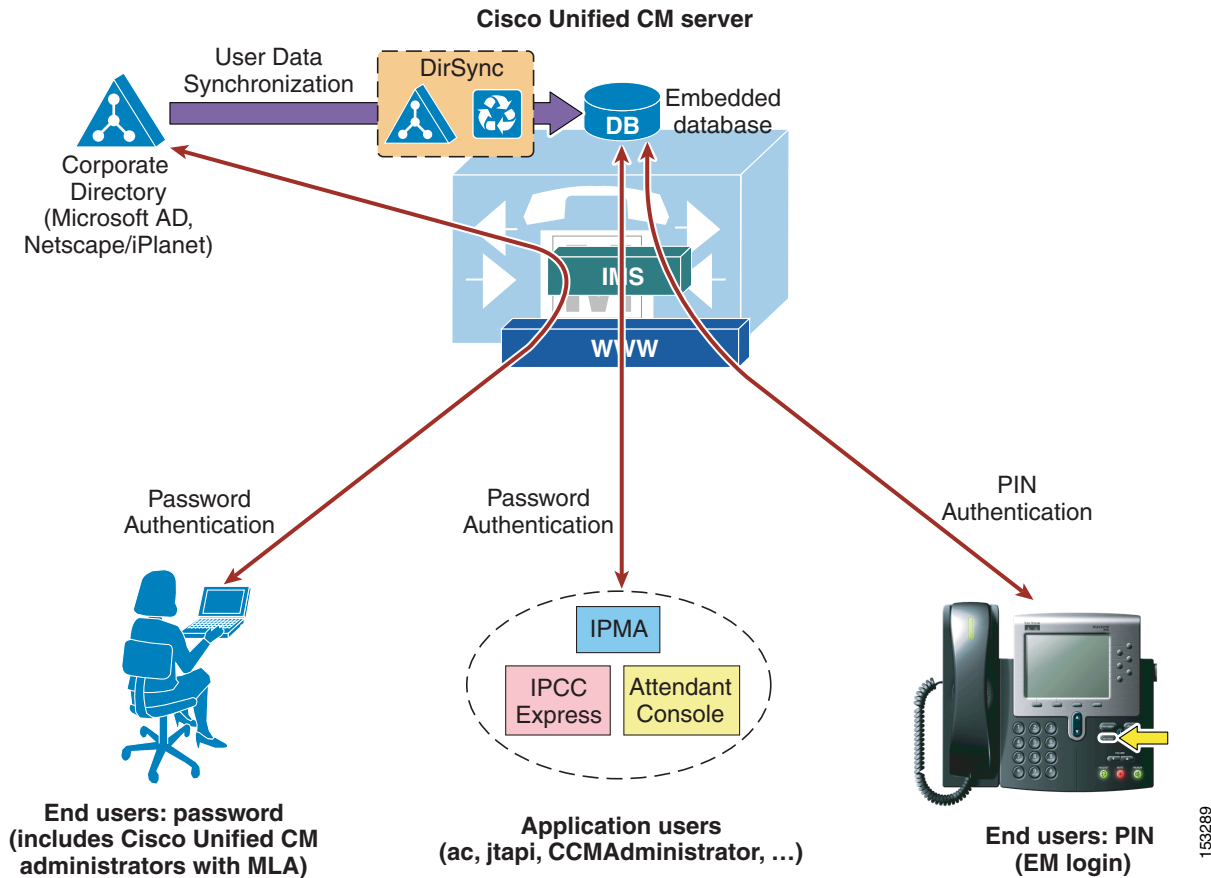
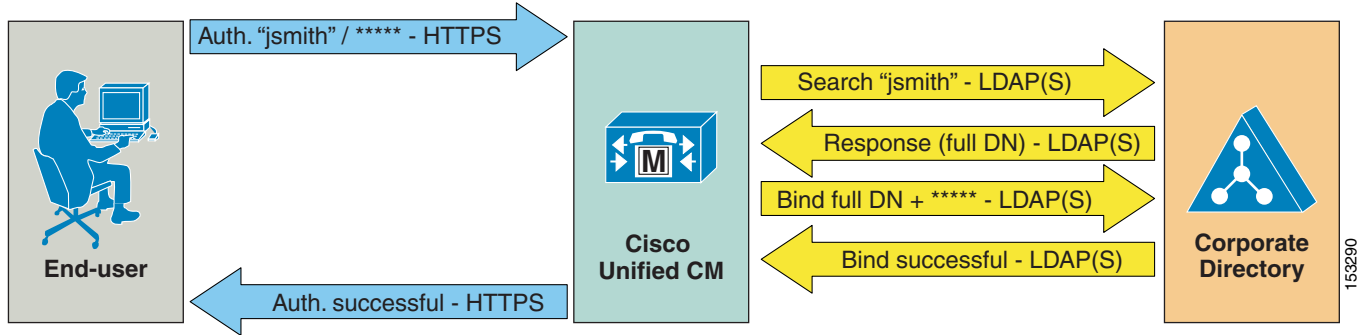
Figure 17-13 Authenticating End User Passwords, Application User Passwords, and End User PINs

Figure 17-14 illustrates the following process, adopted by Unified CM to authenticate an end user against a corporate LDAP directory:

1. First, a user connects to the Unified CM User Options page via HTTPS and attempts to authenticate with a user name and password. In this example, the user name is jsmith.
2. Unified CM then issues an LDAP query for the user name jsmith, using the value specified in the LDAP Search Base on the LDAP Authentication configuration page as the scope for this query. If SLDAP is enabled, this query travels over an SSL connection.
3. The corporate directory server replies via LDAP with the full Distinguished Name (DN) of user jsmith (for example, "cn=jsmith, ou=Users, dc=vse, dc=lab").
4. Unified CM then attempts an LDAP bind using this full DN and the password provided by the user.
5. If the LDAP bind is successful, Unified CM allows the user to proceed to the configuration page requested.

Figure 17-14 Authentication Process

Observe the following design and implementation best-practices when deploying LDAP authentication with Cisco Unified CM 5.x:

- Create an account within the corporate directory to allow Unified CM to connect and authenticate to it. Cisco recommends that you use an account dedicated to Unified CM, with minimum permissions set to "read" all user objects within the desired search base and with a password set to never expire. (When the password for this account changes in the directory, Unified CM must be reconfigured to take the change into account.) If LDAP synchronization is also enabled, you can use the same account for both functions.
- Enable LDAP authentication on Unified CM by specifying the credentials of the aforementioned account under LDAP Manager Distinguished Name and LDAP Password, and by specifying the directory subtree where all the users reside under LDAP User Search Base.
- Configure at least two LDAP servers for redundancy. You can use IP addresses instead of host names to eliminate dependencies on Domain Name System (DNS) availability.
- This method provides single logon functionality to all end users: when they log in to the Unified CM User Options page, they can now use their corporate directory credentials.
- Manage end user passwords from within the corporate directory interface. (Note that the password field is no longer displayed in the Unified CM Administration pages when authentication is enabled.)
- Manage end user PINs from within Unified CM Administration or from the Unified CM User Options page.
- Manage Application User passwords from within Unified CM Administration. (Remember that these virtual users are only for communication with other Cisco IP Communications functions and applications, and they are not associated with real people.)
- Enable single logon for Unified CM administrators by adding their corresponding end user to the Unified CM Super Users user group from within the Unified CM Administration pages. Multiple levels of administrator rights can be defined by creating customized user groups and roles.

Additional Considerations for Microsoft Active Directory

When you enable LDAP authentication with Microsoft Active Directory, Cisco recommends that you configure Unified CM to query a Microsoft Active Directory Global Catalog server for faster response times.

To enable queries against the Global Catalog, simply configure the LDAP Server Information in the LDAP Authentication page to point to the IP address or host name of a Domain Controller that has the Global Catalog role enabled, and configure the LDAP port as 3268.

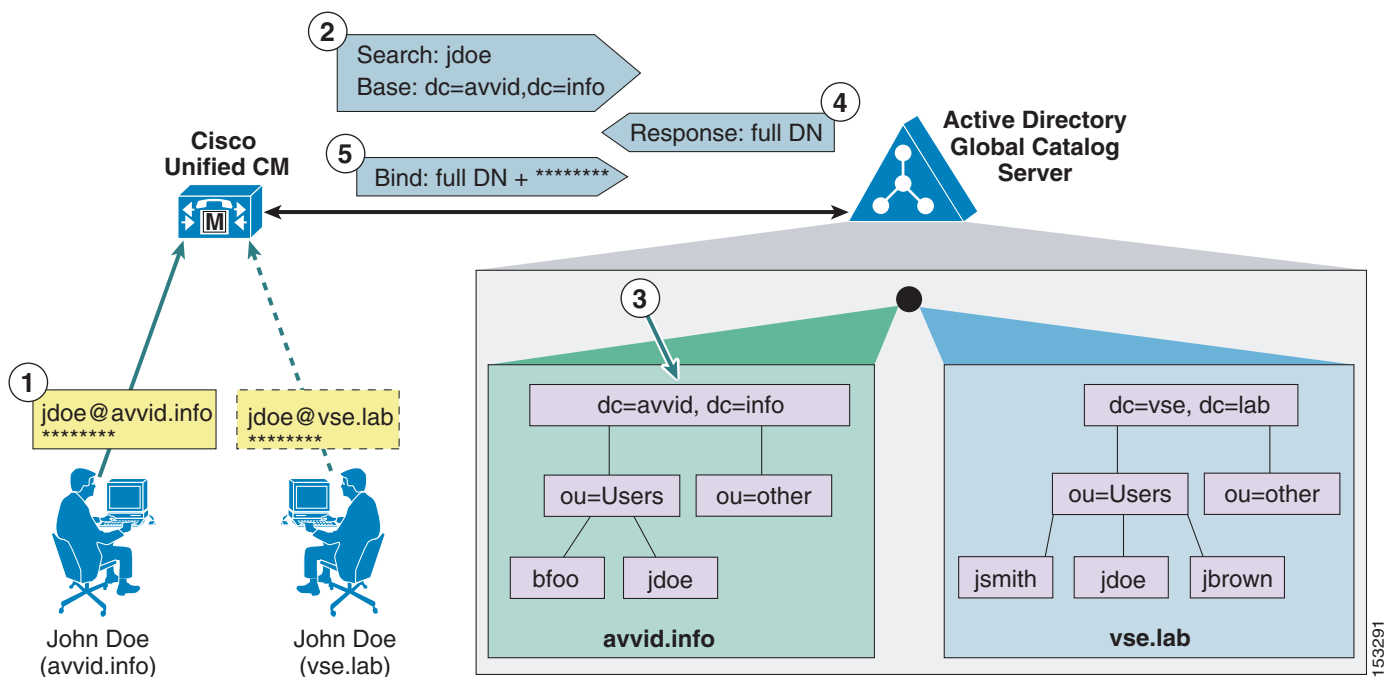
The use of Global Catalog for authentication becomes even more efficient if the users synchronized from Microsoft AD belong to multiple domains, because it allows Unified CM to authenticate users immediately without having to follow referrals. For these cases, point Unified CM to a Global Catalog server and set the LDAP User Search Base to the top of the root domain.

In the case of a Microsoft AD forest that encompasses multiple trees, some additional considerations apply. Because a single LDAP search base cannot cover multiple namespaces, Unified CM must use a different mechanism to authenticate users across these discontinuous namespaces.

As mentioned in the section on [LDAP Synchronization, page 17-11](#), in order to support synchronization with an AD forest that has multiple trees, the UserPrincipalName (UPN) attribute must be used as the user ID within Unified CM. When the user ID is the UPN, the LDAP authentication configuration page within Unified CM Administration does not allow you to enter the LDAP Search Base field, but instead it displays the note, "LDAP user search base is formed using userid information."

In fact, the user search base is derived from the UPN suffix for each user, as shown in [Figure 17-15](#). In this example, a Microsoft Active Directory forest consists of two trees, `avvid.info` and `vse.lab`. Because the same user name may appear in both trees, Unified CM has been configured to use the UPN to uniquely identify users in its database during the synchronization and authentication processes.

Figure 17-15 Authentication with Microsoft AD Forests with Multiple Trees



As shown in [Figure 17-15](#), a user named John Doe exists in both the `avvid.info` tree and the `vse.lab` tree. The following steps illustrate the authentication process for the first user, whose UPN is `jdoe@avvid.info`:

1. The user authenticates to Unified CM via HTTPS with its user name (which corresponds to the UPN) and password.
2. Unified CM performs an LDAP query against a Microsoft Active Directory Global Catalog server, using the user name specified in the UPN (anything before the @ sign) and deriving the LDAP search base from the UPN suffix (anything after the @ sign). In this case, the user name is `jdoe` and the LDAP search base is "`dc=avvid, dc=info`".

3. Microsoft Active Directory identifies the correct Distinguished Name corresponding to the user name in the tree specified by the LDAP query. In this case, "cn=jdoe, ou=Users, dc=avvid, dc=info".
4. Microsoft Active Directory responds via LDAP to Unified CM with the full Distinguished Name for this user.
5. Unified CM attempts an LDAP bind with the Distinguished Name provided and the password initially entered by the user, and the authentication process then continues as in the standard case shown in [Figure 17-14](#).

**Note**

Support for LDAP authentication with Microsoft AD forests containing multiple trees relies exclusively on the approach described above. Therefore, support is limited to deployments where the UPN suffix of a user corresponds to the root domain of the tree where the user resides. If the UPN suffix is disjointed from the actual namespace of the tree, it is not possible to authenticate Unified CM users against the entire Microsoft Active Directory forest. (It is, however, still possible to use a different attribute as user ID and limit the integration to a single tree within the forest.)

User Filtering for Directory Synchronization and Authentication

This section describes the LDAP Query Filter mechanism available in Cisco Unified CM 5.0 and later. To optimize directory synchronization performance, Cisco recommends importing only those directory user accounts that will be assigned IP telephony resources from Unified CM. In many cases, a unique search base is all that is needed. When a unique search base is not available, a custom LDAP filter might be required. The information in the following sections addresses both methods that can be used to optimize directory synchronization.

Optimizing Unified CM Database Synchronization

The Unified CM Database Synchronization feature provides a mechanism for importing a subset of the user configuration data (attributes) from the LDAP store into the Unified CM publisher database. Once synchronization of a user account has occurred, the copy of each user's LDAP account information may then be associated to additional data required to enable specific IP communication features for that user. When authentication is also enabled, the user's credentials are used to bind to the LDAP store for password verification. The end user's password is never stored in the Unified CM database when enabled for synchronization and/or authentication.

User account information is cluster-specific. Each Unified CM publisher server maintains a unique list of those users receiving IP telephony services from that cluster. Synchronization agreements are cluster-specific, and each publisher has its own unique copy of user account information.

Only those users who will be assigned IP telephony resources should be synchronized with Unified CM. The following is a partial list of common reasons why the entire set of users defined in the LDAP directory should not be imported into the Unified CM cluster:

- Importing users who will not be assigned IP telephony resources can increase directory synchronization time.
- Importing users who will not be assigned IP telephony resources can slow Unified CM searches and overall database performance.

- In many cases, the number of user accounts in the LDAP store far exceeds the total user capacity of the Unified CM cluster. Limit synchronization of users to reflect the recommended limits for specific software and hardware combinations as listed in the Cisco Unified CM Release Notes, available at

<http://www.cisco.com>

Unified CM has no enforced limit on the number of accounts that may be added to the system. In all cases, Cisco recommends limiting the number of users to the same quantity as the supported number of endpoints. For example, a full-sized cluster of Cisco MCS-7845 servers can support up to 30,000 endpoints, and the system should be administered to limit the number of users to roughly 30,000. There might be cases where accounts are needed for applications, and some designs might require additional accounts.

Cisco recommends using the control mechanisms described here to minimize the number of user accounts imported, regardless of the LDAP database size. This will improve the speed of the first and subsequent periodic synchronizations and will also improve manageability of the user accounts.

Using the LDAP Structure to Control Synchronization

Many deployments of LDAP directories use the Organizational Unit Name (OU) to organize users into some logical grouping. If the LDAP directory has a structure that organizes users into multiple OUs, then it often is possible to use that structure to control which users are imported. Each individual Unified CM synchronization agreement specifies a single OU. Only those users in the OU are synchronized. When multiple OUs containing users are required in a cluster, then multiple synchronization agreements are required. When an OU contains users that will *not* be assigned IP telephony resources, then Cisco recommends omitting those OUs from the directory synchronization.



CHAPTER 18

IP Telephony Migration Options

Last revised on: September 27, 2007

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 18-1](#)
- [Parallel Cutover, page 18-2](#)

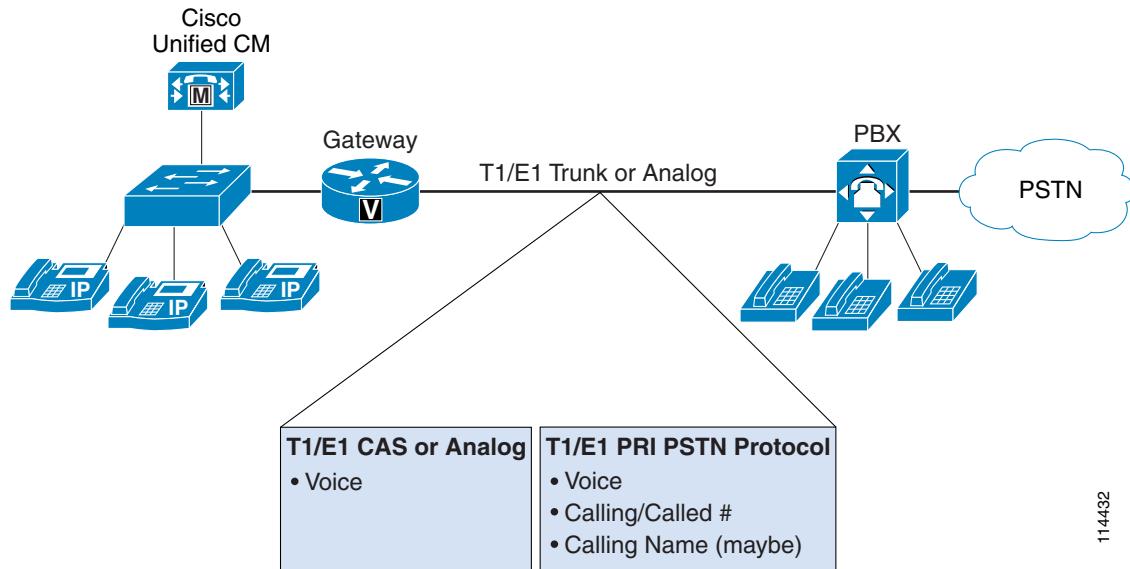
Neither method is right or wrong, and both depend upon individual customer circumstances and preferences to determine which option is most suitable.

This chapter also discusses [The Need for QSIG in Multisite Enterprises, page 18-3](#).

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 Unified Communications Manager (Unified CM) can support either regular PSTN-type PRI or QSIG PRI as well as H.323 and SIP. Of these options, T1/E1 QSIG provides the highest level of feature transparency between the two systems.

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 18-1](#).

Figure 18-1 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 Unified CM because Unified CM can be configured as the "network" side of the connection.

Cisco Unified CM, 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 Multisite Enterprises](#), page 18-3.)

With either PSTN-type PRI or QSIG, the process of phased migration is similar: move subscribers from the PBX to Unified CM 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 Unified CM. During the weekend, new extensions were created in Unified CM 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.

Parallel Cutover

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 parallel cutover has the following advantages over a phased migration:

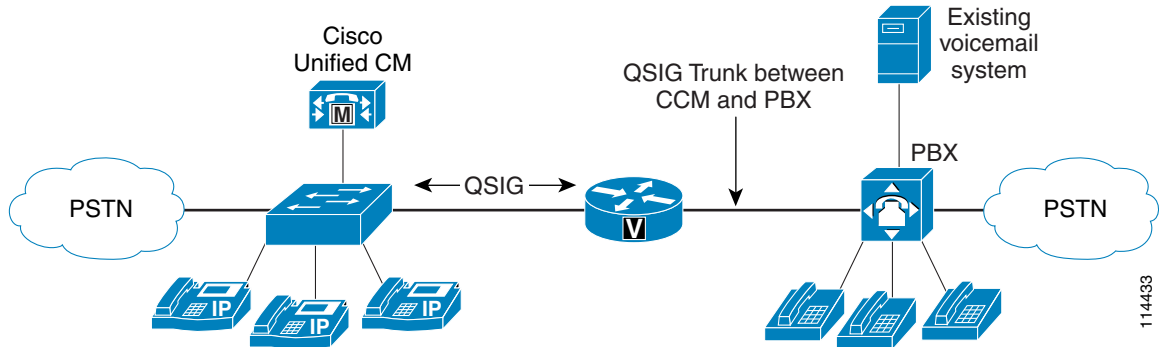
- If something unexpected occurs, the parallel cutover provides a back-out plan that allows you to revert to the PBX system by simply transferring the inbound PSTN trunks from the IP Telephony gateways back to the PBX.
- The parallel cutover allows for verification of the IP Telephony database configuration 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 parallel cutover.

One disadvantage of the parallel 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 Multisite 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 multisite 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 Unified CM Release 3.3 to enable Unified CM to be part of a large enterprise network. (See [Figure 18-2](#).)

Figure 18-2 QSIG Used Between Cisco Unified CM and a PBX

QSIG as implemented in Cisco Unified CM Release 4.1 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
- Divert (by re-route)
- Divert (responding to “check restriction” request)
- Alerting name (on-ringing)
- Path Replacement
- Callback — Call Completion Busy Subscriber (CCBS) and Call Completion No Reply (CCNR)

By supporting QSIG, Unified CM 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 parallel cutover method usually works best in most cases. In addition, large enterprises can improve upon either migration method by using QSIG to enable Unified CM to become part of the enterprise network.

Cisco has a lab facility dedicated to testing interoperability between Unified CM and PBX systems. The results of that testing are made available as Application Notes, which are posted at

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

The Application Notes are updated frequently, and new documents are continuously added to this website. Check the website often to obtain the latest information.



CHAPTER 19

Voice Security

Last revised on: October 30, 2008

This chapter presents guidelines and recommendations for securing IP Telephony networks. Following the guidelines in this chapter does not guarantee a secure environment, nor will it prevent all penetration attacks on a network. You can achieve reasonable security by establishing a good security policy, following that security policy, staying up-to-date on the latest developments in the hacker and security communities, and maintaining and monitoring all systems with sound system administration practices.

The security guidelines presented in this chapter pertain specifically to IP Telephony technology and the voice network. For more information on data network security, refer to the Cisco SAFE Blueprint documentation available at

http://www.cisco.com/en/US/netsol/ns744/networking_solutions_program_home.html

This chapter addresses centralized but not distributed call processing, and it includes clustering over the WAN but not local failover mechanisms such as Survivable Remote Site Telephony (SRST). This chapter assumes that all remote sites have a redundant link to the head-end or local call-processing backup in case of head-end failure. The interaction between Network Address Translation (NAT) and IP Telephony, for the most part, is not addressed here. This chapter also assumes that all networks are privately addressed and do not contain overlapping IP addresses.

What's New in This Chapter

Table 19-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 19-1 *New or Changed Information Since the Previous Release of This Document*

New or Revised Topic	Described in:
Antivirus software on Cisco Security Agent	Cisco Security Agent on the Unified CM and Application Servers , page 19-39
Firewalls and H.323	Firewalls and H.323 , page 19-31
OS hardening	General Server Guidelines , page 19-40

General Security

This section covers general security features and practices that can be used to protect the voice data within a network.

Security Policy

This chapter presumes that your enterprise already has a security policy in place. Cisco Systems recommends that you do not deploy any technology without an associated security policy. The security policy defines which data in your network is sensitive so that it can be protected properly when transported throughout the network. Having this security policy helps you define the security levels required for the types of data traffic that are on your network. Each type of data may or may not require its own security policy.

If no security policy exists for data on the company network, you should create one before enabling any of the security recommendations in this chapter. Without a security policy, there is no way of telling if the security that is enabled in a network is doing what it is designed to accomplish. Without a security policy, there is also no systematic way of enabling security for all the applications and types of data that run in a network.

**Note**

While it is important to adhere to the security guidelines and recommendations presented in this chapter, they alone are not sufficient to constitute a security policy for your company. You must define a corporate security policy before implementing any security technology.

This chapter details the features and functionality of a Cisco Systems network that are available to protect the voice data on a network. It is up to the security policy to define which data to protect, how much protection is needed for that type of data, and which security techniques to use to provide that protection.

One of the more difficult issues with a security policy that includes IP Telephony is combining the security policies that usually exist for both the data network and the traditional voice network. Ensure that all aspects of the integration of the voice data onto the network are secured at the correct level for your security policy or corporate environment.

The basis of a good security policy is defining how important your data is within the network. Once you have ranked the data according to its importance, you can decide how the security levels should be established for each type of data. You can then achieve the correct level of security by using both the network and application features.

In summary, you can use the following process to define a security policy:

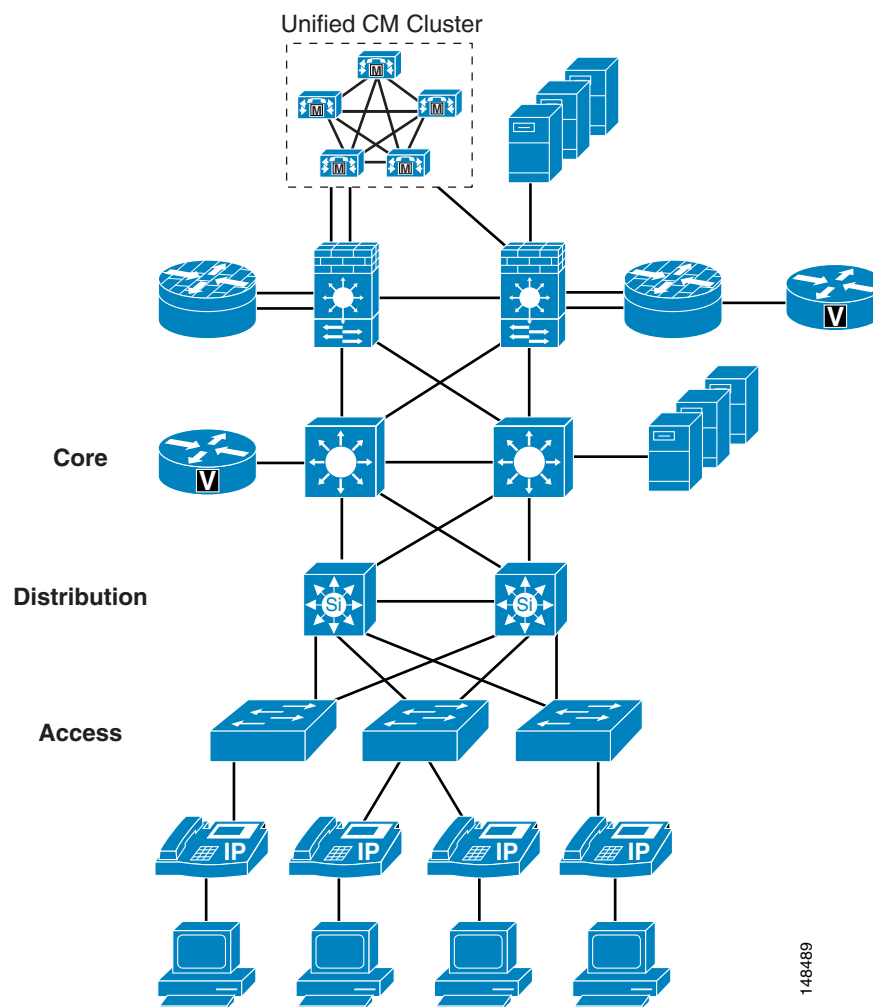
- Define the data that is on the network.
- Define the importance of that data.
- Apply security based on the importance of the data.

Security in Layers

This chapter starts at the phone port that a user can plug a PC into, and it works its way through the network from the phone to the access switch, to the distribution layer, into the core, and then into the data center. (See [Figure 19-1](#).) We build layer upon layer of security, starting at the access port into the network itself. With each feature or function that is discussed, we list advantages and disadvantages that need to be taken into account from the standpoint of your corporate security policy.

For example, [Figure 19-1](#) shows both the advantage and disadvantage of using an IP Telephony network. The voice products can be placed anywhere in a network because they use IP to connect to all those devices. This feature gives a network designer the ability to place the devices where it is both physically and logically easy to deploy IP Telephony applications. But because of this ease of deployment, the security complexity increases because the IP Telephony devices can be placed anywhere in a network as long as they have connectivity.

Figure 19-1 **Layers of Security**



Secure Infrastructure

As the IP Telephony data crosses a network, that data is only as safe and secure as the devices that are transporting the data. Depending on the security level that is defined in your security policy, the security of the network devices might have to be improved or they might already be secure enough for the transportation of IP Telephony traffic.

There are many best-practices within a data network that, if used, will increase the entire security of your network. For example, instead of using Telnet (which sends passwords in clear text) to connect to any of the network devices, use Secure Shell (SSH, the secure form of Telnet) so that an attacker would not be able to see a password in clear text. There are many documents on the Cisco.com website that talk about overall security within a network. Use that documentation along with your security policy to determine what security the infrastructure needs.

Video Infrastructure

The Cisco IOS feature sets that provide the gatekeeper functionality (that is, the IP/H323 and EnterprisePlus/H323 MCU feature sets) support only Telnet and not Secure Shell (SSH). Cisco recommends that you use access control lists (ACLs) to control who is permitted to connect to the routers using Telnet, and that you always try to connect to the gatekeeper from a host that is in a secure segment of the network, because user names and passwords are sent over Telnet in clear text.

Cisco Unified Videoconferencing 3500 Series MCUs and H.320 gateways support Telnet, FTP, HTTP, and SNMP. These IP/VC devices do not support TACACS or RADIUS authentication, and only a limited number of administrative accounts can be configured locally in the device. User names and passwords are sent via clear text in all Telnet, FTP, HTTP, and SNMP communications. Cisco recommends that you access these devices from a host that is in a secure segment of the network. You should also use firewalls, access control lists, Cisco Authentication Proxy, and other Cisco security tools to help protect these devices from unauthorized access.

The following links list just a few of the security documents available on Cisco.com:

- Best Practices for Cisco Switches (login authentication required)
http://cisco.com/en/US/partner/products/hw/switches/ps663/products_tech_note09186a0080094713.shtml
- SAFE: A Security Blueprint for Enterprise Networks
http://www.cisco.com/en/US/prod/collateral/wireless/wirelssw/ps1953/product_implementation_design_guide09186a00800a3016.pdf

Physical Security

Just as a traditional PBX is usually locked in a secure environment, the IP network should be treated in a similar way. Each of the devices that carries IP Telephony traffic is really part of an IP PBX, and normal general security practices should be used to control access to those devices. Once a user or attacker has physical access to one of the devices in a network, all kinds of problems could occur. Even if you have excellent password security and the user or attacker cannot get into the network device, that does not mean that they cannot cause havoc in a network by simply unplugging the device and stopping all traffic.

For more information on general security practices, refer to the documentation at the following locations:

- http://www.cisco.com/en/US/netsol/ns744/networking_solutions_program_home.html
- http://www.cisco.com/web/about/ac123/iqmagazine/archives/q2_2005/addressing_network_security.html

IP Addressing

IP addressing can be critical for controlling the data that flows in and out of the logically separated IP Telephony network. The more defined the IP addressing is within a network, the easier it becomes to control the devices on the network.

As stated in other sections of this document (see [Campus Access Layer, page 3-4](#)), you should use IP addressing based on RFC 1918. This method of addressing allows deployment of a IP Telephony system into a network without redoing the IP addressing of the network. Using RFC 1918 also allows for better control in the network because the IP addresses of the voice endpoints are well defined and easy to understand. If the voice endpoints are all addressed within a 10.x.x.x. network, access control lists (ACLs) and tracking of data to and from those devices are simplified.

Advantages

If you have a well defined IP addressing plan for your voice deployments, it becomes easier to write ACLs for controlling the IP Telephony traffic and it also helps with firewall deployments.

Using RFC 1918 enables you easily to deploy one VLAN per switch, which is a best-practise for campus design, and also enables you to keep the Voice VLAN free of any Spanning Tree Protocol (STP) loops.

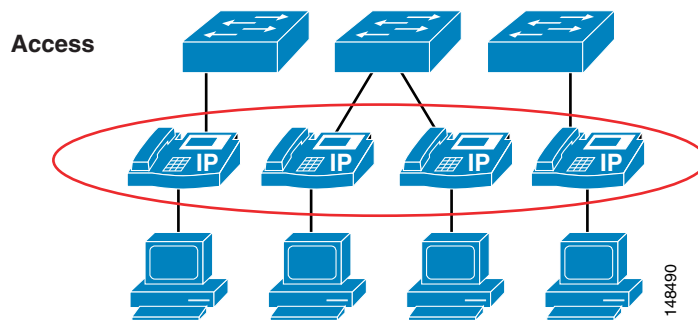
If deployed correctly, route summarization could help to keep the routing table about the same as before the voice deployment, or just slightly larger.

Disadvantages

Routing tables could become large if not designed correctly or if route summarization is not used.

Phone Security

Cisco Unified IP Phones contain built-in features to increase security on a IP Telephony network. These features can be enabled or disabled on a phone-by-phone basis to increase the security of an IP Telephony deployment. Depending on the placement of the phones, a security policy will help determine if these features need to be enabled and where they should be enabled. (See [Figure 19-2](#).)

Figure 19-2 Security at the Phone Level

Before attempting to configure the security features on a phone, check the documentation at the following link to make sure the features are available on that particular phone model:

http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_administration_guide_chapter_09186a00803fe693.html

PC Port on the Phone

The phone has the ability to turn on or turn off the port on the back of the phone, to which a PC would normally be connected. This feature can be used as a control point to access the network if that type of control is necessary.

Depending on the security policy and placement of the phones, the PC port on the back of any given phone might have to be disabled. Disabling this port would prevent a device from plugging into the back of the phone and getting network access through the phone itself. A phone in a common area such as a lobby would typically have its port disabled. Most companies would not want someone to get into the network on a non-controlled port because physical security is very weak in a lobby. Phones in a normal work area might also have their ports disabled if the security policy requires that no device should ever get access to the network through a phone PC port. Depending on the model of phone deployed, Cisco Unified Communications Manager (Unified CM) can disable the PC port on the back of the phone. Before attempting to enable this feature, check the documentation at the following link to verify that this feature is supported on your particular model of Cisco Unified IP Phone:

http://www.cisco.com/en/US/products/hw/phones/ps379/tsd_products_support_series_home.html

Advantages

Disabling the phone PC port allows for phones to be placed in areas where network access from the phone should not be allowed. It controls access to the network that otherwise would be there if the PC port on the back of the phone was enabled.

Disadvantages

For each person who needs to have network access and is approved for access, a separate Ethernet port would be required to provide that person with network access if the PC port on the phone is disabled. A person could still unplug the ethernet jack from the phone and attempt to plug it into another device.

For Cisco Unified Video Advantage to operate properly, the PC port and Video Capabilities must both be enabled. The other settings can safely be disabled.

Gratuitous ARP

Just like any other data device on the network, the phones are vulnerable to traditional data attacks. The phones have features to prevent some of the common data attacks that can occur on a corporate network. One such feature is Gratuitous APR (Gratuitous Address Resolution Protocol, or GARP). This feature helps to prevent man-in-the-middle (MITM) attacks to the phone. A MITM attack involves an attacker who tricks an end station into believing that he is the router and tricks the router into believing that he is the end station. This scheme makes all the traffic between the router and the end station travel through the attacker, thus enabling the attacker to log all of the traffic or inject new traffic into the data conversation.

Gratuitous ARP helps protect the phones from having an attacker capture the signaling and RTP voice streams from the phone if the attacker was able to get onto the voice segment of the network. This feature protects only the phones; it does not protect the rest of the infrastructure from a Gratuitous ARP attack. This feature is of less importance if you are running a Cisco infrastructure because the switch port provides features that protect both the phones and the network gear. For a description of these switch port features see the section on [Switch Port](#), page 19-13.

Advantages

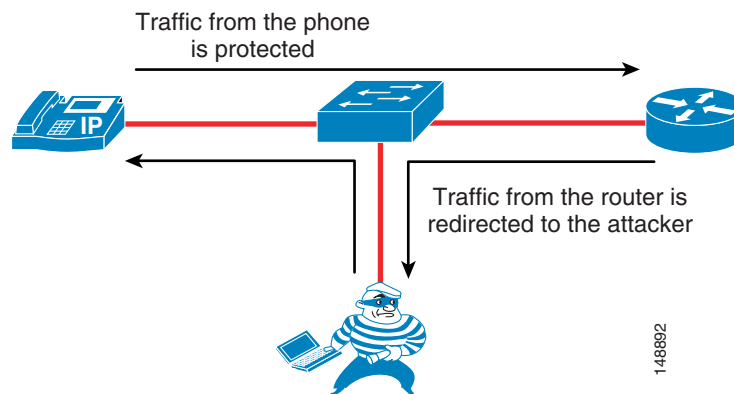
The Gratuitous ARP feature protects the phone from a traditional MITM attack on the signaling and RTP voice streams that are sourced from the phone to the network.

Disadvantages

The downstream signaling and RTP voice streams coming from another phone or coming across the network are not protected by this feature in the phone. Only the data coming from the phone that has this feature enabled is protected. (See [Figure 19-3](#).)

If the default gateway is running Hot Standby Router Protocol (HSRP), if the HSRP configuration uses the burned-in MAC address rather than the virtual MAC address for the default gateway, and if the primary router fails-over to a secondary router that has a new MAC address, the phones could maintain the old MAC address of the default gateway. This scenario could cause an outage for up to 40 minutes. Always use the virtual MAC address in an HSRP environment to avoid this potential problem.

Figure 19-3 *Gratuitous ARP Protects the Phone that Has It but Not Other Traffic*



As shown in [Figure 19-3](#), the traffic from the phone that has Gratuitous ARP is protected, but the attacker could still see the traffic coming from another endpoint because that endpoint might not have the ability to protect the data flow.

PC Voice VLAN Access

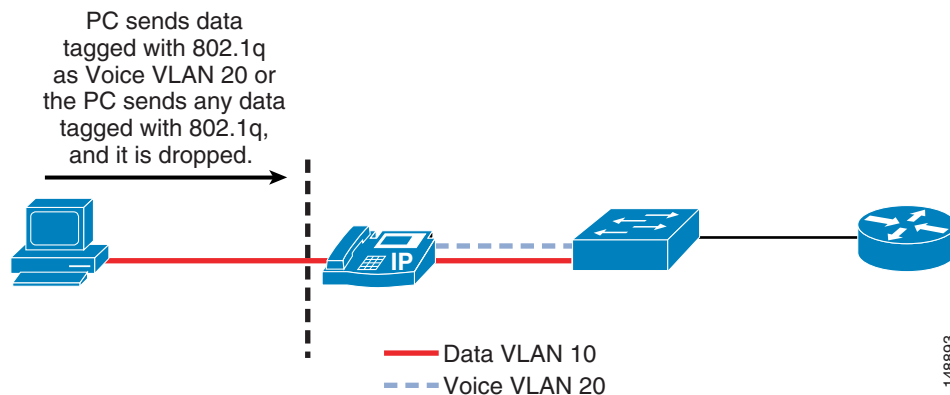
Because there are two VLANs from the switch to the phone, the phone needs to protect the voice VLAN from any unwanted access. The phones can prevent unwanted access into the voice VLAN from the back of the phone. A feature called PC Voice VLAN Access prevents any access to the voice VLAN from the PC port on the back of the phone. When disabled, this feature does not allow the devices plugged into the PC port on the phone to "jump" VLANs and get onto the voice VLAN by sending 802.1q tagged information destined for the voice VLAN to the PC port on the back of the phone. The feature operates one of two ways, depending on the phone that is being configured. On the more advanced phones, the phone will block any traffic destined for the voice VLAN that is sent into the PC port on the back of the phone. In the example shown in Figure 19-4, if the PC tries to send any voice VLAN traffic (with an 802.1q tag of 200 in this case) to the PC port on the phone, that traffic will be blocked. The other way this feature can operate is to block all traffic with an 802.1q tag (not just voice VLAN traffic) that comes into the PC port on the phone.

Currently, 802.1q tagging from an access port is not normally used. If that feature is a requirement for the PC plugged into the port on the phone, you should use a phone that allows 802.1q tagged packets to pass through the phone.

Before attempting to configure the PC Voice VLAN Access feature on a phone, check the documentation at the following link to make sure the feature is available on that particular phone model:

http://www.cisco.com/en/US/products/hw/phones/ps379/tsd_products_support_series_home.html

Figure 19-4 Blocking Traffic to the Voice VLAN from the Phone PC Port



Advantages

The PC Voice VLAN Access feature prevents attackers from sending uncontrolled data into the voice VLAN through the PC port on the back of the phone.

Disadvantages

If the device plugged into the phone is normally allowed to send 802.1q tagged packets, then these packets will be dropped. Most end stations are not allowed to perform this function at the access layer. If that function is considered normal operation within a network, this feature would not allow that function to work.

Web Access

Each Cisco Unified IP Phone has a web server built into it to help with debugging and remote status of the phone for management purposes. The web server also enables the phones to receive applications pushed from Cisco Unified Communications Manager (Unified CM) to the phones. Access to this web server can be enabled or disabled on a phone by means of the Web Access feature in the Unified CM configuration. This setting can be global, or it could be enabled or disabled on a phone-by-phone basis.

Advantages

With Web Access enabled on the phones, the phones can be used to assist in debugging issues with a phone or within the network. If Web Access from the phone is disabled, users or an attacker cannot get information from the phone about the IP Telephony network.

Disadvantages

If Web Access from the phone is disabled, debugging network or IP Telephony issues can be more difficult. If the web server is globally disabled but it is needed to help with debugging, then the administrator for Unified CM will have to enable this feature on the phones. The ability to get to this web page can be controlled by an ACL in the network, leaving network operators with the capability to get to the web page when needed.

With the Web Access feature disabled, the phones will be unable to receive applications pushed to them from Unified CM.

Video Capabilities

For Cisco Unified Video Advantage to operate properly, the PC port and Video Capabilities must both be enabled. The other settings can safely be disabled. The Device Security Mode operates as specified, even when Cisco Unified Video Advantage is in use, but Cisco Unified Video Advantage itself does not support authentication or encryption of the Cisco Audio Session Tunnel (CAST) protocol or its RTP media traffic. When an IP Phone is in Authenticated mode, the Skinny Client Control Protocol (SCCP) signaling between the phone and Unified CM will be authenticated, but the CAST signaling between the phone and Cisco Unified Video Advantage will not be authenticated. Likewise, when a phone is in Encrypted mode, the audio stream between the phones will be encrypted, but the video streams between the Cisco Unified Video Advantage clients will not be encrypted. Users should be notified that the video channel is not encrypted even though an icon on the phone appears to indicate that they are in an encrypted call.

Advantages

The PC port and Video Capabilities are required for Cisco Unified Video Advantage to function correctly.

Disadvantages

Enabling these features could possibly allow communication to the phone from the PC if ACLs are not used to protect the phone.

Settings Access

Each Cisco Unified IP Phone has a network settings page that lists many of the network elements and detailed information that is needed for the phone to operate. This information could be used by an attacker to start a reconnaissance on the network with some of the information that is displayed on the phone's web page. For example, an attacker could look at the settings page to determine the default gateway, the TFTP server, and the Unified CM IP address. Each of these pieces of information could be used to gain access to the voice network or to attack a device in the voice network.

This access can be disabled on a phone-by-phone basis (see [Figure 19-5](#)) to prevent end users or attackers from obtaining the additional information such as Unified CM IP address and TFTP server information.

For more information on the phone settings page, refer to *Cisco Unified IP Phone Authentication and Encryption for Cisco Unified CallManager*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

Figure 19-5 Phone Configuration Page in Unified CM

Product Specific Configuration	
Disable Speakerphone	☐
Disable Speakerphone and Headset	☐
Forwarding Delay*	Disabled
PC Port*	Disabled
Settings Access*	Disabled
Gratuitous ARP*	Disabled
PC Voice VLAN Access*	Disabled
Video Capabilities*	Disabled
Auto Line Select*	Disabled
Web Access*	Disabled

Advantages

With access to the phone settings page disabled, end users and potential attackers are not able to see detailed information about the network and the IP Telephony information used by the voice system. With this feature disabled, some of the information that would be protected includes the IP address of the phone and the Unified CM to which the phone is registered.

Disadvantages

With access to the phone settings page disabled, end users lose the ability to change many of the settings on the phone that they would normally be able to control, such as speaker volume, contrast, and ring type. It might not be practical to use this security feature because of the limitations it places on end users with respect to the phone interface. However, access to the phone settings page does not have to be lost if the administrator chooses to restrict access rather than disable it.

Phone Authentication and Encryption

Unified CM can be configured to provide multiple levels of security to the phones within a voice system, if those phones support those features. Depending on your security policy, phone placement, and phone support, the security can be configured to fit the needs of your company.

For information on which Cisco Unified IP Phone models support specific security features, refer to the documentation available at

http://www.cisco.com/en/US/products/hw/phones/ps379/tsd_products_support_series_home.html

To enable security on the phones and in the Unified CM cluster, refer to the *Cisco Unified Communications Manager Security Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

Advantages

When the security features are properly configured in Unified CM, all supported phones will have the following capabilities:

- Integrity — Does not allow TFTP file manipulation but does allow Transport Layer Security (TLS) signaling to the phones when enabled.
- Authentication — The image for the phone is authenticated from Unified CM to the phone, and the device (phone) is authenticated to Unified CM. All signaling messages between the phone and Unified CM are verified as being sent from the authorized device.
- Encryption — For supported devices, signaling and media can be encrypted to prevent eavesdropping.
- Secure Real-time Transport Protocol (SRTP) — Is supported to Cisco IOS MGCP gateways and, of course, phone-to-phone. Cisco Unity also supports SRTP for voicemail.

Disadvantages

Unified CM supports authentication, integrity, and encryption for calls between two Cisco Unified IP Phones within a single cluster where no media services are used. However, Unified CM does not provide authentication, integrity, or encryption for all devices or phones. To determine if your device supports these features, refer to the documentation available at

http://www.cisco.com/en/US/products/hw/phones/ps379/tsd_products_support_series_home.html

In addition, auto-registration does not work if you configure the cluster for mixed mode, which is required for device authentication. You cannot implement signaling or media encryption if device authentication does not exist in the cluster – that is, if you do not install and configure the Cisco Certificate Trust List (CTL) client. Application Layer Gateways (ALGs) that allow IP Telephony traffic to traverse firewalls and Network Address Translation (NAT) also do not work with signaling encryption. Not all gateways, phones, or conference are supported with encrypted media.

Access Security

This section covers security features at the Access level that can be used to protect the voice data within a network.

Voice and Video VLANs

Before the phone has its IP address, the phone determines which VLAN it should be in by means of the Cisco Discovery Protocol (CDP) negotiation that takes place between the phone and the switch. This negotiation allows the phone to send packets with 802.1q tags to the switch in a "voice VLAN" so that the voice data and all other data coming from the PC behind the phone are separated from each other at Layer 2. Voice VLANs are not required for the phones to operate, but they provide additional separation from other data on the network.

Cisco Unified Video Advantage is a client application running on a PC, but it is also associated with an IP Phone. The PC will likely reside in the data VLAN while the phone will likely reside in the voice VLAN. To associate with the IP Phone, Cisco Unified Video Advantage uses the Cisco Audio Session Tunnel (CAST) protocol, which operates over TCP/IP. Therefore, Cisco Unified Video Advantage will have to communicate through whatever Layer 3 router is configured to route IP packets between the voice and data VLANs. If there are any access control lists or firewalls configured between those VLANs, they will have to be modified to permit the CAST protocol to operate. Fortunately, CAST uses TCP port 4224 in both directions, making this task easier. Cisco Unified Video Advantage communicates with the IP Phone but not with Unified CM, except when Cisco Unified Video Advantage periodically checks with the TFTP server (which could be co-resident on one or more of the Unified CM servers) to download any software updates. Therefore, you must also permit the TFTP protocol between the data VLAN and the TFTP server(s).

Sony and Tandberg SCCP endpoints do not support Cisco Discovery Protocol (CDP) or 802.1Q VLAN ID tagging. Therefore, in a typical environment, these devices will reside in the data VLAN, unless the port has been configured manually to use the voice VLAN as the native VLAN. The Sony and Tandberg endpoints communicate with the TFTP server to download their configurations, with Unified CM for SCCP signaling, and with other endpoints for RTP audio/video media channels. Therefore, the TFTP protocol must be permitted between the data VLAN and the TFTP server(s), TCP port 2000 must be permitted between the data VLAN and the Unified CM server(s), and UDP ports for RTP media must be permitted between the data and voice VLANs.

H.323 clients, Multipoint Control Units (MCUs), and gateways communicate with Unified CM using the H.323 protocol. Unified CM H.323 trunks (such as H.225 and intercluster trunk variants as well as the RASAggregator trunk type) use a random port range rather than the well-known TCP port 1720. Therefore, you must permit a wide range of TCP ports between these devices and the Unified CM servers. MCUs and gateways are considered infrastructure devices, and they typically reside within the datacenter adjacent to the Unified CM servers. H.323 clients, on the other hand, typically reside in the data VLAN.

Cisco Unified Videoconferencing 3500 Series MCUs configured to run in SCCP mode communicate with the TFTP server(s) to download their configuration, with the Unified CM servers for signaling, and with other endpoints for RTP media traffic. Therefore, TFTP must be permitted between the MCU and the TFTP server(s), TCP port 2000 must be permitted between the MCUs and the Unified CM server(s), and UDP ports for RTP media must be permitted between the MCUs and the voice, data, and gateway VLANs.

Advantages

Voice VLANs can be assigned automatically from the switch to the phone, thus allowing for Layer 2 and Layer 3 separations between voice data and all other data on a network. A voice VLAN also allows for a different IP addressing scheme because the separate VLAN can have a separate IP scope at the Dynamic Host Configuration Protocol (DHCP) server.

Applications use the CDP messaging from the phones to assist in locating phones during an emergency phone call. The location of the phone will be much more difficult to determine if CDP is not enabled on the access port to which that phone is attached.

Disadvantages

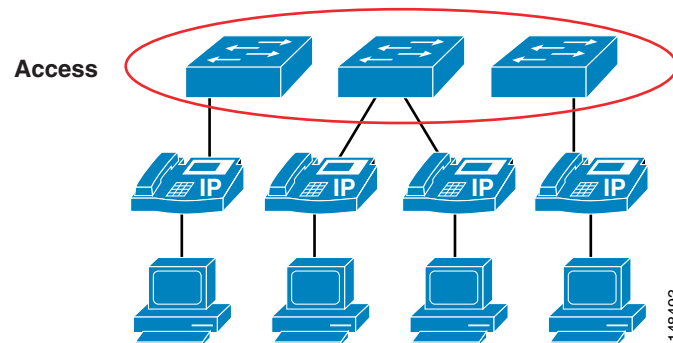
There is a possibility that information could be gathered from the CDP messaging that would normally go to the phone, and that information could be used to discover some of the network. Not all devices that can be used for voice or video with Unified CM are able to use CDP to assist in discovering the voice VLAN.

Switch Port

There are many security features within a Cisco switch infrastructure that can be used to secure a data network. This section describes some of the features that can be used in Cisco Access Switches to protect the IP Telephony data within a network. (See Figure 19-6.) This section does not cover all of the security features available for all of the current Cisco switches, but it does list the most common security feature used across many of the switches that Cisco manufactures. For additional information on the security features available on the particular Cisco gear deployed within your network, refer to the appropriate product documentation available at

<http://www.cisco.com>

Figure 19-6 A Typical Access Layer Design to Which the Phones Attach



Port Security: MAC CAM Flooding

A classic attack on a switched network is a MAC content-addressable memory (CAM) flooding attack. This type of attack floods the switch with so many MAC addresses that the switch does not know which port an end station or device is attached to. When the switch does not know which port a device is attached to, it broadcasts the traffic destined for that device to the entire VLAN. In this way, the attacker is able to see all traffic that is coming to all the users in a VLAN.

To disallow malicious MAC flooding attacks from hacker tools such as *macof*, limit the number of MAC addresses allowed to access individual ports based on the connectivity requirements for those ports. Malicious end-user stations can use *macof* to originate MAC flooding from random-source to random-destination MAC addresses, both directly connected to the switchport or through the IP phone. The *macof* tool is very aggressive and typically can fill a Cisco Catalyst switch content-addressable memory (CAM) table in less than ten seconds. The flooding of subsequent packets that remain unlearned because the CAM table is filled, is as disruptive and insecure as packets on a shared Ethernet hub for the VLAN that is being attacked.

Either port security or dynamic port security can be used to inhibit a MAC flooding attack. A customer with no requirement to use port security as an authorization mechanism would want to use dynamic port security with the number of MAC addresses appropriate to the function attached to a particular port. For example, a port with only a workstation attached to it would want to limit the number of learned MAC addresses to one. A port with a Cisco Unified IP Phone and a workstation behind it would want to set the number of learned MAC addresses to two (one for the IP phone itself and one for the workstation behind the phone) if a workstation is going to plug into the PC port on the phone. This setting in the past has been three MAC addresses, used with the older way of configuring the port in trunk mode. If you use the multi-VLAN access mode of configuration for the phone port, this setting will be two MAC addresses, one for the phone and one for the PC plugged into the phone. If there will be no workstation on the PC port, then the number of MAC addresses on that port should be set to one. These configurations are for a multi-VLAN access port on a switch. The configuration could be different if the port is set to trunk mode (not the recommended deployment of an access port with a phone and PC).

Port Security: Prevent Port Access

Prevent all port access except from those devices designated by their MAC addresses to be on the port. This is a form of device-level security authorization. This requirement is used to authorize access to the network by using the single credential of the device's MAC address. By using port security (in its non-dynamic form), a network administrator would be required to associate MAC addresses statically for every port. However, with dynamic port security, network administrators can merely specify the number of MAC addresses they would like the switch to learn and, assuming the correct devices are the first devices to connect to the port, allow only those devices access to that port for some period of time.

The period of time can be determined by either a fixed timer or an inactivity timer (non-persistent access), or it can be permanently assigned. The feature to permanently assign a MAC address on a Cisco 6000 switch is called *autoconfigure*, and on the Cisco 4500, 2550, 2750, or 2950 switches the feature is called *sticky*. In both cases, the MAC address learned will remain on the port even in the event of a reload or reboot of the switch.

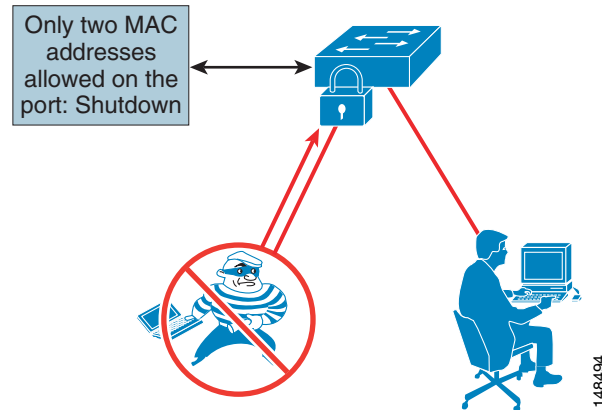
Persistent assignment of MAC addresses via *autoconfigure* or *sticky* can be cleared only by a command. The most common default behavior seen across the Cisco Catalyst switching platforms currently is the non-persistent behavior; the only behavior prior to Cisco CatOS Release 7.6(1) was persistent. No provision is made for device mobility by static port security or persistent dynamic port security. Although it is not the primary requirement, MAC flooding attacks are implicitly prevented by port security configurations that aim to limit access to certain MAC addresses.

From a security perspective, there are better mechanisms for both authenticating and authorizing port access based on userid and/or password credentials rather than using MAC address authorization. MAC addresses alone can easily be spoofed or falsified by most operating systems.

Port Security: Prevent Rogue Network Extensions

Prevent rogue network extensions via hub or wireless access points (APs). Because it limits the number of MAC addresses to a port, port security can also be used as a mechanism to inhibit user extension to the IT-created network. For example, if a user plugs a wireless AP into a user-facing port or data port on a phone with port security defined for a single MAC address, the wireless AP itself would occupy that MAC address and not allow any devices behind it to access the network. (See [Figure 19-7](#).) Generally, a configuration appropriate to stop MAC flooding is also appropriate to inhibit rogue access.

Figure 19-7 Limited Number of MAC Addresses Prevents Rogue Network Extensions



Advantages

Port security prevents an attacker from flooding the CAM table of a switch and from turning any VLAN into a hub that transmits all received traffic to all ports. It also prevents unapproved extensions of the network by adding hubs or switches into the network.

Disadvantages

If the number of MAC addresses is not defined correctly, there is a possibility of denying access to the network or error-disabling the port and removing all devices from the network.

Configuration Example



Note This example configuration is based on a switch running correct code levels to support these features and not running trunk mode to the phone.

The following example illustrates the Cisco IOS commands to configure an access port with dynamic port security, running a phone with a device plugged into the data port on the phone:

```
switchport access vlan 10
switchport mode access
switchport voice vlan 20
switchport port-security
switchport port-security maximum 2
switchport port-security violation restrict
switchport port-security aging time 2
switchport port-security aging type inactivity
```

The commands in the preceding example perform the following functions:

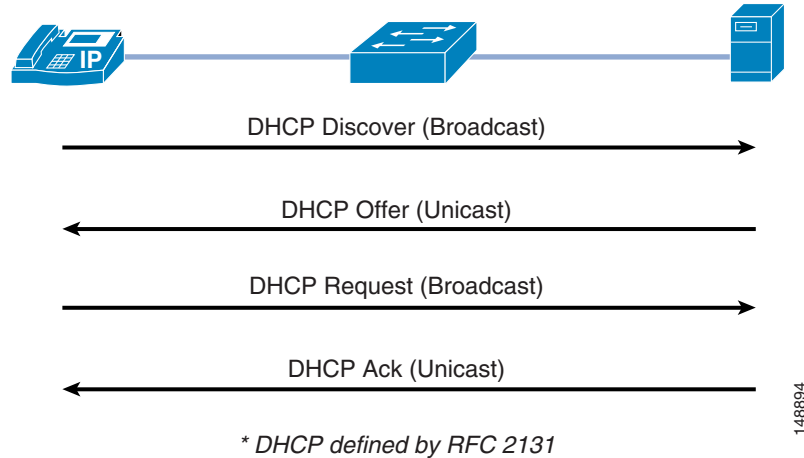
- **switchport port-security *x/x* enable**
This command enables the port security on the specified module/port.
- **switchport port-security violation restrict**
This command is the recommended configuration. The default is to disable the port. If you **restrict** the port, it will learn up to the maximum number of MAC addresses and then stop learning any new MAC addresses. If you have the port on the default setting of **disable** and the maximum number of MAC addresses is exceeded, the port will error-disable and turn off power to the phone. The default timer for the port to re-enable is 5 minutes. Depending on your security policy, it might be preferable to restrict the port and not shut down the phone by disabling the port.
- **switchport port-security aging time 2**
This command sets the amount of time that the MAC address will remain on the port without any traffic from that MAC address. Because of CDP communication between some switches and the phone, the recommended minimum time is 2 minutes.
- **switchport port-security aging type inactivity**
This command defines the type of aging that will be used on the port to time-out the learned MAC address.

DHCP Snooping: Prevent Rogue DHCP Server Attacks

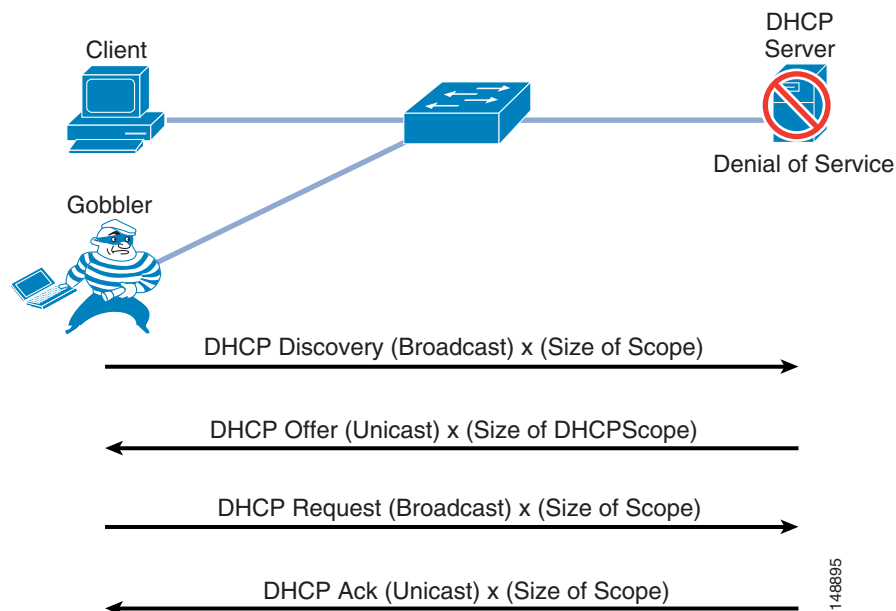
Dynamic Host Configuration Protocol (DHCP) Snooping prevents a non-approved DHCP or rogue DHCP server from handing out IP addresses on a network by blocking all replies to a DHCP request unless that port is allowed to reply. Because most phone deployments use DHCP to provide IP addresses to the phones, you should use the DHCP Snooping feature in the switches to secure DHCP messaging. Rogue DHCP servers can attempt to respond to the broadcast messages from a client to give out incorrect IP addresses, or they can attempt to confuse the client that is requesting an address.

When enabled, DHCP Snooping treats all ports in a VLAN as untrusted by default. An untrusted port is a user-facing port that should never make any reserved DHCP responses. If an untrusted DHCP-snooping port makes a DHCP server response, it will be blocked from responding. Therefore, rogue DHCP servers will be prevented from responding. However, legitimately attached DHCP servers or uplinks to legitimate servers must be trusted.

Figure 19-8 illustrates the normal operation of a network-attached device that requests an IP address from the DHCP server.

Figure 19-8 Normal Operation of a DHCP Request

However, an attacker can request not just a single IP address but all of the IP addresses that are available within a VLAN. (See Figure 19-9.) This means that there would be no addresses for a legitimate device trying to get on the network, and without an IP address the phone cannot connect to Unified CM.

Figure 19-9 An Attacker Can Take All Available IP Addresses on the VLAN**Advantaged**

DHCP Snooping prevents unapproved DHCP servers from being on a network.

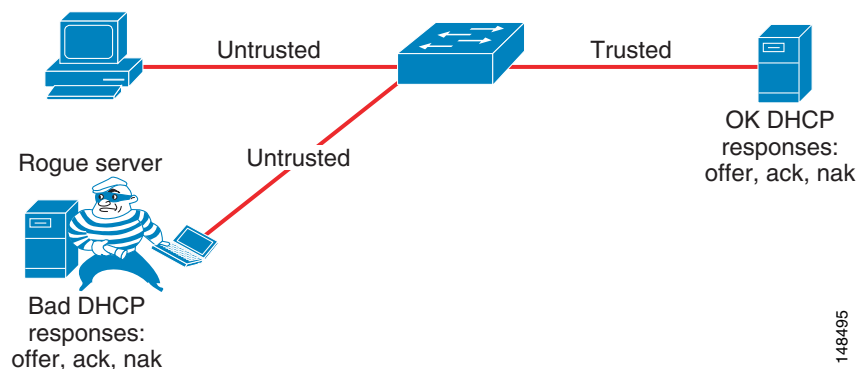
Disadvantages

Incorrect configurations of this feature can deny IP addresses to approved users.

DHCP Snooping: Prevent DHCP Starvation Attacks

DHCP address scope starvation attacks from tools such as Gobbler are used to create a DHCP denial-of-service (DoS) attack. Because the Gobbler tool makes DHCP requests from different random source MAC addresses, you can prevent it from starving a DHCP address space by using port security to limit the number of MAC addresses. (See [Figure 19-10](#).) However, a more sophisticated DHCP starvation tool can make the DHCP requests from a single source MAC address and vary the DHCP payload information. With DHCP Snooping enabled, untrusted ports will make a comparison of the source MAC address to the DHCP payload information and fail the request if they do not match.

Figure 19-10 Using DHCP Snooping to Prevent DHCP Starvation Attacks



Advantages

DHCP Snooping prevents any single device from capturing all the IP addresses in any given scope.

Disadvantages

Incorrect configurations of this feature can deny IP addresses to approved users.

Configuration Example

The following example illustrates the Cisco IOS commands to configure an access port with DHCP Snooping, running a phone with a device plugged into the data port on the phone:

- Global commands


```
ip dhcp snooping vlan 10, 20
no ip dhcp snooping information option
ip dhcp snooping
```
- Interface commands


```
no ip dhcp snooping trust (Default)
ip dhcp snooping limit rate 10 (pps)
ip dhcp snooping trust
```

The global commands in the preceding example perform the following functions:

- **ip dhcp snooping vlan 10, 20**
This command specifies which VLANs have DHCP Snooping enabled.

- **No ip dhcp snooping information option**

This command should be used so that Option 82 information is not required to lease a DHCP address. The Option 82 information must be supported by the DHCP server, but most enterprise servers do not support this feature. Option 82 is supported in Cisco IOS DHCP servers.

- **ip dhcp snooping**

This command enables DHCP Snooping at the global level on the switches.

The interface commands in the preceding example perform the following functions:

- **no ip dhcp snooping trust**

This command sets the interface to not trust any information coming into the port from a DHCP server.

- **ip dhcp snooping limit rate 10**

This command sets the default rate limit that is configured on the interface when DHCP Snooping is initially configured. This value can be changed depending on your security policy.

- **ip dhcp snooping trust**

You must use this command on the port through which DHCP information will be sent from a DHCP server. If you do not trust the port from which the DHCP information is coming, then none of the devices will ever receive a DHCP address. At least one port (access port or trunk port) with the DHCP server on it must be configured to allow this information to get to the clients. This command can also be used to trust any device connected to a port that has a static IP address and that will not use DHCP to get an IP address. Note that the uplink port to the DHCP server, or the trunk port to the DHCP server, will also have to be trusted.

DHCP Snooping: Binding Information

Another function of DHCP Snooping is to record the DHCP binding information for untrusted ports that successfully get IP addresses from the DHCP servers. The binding information is recorded in a table on the Cisco Catalyst switch. The DHCP binding table contains the IP address, MAC address, lease length, port, and VLAN information for each binding entry. The binding information from DHCP Snooping remains in effect for the length of the DHCP binding period set by the DHCP server (that is, the DHCP lease time). The DHCP binding information is used to create dynamic entries for Dynamic ARP Inspection (DAI) to limit ARP responses for only those addresses that are DHCP-bound. The DHCP binding information is also used by the IP source guard to limit sourcing of IP packets to only those addresses that are DHCP-bound.

The following examples show binding information from DHCP Snooping.

- Displaying binding information for Cisco IOS:

```
show ip dhcp snooping binding
-----
MacAddress      IpAddress      Lease(sec)      Type            VLAN  Interface
-----
00:03:47:B5:9F:AD  10.120.4.10    193185          dhcp-snooping   10    FastEthernet3/18
```

- Displaying binding information for Cisco CatOS:

```
ngcs-6500-1> (enable) show dhcp-snooping bindings
MacAddress      IpAddress      Lease(sec)      VLAN      Port
-----
00-10-a4-92-bf-dd  10.10.10.21    41303           10        2/5
```

There is a maximum limit to the number of binding table entries that each type of switch can store for DHCP Snooping. (Refer to the product documentation for your switch to determine this limit.) If you are concerned about the number of entries in your switch's binding table, you can reduce the lease time on the DHCP scope so that the entries in the binding table time-out sooner. The entries remain in the DHCP binding table until the lease runs out. In other words, the entries remain in the DHCP Snooping binding table as long as the DHCP server thinks the end station has that address. They are not removed from the port when the workstation or phone is unplugged.

If you have a Cisco Unified IP Phone plugged into a port and then move it to a different port, you might have two entries in the DHCP binding table with the same MAC and IP address on different ports. This behavior is considered normal operation.

Requirement for Dynamic ARP Inspection

Dynamic Address Resolution Protocol (ARP) Inspection (DAI) is a feature used on the switch to prevent Gratuitous ARP attacks on the devices plugged into the switch and on the router. Although it is similar to the Gratuitous ARP feature mentioned previously for the phones, Dynamic ARP protects all the devices on the LAN, and it is not just a phone feature.

In its most basic function, Address Resolution Protocol (ARP) enables a station to bind a MAC address to an IP address in an ARP cache, so that the two stations can communicate on a LAN segment. A station sends out an ARP request as a MAC broadcast. The station that owns the IP address in that request will give an ARP response (with its IP and MAC address) to the requesting station. The requesting station will cache the response in its ARP cache, which has a limited lifetime. The default ARP cache lifetime for Microsoft Windows is 2 minutes; for Linux, the default lifetime is 30 seconds; and for Cisco IP phones, the default lifetime is 40 minutes.

ARP also makes the provision for a function called Gratuitous ARP. Gratuitous ARP (GARP) is an unsolicited ARP reply. In its normal usage, it is sent as a MAC broadcast. All stations on a LAN segment that receive a GARP message will cache this unsolicited ARP reply, which acknowledges the sender as the owner of the IP address contained in the GARP message. Gratuitous ARP has a legitimate use for a station that needs to take over an address for another station on failure.

However, Gratuitous ARP can also be exploited by malicious programs that want to illegitimately take on the identity of another station. When a malicious station redirects traffic to itself from two other stations that were talking to each other, the hacker who sent the GARP messages becomes the man-in-the-middle. Hacker programs such as ettercap do this with precision by issuing "private" GARP messages to specific MAC addresses rather than broadcasting them. In this way, the victim of the attack does not see the GARP packet for its own address. Ettercap also keeps its ARP poisoning in effect by repeatedly sending the private GARP messages every 30 seconds.

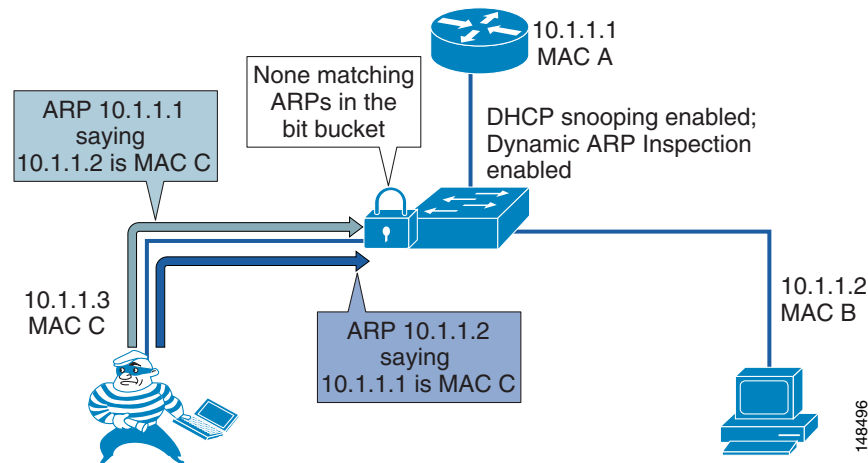
Dynamic ARP Inspection (DAI) is used to inspect all ARP requests and replies (gratuitous or non-gratuitous) coming from untrusted (or user-facing) ports to ensure that they belong to the ARP owner. The ARP owner is the port that has a DHCP binding which matches the IP address contained in the ARP reply. ARP packets from a DAI trusted port are not inspected and are bridged to their respective VLANs.

Using DAI

Dynamic ARP Inspection (DAI) requires that a DHCP binding be present to legitimize ARP responses or Gratuitous ARP messages. If a host does not use DHCP to obtain its address, it must either be trusted or an ARP inspection access control list (ACL) must be created to map the host's IP and MAC address. (See [Figure 19-11](#).) Like DHCP Snooping, DAI is enabled per VLAN, with all ports defined as untrusted by default. To leverage the binding information from DHCP Snooping, DAI requires that DHCP Snooping be enabled on the VLAN prior to enabling DAI. If DHCP Snooping is not enabled before you

enable DAI, none of the devices in that VLAN will be able to use ARP to connect to any other device in their VLAN, including the default gateway. The result will be a self-imposed denial of service to any device in that VLAN.

Figure 19-11 Using DHCP Snooping and DAI to Block ARP Attacks



Because of the importance of the DHCP Snooping binding table to the use of DAI, it is important to back up the binding table. The DHCP Snooping binding table can be backed up to bootflash, File Transfer Protocol (FTP), Remote Copy Protocol (RCP), slot0, and Trivial File Transfer Protocol (TFTP). If the DHCP Snooping binding table is not backed up, the Cisco Unified IP Phones could lose contact with the default gateway during a switch reboot. For example, assume that the DHCP Snooping binding table is not backed up and that you are using Cisco Unified IP Phones with a power adapter instead of line power. When the switch comes back up after a reboot, there will be no DHCP Snooping binding table entry for the phone, and the phone will not be able to communicate with the default gateway unless the DHCP Snooping binding table is backed up and loads the old information before traffic starts to flow from the phone.

Advantages

DAI prevents an attacker from running ARP-based attacks in a network to disrupt or sniff the traffic between people who are adjacent to the attacker at Layer 2.

Disadvantages

Incorrect configurations of this feature can deny network access to approved users. If a device has no entry in the DHCP Snooping binding table, then that device will not be able to use ARP to connect to the default gateway and therefore will not be able to send traffic. If you use static IP addresses, those addresses will have to be entered manually into the DHCP Snooping binding table. If you have devices that do not use DHCP again to obtain their IP addresses when a link goes down (some UNIX or Linux machines behave this way), then you must back up the DHCP Snooping binding table.

Configuration Example

The following example illustrates the Cisco IOS commands to configure access ports with DHCP Snooping and Dynamic ARP Inspection:

- Global commands

```
ip dhcp snooping vlan 10,20 (required)
no ip dhcp snooping information option (required without option 82 dhcp server)
ip dhcp snooping (required)
```

```
ip arp inspection vlan 10,20
ip dhcp snooping database tftp://172.26.168.10/tftpboot/cisco/ngcs-dhcpdb
```

- **Interface commands**

```
ip dhcp snooping trust
ip arp inspection trust
no ip arp inspection trust (default)
ip arp inspection limit rate 15 (pps)
```

The global commands in the preceding example perform the following functions:

- **ip arp inspection vlan 10,20**

This command specifies which VLANs have Dynamic ARP Inspection (DAI) enabled.

- **ip arp inspection trust**

As with **ip dhcp snooping trust**, this command allows a trusted device such as a router to reply to ARP messages. This command must be configured on the port for your router, otherwise the router will not be able to respond to any ARP requests because the router will not be in the DHCP Snooping binding table.

- **no ip arp inspection trust**

This is the default setting for every port in the VLAN. Trust must be enabled.

- **ip arp inspection limit rate 15 (pps)**

This command sets the global default value for the maximum number of packets per second (pps) that are allow for ARP messages on the interfaces. If this value is exceeded, then the interface will be disabled. If this behavior is a concern, you can increase or decrease the limit or set it to **none**.

- **ip dhcp snooping database tftp://172.26.168.10/tftpboot/cisco/ngcs-dhcpdb**

This command backs up the DHCP Snooping binding table to the TFTP server. The DHCP Snooping binding table can be backed up to bootflash, FTP, RCP, slot0, and TFTP.

The interface commands in the preceding example perform the following functions:

- **no ip arp inspection trust**

This command enables DAI on the port and checks all ARPs against the DHCP Snooping binding table.

- **ip arp inspection limit rate 15 (pps)**

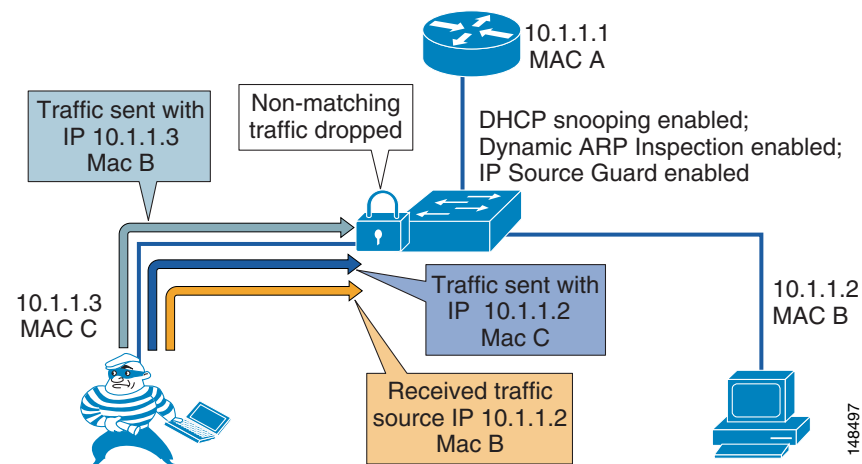
This command specifies the maximum number of packets per second (pps) that are allow for ARP messages on the interface. If the interface sees more than the specified number of ARP messages in a second, it will disable the port. Depending on your security policy, the default value (15 pps) might be the preferred setting. If you do not want to disable the phone when the port receives more than 15 ARP messages in a second, you can set the rate limit to **none**, which will allow the phone to stay up.

IP Source Guard

Beyond ARP spoofing, an attacker can also do IP address spoofing. This method is commonly use to perform DoS attacks on a second party by sending packets through a third party, thus masking the identity of the attacking system. A simple example of this occurs when an attacker pings a third-party system while sourcing the IP address of the second party that is being attacked. The ping response will be directed to the second party from the third-party system. Aggressive SYN-flooding originating from spoofed IP addresses is another common type of attack used to overwhelm a server with TCP half-sessions.

The IP Source Guard (IPSG) feature, when invoked, dynamically creates an ACL based on the contents of the DHCP Snooping binding table. This ACL ensures that traffic is sourced from the IP address issued at DHCP binding time and prevents any traffic from being forwarded by other spoofed addresses. While DHCP Snooping is a prerequisite for IP Source Guard, DAI is not. However, Cisco recommend that you enable DAI in addition to IP Source Guard to prevent ARP-poisoning man-in-the-middle attacks in addition to IP address spoofing. (See [Figure 19-12](#).)

Figure 19-12 Using IP Source Guard to Prevent Address Spoofing



With IP address spoofing, an attacker can impersonate a valid address either by manually changing an address or by running a program design to do address spoofing, such as hping2. Internet worms can use spoofing techniques to disguise their origins.

Configuration Example

The following example illustrates the Cisco IOS commands to configure access ports with IP Source Guard:

- Commands that must be enabled before you enable IP Source Guard


```
ip dhcp snooping vlan 4,104
no ip dhcp snooping information option
ip dhcp snooping
```
- Interface command — This command enables the IP Source Guard without DHCP Option 82.


```
ip verify source vlan dhcp-snooping
```

Additional Information

For additional information on network security, refer to the Cisco documentation at the following locations:

- http://www.cisco.com/en/US/products/hw/switches/ps708/products_white_paper0900aecd8015f0ae.shtml
- http://www.cisco.com/en/US/prod/collateral/wireless/wirelssw/ps1953/product_implementation_design_guide09186a00800a3016.pdf

Quality of Service

Quality of Service (QoS) is a vital part of any security policy for an enterprise network. Even though most people think of QoS as setting the priority of traffic in a network, it also controls the amount of data that is allowed into the network. In the case of Cisco switches, that control point is at the port level when the data comes from the phone to the Ethernet switch. The more control applied at the edge of the network at the access port, the fewer problems will be encountered as the data aggregates in the network.

As mentioned previously in the lobby phone example, you can provide enough flow control of the traffic at the access port level to prevent any attacker from launching a denial-of-service (DoS) attack from that port in the lobby. The configuration for that example was not as aggressive as it could be because the QoS configuration allowed traffic sent to the port to exceed the maximum rate, but the traffic was remarked to the level of scavenger class. Given a more aggressive QoS policy, any amount of traffic that exceeded that maximum limit of the policy could just be dropped at the port, and that "unknown" traffic would never make it into the network. QoS should be enabled across the entire network to give the IP Telephony data high priority from end to end.

For more information on QoS, refer to the chapter on [Network Infrastructure, page 3-1](#), and the *Enterprise QoS Solution Reference Network Design (SRND) Guide* available at

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

Advantages

QoS can be used to control not only the priority of the traffic in the network but also the amount of traffic that can travel through any specific interface. Cisco Smartports templates have been created to assist in deploying voice QoS in a network at the access port level.

Disadvantages

If QoS settings are outside the standard Cisco Smartports template, the configuration can be complex and difficult to manage in large IP Telephony deployments.

Access Control Lists

This section covers access control lists (ACLs) and their uses in protecting voice data.

VLAN Access Control Lists

You can use VLAN access control lists (ACLs) to control data that flows on a network. Cisco switches have the capability of controlling Layers 2 to 4 within a VLAN ACL. Depending on the types of switches in a network, VLAN ACLs can be used to block traffic into and out of a particular VLAN. They can also be used to block intra-VLAN traffic to control what happens inside the VLAN between devices.

If you plan to deploy a VLAN ACL, you should verify which ports are needed to allow the phones to function with each application used in your IP Telephony network. Normally any VLAN ACL would be applied to the VLAN that the phones use. This would allow control at the access port, as close as possible to the devices that are plugged into that access port.

Refer to the following product documentation for information on configuring VLAN ACLs:

- Cisco Catalyst 3750 Switches

http://www.cisco.com/en/US/products/hw/switches/ps5023/products_installation_and_configuration_guides_list.html

- Cisco Catalyst 4500 Series Switches
http://www.cisco.com/en/US/products/hw/switches/ps4324/tsd_products_support_configure.html
- Cisco Catalyst 6500 Series Switches running Cisco IOS
http://www.cisco.com/en/US/products/hw/switches/ps708/products_installation_and_configuration_guides_list.html
- Cisco 6500 Series Switches running Cisco CatOS
http://www.cisco.com/en/US/products/hw/switches/ps708/tsd_products_support_configure.html

The following example represents a VLAN ACL that allows only the traffic for a Cisco 7960 IP Phone to boot and function in a VLAN. (Inline comments indicate the purpose of each line of the ACL.) This example VLAN ACL is for the ports used with Cisco Unified CM Release 4.1. The example uses the following IP address ranges:

- Phones are in the range 10.0.20.*
- Servers are in the range 10.0.10.*
- Gateways are in the range 10.0.30.*
- Default gateways are 10.0.10.2 and 10.0.10.3
- DNS server IP address is 10.0.40.3

**Note**

The ports do change when either the application is updated or the OS is updated, or both. This note applies to all the IP Telephony devices in the network, including phones. To obtain the latest list of ports used by a product, refer to the appropriate documentation for the version of the product that is running on your network. Port usage documentation is available at http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

```
20 permit udp host 10.0.10.2 eq 1985 any
30 permit udp host 10.0.10.3 eq 1985 any
!permit HSRP from the routers
40 permit udp any any eq bootpc
50 permit udp any any eq bootps
!permit DHCP activity
60 permit udp 10.0.10.0 0.0.0.255 range 49152 65535 10.0.20.0 0.0.0.255 eq tftp
70 permit udp 10.0.20.0 0.0.0.255 range 1024 5000 10.0.10.0 0.0.0.255 range 49152 65535
80 permit udp 10.0.10.0 0.0.0.255 range 49152 65535 10.0.20.0 0.0.0.255 range 1024 5000
!permit the tftp traffic from the tftp server and phone
90 permit udp 10.0.10.0 0.0.0.255 range 49152 65535 host 10.0.40.3 eq domain
100 permit udp host 172.19.244.2 eq domain 10.0.10.0 0.0.0.255 range 49152 65535
!permit DNS to and from the phone
110 permit tcp 10.0.10.0 0.0.0.255 range 49152 65535 10.0.20.0 0.0.0.255 eq 2000
120 permit tcp 10.0.20.0 0.0.0.255 eq 2000 10.0.10.0 0.0.0.255 range 49152 65535
!permit signaling to and from the phone.
130 permit udp 10.0.10.0 0.0.0.255 range 16384 32767 10.0.10.0 0.0.0.255 range 16384 32767
140 permit udp 10.0.0.0 0.0.0.255 range 16384 32767 10.0.10.0 0.0.0.255 range 16384
32767
150 permit udp 10.0.10.0 0.0.0.255 range 16384 32767 10.0.0.0 0.0.0.255 range 16384
43767
!permit all phones to send udp to each other
160 permit tcp 10.0.10.0 0.0.0.255 range 49152 65535 10.0.20.0 0.0.0.255 eq www
170 permit tcp 10.0.20.0 0.0.0.255 eq www 10.0.10.0 0.0.0.255 range 49152 65535
180 permit tcp 10.0.20.0 0.0.0.255 range 49152 65535 10.0.10.0 0.0.0.255 eq www
190 permit tcp 10.0.10.0 0.0.0.255 eq www 10.0.20.0 0.0.0.255 range 49152 65535
!permit web access to and from the phone
200 permit Intelligent Contact ManagementP any any
```

```
!allow all icmp - phone to phone, gateway to phone, and NMS to phone
220 permit udp 10.0.30.0 0.0.0.255 rang 16384 32767 10.0.10.0 0.0.0.255 rang 16384 32767
!permit udp to the gateways in the network for pstn access
```

As this example ACL illustrates, the more well-defined the IP addresses are in a network, the easier it is to write and deploy an ACL.

For more details on how to apply VLAN ACLs, refer to the following documentation:

- Cisco Catalyst 3750 Switch
http://www.cisco.com/en/US/products/hw/switches/ps5023/products_installation_and_configuration_guides_list.html
- Cisco Catalyst 4500 Series Switches
http://www.cisco.com/en/US/products/hw/switches/ps4324/tsd_products_support_configure.html
- Cisco Catalyst 6500 Series Switches with Cisco CatOS
http://www.cisco.com/en/US/products/hw/switches/ps708/products_installation_and_configuration_guides_list.html
- Cisco Catalyst 6500 Series Switches with Cisco IOS
http://www.cisco.com/en/US/products/hw/switches/ps708/tsd_products_support_configure.html

Advantages

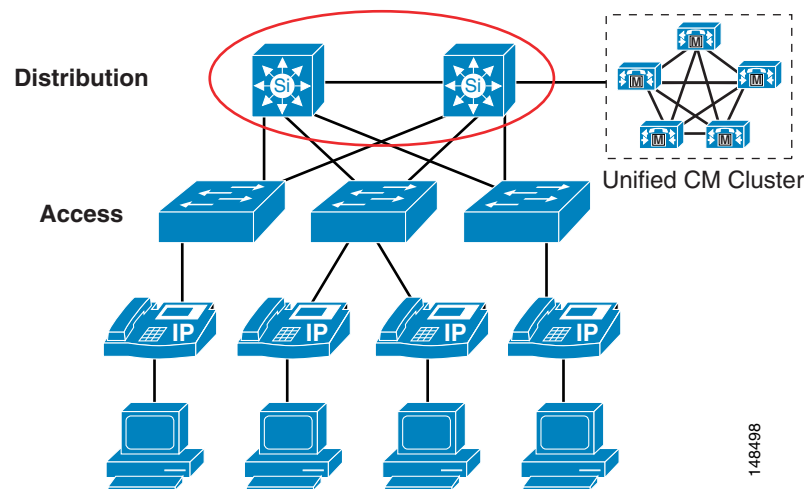
ACLs provide the ability to control the network traffic in and out of a VLAN as well as the ability to control the traffic within the VLAN.

Disadvantages

VLAN ACLs are very difficult to deploy and manage at an access-port level that is highly mobile. Because of these management issues, care should be taken when deploying VLAN ACLs at the access port in the network.

Router Access Control Lists

As with VLAN ACLs, routers have the ability to process both inbound and outbound ACLs by port. The first Layer 3 device is the demarcation point between voice data and other types of data when using voice and data VLANs, where the two types of data are allowed to send traffic to each other. Unlike the VLAN ACLs, router ACLs are not deployed in every access device in your network. Rather, they are applied at the edge router, where all data is prepared for routing across the network. This is the perfect location to apply a Layer 3 ACL to control which areas the devices in each of the VLANs have the ability to access within a network. Layer 3 ACLs can be deployed across your entire network to protect devices from each other at points where the traffic converges. (See [Figure 19-13](#).)

Figure 19-13 Router ACLs at Layer 3

There are many types of ACLs that can be deployed at Layer 3. For descriptions and examples of the most common types, refer to *Configuring Commonly Used IP ACLs*, available (with Cisco partner login required) at

http://cisco.com/en/US/partner/tech/tk648/tk361/technologies_configuration_example09186a0080100548.shtml

Depending on your security policy, the Layer 3 ACLs can be as simple as not allowing IP traffic from the non-voice VLANs to access the voice gateway in the network, or the ACLs can be detailed enough to control the individual ports and the time of the day that are used by other devices to communicate to IP Telephony devices. Assuming that there are no software phones, ACLs can be written to block all traffic (by IP address or IP range) to Unified CMs, voice gateways, phones, and any other voice application that is being used for voice-only services. This method simplifies the ACLs at Layer 3 compared to the ACLs at Layer 2 or VLAN ACLs.

This example uses the following IP address ranges:

- Phones are in the range 10.0.20.*
- IP Telephony servers are in the range 10.0.10.*
- Gateways are in the range 10.0.30.*
- All other devices in the network are in the range 192.168.*.*

```
10 deny ip 192.168.0.0 0.0.255.255 10.0.10.0 0.0.0.255
!deny all non voice devices to the voip servers
20 deny 192.168.0.0 0.0.255.255 10.0.30.0 0.0.0.255
!deny all non voice devices to the voip gateways
30 deny 192.168.0.0 0.0.255.255 10.0.20.0 0.0.0.255
!deny all non voice devices to communicate with the phones ip addresses
```

Advantages

ACLs are much easier to manage and deploy at Layer 3. Layer 3 is one of the first opportunities to apply control to voice data and other non-voice data in the network.

Disadvantages

As the ACLs become more granular and detailed, any changes in port usage in a network could break not only voice but also other applications in the network.

If there are software phones in the network, if web access to the phone is allowed, or if you use the Attendant Console or other applications that need access to the voice VLAN subnets, the ACLs are much more difficult to deploy and control.

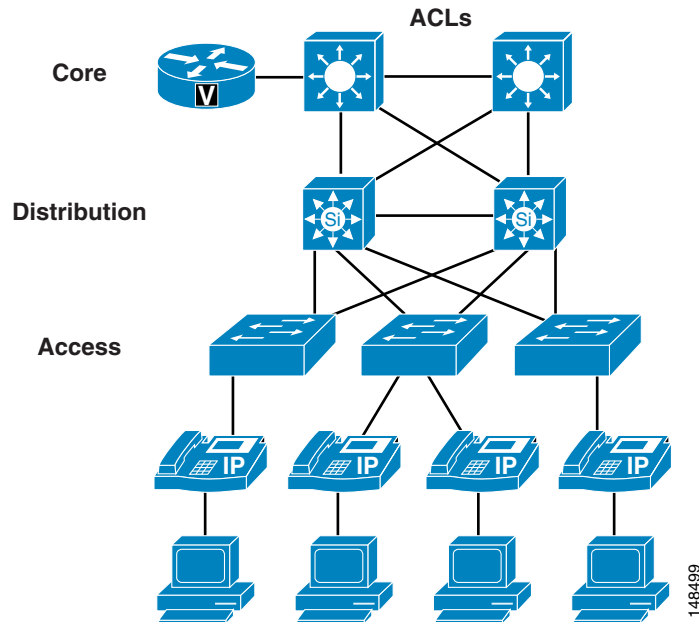
Gateways and Media Resources

Gateways and media resources are devices that convert an IP Telephony call into a PSTN call. When an outside call is placed, the gateway or media resource is one of the few places within an IP Telephony network to which all the voice RTP streams flow.

Because IP Telephony gateways and media resources can be placed almost anywhere in a network, securing an IP Telephony gateway or media resource might be considered more difficult than securing other devices, depending on your security policy. However, depending on which point trust is established in the network, the gateways and media resources can be quite easy to secure. Because of the way the gateways and media resources are controlled by Unified CM, if the path that the signaling takes to the gateway or media resource is in what is considered a secure section of the network, a simple ACL can be used to control signaling to and from the gateway or media resource. If the network is not considered secure between the gateways (or media resources) and where the Unified CMs are located (such as when a gateway is located at a remote branch), the infrastructure can be used to build IPSec tunnels to the gateways and media resources to protect the signaling. In most networks there would most likely be a combination of the two approaches (ACL and IPSec) to secure those devices.

For H.323 videoconferencing devices, an ACL can be written to block port 1720 for H.225 trunks from any H.323 client in the network. This method would block users from initiating an H.225 session with each other directly. Cisco devices might use different ports for H.225, so refer to the product documentation for your equipment to see which port is used. If possible, change the port to 1720 so that only one ACL is needed to control signaling.

Because we use QoS at the edge of the network, if an attacker can get into the voice VLAN and determine where the gateways and media resources are, QoS at the port would limit how much data the attacker would be able to send to the gateway or media resource. (See [Figure 19-14](#).)

Figure 19-14 Securing Gateways and Media Resources with IPSec, ACLs, and QoS

Some gateways and media resources support Secure RTP (SRTP) to the gateways and media resources from the phones, if the phone is enabled for SRTP. To determine if a gateway or media resource supports SRTP, refer to the appropriate product documentation at

<http://www.cisco.com>

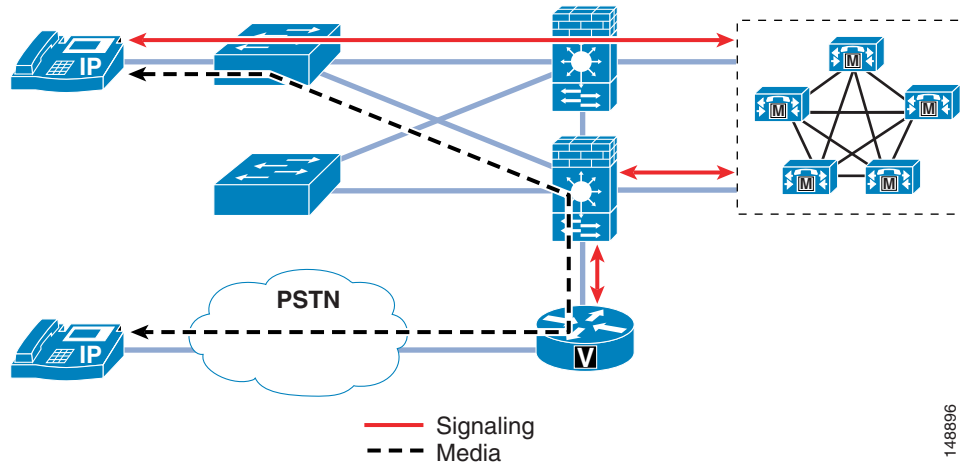
For more information on IPSec tunnels, refer to the *Site-to-Site IPSec VPN Solution Reference Network Design (SRND)*, available at

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

Putting Firewalls Around Gateways

Some very interesting issues arise from placing firewalls between a phone making a call and the gateway to the PSTN network. Stateful firewalls look into the signaling messages between Unified CM, the gateway, and the phone, and they open up a pinhole for the RTP streams to allow the call to take place. To do the same thing with a normal ACL, the entire port ranges used by the RTP streams would have to be open to the gateway.

There are two ways to deploy gateways within a network: behind a firewall and in front of a firewall. If you place the gateway behind a firewall, all the media from the phones that are using that gateway have to flow through the firewall, and additional CPU resources are required to run those streams through the firewall. In turn, the firewall adds control of those streams and protects the gateway from denial-of-service attacks. (See Figure 19-15.)

Figure 19-15 Gateway Placed Behind a Firewall

The second way to deploy the gateway is outside the firewall. Because the only type of data that is ever sent to the gateway from the phones is RTP streams, the access switch's QoS features control the amount of RTP traffic that can be sent to that gateway. The only thing that Unified CM sends to the gateway is the signaling to set up the call. If the gateway is put in an area of the network that is trusted, the only communication that has to be allowed between Unified CM and the gateway is that signaling. (See [Figure 19-15](#).) This method of deployment decreases the load on the firewall because the RTP streams are not going through the firewall.

Advantages

Unlike an ACL, most firewall configurations will open only the RTP stream port that Unified CM has told the phone and the gateway to use between those two devices as long as the signaling goes through the firewall. The firewall also has additional features for DoS attacks and Cisco Intrusion Detection System (IDS) signatures to look at interesting traffic and determine if any attackers are doing something they should not be doing.

Disadvantages

As stated in the section on [Firewalls](#), [page 19-31](#), when a firewall is looking at all the signaling and RTP streams from phones to a gateway, capacity could be an issue. Also, if data other than voice data is running through the firewall, CPU usage must be monitored to make sure that the firewall does not affect the calls that are running through the firewall.

In active or standby mode, the lowest setting for the failover timer is 3 seconds for the Cisco Adaptive Security Appliance (ASA) and Cisco Private Internet Exchange (PIX), and the minimum timer setting for the Cisco Firewall Services Module (FWSM) is also 3 seconds. The firewalls will fail-over quickly once the standby unit has determined that it needs to take over. The state of the data running through the primary firewall will be passed to the failover unit if stateful failover is configured, so everything that was taking place before the failover will be maintained. But there is a possibility that, in the case of a complete failure of the primary unit or connectivity to that unit, at least 3 seconds of traffic would not pass to the gateway for the ASA or PIX, or 3 seconds of traffic for the FWSM. In other words, if there is some type of failure that forces the firewalls to fail-over, there would be at least a 3-second stoppage of RTP streams.

Firewalls and H.323

H.323 utilizes H.245 for setting up the media streams between endpoints, and for the duration of that call the H.245 session remains active between Unified CM and the H.323 gateway. Subsequent changes to the call flow are done via H.245.

By default, a Cisco firewall tracks the H.245 session and the associated RTP streams of calls, and it will time-out the H.245 session if no RTP traffic crosses the firewall for longer than 5 minutes. For topologies where at least one H.323 gateway and the other endpoints are all on one side of the firewall, the firewall will not see the RTP traffic. After 5 minutes, the H.245 session will be blocked by the firewall, which stops control of that stream but does not affect the stream itself. For example, no supplementary services will be available.

To make this topology function, you may change the default behavior of the firewall by using the command **timeout h323** [*hr:min:sec*] to set the value high enough so that the maximum anticipated call duration is specified. For example, you might want to increase the duration for a mobile agent for whom a permanent call is set up first thing in the morning and exists for the agent's entire shift. In that case, the duration would be set to a value longer than the agent's shift.

Advantages

The advantage of the configuration change from default is that it allows H.323 to not lose any call functionality when all endpoints are on the same side of the firewall.

Disadvantages

The timeout feature increases protection of the call agents that are behind the firewall, but increasing the timeout reduces that value of the feature.

Firewalls

Firewalls can be used in conjunction with ACLs to protect the voice servers and the voice gateways from devices that are not allowed to communicate with IP Telephony devices. Because of the dynamic nature of the ports used by IP Telephony, having a firewall does help to control opening up a large range of ports needed for IP Telephony communications. Given the complexities that firewalls introduce into a network design, you must take care in placing and configuring the firewalls and the devices around the firewalls to allow the traffic that is considered correct to pass while blocking the traffic that needs to be blocked.

IP Telephony networks have unique data flows. The phones use a client/server model for signaling for call setup, and Unified CM controls the phones through that signaling. The data flows for the IP Telephony RTP streams are more like a peer-to-peer network, and the phones or gateways talk directly to each other via the RTP streams. If the signaling flows do not go through the firewall so that the firewall can inspect the signaling traffic, the RTP streams could be blocked because the firewall will not know which ports need to be opened to allow the RTP streams for a conversation.

A firewall placed in a correctly designed network can force all the data through that device, so capacities and performance need to be taken into account. Performance includes the amount of latency, which can be increased by a firewall if the firewall is under high load or even under attack. The general rule in an IP Telephony deployment is to keep the CPU usage of an FWSM, ASA, or PIX to less than 60% for normal usage. If the CPU runs over 60%, it increases the chance of impacting IP phones, call setup, and registration. If the CPU usage stays at a sustained level above 60%, the registered IP phones will be affected, quality of calls in progress will degrade, and call setup for new calls will suffer. In the worst case, if the sustained CPU usage stays above 60%, phones will start to unregister. When this happens, they will attempt to re-register with Unified CM, thus increasing the load on the firewalls even more. If

this were to happen, the effect would be a rolling blackout of phones unregistering and attempting to re-register with Unified CM. Until the CPU usage of the firewall decreases to under 60% sustained load, this rolling blackout would continue and most (if not all) of the phones would be affected. If you are currently using a Cisco firewall in your network, you should monitor the CPU usage carefully when adding IP Telephony traffic to your network so that you do not adversely affect that traffic.

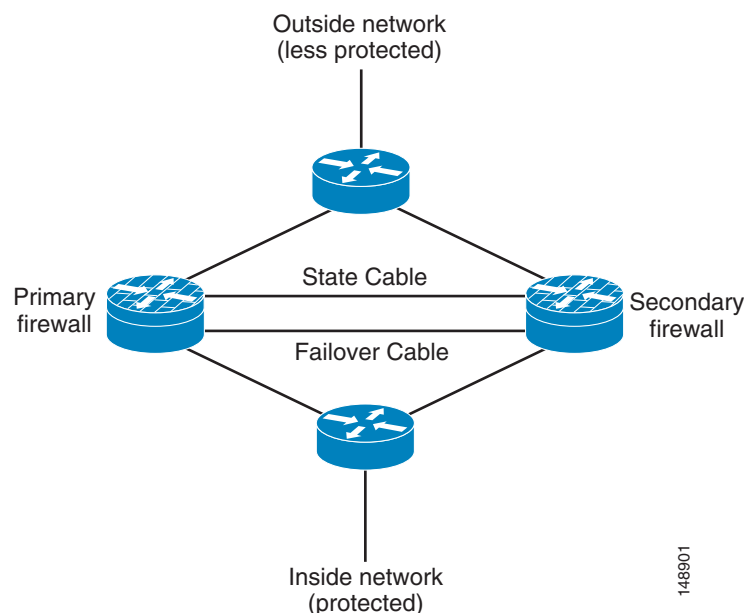
There are many ways to deploy firewalls. This section concentrates on the ASA, PIX, and FWSM in the active/standby mode in both routed and transparent scenarios. Each of the configurations in this section is in single-context mode within the voice sections of the firewall configurations.

All of the Cisco firewalls can run in either multiple-context or single-context mode. In single-context mode, the firewall is a single firewall that controls all traffic flowing through it. In multiple-context mode, the firewalls can be turned into many virtual firewalls. Each of these contexts or virtual firewalls have their own configurations and can be controlled by different groups or administrators. Each time a new context is added to a firewall, it will increase the load and memory requirements on that firewall. When you deploy a new context, make sure that the CPU requirements are met so that voice RTP streams are not adversely affected.

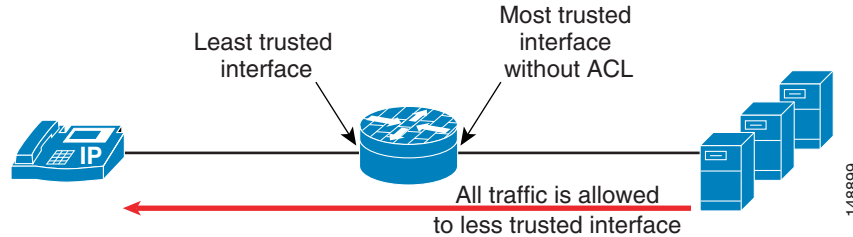
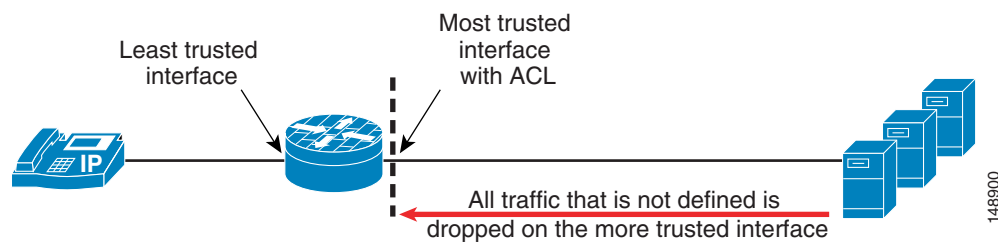
Functionality Differences of the ASA or PIX and the FWSM

Figure 19-16 shows a logical representation of redundant firewalls in a network. Placement is the same for both routed configurations and transparent configurations.

Figure 19-16 Redundant Routed or Transparent Firewalls



Both the Cisco Adaptive Security Appliance (ASA) and the Cisco Private Internet Exchange (PIX) operate in a different manner than the Cisco Firewall Services Modules (FWSM). Within an ASA or PIX, as long as there are no ACLs on a more trusted interface, all traffic from that interface is trusted and allowed out to a less-trusted interface. (See Figure 19-17.) For example, all traffic from the inside or data center interface of an ASA will be allowed to go out to the outside interface of the ASA. Once any ACL is applied to the more trusted interface on a ASA/PIX, all other traffic is denied (DENY) and the firewall will then function very much like the FWSM. (See Figure 19-18.)

Figure 19-17 *Functionality of a Cisco ASA or PIX***Figure 19-18** *Functionality of a Cisco FWSM***Overall Advantages of Firewalls**

A firewall provides a security control point in the network for applications that run over the network. A firewall also provides dynamic opening of ports for IP Telephony conversations if that traffic is running through the firewall.

Using its Application Layer Gateway (ALG) capability, the firewall can inspect the traffic that runs through it to determine if that traffic is really the type of traffic that the firewall is expecting. For example, does the HTTP traffic really look like HTTP traffic, or is it an attack? If it is an attack, then drop that packet and do not allow that packet to get to the HTTP server behind the firewall.

Overall Disadvantages of Firewalls

Not all IP Telephony application servers or applications are supported through a firewall. Some applications that are not supported with firewalls or with an ALG in the firewall include Cisco Unity voicemail servers, Attendant Console, Cisco Unified Contact Center Enterprise, and Cisco Unified Contact Center Express. ACLs can be written for these applications to allow traffic to flow through a firewall.

Versions of Cisco FWSM prior to version 3.0 do not support SCCP fragmentation. If an SCCP packet is fragmented from a phone, from Unified CM, or from a gateway to another IP Telephony device, the fragmented packet will not be allowed through the FWSM. In cases where fragmentation occurs with an FWSM running version 2.x code, an ACL should be used without the ALG feature of the firewall for the signaling traffic. This configuration will allow the signaling traffic through the FWSM but will not do packet inspection as the signaling goes through the firewall.

If there are other applications that use the same port as SCCP (TCP 2000), those applications could be affected by the SCCP inspection. All traffic that is going to the SCCP TCP port will be inspected to see if it is SCCP traffic. If it is not SCCP traffic, it will be dropped.

To determine if the applications running on your network are supported with the version of firewall in the network or if ACLs have to be written, refer to the appropriate application documentation available at

<http://www.cisco.com>

Routed ASA and PIX

The ASA or PIX firewall in routed mode acts as a router between connected networks, and each interface requires an IP address on a different subnet. In single-context mode, the routed firewall supports Open Shortest Path First (OSPF) and Routing Information Protocol (RIP) in passive mode. Multiple-context mode supports static routes only. ASA version 8.x also supports Enhanced Interior Gateway Routing Protocol (EIGRP). Cisco recommends using the advanced routing capabilities of the upstream and downstream routers instead of relying on the security appliance for extensive routing needs. For more information on the routed mode, refer to the *Cisco Security Appliance Command Line Configuration Guide*, available at

http://www.cisco.com/en/US/products/ps6120/products_installation_and_configuration_guides_list.html

Advantages

The routed ASA or PIX firewall supports QoS, NAT, and VPN termination to the box, which are not supported in the transparent mode (see [Transparent ASA and PIX](#), page 19-34).

[Figure 19-16](#) shows the logical placement of firewalls for both routed and transparent configurations in active standby mode. With the routed configuration, each interface on the ASA or PIX would have an IP address. In the transparent mode, there would be no IP address on the interfaces other than the IP address to manage the ASA or PIX remotely.

Disadvantages

Unlike with transparent mode, the device can be seen in the network and, because of that, it can be a point of attack. Placing a routed ASA or PIX firewall in a network changes the network routing because some of the routing can be done by the firewall. IP addresses must also be available for all the interfaces on the firewall that are going to be used, so changing the IP addresses of the routers in the network might also be required. If a routing protocol or RSVP is to be allowed through the ASA or PIX firewall, then an ACL will have to be put on the inside (or most trusted) interface to allow that traffic to pass to the outside (or lesser trusted) interfaces. That ACL must also define all other traffic that will be allowed out of the most trusted interface.

Transparent ASA and PIX

The ASA or PIX firewall can be configured to be a Layer 2 firewall (also known as "bump in the wire" or "stealth firewall"). In this configuration, the firewall does not have an IP address (other than for management purposes), and all of the transactions are done at Layer 2 of the network. Even though the firewall acts as a bridge, Layer 3 traffic cannot pass through the security appliance unless you explicitly permit it with an extended access list. The only traffic allowed without an access list is Address Resolution Protocol (ARP) traffic.

Advantages

This configuration has the advantage that an attacker cannot see the firewall because it is not doing any dynamic routing. Static routing is required to make the firewall work even in transparent mode.

This configuration also makes it easier to place the firewall into an existing network because routing does not have to change for the firewall. It also makes the firewall easier to manage and debug because it is not doing any routing within the firewall. Because the firewall is not processing routing requests, the performance of the firewall is usually somewhat higher with **inspect** commands and overall traffic than the same firewall model and software that is doing routing.

Disadvantages

With transparent mode, you are unable to use NAT on the firewall. If you are going to pass data for routing, you will also have to define the ACLs both inside and outside the firewall to allow traffic, unlike with the same firewall in routed mode. Cisco Discovery Protocol (CDP) traffic will not pass through the device even if it is defined. Each directly connected network must be on the same subnet. You cannot share interfaces between contexts; if you plan on running multiple-context mode, you will have to use additional interfaces. You must define all non-IP traffic, such as routing protocols, with an ACL to allow that traffic through the firewall. QoS is not supported in transparent mode. Multicast traffic can be allowed to go through the firewall with an extended ACL, but it is not a multicast device. In transparent mode, the firewall does not support VPN termination other than for the management interface.

If a routing protocol or RSVP is to be allowed through the ASA or PIX firewall, then an ACL will have to be put on the inside (or most trusted) interface to allow that traffic to pass to the outside (or lesser trusted) interfaces. That ACL must also define all other traffic that will be allowed out of the most trusted interface.

For more information on the transparent mode, refer to the *Cisco Security Appliance Command Line Configuration Guide*, available at

http://www.cisco.com/en/US/products/ps6120/products_installation_and_configuration_guides_list.html

ASA and PIX Configuration Example

The following configuration example lists the ports and **inspect** commands that are used to make the firewalls work with voice for ASA and PIX software Release 7.04. This is an example only, and you should review the ports list from all the applications that are used in your network before deploying any firewall. This configuration example shows only the voice sections.

```
!
!
object-group service remote-access tcp
  description remote access
  !Windows terminal
  port-object range 3389 3389
  !VNC
  port-object range 5800 5800
  !VNC
  port-object range 5900 5900
  port-object range 8080 8080
  port-object eq ssh
  !SSH
  port-object eq ftp-data
  !FTP data transport
  port-object eq www
  !HTTP Access
  port-object eq ftp
  !FTP
  port-object eq https
  !HTTPS Access
object-group service voice-protocols tcp
  description TCP voice protocols
```

```

CTI/QBE
port-object range 2428 2428
!SIP communication
port-object eq ctique
!SCCP
port-object range 2000 2000
!Secure SCCP
port-object range 2443 2443
object-group service voice-protocols-udp udp
!TFTP
port-object eq tftp
!MGCP Signaling
port-object range 2427 2427
!DNS
port-object eq domain
!RAS
port-object range 1719 1719
!SIP

!Object Group applied for remote-access
access-list OUTSIDE extended permit tcp any any object-group remote-access
!Object Group applied for voice-protocols-tcp
access-list OUTSIDE extended permit tcp any any object-group voice-protocols-tcp
!Object Group applied for voice-protocols-udp
access-list OUTSIDE extended permit udp any any object-group voice-protocols-udp
! Object Group applied for remote-access
access-list inside_access_in extended permit tcp any any object-group remote-access
! Object Group applied for voice-protocols-tcp
access-list inside_access_in extended permit tcp any any object-group voice-protocols-tcp
! Object Group applied for voice-protocols-udp
access-list inside_access_in extended permit udp any any object-group voice-protocols-udp

!Failover config
ip address 172.19.245.3 255.255.255.248 standby 172.19.245.4
failover
failover lan unit primary
failover lan interface failover GigabitEthernet0/3
failover polltime unit 1 holdtime 3
!Lowest and fastest setting for failover
failover polltime interface 3
failover link failover_state GigabitEthernet0/2
failover interface ip failover 192.168.1.1 255.255.255.0 standby 192.168.1.2
failover interface ip failover_state 192.168.0.1 255.255.255.0 standby 192.168.0.2

!
!Default inspection with inspections enabled
class-map inspection_default
match default-inspection-traffic
!
!
policy-map global_policy
class inspection_default
inspect h323 h225
inspect h323 ras
inspect skinny
inspect sip
inspect tftp
inspect mgcp

```

FWSM Routed Mode

In routed mode, the FWSM is considered to be a router hop in the network. It performs NAT between connected networks and can use OSPF or passive RIP (in single-context mode). Routed mode supports up to 256 interfaces per context or in single mode, with a maximum of 1000 interfaces divided between all contexts.

Advantages

As a routed device in your network, the FWSM supports routing and all other features that are not available in transparent mode.

Disadvantages

Unlike the transparent mode, the routed device is visible in the network and, because of that, it can be a point of attack. To place the device in a network, IP addressing and routing must be change.

FWSM Transparent Mode

In transparent mode, the FWSM acts like a "bump in the wire," or a "stealth firewall," and is not a router hop. The FWSM connects the same network on its inside and outside interfaces, but each interface must be on a different VLAN. No dynamic routing protocols or NAT are required. However, like routed mode, transparent mode also requires ACLs to allow traffic to pass through. Transparent mode can also optionally use EtherType ACLs to allow non-IP traffic. Transparent mode supports only two interfaces, an inside interface and an outside interface.

You might use a transparent firewall to simplify your network configuration. Transparent mode is also useful if you want the firewall to be invisible to attackers. You can also use a transparent firewall for traffic that would otherwise be blocked in routed mode. For example, a transparent firewall can allow multicast streams using an EtherType ACL.

Advantages

This configuration has the advantage that an attacker cannot see the firewall because it is not doing any routing. This configuration also makes it easier to place the firewall into an existing network because routing does not have to change for the firewall. It also makes the firewall easier to manage and debug because it is not doing any routing within the firewall. You can also bridge non-IP traffic and IP multicast traffic, static ARP inspection, and MAC move detection and static MAC.

Disadvantages

To avoid loops when you use failover in transparent mode, you must use switch software that supports Bridge Port Data Unit (BPDU) forwarding, and you must configure the FWSM to allow BPDUs. Transparent mode does not support NAT, dynamic routing, or a unicast reverse path forwarding (RPF) check. There is no NAT 0 with the FWSM in transparent mode.

FWSM Configuration Example

The following configuration example lists the ports and **inspect** commands that are used to make the firewalls work with voice for FWSM software Release 2.3.x. This is only an example, and you should review the ports list from all the applications that are used in your network before deploying any firewall. This configuration example shows only the voice sections.

```

fixup protocol h323 H225 1720
!Enable fixup h3232 h225

fixup protocol h323 ras 1718-1719
!Enable fixup h323 RAS

fixup protocol mgcp 2427
!Enable fixup mgcp

fixup protocol skinny 2000
!Enable fixup

fixup protocol tftp 69
!Enable fixup

object-group service VoiceProtocols tcp
  description Unified CM Voice protocols
  port-object eq ctigbe
  port-object eq 2000
  port-object eq 3224
  port-object eq 2443
  port-object eq 2428
  port-object eq h323
!Defining the ports for TCP voice

object-group service VoiceProtocolsUDP udp
  description UDP based Voice Protocols
  port-object range 2427 2427
  port-object range 1719 1719
  port-object eq tftp
!Defining the ports for UDP voice

object-group service RemoteAccess tcp
  description Remote Acces
  port-object range 3389 3389
  port-object range 5800 5809
  port-object eq ssh
  port-object range 5900 5900
  port-object eq www
  port-object eq https
!Defining remote access TCP ports

access-list inside_nat0_outbound extended permit ip any any
!

access-list phones_access_in extended permit tcp any any object-group RemoteAccess log
notifications interval 2
access-list phones_access_in extended permit tcp any any object-group VoiceProtocols log
notifications interval 2
access-list phones_access_in extended permit udp any any object-group VoiceProtocolsUDP
log notifications interval 2
access-list phones_access_in extended deny ip any any log notifications interval 2
access-list outside_access_in extended permit tcp any any object-group VoiceProtocols log
notifications interval 2
access-list outside_access_in extended permit tcp any any object-group RemoteAccess log
notifications interval 2
access-list outside_access_in extended permit udp any any object-group VoiceProtocolsUDP
log notifications interval 2
!Access lists applying the object groups defined above for inside and outside interfaces

access-list outside_access_in extended deny ip any any log notifications interval 2
access-list inside_access_in extended deny ip any any
!Deny all other traffic

```

```

access-list phones_nat0_outbound extended permit ip any any
!

failover
failover lan unit primary
failover lan interface flin vlan 4050
failover polltime unit 1 holdtime 5
failover polltime interface 15
!Failover config - 15 seconds
failover interface-policy 50%
failover link flin vlan 4051
failover interface ip flin 1.1.1.1 255.255.255.252 standby 1.1.1.2
failover interface ip flin 1.1.1.5 255.255.255.252 standby 1.1.1.6
nat (inside) 0 access-list inside_nat0_outbound_V1
access-group outside_access_in in interface outside
access-group inside_access_in in interface inside

```

Data Center

Within the data center, the security policy should define what security is needed for the IP Telephony applications servers. Because the Cisco Unified Communications servers are based on IP, the security that you would put on any other time-sensitive data within a data center could be applied to those servers as well.

If clustering over the WAN is being used between data centers, any additional security that is applied both within and between those data centers has to fit within the maximum round-trip time that is allowed between nodes in a cluster. If your current security policy for application servers within your network covers the IP Telephony servers from Cisco, then you should use that security. You can also use any of the infrastructure security that is already deployed.

To design appropriate data center security for your data applications, Cisco recommends following the guidelines presented in the *Data Center Networking: Server Farm Security SRND (Server Farm Security in the Business Ready Data Center Architecture)*, available at

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

Applications Servers

For a list of the Unified CM security features and how to enable them, refer to the *Cisco Unified Communications Manager Security Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

Before enabling any of the Unified CM security features, verify that they will satisfy the security requirements specified in your enterprise security policy for these types of devices in a network.

Cisco Security Agent on the Unified CM and Application Servers

The Cisco Security Agent is used on most of the application servers that Cisco uses to provide IP Telephony and IP Telephony services. The Cisco Security Agent software is Host Intrusion Prevention software that looks at the behavior of the traffic to and from the server, and the way the applications are running on that server, to determine if everything is working correctly. If something is considered

abnormal, then the Cisco Security Agent software prevents that activity from happening. For example, if a virus is trying to install a software package on a Unified CM and that is something that has never happened before, the virus would be prevented from installing.

Unmanaged Cisco Security Agent

Cisco developed a default Cisco Security Agent policy for its servers that allows all the correct things needed for that IP Telephony server to function, while also preventing known and unknown attacks from affecting the IP Telephony servers. This software protects the application and the operating system from both viruses and worm attacks. To get the maximum protection from these types of intrusions, ensure that the newest version of the Cisco Security Agent software is always installed on your servers. With the unmanaged agent installed on your servers, you will be able to see the logs of attacks on only the system where the agent is installed. You will have to log into each system to check the log files that might be there because of some type of alarm that has occurred. The unmanaged Cisco Security Agent is installed by default when you install Unified CM.

Advantages

The unmanaged Cisco Security Agent protects each system from both known and unknown attacks, worms, and viruses.

Disadvantages

When run in unmanaged mode, Cisco Security Agent does not correlate alarms, and you have to access each system individually to see the log files on that system.

Managed Cisco Security Agent



Note

Cisco Unified CM 5.x currently does *not* support managed Cisco Security Agent capability.

Antivirus



Note

Cisco Unified CM 5.x currently does *not* support the use of antivirus software.

General Server Guidelines

Your Unified CM and other IP Telephony application servers should not be treated as normal servers. Anything you do while configuring the system could affect calls that are trying to be placed or that are in progress. As with any other business-class application, major configuration changes should be done within maintenance windows to keep from disrupting phone conversations.

Standard security policies for application servers might not be adequate for IP Telephony servers. Unlike email servers or web servers, voice servers will not allow you to refresh a screen or re-send a message. The voice communications are real-time events. Any security policy for IP Telephony servers should ensure that work that is not related to configuring or managing the voice systems is not done on the IP Telephony servers at any time. Activities that might be considered normal on application servers within a network (for example, surfing the internet) should not take place on the IP Telephony servers.

In addition, Cisco provides a well defined patch system for the IP Telephony servers, and it should be applied based on the patch policy within your IT organization. You should not patch the system normally using the OS vendor's patch system unless it is approved by Cisco Systems. All patches should be downloaded from Cisco or from the OS vendor as directed by Cisco Systems, and applied according to the patch installation process.

Various software patches are available at the following location:

<http://www.cisco.com/kobayashi/sw-center/sw-voice.shtml>

**Note**

A Cisco.com login account is required for access to this link.

The above site also contains a notification tool that will email you when a critical patch must be applied to an IP Telephony server.

Advantages

General server security practices help to mitigate viruses and worms if the application server is treated like a PBX and not like other application servers.

Disadvantages

When the additional security features are configured, some of the Unified CM functionality might be reduced. Also, additional care is needed during upgrades because some of the services that are disabled for the additional security will have to be enabled for a successful upgrade.

Deployment Examples

This section presents examples of what could be done from a security perspective for a lobby phone and a firewall deployment. A good security policy should be in place to cover deployments similar to these types.

Lobby Phone Example

The example in this section illustrates one possible way to configure a phone and a network for use in an area with low physical security, such as a lobby area. None of the features in this example are required for a lobby phone, but if your security policy states more security is needed, then you could use the features listed in this example.

Because you would not want anyone to gain access to the network from the PC port on the phone, you should disable the PC port on the back of the phone to limit network access (see [PC Port on the Phone, page 19-6](#)). You should also disable the settings page on the phone so that potential attackers cannot see the IP addresses of the network to which the lobby phone is connected (see [Settings Access, page 19-10](#)). The disadvantage of not being able to change the settings on the phone usually will not matter for a lobby phone.

Because there is very little chance that a lobby phone will be moved, you could use a static IP address for that phone. A static IP address would prevent an attacker from unplugging the phone and then plugging into that phone port to get a new IP address (see [IP Addressing, page 19-5](#)). Also, if the phone is unplugged, the port state will change and the phone will no longer be registered with Unified CM. You can track this event in just the lobby phone ports to see if someone is trying to attach to the network.

Using static port security for the phone and not allowing the MAC address to be learned would mean that an attacker would have to change his MAC address to that of the phone, if he were able to discover that address. Dynamic port security could be used with an unlimited timer to learn the MAC address (but never unlearn it), so that it would not have to be added. Then the switchport would not have to be changed to clear that MAC address unless the phone is changed. The MAC address is listed in a label on the bottom of the phone. If listing the MAC address is considered a security issue, the label can be removed and replaced with a "Lobby Phone" label to identify the device. (See [Switch Port, page 19-13](#).)

A single VLAN could be used and Cisco Discovery Protocol (CDP) could be disabled on the port so that attackers would not be able to see any information from the Ethernet port about that port or switch to which it is attached. In this case, the phone would not have a CDP entry in the switch for E911 emergency calls, and each lobby phone would need either a label or an information message to local security when an emergency number is dialed.

A static entry in the DHCP Snooping binding table could be made because there would be no DHCP on the port (see [DHCP Snooping: Prevent Rogue DHCP Server Attacks, page 19-16](#)). Once the static entry is in the DHCP Snooping binding table, Dynamic ARP Inspection could be enabled on the VLAN to keep the attacker from getting other information about one of the Layer 2 neighbors on the network (see [Requirement for Dynamic ARP Inspection, page 19-20](#)).

With a static entry in the DHCP Snooping binding table, IP Source Guard could be used (see [IP Source Guard, page 19-22](#)). If an attacker got the MAC address and the IP address and then started sending packets, only packets with the correct IP address could be sent.

A VLAN ACL could be written to allow only the ports and IP addresses that are needed for the phones to operate (see [VLAN Access Control Lists, page 19-24](#)). The following example contains a very small ACL that can be applied to a port at Layer 2 or at the first Layer 3 device to help control access into the network (see [Router Access Control Lists, page 19-26](#)). This example is based on a Cisco 7960 IP Phone being used in a lobby area, without music on hold to the phone or HTTP access from the phone.

This example uses the following IP address ranges:

- The lobby phone has an IP address of 10.0.40.5
- The Unified CM cluster uses the address range 10.0.20.*
- The DNS server has an IP address of 10.0.30.2
- The HSRP routers have IP addresses 10.0.10.2 and 10.0.10.3
- Other phones in the network use IP addresses in the range 10.0.*.*

```
10 permit icmp any any
! Allow all icmp - phone to phone, gateway to phone and NMS to phone

20 permit udp host 10.0.10.2 eq 1985 any
!Allow HSRP information in, do not allow out

30 permit udp host 10.0.10.3 eq 1985 any
! Allow in from HSRP neighbor

40 permit udp host 10.0.40.5 range 49152 65535 10.0.20.0 0.0.0.255 eq tftp
! Using ip host from ephemeral port range from phone to the TFTP server port 69 (start of tftp)

50 permit udp 10.0.20.0 0.0.0.255 range 1024 5000 host 10.0.40.5 range 49152 65535
!Using IP subnet from TFTP server with ephemeral port range to ip host and ephemeral port range for phone

60 permit udp host 10.0.40.5 range 49152 65535 10.0.20.0 0.0.0.255 range 1024 5000
! Using host from phone to TFTP server with ephemeral port range to ip range and ephemeral port range for TFTP (continue the TFTP conversation)
```

```

70 permit udp host 10.0.40.5 range 49152 65535 host 10.0.30.2 eq domain
! Using IP host and ephemeral port range from phone to DNS server host

80 permit udp host 10.0.30.2 eq domain host 10.0.40.5 range 49152 65535
! Using IP from DNS server to phone host ip and ephemeral port range

90 permit tcp 10.0.40.5 range 49152 65535 10.0.20.0 0.0.0.255 eq 2000
! Using IP host and ephemeral port range from phone to Unified CM cluster for SCCP

100 permit tcp 10.0.20.0 0.0.0.255 eq 2000 host 10.0.40.5 range 49152 65535
! Using IP range and SCCP port to phone IP host and ephemeral port range

110 permit udp 10.0.0.0 0.0.255.255 range 16384 32767 host 10.0.40.5 range 16384 32767
! Using IP range and ephemeral port range from all phones or gateways outside a vlan to
the IP host to phone

120 permit udp host 10.0.40.5 range 16384 32767 10.0.0.0 0.0.255.255 range 16384 43767
! Using IP host and ephemeral port range from vlan to all other phones or gateways

130 permit udp host 172.19.244.3 range 1024 5000 host 10.0.40.5 eq snmp
!From IP host of NMS server and ephemeral port range (Different for Windows vs Sun) to IP
host of phones and SNMP port (161)

140 permit udp host 10.0.40.5 eq snmp host 172.19.244.3 range 1024 5000
! From IP host of phone with SNMP port (161) to IP host of NMS server and ephemeral port
range

```

Basic QoS Example for a Lobby Phone

Set the voice stream to G.729 and use QoS to limit the amount of traffic that can be sent into the port (see [Quality of Service, page 19-24](#)). Even if the QoS maximum is exceeded, the traffic will be reset to CS1, or Scavenger Class, which is the lowest priority traffic in most enterprise networks.

```

CAT2970(config)#mls qos map policed-dscp 0 24 46 to 8
! Excess traffic marked 0 or CS3 or EF will be remarked to CS1
CAT2970(config)#
CAT2970(config)#class-map match-all LOBBY-VOICE
CAT2970(config-cmap)# match access-group name LOBBY-VOICE
CAT2970(config-cmap)#class-map match-all LOBBY-SIGNALING
CAT2970(config-cmap)# match access-group name LOBBY-SIGNALING
CAT2970(config-cmap)#exit
CAT2970(config)#
CAT2970(config)#policy-map LOBBY-PHONE
CAT2970(config-pmap)#class LOBBY-VOICE
CAT2970(config-pmap-c)# set ip dscp 46 ! Lobby phone VoIP is marked to DSCP EF
CAT2970(config-pmap-c)# police 48000 8000 exceed-action policed-dscp-transmit
! Out-of-profile Lobby voice traffic (g.729) is marked down to Scavenger (CS1)
CAT2970(config-pmap-c)#class LOBBY-SIGNALING
CAT2970(config-pmap-c)# set ip dscp 24 ! Signaling is marked to DSCP CS3
CAT2970(config-pmap-c)# police 32000 8000 exceed-action policed-dscp-transmit
! Out-of-profile Signaling traffic is marked down to Scavenger (CS1)
CAT2970(config-pmap-c)#class class-default
CAT2970(config-pmap-c)# set ip dscp 0
CAT2970(config-pmap-c)# police 56000 8000 exceed-action policed-dscp-transmit
! Out-of-profile data traffic is marked down to Scavenger (CS1)
CAT2970(config-pmap-c)# exit
CAT2970(config-pmap)#exit
CAT2970(config)#
CAT2970(config)#interface GigabitEthernet0/1

```

```

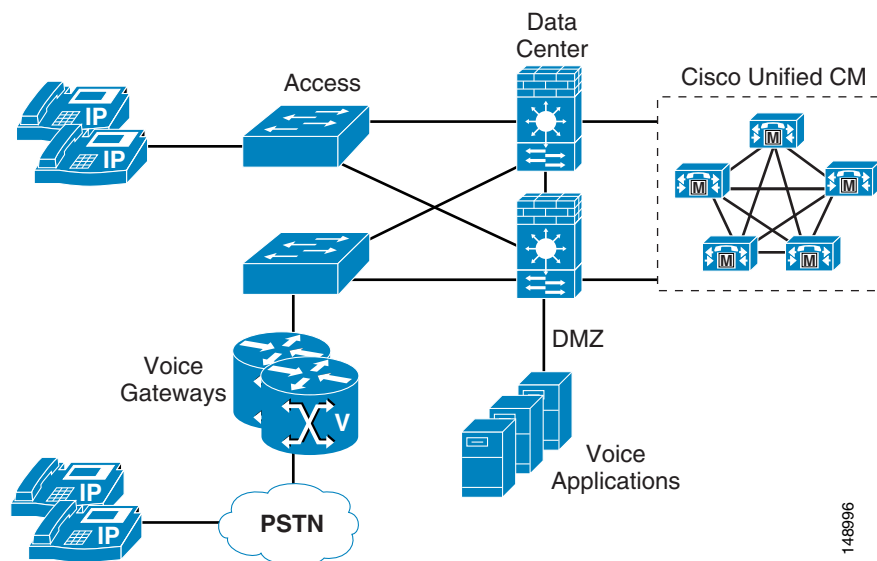
CAT2970(config-if)# service-policy input LOBBY-PHONE ! Applies policy to int
CAT2970(config-if)#exit
CAT2970(config)#
CAT2970(config)#ip access list extended LOBBY-VOICE
CAT2970(config-ext-nacl)# permit udp any any range 16384 32767 ! VoIP ports
CAT2970(config-ext-nacl)#
CAT2970(config-ext-nacl)#ip access list extended LOBBY-SIGNALING
CAT2970(config-ext-nacl)# permit tcp any any range 2000 2002 ! SCCP ports
CAT2970(config-ext-nacl)#end
CAT2970#

```

Firewall Deployment Example (Centralized Deployment)

The example in this section is one way that firewalls could be deployed within the data center, with Unified CMs behind them (see [Figure 19-19](#)). In this example, the Unified CMs are in a centralized deployment, single cluster with all the phones outside the firewalls. Because the network in this deployment already contained firewalls that are configured in routed mode within the corporate data center, the load was reviewed before the placement of gateways was determined. After reviewing the average load of the firewall, it was decided that all the RTP streams would not transverse the firewall in order to keep the firewalls under the 60% CPU load (see [Putting Firewalls Around Gateways](#), [page 19-29](#)). The gateways are placed outside the firewalls, and ACLs within the network are used to control the TCP data flow to and from the gateways from the Unified CMs. An ACL is also written in the network to control the RTP streams from the phones because the IP addresses of the phones are well defined (see [IP Addressing](#), [page 19-5](#)). The voice applications servers are placed within the demilitarized zone (DMZ), and ACLs are used at the firewalls to control access to and from the Unified CMs and to the users in the network. This configuration will limit the amount of RTP streams through the firewall using inspects, which will minimize the impact to the firewalls when the new voice applications are added to the existing network.

Figure 19-19 Firewall Deployment Example



Conclusion

This chapter did not cover all of the security that could be enabled to protect the voice data within your network. The techniques presented here are just a subset of all the tools that are available to network administrators to protect all the data within a network. On the other hand, even these tools do not have to be enabled within a network, depending on what level of security is required for the data within the network overall. Choose your security methods wisely. As the security within a network increases, so do the complexity and troubleshooting problems. It is up to each enterprise to define both the risks and the requirements of its organization and then to apply the appropriate security within the network and on the devices attached to that network.



CHAPTER 20

IP Telephony Endpoints

Last revised on: September 27, 2007

This chapter summarizes various types of IP Telephony endpoints along with their features and QoS recommendations. The IP Telephony endpoints can be categorized into the following major types:

- [Analog Gateways, page 20-2](#)
- [Cisco Unified IP Phones, page 20-6](#)
- [Software-Based Endpoints, page 20-10](#)
- [Wireless Endpoints, page 20-13](#)
- [Cisco IP Conference Station, page 20-19](#)
- [Video Endpoints, page 20-20](#)
- [Third-Party SIP IP Phones, page 20-25](#)

The sections listed above provide detailed information about each endpoint type. In addition, the section on [QoS Recommendations, page 20-25](#), lists generic QoS configurations, and the [Endpoint Features Summary, page 20-37](#), lists all the endpoint features.

The following list summarizes high-level recommendations for selecting IP Telephony endpoints:

- For low-density analog connections, use the Cisco Analog Telephone Adapter (ATA) or low-density analog interface module.
- For medium to high-density analog connections, use the high-density analog interface module, Cisco Communication Media Module (CMM) with 24-FXS port adapter, Catalyst 6500 24-FXS analog interface module, Cisco VG224, or Cisco VG248.
- For telephony users with limited call features who generate small amounts of traffic, use the Cisco Unified IP Phones 7902G, 7905G, 7906G, 7911G, 7912G, or 7912G-A.
- For transaction-type telephony users who generate a medium amount of traffic, use Cisco Unified IP Phones 7940G, 7941G, or 7941G-GE.
- For managers and administrative assistants who generate medium to heavy telephony traffic, use Cisco Unified IP Phones 7960G, 7961G, or 7961G-GE.
- For executives with extensive call features who generate high amounts of telephony traffic, use Cisco Unified IP Phones 7970G or 7971G-GE.
- For mobile workers and telecommuters, use Cisco IP Communicator.
- For users who need a mobile IP phone, use the Cisco Unified Wireless IP Phone 7920.

- For making video calls, use Cisco Unified Video Advantage associated with a Cisco Unified IP Phone or Cisco IP Communicator, the Cisco IP Video Phone 7985G, or Sony and Tandberg SCCP endpoints.
- For accessing voice, video, document sharing, and presence information from a single integrated interface, use Cisco Unified Personal Communicator.
- For formal conferencing environments, use the Cisco Unified IP Conference Station 7936.

Analog Gateways

Analog gateways include router-based analog interface 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. An 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 Interface Module

Cisco router-based analog interface modules include low-density interface modules (NM-1V, NM-2V, NM-HD-1V, NM-HD-2V, NM-HD-2VE, NM-HDV2, NM-HDV2-1T1/E1, and NM-HDV2-2T1/E1) and high-density interface modules (NM-HDA-4FXS and EVM-HD-8FXS/DID). Cisco analog interface 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 interface modules are best suited for connecting low- to high-density analog devices to the IP network with limited call features.

Low-Density Analog Interface Module

The low-density analog interface modules include the NM-1V, NM-2V, NM-HD-1V, NM-HD-2V, NM-HD-2VE, NM-HDV2, NM-HDV2-1T1/E1, and NM-HDV2-2T1/E1. The NM-1V and NM-2V contain one or two interface cards (VIC). The interface cards include: two-port FXS VIC (VIC-2FXS); two-port FXO VIC (VIC-2FXO, VIC-2FXO-M1/M2/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). The NM-1V and NM-2V can serve up to two and four FXS connections, respectively.



Note

The NM-1V and NM-2V are not supported on the Cisco 2800 and 3800 Series platforms. On the Cisco 2800 and 3800 Series platforms, the voice interface cards are supported in the on-board High-Speed WIC slots, including the VIC-2DID, VIC4-FXS/DID, VIC2-2FXO, VIC-2-4FXO, VIC2-2FXS, VIC2-2E/M, and VIC2-2BRI-NT/TE.

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 4, 8, and 8 FXS or FXO connections, respectively. The NM-HDV2, NM-HDV2-1T1/E1, and NM-HDV2-2T1/E1 can be fitted with either digital T1/E1 or analog/BRI voice interface cards, with up to 4 FXS or FXO connections. The difference among these three interface modules is that the NM-HDV2-1T1/E1 has one built-in T1/E1 port while the NM-HDV2-2T1/E1 has two built-in T1/E1 ports.

The voice interface cards include: 2-port and 4-port FXS VICs (VIC2-2FXS and VIC-4FXS/DID); 2-port and 4-port FXO VICs (VIC2-2FXO and VIC2-4FXO); 2-port Direct Inward Dial VIC (VIC-2DID); 2-port E&M VIC (VIC2-2E/M); and 2-port BRI VIC (VIC2-2BRI-NT/TE). The voice/WAN interface cards include: 1-port and 2-port RJ-48 multiflex trunk (MFT) T1/E1 VWICs for both voice and WAN connections (VWIC-1MFT-T1, VWIC-2MFT-T1, VWIC-2MFT-T1-DI, VWIC-1MFT-E1, VWIC-2MFT-E1, VWIC-2MFT-E1-DI, VWIC-1MFT-G703, VWIC-2MFT-G703, VWIC2-1MFT-T1/E1, VWIC2-2MFT-T1/E1, VWIC2-1MFT-G703, and VWIC2-2MFT-G703). The G.703 interface cards are primarily for data connectivity but can in some cases be configured to support voice applications.

High-Density Analog Interface Module

The high-density analog interface module includes the NM-HDA-4FXS and EVM-HD-8FXS/DID. The NM-HDA-4FXS has four on-board FXS ports and room for two expansion modules from the following options:

- EM-HDA-8FXS: An 8-port FXS interface card
- EM-HDA-4FXO/EM2-HDA-4FXO: A 4-port FXO interface card

The NM-HDA-4FXS provides up to 12 analog ports (4 FXS and 8 FXO) with four built-in FXS ports and two EM-HDA-4FXO or EM2-HDA-4FXO extension modules, or 16 analog ports (12 FXS and 4 FXO) with four built-in FXS ports and one EM-HDA-8FXS, and one EM-HDA-4FXO or EM2-HDA-4FXO extension module. A configuration using two 8-port FXS extension modules is not supported. The NM-HDA also has a connector for a daughter module (DSP-HDA-16) that provides additional DSP resources to serve an additional 8 high-complexity calls or 16 medium-complexity calls.



Note

The EM2-HDA-4FXO supports the same density and features as the EM-HDA-FXO, but it provides enhanced features including longer loop length support of up to 15,000 feet and improved performance under poor line conditions when used in ground-start signaling mode.

The EVM-HD-8FXS/DID provides eight individual ports on the baseboard module and can be configured for FXS or DID signaling. In addition, the EVM-HD-8FXS/DID has room for two expansion modules from the following options:

- EM-HDA-8FXS: An 8-port FXS interface card
- EM-HDA-6FXO: A 6-port FXO interface card
- EM-HDA-3FXS/4FXO: A 3-port FXS and 4-port FXO interface card
- EM-4BRI-NT/TE: A 4-port BRI interface card

These extension modules can be used in any combination and provide for configurations of up to 24 FXS ports per EVM-HD-8FXS/DID.

Supported Platforms and Cisco IOS Requirements for Analog Interface Modules

The supported platforms for Cisco analog interface modules are the Cisco 2600, 2800, 3600, 3700, and 3800 Series. [Table 20-1](#) lists the maximum number of interface modules supported per platform, and [Table 20-2](#) lists the minimum Cisco IOS software version required.

Table 20-1 Maximum Number of Analog Interface Modules Supported per Platform

Platform	Maximum Number of Interface Modules Supported				
	NM-1V, -2V	NM-HDA-4FXS	EVM-HD	NM-HD-1V, -2V, -2VE	NM-HDV2, -1T1/E1, -2T1/E1
Cisco 2600XM	1	1	No	1	1
Cisco 2691	1	1	No	1	1
Cisco 3640	3	3	No	3	No
Cisco 3660	6	6	No	6	No
Cisco 3725	2	2	No	2	2
Cisco 3745	4	4	No	4	4
Cisco 2811	No	1	1	1	1
Cisco 2821	No	1	1	1	1
Cisco 2851	No	1	1	1	1
Cisco 3825	No	2	1	2	2
Cisco 3845	No	4	2	4	4

Table 20-2 Minimum Cisco IOS Requirements for Analog Interface Modules

Platform	Minimum Cisco IOS Software Release Required				
	NM-1V, -2V	NM-HDA-4FXS	EVM-HD	NM-HD-1V, -2V, -2VE	NM-HDV2, -1T1/E1, -2T1/E1
Cisco 2600XM	12.2(8)T	12.2(8)T	No	12.3.4T	12.3(7)T
Cisco 2691	12.2(8)T	12.2(8)T	No	12.3.4T	12.3(7)T
Cisco 3640	12.0(1)T or later	12.2(8)T or later	No	12.3.4T	No
Cisco 3660	12.0(1)T or later	12.2(8)T or later	No	12.3.4T	No
Cisco 3725	12.2(8)T or later	12.2(8)T	No	12.3.4T	12.3(7)T
Cisco 3745	12.2(8)T or later	12.2(8)T	No	12.3.4T	12.3(7)T
Cisco 2811	No	12.3.8T4	12.3.8T4	12.3.8T4	12.3.8T4
Cisco 2821	No	12.3.8T4	12.3.8T4	12.3.8T4	12.3.8T4
Cisco 2851	No	12.3.8T4	12.3.8T4	12.3.8T4	12.3.8T4
Cisco 3825	No	12.3(11)T	12.3(11)T	12.3(11)T	12.3(11)T
Cisco 3845	No	12.3(11)T	12.3(11)T	12.3(11)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), 6-port T1 interface port adapter (WS-SVC-CMM-6T1), 6-port E1 interface port adapter (WS-SVC-CMM-6E1), and conference/transcoding port adapter (WS-SVC-CMM-ACT). [Table 20-3](#) lists the minimum software requirements for the compatible port adapters.

Table 20-3 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	12.3(8)XY	12.3(8)XY	12.3(8)XY	12.3(8)XY
CatOS Release	7.3(1)	7.3(1)	7.3(1)	7.6.8
Native IOS Release	12.1(15)E	12.1(14)E	12.1(13)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.



Note

The WS-X6624 FXS analog interface module is no longer available for sale.

Cisco VG224 Gateway

The Cisco VG224 analog gateway is a Cisco IOS high-density 24-port gateway for connecting analog devices to the IP Telephony network. In Cisco IOS Release 12.4(2)T and later, the Cisco VG224 can act as an Session Initiation Protocol (SIP), Skinny Client Control Protocol (SCCP), Media Gateway Control Protocol (MGCP), or H.323 endpoint with Cisco Unified Communications Manager (Unified CM) and re-home to a Survivable Remote Site Telephone (SRST) router in failover scenarios. The Cisco VG224 supports Cisco Unified CM Release 3.1 and later. The Cisco VG224 also supports modem pass-through, modem relay, fax pass-through, and fax relay.

Cisco VG248 Gateway

The Cisco VG248 is a high-density, 48 port, Skinny Client Control Protocol (SCCP) gateway for connecting analog devices such as analog phones, fax machines, modems and speakerphones to an enterprise Cisco Unified CM (Release 3.1 and later) and voice network. The Cisco VG248 also supports Unified CM integration with legacy voicemail systems and PBXs compatible with Simplified Message Desk Interface (SMDI), NEC Message Center Interface (MCI), or Ericsson voicemail protocols. The Cisco VG248 supports failover to Survivable Remote Site Telephone (SRST).

Cisco ATA 186 and 188

The Cisco Analog Telephone Adaptor (ATA) 186 or 188 can connect two analog devices to the IP telephony network, and it is the best suited for low-density analog devices connecting to the IP 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 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 a SIP client that registers with the SIP proxy 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. Cisco Unified CM 5.x does not have native SIP support for the Cisco ATA 186 or 188.

Cisco Unified IP Phones

The Cisco IP phone portfolio includes basic IP phones, business IP phones, manager IP Phones, and executive IP phones.

Cisco Basic IP Phones

The Cisco basic IP phone is best suited for low-traffic users with limited call features and budget requirements. The basic IP phones include Cisco Unified IP Phone 7902G, 7905G, 7906G, 7911G, and 7912G.

Cisco Unified IP Phone 7902G

The Cisco Unified 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 Unified IP Phone 7902G does not have any liquid crystal display (LCD) screen. The Cisco Unified IP Phone 7902G supports SCCP but has no support for SIP.

Cisco Unified IP Phone 7905G

The Cisco Unified 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 Unified IP Phone 7905G supports SCCP and SIP; however, the features and user interface (UI) are not consistent across the two call signaling protocols.

Cisco Unified IP Phone 7906G

The Cisco Unified IP Phone 7906G supports a single line and it has a single 10/100 BaseT Ethernet port on the back of the phone. The speaker operates in one-way listen mode only. Power is supplied via IEEE 802.3af, Cisco inline power, or local power through a power adaptor (CP-PWR-CUBE-3). The Cisco

Unified IP Phone 7906G supports SCCP and SIP and is one of the phones included in the enhanced architecture of Cisco Desktop IP Phones. This architecture provides for feature and UI consistency across the Cisco Desktop IP Phones regardless of call signaling protocol. The end-user experience for a supported feature will behave consistently whether using SCCP or SIP call control signaling.

Cisco Unified IP Phone 7911G

The Cisco Unified IP Phone 7911G supports only a single line and it has two 10/100 BaseT Ethernet connections. The speaker operates in one-way listen mode only. Power is supplied via IEEE 802.3af, Cisco inline power, or local power through a power adaptor (CP-PWR-CUBE-3). The Cisco Unified IP Phone 7911G supports SCCP and SIP and is one of the phones included in the enhanced architecture of Cisco Desktop IP Phones. This architecture provides for feature and UI consistency across the Cisco Desktop IP Phones regardless of call signaling protocol. The end-user experience for a supported feature will behave consistently whether using SCCP or SIP call control signaling.

Cisco Unified IP Phone 7912G

The Cisco Unified IP Phone 7912G supports only a single line and it has two 10/100 BaseT Ethernet connections. The speaker operates in one-way listen mode only. The Cisco Unified IP Phone 7912G supports SCCP and SIP; however, the features and user interface (UI) are not consistent across the two call signaling protocols.

**Note**

The initial version of the Cisco Unified IP Phone 7912G is no longer available for sale. The initial version of the Cisco Unified IP Phone 7912G is now replaced by the Cisco Unified IP Phone 7912G-A, which offers identical features but has an enhanced Ethernet switch.

Cisco Business IP Phones

The Cisco business IP phone is best suited for the transaction-type worker with medium telephony traffic use and extensive call features, such as speakers, headset, and so forth. The business IP phones include Cisco Unified IP Phone 7940G, 7941G, and 7941G-GE.

Cisco Unified IP Phone 7940G

The Cisco Unified IP Phone 7940G can have up to two directory numbers and includes two 10/100 BaseT Ethernet connections. The Cisco Unified IP Phone 7940G supports SCCP and SIP; however, the features and user interface (UI) are not consistent across the two call signaling protocols. For example, the Cisco Unified IP Phone 7940G using SCCP has full security capability, whereas SIP does not have any previously implemented security features. The Cisco Unified IP Phone 7940G using SCCP is compatible with the Cisco Unified Video Advantage video-enabled endpoint to make video calls, whereas the Cisco Unified IP Phone 7940G using SIP has no video support. For a complete list of supported features, see the [Endpoint Features Summary, page 20-37](#).

Cisco Unified IP Phone 7941G

The Cisco Unified IP Phone 7941G can have up to two directory numbers and includes two 10/100 BaseT Ethernet connections. The Cisco Unified IP Phone 7941G supports SCCP and SIP and is included in the enhanced architecture of Cisco Unified IP Phones. This architecture provides for feature and UI consistency across the Cisco IP phones regardless of call signaling protocol. The end-user experience for a supported feature will behave consistently whether using SCCP or SIP call control signaling.

There are a few features that are not supported with SIP that are supported with SCCP. For example, the Cisco Unified IP Phone 7941G using SCCP is compatible with the Cisco Unified Video Advantage video-enabled endpoint to make video calls, whereas SIP has no video support. The Cisco Unified IP Phone 7941G using SCCP supports Tone on Hold, whereas SIP has no Tone on Hold support. For a complete list of supported features, see the [Endpoint Features Summary, page 20-37](#).

This phone includes a higher-resolution, 4-bit grayscale display for enhancing feature usage and Extensible Markup Language (XML) applications, as well as for enabling support for double-byte languages. Power is supplied via IEEE 802.3af, Cisco inline power, or local power through a power adaptor (CP-PWR-CUBE-3).

Cisco Unified IP Phone 7941G-GE

The Cisco Unified IP Phone 7941G-GE can have up to two directory numbers and is the equivalent of the Cisco Unified IP Phone 7941G with the exception that it includes two 10/100/1000 BaseT Ethernet connections. The addition of gigabit throughput capability allows for high bit-rate and bandwidth-intensive applications on a co-located PC.

Cisco Manager IP Phones

The Cisco manager IP phone is best suited for managers and administrative assistants with medium to heavy telephony traffic use and extensive call features such as speakers, headset, and so forth. The business IP phones include Cisco Unified IP Phone 7960G, 7961G, and 7961G-GE.

Cisco Unified IP Phone 7960G

The Cisco Unified IP Phone 7960G can have up to six directory numbers and includes two 10/100 BaseT Ethernet connections. The Cisco Unified IP Phone 7960G supports SCCP and SIP; however, the features and user interface (UI) are not consistent across the two call signaling protocols. For example, the Cisco Unified IP Phone 7960G using SCCP has full security capability, whereas SIP does not have any previously implemented security features. The Cisco Unified IP Phone 7960G using SCCP is compatible with the Cisco Unified Video Advantage video-enabled endpoint to make video calls, whereas the Cisco Unified IP Phone 7960G using SIP has no video support. The Cisco Unified IP Phone 7960G using SCCP supports the Cisco Unified IP Phone Expansion Module 7914, whereas SIP has no support for the expansion module. For a complete list of supported features, see the [Endpoint Features Summary, page 20-37](#).

Cisco Unified IP Phone 7961G

The Cisco Unified IP Phone 7961G can have up to six directory numbers and includes two 10/100 BaseT Ethernet connections. The Cisco Unified IP Phone 7961G supports SCCP and SIP and is included in the enhanced architecture of the Cisco Unified IP Phones. This architecture provides for

feature and UI consistency across the Cisco IP phones regardless of call signaling protocol. The end-user experience for a supported feature will behave consistently whether using SCCP or SIP call control signaling.

There are a few features that are not supported with SIP that are supported with SCCP. For example, the Cisco Unified IP Phone 7961G using SCCP is compatible with the Cisco Unified Video Advantage video-enabled endpoint to make video calls, whereas SIP has no video support. The Cisco Unified IP Phone 7961G using SCCP supports Tone on Hold, whereas SIP has no Tone on Hold support. The Cisco Unified IP Phone 7961G using SCCP supports the Cisco Unified IP Phone Expansion Module 7914, whereas SIP has no support for the expansion module. For a complete list of supported features, see the [Endpoint Features Summary, page 20-37](#).

This phone includes a higher-resolution, 4-bit grayscale display for enhancing feature usage and Extensible Markup Language (XML) applications, as well as for enabling support for double-byte languages. Power is supplied via IEEE 802.3af, Cisco inline power, or local power through a power adaptor (CP-PWR-CUBE-3).

Cisco Unified IP Phone 7961G-GE

The Cisco Unified IP Phone 7961G-GE can have up to six directory numbers and is the equivalent of the Cisco Unified IP Phone 7961G with the exception that it includes two 10/100/1000 BaseT Ethernet connections. The addition of gigabit throughput capability allows for high bit-rate and bandwidth-intensive applications on a co-located PC.

Cisco Executive IP Phones

The Cisco executive IP phone is best suited for the executive high-traffic user with extensive call features. The executive IP phones include the Cisco Unified IP Phone 7970G and 7971G-GE.

Cisco Unified IP Phone 7970G

The Cisco Unified IP Phone 7970G can have up to eight directory numbers, has a high-resolution color touch screen, and has more access keys on the phone compared to other Cisco Unified IP Phones. The Cisco Unified IP Phone 7970G supports both SCCP and SIP and is included in the enhanced architecture of Cisco Desktop IP Phones. This architecture provides for feature and UI consistency across the Cisco Desktop IP Phones regardless of call signaling protocol. The end-user experience for a supported feature will behave consistently whether using SCCP or SIP call control signaling.

There are a few features that are not supported with SIP that are supported with SCCP. For example, the Cisco Unified IP Phone 7970G using SCCP is compatible with the Cisco Unified Video Advantage video-enabled endpoint to make video calls, whereas SIP has no video support. The Cisco Unified IP Phone 7970G using SCCP supports Tone on Hold, whereas SIP has no Tone on Hold support. The Cisco Unified IP Phone 7970G using SCCP supports the Cisco Unified IP Phone Expansion Module 7914, whereas SIP has no support for the expansion module. For a complete list of supported features, see the [Endpoint Features Summary, page 20-37](#).

The Cisco Unified IP Phone 7970G includes a high-resolution color touch screen for enhancing feature usage and Extensible Markup Language (XML) applications, as well as for enabling support for double-byte languages. Power is supplied via IEEE 802.3af, Cisco inline power, or local power through a power adaptor (CP-PWR-CUBE-3). For the Cisco Unified IP Phone 7970G to have full display brightness, the external power adaptor (CP-PWR-CUBE-3) must be used for both Cisco inline power and IEEE 802.3af Power over Ethernet (PoE).

Cisco 7971G-GE

The Cisco Unified IP Phone 7971G-GE can have up to eight directory numbers and is the equivalent of the Cisco Unified IP Phone 7970G with the exception that it includes two 10/100/1000 BaseT Ethernet connections. The addition of gigabit throughput capability allows for high bit-rate and bandwidth-intensive applications on a co-located PC.

**Note**

In addition to using the inline power from the access switch or local wall power, a Cisco Unified IP Phone can also be supplied power by a power injector, Promax. Promax connects Cisco Unified IP Phones to Cisco switches that do not support inline power or to non-Cisco switches. Promax is compatible with all Cisco Unified IP Phones, and it supports both Cisco PoE and IEEE 802.3af PoE. It has two 10/100/1000 Base-T Ethernet ports. One Ethernet port connects to the switch access port and the other connects to the Cisco Unified IP Phone.

Cisco Unified IP Phone Expansion Module 7914

The Cisco Unified IP Phone Expansion Module 7914 is for administrative assistants and others who need to determine the status of a number of lines beyond the current line capability of the phone.

The Cisco Unified IP Phone Expansion Module 7914 extends the capability of the Cisco Unified IP Phone 7960G, 7961G, 7961G-GE, 7970G, or 7971G-GE with additional buttons and an LCD. The Cisco Unified IP Phone Expansion Module 7914 provides 14 buttons per module, and the Cisco Unified IP Phones 796xG and 797xG can support up to two Cisco Unified IP Phone Expansion Modules. If the IP phone uses Cisco inline power or IEEE802.3af PoE, then the Cisco Unified IP Phone Expansion Module 7914 requires the use of an external power adaptor (CP-PWR-CUBE-3).

Software-Based Endpoints

Software-based endpoints include Cisco Unified Personal Communicator and Cisco IP Communicator. A software-based endpoint is an application installed on a client PC, and it registers with (and is controlled by) Unified CM.

Cisco Unified Personal Communicator

Cisco Unified Personal Communicator is a software application based on Microsoft Windows or Macintosh, and it integrates a wide variety of communications applications and services into a single desktop application to help people communicate effectively. It lets users access a variety of powerful communications tools, including voice, video, call management, presence, and web conferencing. The integrated applications include Cisco Unified Communications Manager (Unified CM), Cisco Unified Presence, Cisco Unity Connection, Cisco Unified MeetingPlace Express, and the Lightweight Directory Access Protocol (LDAP) version 3 (v3) server. This section summarizes the following design considerations that apply when using Cisco Unified Personal Communicator with Unified CM:

- [Maximum Cisco Unified Personal Communicator Configuration Limits, page 20-11](#)
- [Codec Selection, page 20-11](#)
- [Call Admission Control, page 20-11](#)

For more information on Cisco Unified Personal Communicator, see the chapter on [Cisco Unified Presence, page 21-1](#).

Maximum Cisco Unified Personal Communicator Configuration Limits

Regardless of the device limits allowed per server, there are maximum limits on the number CTI devices you can configure in Unified CM. The CTI device limits as they apply to Cisco Unified Personal Communicator are as follows:

- Maximum of 800 Cisco Unified Personal Communicators per Cisco Media Convergence Server (MCS) 7825 or 7835; maximum of 3,200 Cisco Unified Personal Communicators per cluster of MCS 7825 or 7835 servers.
- Maximum of 2,500 Cisco Unified Personal Communicators per Cisco Media Convergence Server (MCS) 7845; maximum of 10,000 Cisco Unified Personal Communicators per cluster of MCS 7845 servers.

The following assumptions apply to the preceding maximum Cisco Unified Personal Communicator limits:

- Each Cisco Unified Personal Communicator is processing an estimated six or fewer busy hour call attempts (BHCA).
- No other CTI applications requiring CTI devices are configured in the Unified CM cluster.

Codec Selection

Audio codec support on Cisco Unified Personal Communicator includes G.711 and G.729a, and its video codec support includes H.263 and H.264. The codec selection can be done by configuring the region in which Cisco Unified Personal Communicator is located. Cisco recommends G.729a low-bandwidth codec configurations for calls across the WAN.

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 Unified Personal Communicator uses the locations mechanism configured in Unified CM for centralized call processing deployments, or it uses the RSVP mechanism in non-centralized call processing deployments. For more information on call admission control with Unified CM locations, see the chapter on [Call Admission Control, page 9-1](#).

Call admission control based on locations or RSVP is effective in managing call bandwidth as long as the Cisco Unified Personal Communicator is mobile within only a single Unified CM location. However, call admission control becomes problematic if the Cisco Unified Personal Communicator is moved between Unified CM locations. For more information on mobility, refer to [Device Mobility and Unified CM, page 20-18](#).

Cisco IP Communicator

Cisco IP Communicator is a Microsoft Windows-based application that endows computers with the functionality of IP phones. This application enables high-quality voice calls on the road, in the office, or from wherever users can access the corporate network. It is an ideal solution for remote users and

telecommuters. Cisco IP Communicator is easy to deploy and features some of the latest technology and advancements available with IP communications today. This section summarizes the following design considerations that apply when using Cisco IP Communicator with Unified CM:

- [Maximum IP Communicator Configuration Limits, page 20-12](#)
- [Codec Selection, page 20-12](#)
- [Call Admission Control, page 20-13](#)

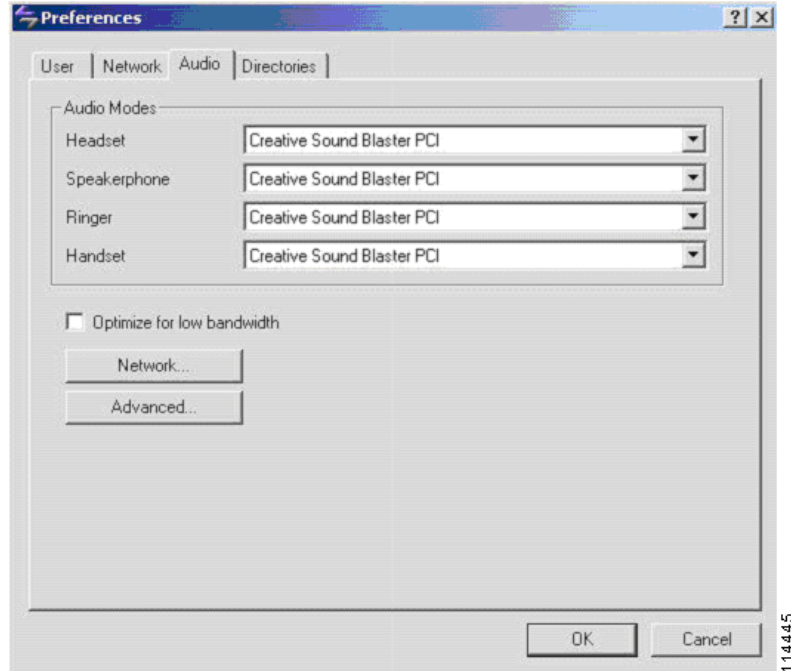
Maximum IP Communicator Configuration Limits

Because Cisco IP Communicator is a standalone device that supports both SCCP and SIP, 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 [Deployment Models, page 2-1](#), for details.

The end-user experience for a supported feature is the same whether using SCCP or SIP call control signaling. There are a few features that are not supported with SIP that are supported with SCCP. For example, Cisco IP Communicator using SCCP is compatible with the Cisco Unified Video Advantage video-enabled endpoint for making video calls, whereas SIP has no video support. In addition, Cisco IP Communicator using SCCP supports Tone on Hold, whereas SIP has no Tone on Hold support. For a complete list of supported features, see the [Endpoint Features Summary, page 20-37](#).

Codec Selection

The 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. The 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 20-1](#)). 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. Cisco IP Communicator can make video calls when it is used together with Cisco Unified Video Advantage. For more information on the video codec support of Cisco Unified Video Advantage, see [Video Telephony Endpoints, page 20-34](#).

Figure 20-1 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, the Cisco IP Communicator uses the *locations* mechanism configured in Unified CM for centralized call processing deployments or it uses the RSVP mechanism in non-centralized call processing deployments. Refer to the chapter on [Call Admission Control, page 9-1](#), for more information on call admission control with Unified CM locations.

Locations and RSVP-based call admission control are effective in managing call bandwidth as long as the Cisco IP Communicator is mobile within only a single Unified CM location. However, call admission control becomes problematic if the Cisco IP Communicator is moved between Unified CM locations. For more information on mobility, refer to [Device Mobility and Unified CM, page 20-18](#).

Wireless Endpoints

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 an area where traditional wired phones are undesirable or problematic. (Refer to [Wireless LAN Infrastructure, page 3-63](#), for more information about wireless network design.)

Cisco offers the following two Voice over WLAN (VoWLAN) IP phones:

- Cisco Unified Wireless IP Phone 7920
- Cisco Unified Wireless IP Phone 7921G

Both are hardware-based phones with built-in radio antenna. The Cisco Unified Wireless IP Phone 7920 enables 802.11b wireless LAN connectivity to the network, while the Cisco Unified Wireless IP Phone 7921G enables 802.11b, 802.11g, or 802.11a connectivity to the network. These phones register with Unified CM using Skinny Client Control Protocol (SCCP), just like the hardware-based phones and Cisco IP Communicator. For more information about these phones, refer to the appropriate phone documentation available at

<http://www.cisco.com>

Site Survey

Before deploying the Cisco Unified Wireless IP Phones, 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 Unified Wireless IP Phones (accessed via **Menu > Network Config > Site Survey** on the 7920 and via **Settings > Status > Site Survey** on the 7921G) 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 Unified Wireless IP Phone 7920 and 7921G 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 Unified Wireless IP Phone must first use one of the following authentication methods to associate and communicate with the AP:

- Extensible Authentication Protocol-Flexible Authentication via Secure Tunneling (EAP-FAST)

This method allows the Cisco Unified Wireless IP Phone to be authenticated to the AP via 802.1X with a user name and password once a secure authenticated tunnel is established between the client and an EAP-compliant Remote Authentication, Authorization, and Accounting server via Protected Access Credential (PAC). Upon authentication, traffic to and from the wireless device is encrypted using TKIP or WEP. Using the 802.1X authentication method and the PAC authenticated tunnel exchange 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.

- Wi-Fi Protected Access (WPA)

This method allows the Cisco Unified Wireless IP Phone to be authenticated to the AP via 802.1X with a user name and password. Upon authentication, traffic to and from the wireless device is encrypted using Temporal Key Integrity Protocol (TKIP). Using the 802.1X authentication 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.

- Wi-Fi Protected Access 2 (WPA2)

This method is the 802.11i enhanced version of WPA, which uses Advanced Encryption Standards (AES) rather than TKIP for encrypting traffic to and from the wireless device. This method is supported only on the Cisco Unified Wireless IP Phone 7921G.

- Wi-Fi Protected Access Pre-Shared Key (WPA-PSK)

This method allows the Cisco Unified Wireless IP Phone to be authenticated to the AP via the configuration of a shared key on the Cisco Unified Wireless IP Phone and the AP. Upon authentication, traffic to and from the wireless device is encrypted using TKIP. This method of authentication is not recommended for enterprise deployments.

- Wi-Fi Protected Access 2 Pre-Shared Key (WPA2-PSK)

This method is the 802.11i enhanced version of WPA-PSK, which uses AES rather than TKIP for encrypting traffic to and from the wireless device. This method is supported only on the Cisco Unified Wireless IP Phone 7921G.

- Cisco Centralized Key Management (Cisco CKM)

This method allows the Cisco Unified Wireless IP Phone to be authenticated to the AP via 802.1x with a user name and password. Upon authentication, traffic to and from the wireless device is encrypted using either WEP 128 or TKIP. The 802.1X authentication method requires an EAP-compliant RADIUS authentication server such as the Cisco ACS, which provides access to a user database for the initial authentication request. Subsequent authentication requests are validated via the wireless domain service (WDS) at the AP, which shortens re-authentication times and ensures fast, secure roaming.

- Cisco LEAP

This method allows the Cisco Unified Wireless IP Phone and AP to be authenticated mutually based on a user name and password. Upon authentication, the dynamic key is generated and used for encrypting traffic between the Cisco Unified Wireless IP Phone 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.

- Shared Key

This method involves the configuration of static 10 (40-bit) or 26 (128-bit) character keys on the Cisco Unified Wireless IP Phone 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 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

The capacity of each AP depends on a number of factors including the AP radio types, associated client radio types, enabled data rates, and channel utilization.

Given an 802.11b-only AP with 802.11b clients and a data rate of 11 Mbps, the AP can support a maximum of seven active G.711 voice streams or eight G.729 streams. 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.

When using 802.11a at a data rate of 54 Mbps, the maximum number of active voice streams increases to between 14 and 18 per AP.

For 802.11g environments with a data rate of 54 Mbps, in theory the maximum number of active voice streams also increases to between 14 and 18 per AP. However, because most 802.11g environments are mixed and include 802.11b clients (and therefore 11 Mbps data rates) as well as 802.11g clients, capacity is typically significantly lower with a maximum of 8 to 12 active voice streams per AP.

Regardless of 802.11 radio type, call capacity can be diminished significantly if there is heavy channel utilization due to data traffic.

For additional information about call capacity, radio types, and data rates, refer to the VoWLAN design recommendations in the latest version of the *Enterprise Mobility Design Guide*, available at

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

**Note**

A call between two phones associated to the same AP counts as two active voice streams.

Based on these active call capacity limits, and using Erlang ratios, you can calculate the number of Cisco Unified Wireless IP Phones that each AP can support. For example, given an 802.11b AP with 802.11b clients and a typical user-to-call capacity ratio of 3:1, a single AP can support 21 to 24 Cisco Unified Wireless IP Phones, depending on whether the codec used is G.711 or G.729. As another example, given an 802.11a AP with an 802.11a client at a data rate of 54 Mbps and user-to-call capacity ratio of 3:1, a single AP can support 42 to 54 Cisco Unified Wireless IP Phone 7921Gs. However, these numbers do not take into consideration the possibility that other Cisco Unified Wireless IP Phones could roam to the AP, so a lower number of phones per AP might be more realistic.

These capacities are based on voice activity detection (VAD) being 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 Unified CM. (It is referred to as Silence Suppression in Unified CM.) Thus, if VAD is enabled for the Cisco Unified Wireless IP Phone 7920, then it will be enabled for all devices in the Unified CM 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 3-49](#).

Phone Configuration

For information on configuring the Cisco Unified Wireless IP Phones, refer to the *Cisco Unified Wireless IP Phone 7920 Administration Guide* and the *Cisco Unified Wireless IP Phone 7921G Administration Guide*, available at

<http://www.cisco.com>

Roaming

Cisco Unified Wireless IP Phones are 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 Unified Wireless IP Phones, the phone roams to a new AP for the first time.

- If the Cisco Unified Wireless IP Phone 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 Unified Wireless IP Phone 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 Unified Wireless IP Phone is changed, the phone must roam to re-associate with an AP.

In trying to determine eligible AP roam targets for 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 Basic 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.
- **Wi-Fi Multimedia Traffic Specification (WMM TSPEC)**
WMM TSPEC is an 802.11e QoS mechanism that assists wireless IP phone roaming by enabling the phone to request bandwidth and priority treatment via a TSPEC indication while roaming to determine if the new AP can handle the phones bandwidth needs based on current utilization. TSPEC is supported only by the Cisco Unified Wireless IP Phone 7921G.

When devices roam at Layer 3, they move from one AP to another AP across native VLAN boundaries. When the WLAN network infrastructure consists of autonomous APs, the Cisco Catalyst 6500 Series Wireless LAN Services Module (WLSM) allows the Cisco Unified Wireless IP Phone to keep its IP address and roam at Layer 3 while still maintaining an active call. Seamless Layer 3 roaming occurs only when the client is roaming within the same mobility group. For details about the Cisco WLSM and Layer 3 roaming, refer to the Cisco WLSM product documentation available at

<http://www.cisco.com>

Seamless Layer 3 roaming for clients across a lightweight access point infrastructure is accomplished by WLAN controllers that use dynamic interface tunneling. Cisco Unified Wireless IP Phones that roam across WLAN controllers and VLANs can keep their IP address when using the same SSID and therefore can maintain an active call.

With stronger authentication methods such as WPA and EAP, the number of information exchanges increases and causes more delay during roaming. To avoid additional delays, use Cisco Centralized Key Management (Cisco CKM) to manage authentication. With Cisco CKM, whether at Layer 2 or Layer 3, roaming can occur without any perceptible delay. Cisco CKM also takes some of the load off the Access Control Server (ACS) by reducing the number of authentication requests that must be sent to the ACS.

**Note**

In dual-band WLANs (those with both 2.4 GHz and 5 GHz bands), it is possible to roam between 802.11b/g and 802.11a with the same SSID, provided the client is capable of supporting both bands. However, this can cause gaps in the voice path. In order to avoid these gaps, use only one band for voice communications.

AP Call Admission Control

Call admission control mechanisms in Unified CM or in a gatekeeper can control WAN bandwidth utilization and provide QoS for existing calls, but both mechanisms are applied 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 as Cisco Unified Wireless IP Phones, there must also be a call admission control mechanism at the AP level because these wireless devices may roam from one AP to another.

Cisco APs and wireless voice clients have two mechanisms that are used for call admission control:

- QoS Basic Service Set (QBSS)

QBSS 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 is not a true call admission control mechanism because it does not guarantee that calls will retain proper QoS or that there is enough bandwidth to handle the call. When a Cisco Unified Wireless 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 oversubscription of the available bandwidth on that AP.

- Wi-Fi Multimedia Traffic Specification (WMM TSPEC)

WMM TSPEC is the QoS mechanism that enables WLAN clients to provide an indication of their bandwidth and QoS requirements so that APs can react to those requirements. When a client is preparing to make a call, it sends an Add Traffic Stream (ADDTS) message to the AP with which it is associated, indicating TSPEC. The AP can then accept or reject the ADDTS request based on whether bandwidth and priority treatment are available. If the call is rejected, the phone will receive a Network Busy message. When roaming, mid-call clients supporting TSPEC will send a ADDTS message to the new AP as part of the association process to ensure that there is available bandwidth for priority treatment. If there is not enough bandwidth, the roam can be load-balanced to a neighboring AP if one is available.

Cisco Unified Wireless IP Phone 7920s support only QBSS, so this is the only mechanism that can be used for call admission control with these devices. However, the Cisco Unified Wireless IP Phone 7921Gs support both QBSS and TSPEC. (TSPEC takes precedence over QBSS.) Therefore call admission control with the Cisco Unified Wireless IP Phone 7921G, when using TSPEC, is more accurate and eliminates the possibility of priority bandwidth oversubscription on the AP.

Device Mobility and Unified CM

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 Unified CM locations-based call admission control

When a wireless IP phone roams from one location to another, there is currently no dynamic mechanism in Unified CM 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 Unified CM 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 Unified CM 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 Unified CM 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.

The Cisco Unified IP Conference Station 7936 has an external speaker and three built-in microphones. The Cisco Unified IP Conference Station 7936 requires Cisco Unified CM Release 3.3(3) SR3 or later. The Cisco Unified 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.

Video Endpoints

Cisco Unified CM 5.x supports the following types of video-enabled endpoints:

- Cisco Unified Video Advantage associated with a Cisco Unified IP Phone 7911, 7940, 7941, 7960, 7961, 7970, or 7971, or with Cisco IP Communicator, running Skinny Client Control Protocol (SCCP)
- Cisco IP Video Phone 7985
- Tandberg 2000 MXP, 1500 MXP, 1000 MXP, 770 MXP, 550 MXP, T-1000, and T-550 models running SCCP
- Sony PCS-1, PCS-TL30, and PCS-TL50 models running SCCP
- H.323 and SIP clients (Polycom, Sony, PictureTel, EyeBeam, Tandberg, VCON, VTEL, Microsoft NetMeeting, and others)

SCCP Video Endpoints

SCCP video endpoints register directly with Unified CM and download their configurations via Trivial File Transfer Protocol (TFTP). They support many features and supplementary services, including hold, transfer, conference, park, pickup and group pickup, music on hold, shared line appearances, mappable softkeys, call forwarding (busy, no answer, and unconditional), and much more.

Cisco Unified Video Advantage

Cisco Unified Video Advantage is a Windows-based application and USB camera that you can install on a Windows 2000 or Windows XP personal computer. When the PC is physically connected to the PC port on a Cisco Unified IP Phone 7911, 7940, 7941, 7960, 7961, 7970, or 7971 running the Skinny Client Control Protocol, the Cisco Unified Video Advantage application "associates" with the phone, thus enabling users to operate their phones as they always have but now with the added benefit of video. In Cisco Unified Video Advantage Release 2.0, this association can also be to Cisco IP Communicator running SCCP on the same PC.

The system administrator can control which IP Phones allow this association to take place by toggling the **Video Capabilities: Enabled/Disabled** setting on the IP Phone configuration page in Unified CM Administration. When this feature is enabled, an icon representing a camera appears in the bottom right-hand corner of the IP Phone display. By default, Cisco Unified Video Advantage is disabled. You can also use the Bulk Administration Tool to modify this setting on many phones at once. Note that the **PC Port: Enabled/Disabled** setting must also be enabled for Cisco Unified Video Advantage to work with a hardware IP Phone; however, the **PC Access to Voice VLAN** setting does not have to be enabled.

To achieve the association with a hardware IP Phone, Cisco Unified Video Advantage installs a Cisco Discovery Protocol (CDP) driver onto the Ethernet interface of the PC. CDP enables the PC and the hardware IP Phone to discover each other automatically, which means that the user does not have to configure anything on the PC or the hardware IP Phone in order for Cisco Unified Video Advantage to work. The user can, therefore, plug the PC into any hardware IP Phone that is video-enabled and automatically associate with it. (See [Figure 20-2](#).)

Cisco Unified Video Advantage 2.0 does not rely on CDP to discover the presence of Cisco IP Communicator running SCCP on the same PC. Instead, it listens for a private Windows message sent from the Cisco IP Communicator process. If Cisco IP Communicator is discovered, the association process works exactly as it does for a hardware IP phone. (See [Figure 20-3](#).)

**Note**

When you install Cisco Unified Video Advantage, the CDP packet drivers install on all Ethernet interfaces of the PC. If you add a new network interface card (NIC) or replace an old NIC with a new one, you must reinstall Cisco Unified Video Advantage so that the CDP drivers also install on the new NIC.

Figure 20-2 Cisco Unified Video Advantage Operational Overview

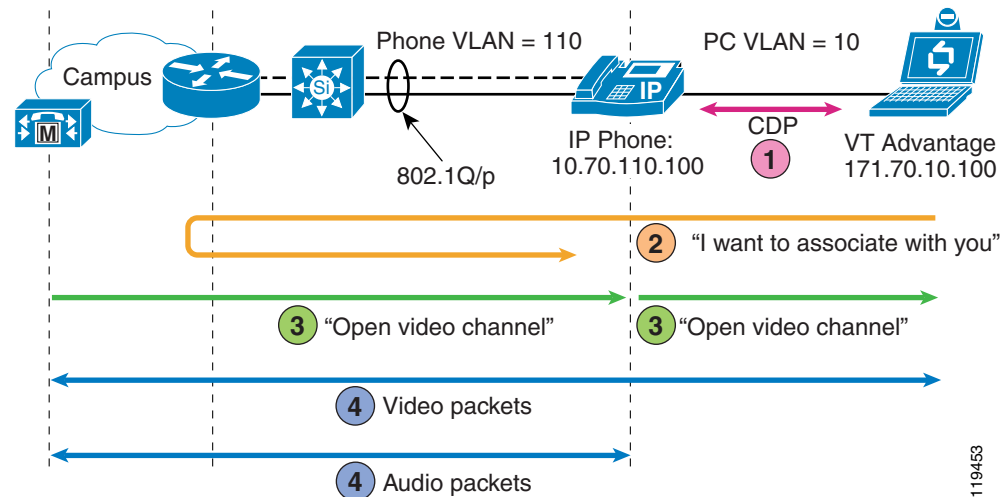


Figure 20-2 illustrates the following events:

1. The IP Phone and PC exchange Cisco Discovery Protocol (CDP) messages. The phone begins listening for PC association packets on TCP port 4224 from the IP address of its CDP neighbor.
2. The PC initiates association messages to the phone over TCP/IP. Association packets are routed up to the Layer-3 boundary between VLANs. Firewalls and/or access control lists (ACLs) must permit TCP port 4224.
3. The phone acts as an SCCP proxy between Cisco Unified Video Advantage and Unified CM. Unified CM tells the phone to open video channels for the call, and the phone proxies those messages to the PC.
4. The phone sends/receives audio, and the PC sends/receives video. Both audio and video traffic are marked DSCP AF41. Video traffic uses UDP port 5445.

Figure 20-3 Cisco IP Communicator Associating with Cisco Unified Video Advantage

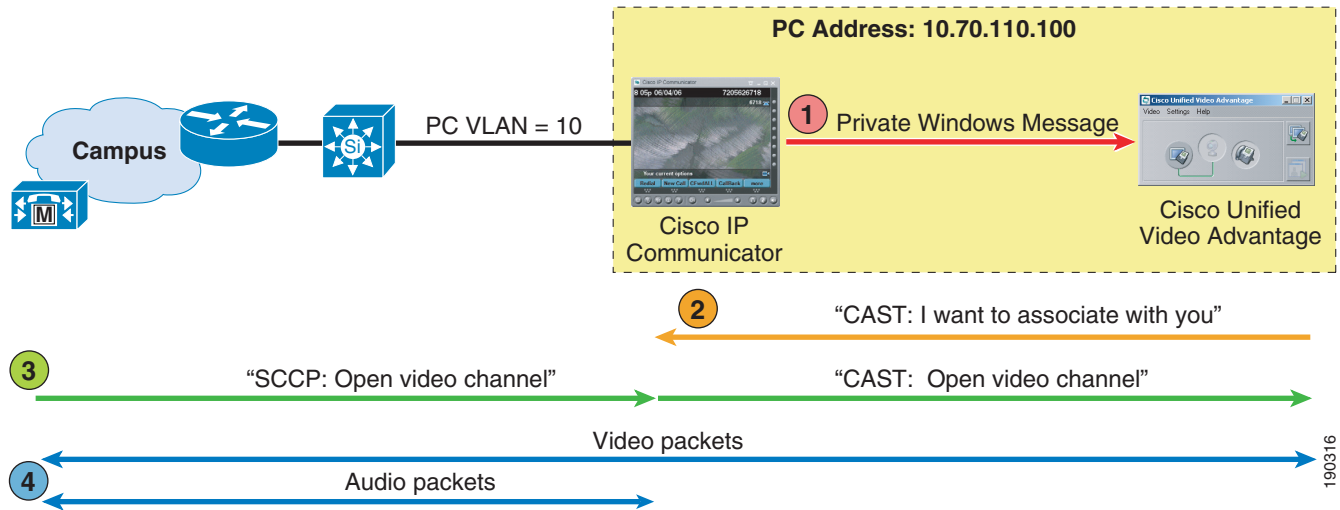


Figure 20-3 illustrates the following events:

1. Cisco IP Communicator sends a private Windows message to Cisco Unified Video Advantage. The message includes the IP address of Cisco IP Communicator and the port number for CAST messages.
2. Cisco Unified Video Advantage initiates CAST messages to Cisco IP Communicator over TCP/IP. CAST messages do not leave the PC because it is a connected address.
3. Cisco IP Communicator acts as an SCCP proxy between Cisco Unified Video Advantage and Unified CM. Unified CM tells IP Communicator to open video channels for the call, and IP Communicator proxies those messages to Cisco Unified Video Advantage via CAST protocol.
4. Cisco IP Communicator sends/receives audio, and Cisco Unified Video Advantage sends/receives video. Both audio and video traffic are marked DSCP AF41. Video traffic uses UDP port 5445.

When a call is made using Cisco Unified Video Advantage, the audio is handled by the IP Phone while the video is handled by the PC. There is no synchronization mechanism between the two devices, so QoS is essential to minimize jitter, latency, fragmented packets, and out-of-order packets.

When using a hardware IP Phone, the phone resides in the voice VLAN while the PC resides in the data VLAN, which means that there must be a Layer-3 routing path between the voice and data VLANs in order for the association to occur. If there are access control lists (ACLs) or firewalls between these VLANs, they must be configured to permit the association protocol (which uses TCP port 4224 in both directions) to pass. When using Cisco IP Communicator, this communication happens internal to the PC, and there are no Layer-3 boundaries to cross.

Cisco Unified Video Advantage supports Differentiated Services Code Point (DSCP) traffic classifications. Unified CM specifies the DSCP value in the SCCP messages it sends to the phone. When the IP Phone makes an audio-only call, it marks its SCCP control traffic as DSCP CS3 and its audio RTP media traffic as DSCP EF. However, when the IP Phone makes a video call, it marks its SCCP control traffic as DSCP CS3 and its audio RTP media traffic as DSCP AF41, and the Cisco Unified Video Advantage application marks its video RTP media traffic as DSCP AF41 as well. Both the IP Phone and the Cisco Unified Video Advantage application mark their "association" protocol messages as DSCP CS3 because it is considered to be signaling traffic and is grouped with all other signaling traffic such as SCCP.

**Note**

Cisco Unified CM Release 4.0 added security features to the Cisco Unified IP Phone 7970 and 7971 to enable it to use Transport Layer Security (TLS) and Secure RTP (SRTP) to authenticate and encrypt signaling and audio media traffic. The association protocol does not use this authentication or encryption, nor are the video RTP media streams encrypted. However, the SCCP signaling and the audio RTP media streams are encrypted if they are so configured.

**Note**

Do not set the voice VLAN equal to the data VLAN because doing so can cause issues with connectivity.

Cisco Unified Video Advantage, like any other application that runs on a PC, does have an impact on system performance, which you should take into consideration. Cisco Unified Video Advantage 1.0 supports two types of video codecs: H.263 and the Cisco VT Camera Wideband Video Codec. Cisco Unified Video Advantage 2.0 also supports two types of codecs: H.263 and H.264. The Cisco VT Camera Wideband Video Codec places the least demand on the PC but the most demand on the network. H.263 places a lower demand on the network but a higher demand on the PC. Finally, H.264 places the least demand on the network but the highest demand on the PC. Therefore, if your network has plenty of available bandwidth, you can use the Cisco VT Camera Wideband Video Codec and save on PC CPU and memory resources.

The H.263 and H.264 codec supports a range of speeds up to 1.5 Mbps. In summary, customers must balance PC performance with network utilization when deploying Cisco Unified Video Advantage.

System Requirements

For detailed PC requirements, refer to the *Cisco Unified Video Advantage Data Sheet*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps5662/products_data_sheet0900aecd8044de04.html

Cisco IP Video Phone 7985G

The Cisco IP Video Phone 7985G is a personal desktop video phone. Unlike Cisco Unified Video Advantage, which is an application that runs on a PC, the Cisco IP Video Phone 7985G is a standalone phone with integrated video features. The phone has an 8.4 inch color LCD screen and an embedded video camera for making video calls. The phone supports up to eight line appearances, has two 10/100 BaseT Ethernet connections, and has buttons for Directories, Messages, Settings, and Services. Like other Cisco Unified IP Phones, the Cisco IP Video Phone 7985G uses CDP to learn VLAN and CoS information from the attached switch to use in 802.1p/q markings.

Codecs Supported by Cisco Unified Video Advantage and Cisco IP Video Phone 7985G

Table 20-4 lists the codecs supported by Cisco Unified Video Advantage and Cisco IP Video Phone 7985G

Table 20-4 **Codecs Supported by Cisco Unified Video Advantage and Cisco IP Video Phone 7985G**

Codec or Feature	Cisco Unified Video Advantage	Cisco IP Video Phone 7985G
H.264	Yes in Release 2.0	Yes
H.263	Yes	Yes

Table 20-4 **Codecs Supported by Cisco Unified Video Advantage and Cisco IP Video Phone 7985G (continued)**

Codec or Feature	Cisco Unified Video Advantage	Cisco IP Video Phone 7985G
H.261	No	Yes
G.711	Yes	Yes
G.722	No	Yes
G.722.1	No	No
G.723.1	No	No
G.728	No	No
G.729	Yes	Yes
Cisco Wideband	Yes in Release 1.0	No
Maximum Bandwidth	7 Mbps for Release 1.0 and 1.5 Mbps for Release 2.0	768 kbps
Video Resolution	CIF, QCIF	NTSC: 4SIF, SIF PAL: 4CIF, QCIF, SQCIF

Third-Party SCCP Video Endpoints

Two manufacturers of video endpoints, Sony and Tandberg, currently have products that support the Cisco Skinny Client Control Protocol (SCCP). SCCP on both the Sony and Tandberg endpoints is modeled after SCCP on the Cisco Unified IP Phone 7940. Most features found on the Cisco Unified IP Phone 7940 user interface are also supported on the Sony endpoints as well as the Tandberg endpoints, including multiple line appearances, softkeys, and buttons for Directories, Messages, Settings, Services, and so forth. The Sony and Tandberg endpoints also support the Option 150 field in DHCP to discover the IP address of the TFTP server, and they download their configurations from the TFTP server. However, software upgrades of the Sony and Tandberg endpoints are not done via TFTP. Instead, the customer must manually upgrade each endpoint using tools provided by the vendor. (Tandberg uses an FTP method, while Sony uses FTP or a physical memory stick.) The Sony and Tandberg endpoints register with up to three Unified CM servers and will fail-over to its secondary or tertiary servers if its primary server becomes unreachable.

While the Sony and Tandberg endpoints support softkey functionality similar to that of the Cisco Unified IP Phone 7940 and 7960, the exact feature support differs between vendor and model. Check the manufacturers' documentation for supported features. Features that are currently known to be missing on some platforms include:

- Messages button
- Directories (placed calls, received calls, missed calls, and corporate directory)
- Settings and Services buttons
- Some XML services (such as Extension Mobility and Berbee's InformaCast)

Because the Sony and Tandberg endpoints use SCCP, dialing a video call from an endpoint is similar to dialing an audio call from a Cisco Unified IP Phone. If users are familiar with Cisco Unified IP Phones, they should also find the Sony and Tandberg endpoints very intuitive to use. The main difference in the user interface is that the Sony and Tandberg endpoints do not have a button keypad or a handset like those on a phone. Instead, the remote control is used to access features and to dial numbers on the Sony and Tandberg endpoints.

**Note**

Sony and Tandberg endpoints do not support the Cisco Discovery Protocol (CDP) or IEEE 802.Q/p. Therefore, you must manually configure the VLAN ID and Quality of Service trust boundary on the ethernet switch to which they are attached. (For more details, see [Network Infrastructure, page 3-1.](#))

Codecs Supported by Sony and Tandberg SCCP Endpoints

Codec support for the third-party SCCP endpoints varies by vendor, model, and software version. Check the vendors' product documentation for the supported codecs.

Third-Party SIP IP Phones

Third-party phones have specific local features that are independent of the call control signaling protocol, such as features access buttons (fixed or variable). Basic SIP RFC support allows for certain desktop features to be the same as Cisco Unified IP Phones and also allows for interoperability of certain features. However, these third-party SIP phones do not provide the full feature functionality of Cisco Unified IP Phones.

Cisco is working with key third-party vendors who are part of the Cisco Technology Development Partner Program and who are developing solutions that leverage the new Unified CM and Cisco Unified Communications Manager Express (Unified CME) SIP capabilities. Vendors include IPCelerate (unified client for education space), RIM (Blackberry 7270 wireless LAN handsets) and IP blue (Softphone). Cisco has also worked with third-party vendor Grandstream to test their Grandstream GXP 2000 to ensure interoperability.

Cisco is also participating in an independent third party testing and interoperability verification process being offered by tekVizion. This independent service provided by tekVizion has been established to enable third-party vendors to test and verify the interoperability of their endpoints with Unified CM and Unified CME.

For more information on Cisco's line-side SIP interoperability and third-party verification, visit <http://www.cisco.com>.

QoS Recommendations

This sections 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/designzone>

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 SUP3, 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 IP 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 20-25](#).

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, 7906, 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, 7906, 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, 3560, 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)# permit tcp Voice_IP_Subnet Subnet_Mask any range 2000 2002 dscp
Af31
CAT2970(config-ext-nacl)# permit udp Voice_IP_Subnet Subnet_Mask any eq 5060 dscp cs3
CAT2970(config-ext-nacl)# permit tcp Voice_IP_Subnet Subnet_Mask any range 5060 5061 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 26 46 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 policed-dscp-transmit
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 policed-dscp-transmit
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 policed-dscp-transmit
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 26 46 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)# permit tcp Voice_IP_Subnet Subnet_Mask any range 2000 2002 dscp
cs3
Af31
CAT4500(config-ext-nacl)# permit udp Voice_IP_Subnet Subnet_Mask any eq 5060 dscp cs3
CAT4500(config-ext-nacl)# permit tcp Voice_IP_Subnet Subnet_Mask any range 5060 5061 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, 26, 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 24 aggregate VVLAN-CALL-SIGNALING tcp
Voice_IP_Subnet Wildcard_bits any range 5060 5061
CAT6500> (enable) set qos acl ip IPPHONE-PC dscp 24 aggregate VVLAN-CALL-SIGNALING udp
Voice_IP_Subnet Wildcard_bits any eq 5060
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)

```

**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.

Software-Based Endpoints

Cisco Unified Personal Communicator and Cisco IP Communicator with Cisco Unified Video Advantage are both voice and video capable, which presents two challenges when using the ACL and policy map for packet classification and DSCP re-marking. First, Cisco Unified Personal Communicator uses the same IP address and UDP port range to source voice and video streams. The ACL that is based on IP address and port number is not granular enough to differentiate a voice call from a video call in order to apply appropriate DSCP re-marking. Second, Cisco IP Communicator uses the same IP address and UDP port range to source its voice packets. Similarly, the ACL is not granular enough to differentiate the voice stream of an audio-only call from the voice stream of a video call. Therefore, using the ACL and policy-map for packet classification and DSCP re-marking is not a feasible QoS solution for software-based endpoints.

Because both Cisco Unified Personal Communicator and Cisco IP Communicator with Cisco Unified Video Advantage mark their signaling and media packets correctly as they ingress the network, Cisco recommends configuring the policy map to trust the DSCP marking of incoming traffic and apply traffic policing and rate limiting. The following sections list the configuration commands for Cisco Unified Personal Communicator and Cisco IP Communicator on the most commonly deployed Cisco Catalyst switches.

**Note**

The Cisco Catalyst 2950 Series switches are not recommended for software-based endpoint QoS implementations because 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, 3560, or 3750

```
CAT2970 (config)#mls qos
CAT2970 (config)#mls qos map policed-dscp 0 24 26 46 to 8
CAT2970 (config)#
CAT2970 (config)#class-map match-all SOFTWARE-BASED-ENDPOINT-VOICE
CAT2970 (config-cmap)# match access-group name SOFTWARE-BASED-ENDPOINT-VOICE
CAT2970 (config-cmap)#class-map match-all SOFTWARE-BASED-ENDPOINT-VIDEO
CAT2970 (config-cmap)# match access-group name SOFTWARE-BASED-ENDPOINT-VIDEO
CAT2970 (config-cmap)#class-map match-all SOFTWARE-BASED-ENDPOINT-SIGNALING
CAT2970 (config-cmap)# match access-group name SOFTWARE-BASED-ENDPOINT-SIGNALING
CAT2970 (config-cmap)#exit
CAT2970 (config)#
CAT2970 (config)#policy-map SOFTWARE-BASED-ENDPOINT
CAT2970 (config-pmap-c)#class SOFTWARE-BASED-ENDPOINT-VOICE
CAT2970 (config-pmap-c)# police 128000 8000 exceed-action drop
CAT2970 (config-pmap-c)#class SOFTWARE-BASED-ENDPOINT-VIDEO
CAT2970 (config-pmap-c)# police 5000000 8000 exceed-action policed-dscp-transmit
CAT2970 (config-pmap-c)#class SOFTWARE-BASED-ENDPOINT-SIGNALING
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 dvlan_id
CAT2970 (config-if)# switchport mode access
CAT2970 (config-if)# service-policy input SOFTWARE-BASED-ENDPOINT
CAT2970 (config-if)# exit
CAT2970 (config)#ip access-list extended SOFTWARE-BASED-ENDPOINT-SIGNALING
CAT2970 (config-ext-nacl)#permit ip PC_Subnet Source wildcard bits any dscp 24
CAT2970 (config-ext-nacl)#exit
CAT2970 (config)#ip access-list extended SOFTWARE-BASED-ENDPOINT-VIDEO
CAT2970 (config-ext-nacl)#permit ip PC_Subnet Source wildcard bits any dscp 34
CAT2970 (config-ext-nacl)#exit
CAT2970 (config)#ip access-list extended SOFTWARE-BASED-ENDPOINT-VOICE
CAT2970 (config-ext-nacl)# permit ip PC_Subnet Source wildcard bits any dscp 46
CAT2970 (config-ext-nacl)#exit
CAT2970 (config)#exit

```

Cisco 3550

```

3550 (config)#class-map match-all SOFTWARE-BASED-ENDPOINT-VOICE
3550 (config-cmap)#match access-group name SOFTWARE-BASED-ENDPOINT-VOICE
3550 (config-cmap)#class-map match-all SOFTWARE-BASED-ENDPOINT-VIDEO
3550 (config-cmap)# match access-group name SOFTWARE-BASED-ENDPOINT-VIDEO
3550 (config-cmap)#class-map match-all SOFTWARE-BASED-ENDPOINT-SIGNALING
3550 (config-cmap)# match access-group name SOFTWARE-BASED-ENDPOINT-SIGNALING
3550 (config-cmap)#exit
3550 (config)#
3550 (config)#policy-map SOFTWARE-BASED-ENDPOINT
3550 (config-pmap)#class SOFTWARE-BASED-ENDPOINT-VOICE
3550 (config-pmap)# police 128000 8000 exceed-action drop
3550 (config-pmap)#class SOFTWARE-BASED-ENDPOINT-VIDEO
3550 (config-pmap)# police 5000000 8000 exceed-action policed-dscp-transmit
3550 (config-pmap)#class SOFTWARE-BASED-ENDPOINT-SIGNALING
3550 (config-pmap)# police 32000 8000 exceed-action policed-dscp-transmit
3550 (config-pmap)#class class-default
3550 (config-pmap)# set ip dscp 0
3550 (config-pmap)# police 5000000 8000 exceed-action policed-dscp-transmit
3550 (config-pmap)# exit
3550 (config)#exit
3550 (config)#
3550 (config)#interface FastEthernet interface_id
3550 (config-if)# switchport access vlan dvlan_id
3550 (config-if)# switchport mode access
3550 (config-if)# service-policy input SOFTWARE-BASED-ENDPOINT
3550 (config-if)# exit
3550 (config)#ip access-list extended SOFTWARE-BASED-ENDPOINT-SIGNALING
3550 (config-ext-nacl)#permit ip PC_Subnet Source wildcard bits any dscp 24
3550 (config-ext-nacl)#exit
3550 (config-if)# ip access-list extended SOFTWARE-BASED-ENDPOINT-VIDEO
3550 (config-ext-nacl)#permit ip PC_Subnet Source wildcard bits any dscp 34
3550 (config-ext-nacl)#exit
3550 (config-if)# ip access-list extended SOFTWARE-BASED-ENDPOINT-VOICE
3550 (config-ext-nacl)# permit ip PC_Subnet Source wildcard bits any dscp 46
3550 (config-ext-nacl)#exit
3550 (config)#exit

```

Cisco 6500

```

CAT6500> (enable) set qos enable
CAT6500> (enable) set qos policed-dscp-map 0, 24, 26, 34, 46:8
CAT6500> (enable)

```

```

CAT6500> (enable) set qos policer aggregate SOFTWARE-BASED-ENDPOINT-VOICE rate 128 burst
8000 drop
CAT6500> (enable) set qos policer aggregate SOFTWARE-BASED-ENDPOINT-VIDEO rate 5000 burst
8000 policed-dscp
CAT6500> (enable) set qos policer aggregate SOFTWARE-BASED-ENDPOINT-SIGNAL rate 32 burst
8000 policed-dscp
CAT6500> (enable) set qos policer aggregate SOFTWARE-BASED-ENDPOINT-DEFAULT rate 5000
burst 8000 policed-dscp
CAT6500> (enable)
CAT6500> (enable) set qos acl ip SOFTWARE-BASED-ENDPOINT trust-dscp aggregate
SOFTWARE-BASED-ENDPOINT-VOICE ip PC_Subnet Source wildcard bits any dscp-field 46
CAT6500> (enable) set qos acl ip SOFTWARE-BASED-ENDPOINT trust-dscp aggregate
SOFTWARE-BASED-ENDPOINT-VIDEO ip PC_Subnet Source wildcard bits any dscp-field 34
CAT6500> (enable) set qos acl ip SOFTWARE-BASED-ENDPOINT trust-dscp aggregate
SOFTWARE-BASED-ENDPOINT-SIGNAL ip PC_Subnet Source wildcard bits any dscp-field 24
CAT6500> (enable) set qos acl ip SOFTWARE-BASED-ENDPOINT dscp 0 aggregate
SOFTWARE-BASED-ENDPOINT-DEFAULT any
CAT6500> (enable) commit qos acl SOFTWARE-BASED-ENDPOINT
CAT6500> (enable) set vlan dvlan_id mod/port
CAT6500> (enable) set port qos mod/port trust untrusted
CAT6500> (enable) set qos acl map SOFTWARE-BASED-ENDPOINT mod/port

```

Cisco Unified Wireless IP Phones

By default, the Cisco Unified Wireless IP Phones marks its SCCP signaling messages using a Per-Hop Behavior (PHB) value of CS3 or a Differentiated Services Code Point (DSCP) value of 24 (this corresponds to a ToS value of 0x60), 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 Unified Wireless IP Phones 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.

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 Unified Wireless IP Phones 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-cmap) #
CAT3550 (config-cmap) #class-map match-all VOICE-SIGNALING
CAT3550 (config-cmap) #match ip dscp 24
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-pmap-c)#policy-map INGRESS-QOS
CAT3550(config-pmap-c)#class INGRESS-VVLAN-VOICE
CAT3550(config-pmap-c)#set ip dscp 46
CAT3550(config-pmap-c)#
CAT3550(config-pmap-c)#class INGRESS-VVLAN-VOICE-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 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 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)
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-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)

```

Video Telephony Endpoints

This section discusses how the following types of endpoint devices classify traffic:

- [Cisco Unified Video Advantage with a Cisco Unified IP Phone](#), page 20-35
- [Cisco IP Video Phone 7985G](#), page 20-35
- [Sony and Tandberg SCCP Endpoints](#), page 20-35
- [H.323 and SIP Video Endpoints](#), page 20-36

Cisco Unified Video Advantage with a Cisco Unified IP Phone

The Cisco Unified Video Advantage application residing on the user's PC supports the classification of video packets using DSCP and, therefore, only at Layer 3. The current best practices for Cisco Unified Communications design recommend that the upstream Ethernet switch to which the phone is attached should be configured to trust the 802.1p CoS from the phone. Because the PC packets are unlikely to have an 802.1Q tag, they are unable to support 802.1p CoS bits. This lack of 802.1p support from the PC leaves the following possible options for providing QoS for Cisco Unified Video Advantage:

Option 1

If your current QoS model extends trust to the IP Phone, then the voice and signaling packets will be correctly marked as they ingress the network. With an additional ACL on the port to match UDP port 5445, the video media channel will also be classified to PHB AF41. Without this ACL, the video media would be classified Best Effort and would incur poor image quality and lip-sync issues. The same ACL could also be used to match the CAST connection between the Cisco Unified Video Advantage PC and the IP Phone, which uses TCP port 4224 (classifying it as CS3), although the benefit of doing so is minimal. The signaling packets from the PC, which is on the data VLAN, are returned over the same high-speed port onto the voice VLAN, therefore they are highly unlikely to encounter any congestion.

Option 2

The *Enterprise QoS Solution Reference Network Design Guide, Version 3.1* (available at <http://www.cisco.com/go/designzone>) presents another method. This alternative method recommends changing the port to trust the DSCP of incoming traffic instead of trusting CoS, and then running the incoming packets through a series of Per-Port/Per-VLAN Access Control Lists that match packets based on their TCP/UDP ports (along with other criteria) and police them to appropriate levels. For instance, Cisco Unified Video Advantage will mark its video packets with DSCP AF41, with the switch port set to trust DSCP. The packet will run through an ACL that matches it based on the fact that it is using UDP port 5445, is marked with DSCP AF41, and is coming in on the data VLAN. This ACL will then be used in a class map or policy map to trust the DSCP and police the traffic to N kbps (where N is the amount of video bandwidth you want to allow per port). Similar ACLs and policers will be present for the voice and signaling packets from the IP Phone in the voice VLAN.

Cisco IP Video Phone 7985G

Like many other Cisco Unified IP Phones, the Cisco IP Video Phone 7985G supports 802.1p/Q tagging for traffic originating from the phone and, because the Cisco IP Video Phone 7985G has a second Ethernet interface for PC access, traffic originating from attached devices as well. The current best practices for Cisco Unified Communications design recommend that the upstream Ethernet switch to which the phone is attached should be configured to trust the 802.1p CoS from the phone. Cisco recommends that trust not be extended to the PC port of the phone and, if the switch supports it, that you configure policers to limit the maximum amount of voice, video, and signaling traffic.

Sony and Tandberg SCCP Endpoints

Sony and Tandberg SCCP endpoints correctly mark their media and signaling packets at Layer 3 using DSCP. They do not, however, support 802.1Q and are therefore unable to classify using 802.1p CoS. If you use the UDP and TCP port-matching option, you would be able to classify the SCCP signaling correctly as CS3 and the video media as AF41; however, you would be unable to tell when a UDP port is being used in a voice-only call and should therefore be classified as EF. In such a case, the call

admission control mechanisms would not be able to account for the bandwidth correctly. To avoid this situation, there is only one viable option for how to classify and trust traffic from a Sony or Tandberg endpoint:

Option 1

Trust DSCP on the port used by the Sony or Tandberg endpoint. If the switch allows it, configure policers to limit the maximum amount of EF, AF41, and CS3 traffic that can be received on that port. Any other device plugged into that port should not necessarily be trusted, even if its packets are classified using DSCP. This option may be acceptable if the Sony or Tandberg system is a permanent installation in an office or small conference room.

Because the Sony or Tandberg device does not support CDP, the VLAN placement of this endpoint requires manual modification if the requirement is to place it in the voice VLAN. The advantage of placing the endpoint directly in the voice VLAN is that it can be treated like any other IP Telephony endpoint in the system. The disadvantage is that the port might pose a security risk because it provides direct access to the voice VLAN. Alternatively, you can leave the Sony or Tandberg endpoint in the data VLAN, but you will have to provision access between the data and voice VLANs to permit SCCP signaling to Unified CM and to allow the UDP media streams to pass between the data and voice VLANs during voice or video calls.

H.323 and SIP Video Endpoints

This type of endpoint is potentially the most challenging from a QoS perspective due to the wide range of H.323 and SIP video endpoints, the variation in implementations, and the feature sets. There are two main QoS options for these endpoints; the first relies on the H.323 or SIP video endpoint to correctly mark all the traffic, and the second relies on detailed knowledge of the TCP and UDP ports used.

Option 1

If the endpoint correctly marks the media and signaling traffic (signaling should include SIP, H.225, H.245, and RAS), you could trust the classifications. Because it is unlikely that the endpoint supports 802.1Q (and therefore 802.1p CoS), you will probably have to use IP Precedence or DSCP in this case. The choice of classification type depends on the specific vendor, model, and software version.



Note

It is highly unlikely that an H.323 or SIP endpoint will mark its packets correctly.

Option 2

Using a combination of either source, destination, or both TCP and UDP port numbers (possibly including IP addresses as well), you could define an ACL that matches and classifies the traffic correctly. In addition, Cisco recommends that you also apply policers to limit the amount of each class of traffic that is admitted to the network. This option has the same potential as Option 1 for classifying voice-only calls incorrectly.

Endpoint Features Summary

The following tables summarize the features supported by the various endpoint devices discussed in this chapter:

- [Table 20-5](#) summarizes the Cisco Unified Communications features for Cisco analog gateways.
- [Table 20-6](#) summarizes the features for Cisco Basic IP Phones with Skinny Client Control Protocol (SCCP).
- [Table 20-7](#) summarizes the features for Cisco Basic IP Phones with Session Initiation Protocol (SIP).
- [Table 20-8](#) summarizes the features for Cisco Business, Manager, and Executive IP Phones with SCCP.
- [Table 20-9](#) summarizes the features for Cisco Business, Manager, and Executive IP Phones with SIP.
- [Table 20-10](#) summarizes the features for specialized endpoints, including Cisco Unified IP Phones 7920, 7921G, 7936, and 7985G.
- [Table 20-11](#) summarizes the features for software-based devices, including Cisco Unified Personal Communicator and Cisco IP Communicator.

Table 20-5 *Cisco Analog Gateway Features*

Feature	Analog Interface Cards	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	24 ⁴	72	24	24	48	2
Caller ID	Y	N	N	Y	Y	Y
Call Waiting	N	N	N	N	Y	Y
Caller ID on Call Waiting	N	N	N	N	Y	Y
Call Hold	N	N	N	Y ⁵	N	Y
Call Transfer	N	N	N	Y ⁵	Y	Y
Call Forward	N	N	N	N	Y ⁶	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 ⁷	Y ⁷
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 ⁸
Message Waiting Indicator (MWI)	N	N	N	N	Y	Y ⁸

Table 20-5 Cisco Analog Gateway Features (continued)

Feature	Analog Interface Cards	Ws-svc-cmm-24fxs	Ws-x6624-fxs	VG224	VG248	ATA 186 and 188
Survivable Remote Site Telephony (SRST) Support	N	N	N	Y	Y	Y
Music on Hold (MoH)	Y	Y	Y	N	Y	Y
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	Y
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	0xB8
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.722	N	N	N	N	N	N
G.723	Y	Y	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

1. Two 10/100 Base-T.

2. One 10/100 Base-T.

3. Two 10/100 Base-T for ATA 188; one 10 Base-T for ATA 186.

4. The EVM-HD-8FXS/DID provides eight ports on the baseboard and can be configured for FXS or DID signaling. In addition, it has room for two EM-HDA-8FXS as extension modules.

5. H.323 and SIP call control.
6. Call Forward All.
7. Last Number Redial.
8. Only on SCCP and SIP version.
9. Supported on VG248 version 1.2 or later.
10. It marks MGCP signaling on UDP port 2427, but it marks the MGCP keep-alive packets as best-effort on TCP port 2428.
11. Fax pass-through and fax relay.
12. Fax pass-through.
13. Unified CM does not support MGCP with the ATA.

Table 20-6 Cisco Basic IP Phones with SCCP

Feature	7902G	7905G	7906G	7910G	7910 +SW	7911G	7912G/G-A
Ethernet Connection	Y ¹	Y ¹	Y ²	Y ¹	Y ³	Y ³	Y ³
Ethernet Switch (PC port)	N	N	Y	N	Y	Y	Y ⁴
Cisco Power-Over-Ethernet (PoE)	Y	Y	Y	Y	Y	Y	Y
IEEE 802.3af Power-Over-Ethernet (PoE)	N	N	Y	N	N	Y	N
Localization	N	Y	Y	N	N	Y	Y
Directory Number	1	1	1	1	1	1	1
Maximum number of calls per line	200	200	200	200	200	200	200
Liquid Crystal Display	N	Y	Y	Y	Y	Y	Y
Caller ID	N	Y	Y	Y	Y	Y	Y
Call Waiting	N	Y	Y	Y	Y	Y	Y
Caller ID on Call Waiting	N	Y	Y	Y	Y	Y	Y
Call Hold	Y	Y	Y	Y	Y	Y	Y
Blind Transfer	N	N	N	N	N	N	N
Early-attended Transfer	Y	Y	Y	Y	Y	Y	Y
Consultative Transfer	Y	Y	Y	Y	Y	Y	Y
Call Forward	Y	Y	Y	Y	Y	Y	Y
Auto-Answer	N	Y ⁵	Y ⁵	N	N	Y ⁵	Y ⁵
Ad Hoc Conference	Y	Y	Y	Y	Y	Y	Y
Meet-Me Conference	N	Y	Y	Y	Y	Y	Y
Call Pickup	N	Y	Y	Y	Y	Y	Y
Group Pickup	N	Y	Y	Y	Y	Y	Y
Redial	Y ⁶	Y ⁶	Y ⁶	Y ⁶	Y ⁶	Y ⁶	Y ⁶
Speed Dial	Y	Y	Y	Y	Y	Y	Y
On-hook Dialing	N	Y	Y	Y	Y	Y	Y
Voice Mail Access	Y	Y	Y	Y	Y	Y	Y
Message Waiting Indicator (MWI)	Y	Y	Y	Y	Y	Y	Y
Video call	N	N	N	N	N	N	N

Table 20-6 Cisco Basic IP Phones with SCCP (continued)

Feature	7902G	7905G	7906G	7910G	7910 +SW	7911G	7912G/G-A
Survivable Remote Site Telephony (SRST) Support	Y	Y	Y	Y	Y	Y	Y
Unicast MoH	Y	Y	Y	Y	Y	Y	Y
Multicast MoH	Y	Y	Y	Y	Y	Y	Y
Tone on Hold	Y	Y	Y	Y	Y	Y	Y
Speaker	N	Y ⁵	Y ⁵	Y ⁵	Y ⁵	Y ⁵	Y ⁵
Headset Jack	N	N	N	N	N	N	N
Mute	N	N	N	Y	Y	N	N
Multilevel Precedence and Preemption (MLPP)	Y	Y	Y	Y	Y	Y	Y
Barge	N	N	N	N	N	N	Y
cBarge	N	Y	Y	N	N	Y	Y
Disable Gratuitous Address Resolution Protocol (GARP)	Y	Y	Y	Y	Y	Y	Y
Signaling and Media Encryption	N	N	N	N	N	N	N
Signaling Integrity	N	N	N	N	N	N	N
Manufacturing-Installed Certificate (X.509v3)	N	N	N	N	N	N	N
Field-Installed Certificate	N	N	N	N	N	N	N
Third-Party XML Service	N	Y	Y	N	N	Y	Y
External Microphone and Speaker	N	N	N	N	N	N	N
Dial plan	N	N	N	N	N	N	N
Signaling Packet ToS Value Marking	0x60	0x60	0x60	0x60	0x60	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8
G.711	Y	Y	Y	Y	Y	Y	Y
G.722	N	N	N	N	N	N	N
G.723	N	Y	N	N	N	N	N
G.726	N	Y	N	N	N	N	N
G.729	Y	Y	Y	Y	Y	Y	Y
Wideband Audio	N	N	N	N	N	N	N
Wideband Video	N	N	N	N	N	N	N
Voice Activity Detection (VAD)	Y	Y	Y	Y	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y	Y	Y	Y	Y
DTMF - H.245	N	N	N	N	N	N	N
DTMF - SCCP	Y	Y	Y	Y	Y	Y	Y
DTMF - RFC2833	N	N	Y	N	N	Y	N

Table 20-6 Cisco Basic IP Phones with SCCP (continued)

Feature	7902G	7905G	7906G	7910G	7910 +SW	7911G	7912G/G-A
DTMF - KPML	N	N	N	N	N	N	N
DTMF - Unsolicited Notify	N	N	N	N	N	N	N

1. One 10 Base-T.
2. One 10/100 Base-T.
3. Two 10/100 Base-T.
4. The Cisco Unified IP Phone 7912G-A has an enhanced version of Ethernet switch.
5. One-way audio monitor mode.
6. Last Number Redial.

Table 20-7 Cisco Basic IP Phones with SIP

Feature	7905G	7906G	7911G	7912G/G-A
Ethernet Connection	Y ¹	Y ²	Y ³	Y ²
Ethernet Switch (PC port)	N	Y	Y	Y ⁴
Cisco Power-Over-Ethernet (PoE)	Y	Y	Y	Y
IEEE 802.3af Power-Over-Ethernet (PoE)	N	Y	Y	N
Localization	N	Y	Y	N
Directory Number	1	1	1	1
Maximum number of calls per line	2	50	50	2
Liquid Crystal Display	Y	Y	Y	Y
Caller ID	Y	Y	Y	Y
Call Waiting	Y	Y	Y	Y
Caller ID on Call Waiting	Y	Y	Y	Y
Call Hold	Y	Y	Y	Y
Blind Transfer	Y	Y	Y	Y
Early-attended Transfer	N	Y	Y	N
Consultative Transfer	Y	Y	Y	Y
Call Forward	Y ⁵	Y	Y	Y ⁵
Auto-Answer	N	Y ⁶	Y ⁶	N
Ad Hoc Conference	Y	Y	Y	Y
Meet-Me Conference	N	Y	Y	N
Call Pickup	N	Y	Y	N
Group Pickup	N	Y	Y	N
Redial	Y ⁷	Y	Y	Y ⁷
Speed Dial	Y ⁸	Y	Y	Y ⁸
On-hook Dialing	Y	Y	Y	Y
Voice Mail Access	Y	Y	Y	Y
Message Waiting Indicator (MWI)	Y	Y	Y	Y

Table 20-7 Cisco Basic IP Phones with SIP (continued)

Feature	7905G	7906G	7911G	7912G/G-A
Video call	N	N	N	N
Survivable Remote Site Telephony (SRST) Support	Y	Y	Y	Y
Unicast MoH	Y	Y	Y	Y
Multicast MoH	N	Y	Y	N
Tone on Hold	N	N	N	N
Speaker	Y ⁶	Y ⁶	Y ⁶	Y ⁶
Headset Jack	N	N	N	N
Mute	N	Y	Y	N
Multilevel Precedence and Preemption (MLPP)	N	N	N	N
Barge	N	Y	Y	N
cBarge	N	Y	Y	N
Disable Gratuitous Address Resolution Protocol (GARP)	Y	Y	Y	Y
Signaling and Media Encryption	N	Y	Y	N
Signaling Integrity	N	Y	Y	N
Manufacturing-Installed Certificate (X.509v3)	N	Y	Y	N
Field-Installed Certificate	N	N	N	N
Third-Party XML Service	N	Y	Y	N
External Microphone and Speaker	N	N	N	N
Dial plan	Y	Y	Y	Y
Signaling Packet ToS Value Marking	0x60	0x60	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0xB8
G.711	Y	Y	Y	Y
G.722	N	N	N	N
G.723	N	Y	Y	N
G.726	N	N	N	N
G.729	Y ⁹	Y ⁹	Y ⁹	Y ⁹
Wideband Audio	N	N	N	N
Wideband Video	N	N	N	N
Voice Activity Detection (VAD)	Y	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y	Y
DTMF - H.245	N	N	N	N
DTMF - SCCP	N	N	N	N
DTMF - RFC2833	Y	Y	Y	Y

Table 20-7 Cisco Basic IP Phones with SIP (continued)

Feature	7905G	7906G	7911G	7912G/G-A
DTMF - KPML	N	Y	Y	N
DTMF - Unsolicited Notify	N	N	N	N

1. One 10 Base-T.
2. One 10/100 Base-T.
3. Two 10/100 Base-T.
4. The Cisco Unified IP Phone 7912G-A has an enhanced version of Ethernet switch.
5. For the Cisco Unified IP Phone 7905G and 7912G with SIP, if CFWDALL is configured on the phone, the phone must be in service to make CFWDALL work because Unified CM has no knowledge of the configuration on the phone. This behavior is different from an SCCP phone, which can be out of service and CFWDALL will still work. If CFWDALL is enabled on the Unified CM User page, Unified CM will handle this change but there will be no status line on the phone to indicate that the call is forwarded. CFWDALL configuration on the Unified CM User page overrides the configuration on the phone.
6. One-way audio monitor mode.
7. Last Number Redial.
8. Speed dial can be configured only on the phone for these models.
9. These IP phone models support G.729b and G.729ab in receiving mode only.

Table 20-8 Cisco Business, Manager, and Executive IP Phones with SCCP

Feature	7940G	7941G/G-GE	7960G	7961G/G-GE	7970G	7971G-GE
Ethernet Connection	Y ¹	Y ²	Y ¹	Y ³	Y ¹	Y ⁴
Ethernet Switch (PC port)	Y	Y	Y	Y	Y	Y
Cisco Power-Over-Ethernet (PoE)	Y	Y	Y	Y	Y	N
IEEE 802.3af Power-Over-Ethernet (PoE)	N	Y ⁵	N	Y ⁶	Y	Y ⁷
Localization	Y	Y	Y	Y	Y	Y
Directory Number	2	2	6	6	8	8
Maximum number of calls per line	200	200	200	200	200	200
Liquid Crystal Display	Y	Y	Y	Y	Y	Y
Caller ID	Y	Y	Y	Y	Y	Y
Call Waiting	Y	Y	Y	Y	Y	Y
Caller ID on Call Waiting	Y	Y	Y	Y	Y	Y
Call Hold	Y	Y	Y	Y	Y	Y
Blind Transfer	N	N	N	N	N	N
Early-Attended Transfer	Y	Y	Y	Y	Y	Y
Consultative Transfer	Y	Y	Y	Y	Y	Y
Call Forward	Y	Y	Y	Y	Y	Y
Auto-Answer	Y	Y	Y	Y	Y	Y
Ad Hoc Conference	Y	Y	Y	Y	Y	Y
Meet-Me Conference	Y	Y	Y	Y	Y	Y
Call Pickup	Y	Y	Y	Y	Y	Y

Table 20-8 Cisco Business, Manager, and Executive IP Phones with SCCP (continued)

Feature	7940G	7941G/G-GE	7960G	7961G/G-GE	7970G	7971G-GE
Group Pickup	Y	Y	Y	Y	Y	Y
Redial	Y ⁸	Y ⁸	Y ⁸	Y ⁸	Y ⁸	Y ⁸
Speed Dial	Y	Y	Y	Y	Y	Y
On-hook Dialing	Y	Y	Y	Y	Y	Y
Voice Mail Access	Y	Y	Y	Y	Y	Y
Message Waiting Indicator (MWI)	Y	Y	Y	Y	Y	Y
Video call	Y	Y	Y	Y	Y	Y
Survivable Remote Site Telephony (SRST) Support	Y	Y	Y	Y	Y	Y
Unicast MoH	Y	Y	Y	Y	Y	Y
Multicast MoH	Y	Y	Y	Y	Y	Y
Tone on Hold	Y	Y	Y	Y	Y	Y
Speaker	Y	Y	Y	Y	Y	Y
Headset Jack	Y	Y	Y	Y	Y	Y
Mute	Y	Y	Y	Y	Y	Y
Multilevel Precedence and Preemption (MLPP)	Y	Y	Y	Y	Y	Y
Barge	Y	Y	Y	Y	Y	Y
cBarge	Y	Y	Y	Y	Y	Y
Disable Gratuitous Address Resolution Protocol (GARP)	Y	Y	Y	Y	Y	Y
Signaling and Media Encryption	Y	Y	Y	Y	Y	Y
Signaling Integrity	Y	Y	Y	Y	Y	Y
Manufacturing-Installed Certificate (X.509v3)	N	Y	N	Y	Y	Y
Field-Installed Certificate	Y	N	Y	N	N	N
Third-Party XML Service	Y	Y	Y	Y	Y	Y
External Microphone and Speaker	Y	Y	Y	Y	Y	Y
Dial Plan	N	N	N	N	N	N
Signaling Packet ToS Value Marking	0x60	0x60	0x60	0x60	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8
G.711	Y	Y	Y	Y	Y	Y
G.722	N	Y	N	Y	Y	Y

Table 20-8 Cisco Business, Manager, and Executive IP Phones with SCCP (continued)

Feature	7940G	7941G/G-GE	7960G	7961G/G-GE	7970G	7971G-GE
G.723	N	N	N	N	N	N
G.726	N	N	N	N	N	N
G.729	Y	Y	Y	Y	Y	Y
Wideband Audio	N	N	N	N	N	N
Wideband Video	N	N	N	N	N	N
Voice Activity Detection (VAD)	Y	Y	Y	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y	Y	Y	Y
DTMF - H.245	N	N	N	N	N	N
DTMF - SCCP	Y	Y	Y	Y	Y	Y
DTMF - RFC2833	N	Y	N	Y	Y	Y
DTMF - KPML	N	N	N	N	N	N
DTMF - Unsolicited Notify	N	N	N	N	N	N

1. Two 10/100 Base-T.
2. The Cisco Unified IP Phone 7941G has two 10/100 Mbps Ethernet connections, and the Cisco Unified IP Phone 7941G-GE has two 10/100/1000 Mbps Ethernet connections.
3. The Cisco Unified IP Phone 7961G has two 10/100 Mbps Ethernet connections, and the Cisco Unified IP Phone 7961G-GE has two 10/100/1000 Mbps Ethernet connections.
4. The Cisco Unified IP Phone 7970G has two 10/100 Mbps Ethernet connections, and the Cisco Unified IP Phone 7971G-GE has two 10/100/1000 Mbps Ethernet connections.
5. The Cisco Unified IP Phone 7941G supports both Cisco Prestandard Power over Ethernet (PoE) and IEEE 802.3af PoE, and the Cisco Unified IP Phone 7941G-GE supports only IEEE 802.3af PoE.
6. The Cisco Unified IP Phone 7961G supports both Cisco Prestandard PoE and IEEE 802.3af PoE, and the Cisco Unified IP Phone 7961G-GE supports only IEEE 802.3af PoE.
7. The Cisco Unified IP Phone 7971G-GE supports only IEEE 802.3af PoE.
8. Last Number Redial.

Table 20-9 Cisco Business, Manager, and Executive IP Phones with SIP

Feature	7940G	7941G/G-GE	7960G	7961G/G-GE	7970G	7971G-GE
Ethernet Connection	Y ¹	Y ²	Y ¹	Y ³	Y ¹	Y ⁴
Ethernet Switch (PC port)	Y	Y	Y	Y	Y	Y
Cisco Power-Over-Ethernet (PoE)	Y	Y ⁵	Y	Y ⁶	Y	N
IEEE 802.3af Power-Over-Ethernet (PoE)	N	Y	N	Y	Y	Y ⁷
Localization	N	Y	N	Y	Y	Y
Directory Number	2	2	6	6	8	8
Maximum number of calls per line	2	50	2	50	50	50
Liquid Crystal Display	Y	Y	Y	Y	Y	Y
Caller ID	Y	Y	Y	Y	Y	Y
Call Waiting	Y	Y	Y	Y	Y	Y
Caller ID on Call Waiting	Y	Y	Y	Y	Y	Y

Table 20-9 *Cisco Business, Manager, and Executive IP Phones with SIP (continued)*

Feature	7940G	7941G/G-GE	7960G	7961G/G-GE	7970G	7971G-GE
Call Hold	Y	Y	Y	Y	Y	Y
Blind Transfer	Y	Y	Y	Y	Y	Y
Early-Attended Transfer	Y	Y	Y	Y	Y	Y
Consultative Transfer	Y	Y	Y	Y	Y	Y
Call Forward	Y ⁸	Y	Y ⁸	Y	Y	Y
Auto-Answer	Y ⁹	Y ¹⁰	Y ⁹	Y	Y	Y
Ad Hoc Conference	Y ¹¹	Y	Y ¹¹	Y	Y	Y
Meet-Me Conference	N	Y	N	Y	Y	Y
Call Pickup	N	Y	N	Y	Y	Y
Group Pickup	N	Y	N	Y	Y	Y
Redial	Y ¹²	Y ¹²	Y ¹²	Y ¹²	Y ¹²	Y ¹²
Speed Dial	Y ¹³	Y	Y ¹³	Y	Y	Y
On-hook Dialing	N	Y	N	Y	Y	Y
Voice Mail Access	Y	Y	Y	Y	Y	Y
Message Waiting Indicator (MWI)	Y	Y	Y	Y	Y	Y
Video call	N	N	N	N	N	N
Survivable Remote Site Telephony (SRST) Support	Y	Y	Y	Y	Y	Y
Unicast MoH	Y	Y	Y	Y	Y	Y
Multicast MoH	Y	Y	Y	Y	Y	Y
Tone on Hold	N	N	N	N	N	N
Speaker	Y	Y ¹⁰	Y	Y	Y	Y
Headset Jack	Y	Y	Y	Y	Y	Y
Mute	Y	Y	Y	Y	Y	Y
Multilevel Precedence and Preemption (MLPP)	N	N	N	N	N	N
Barge	N	Y	N	Y	Y	Y
cBarge	N	Y	N	Y	Y	Y
Disable Gratuitous Address Resolution Protocol (GARP)	Y	Y	Y	Y	Y	Y
Signaling and Media Encryption	N	Y	N	Y	Y	Y
Signaling Integrity	N	Y	N	Y	Y	Y
Manufacturing-Installed Certificate (X.509v3)	N	Y	N	Y	Y	Y
Field-Installed Certificate	N	N	N	N	N	N
Third-Party XML Service	Y ¹⁴	Y	Y ¹⁴	Y	Y	Y
External Microphone and Speaker	Y	Y	Y	Y	Y	Y

Table 20-9 Cisco Business, Manager, and Executive IP Phones with SIP (continued)

Feature	7940G	7941G/G-GE	7960G	7961G/G-GE	7970G	7971G-GE
Dial Plan	Y	Y	Y	Y	Y	Y
Signaling Packet ToS Value Marking	0x60	0x60	0x60	0x60	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0xB8	0xB8	0xB8
G.711	Y	Y	Y	Y	Y	Y
G.722	N	Y	N	Y	Y	Y
G.723	Y	Y	Y	Y	Y	Y
G.726	N	N	N	N	N	N
G.729	Y ¹⁵	Y ¹⁵	Y ¹⁵	Y ¹⁵	Y ¹⁵	Y ¹⁵
Wideband Audio	N	N	N	N	N	N
Wideband Video	N	N	N	N	N	N
Voice Activity Detection (VAD)	Y	Y	Y	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y	Y	Y	Y
DTMF - H.245	N	N	N	N	N	N
DTMF - SCCP	N	N	N	N	N	N
DTMF - RFC2833	Y	Y	Y	Y	Y	Y
DTMF - KPML	N	Y	N	Y	Y	Y
DTMF - Unsolicited Notify	N	N	N	N	N	N

- Two 10/100 Base-T.
- The Cisco Unified IP Phone 7941G has two 10/100 Mbps Ethernet connections, and the Cisco Unified IP Phone 7941G-GE has two 10/100/1000 Mbps Ethernet connections.
- The Cisco Unified IP Phone 7961G has two 10/100 Mbps Ethernet connections, and the Cisco Unified IP Phone 7961G-GE has two 10/100/1000 Mbps Ethernet connections.
- The Cisco Unified IP Phone 7970G has two 10/100 Mbps Ethernet connections, and the Cisco Unified IP Phone 7971G-GE has two 10/100/1000 Mbps Ethernet connections.
- The Cisco Unified IP Phone 7941G supports both Cisco Prestandard Power over Ethernet (PoE) and IEEE 802.3af PoE, and the Cisco Unified IP Phone 7941G-GE supports only IEEE 802.3af PoE.
- The Cisco Unified IP Phone 7961G supports both Cisco Prestandard PoE and IEEE 802.3af PoE, and the Cisco Unified IP Phone 7961G-GE supports only IEEE 802.3af PoE.
- The Cisco Unified IP Phone 7971G-GE supports only IEEE 802.3af PoE.
- For the Cisco Unified IP Phone 7905, 7912, 7940, or 7960 with SIP, if CFWDALL is configured on the phone, the phone must be in service to make CFWDALL work because Unified CM has no knowledge of the configuration on the phone. This behavior is different from an SCCP phone, which can be out of service and CFWDALL will still work. If CFWDALL is enabled on the Unified CM User page, Unified CM will handle this change but there will be no status line on the phone to indicate that the call is forwarded. CFWDALL configuration on the Unified CM User page overrides the configuration on the phone.
- This feature can be configured locally on the phone.
- One-way audio monitor mode.
- The Cisco Unified IP Phone 7940G and 7960G with IP support only local mixing for ad hoc conferences and up to three parties in the conference.
- Last Number Redial.
- Speed dial can be configured only on the phone.
- With limited support.
- These IP phone models support G.729b and G.729ab in receiving mode only.

Table 20-10 Specialized Endpoints

Feature	7920	7921G	7936	7985G
Ethernet Connection	N	N	Y ¹	Y ²
Ethernet Switch (PC port)	N	N	N	Y
Cisco Power-Over-Ethernet (PoE)	N	N	N	N
IEEE 802.3af Power-Over-Ethernet (PoE)	N	N	N	Y
Localization	Y	Y	N	Y
Directory Number	12	6	1	2
Max number of calls per line	2	2	2	100
Liquid Crystal Display	Y	Y	Y	Y
Caller ID	Y	Y	Y	Y
Call Waiting	Y	Y	Y	Y
Caller ID on Call Waiting	Y	Y	Y	Y
Call Hold	Y	Y	Y	Y
Blind Transfer	N	N	N	N
Early-Attended Transfer	Y	Y	Y	Y
Consultative Transfer	Y	Y	Y	Y
Call Forward	Y	Y	Y	Y
Auto-Answer	N	Y	N	Y
Ad Hoc Conference	Y	Y	Y	Y
Meet-Me Conference	Y	Y	Y	Y
Call Pickup	Y	Y	Y	Y
Group Pickup	Y	Y	Y	Y
Redial	Y ³	Y ³	Y	Y
Speed Dial	Y	Y	N	Y
On-hook Dialing	Y	Y	Y	Y
Voice Mail Access	Y	Y	N	Y
Message Waiting Indicator (MWI)	Y	Y	N	Y
Video call	N	N	N	Y
Survivable Remote Site Telephony (SRST) Support	Y	Y	Y	Y ⁴
Unicast MoH	Y	Y	Y	Y
Multicast MoH	Y	Y	Y	N
Tone on Hold	Y	Y	Y	Y
Speaker	N	Y	Y	Y
Headset Jack	Y	Y	N	Y
Mute	Y	Y	Y	Y
Multilevel Precedence and Preemption (MLPP)	N	Y	N	Y

Table 20-10 Specialized Endpoints (continued)

Feature	7920	7921G	7936	7985G
Barge	N	Y	N	Y
cBarge	N	Y	N	Y
Disable Gratuitous Address Resolution Protocol (GARP)	N	Y	N	N
Signaling and Media Encryption	Y	Y	N	N
Signaling Integrity	N	Y	N	N
Manufacturing-Installed Certificate (X.509v3)	N	Y	N	N
Field-Installed Certificate	N	Y	N	N
Third-Party XML Service	Y	Y	N	N
External Microphone and Speaker	N	Y	N	N
Signaling Packet ToS Value Marking	0x60	0x60	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8	0x88
G.711	Y	Y	Y	Y
G.722	N	N	N	Y
G.723	N	N	N	N
G.726	N	N	N	N
G.729	Y	Y	Y	Y
Wideband Audio	N	N	N	N
Wideband Video	N	N	N	N
H.261	N	N	N	Y
H.263	N	N	N	Y
H.263+	N	N	N	Y
H.264	N	N	N	Y
Voice Activity Detection (VAD)	Y	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y	Y
DTMF - H.245	N	N	N	N
DTMF - SCCP	Y	Y	Y	Y
DTMF - RFC2833	N	N	N	N

1. One 10/100 Base-T.
2. Two 10/100 Base-T.
3. Last Number Redial.
4. Only audio supported on SRST.

Table 20-11 Software-Based Endpoint Features

Feature	Unified Personal Communicator	IP Communicator with SCCP	IP Communicator with SIP
Directory Number	1	8	8
Caller ID	Y	Y	Y
Call Waiting	Y	Y	Y
Caller ID on Call Waiting	Y	Y	Y
Call Hold	Y	Y	Y
Call Transfer	N ¹	Y	Y
Call Forward	N	Y	Y
Auto-Answer	Y	Y	Y
Ad Hoc Conference	Y ²	Y	Y
Meet-Me Conference	N ³	Y	N
Web Conference	Y	N	N
Call Pickup	N	Y	Y
Group Pickup	N	Y	Y
Redial	Y ⁴	Y ⁴	Y ⁴
Speed Dial	N ⁵	Y	Y
On-hook Dialing	Y	Y	Y
Voice Mail Access	Y	Y	Y
Message Waiting Indicator (MWI)	N	Y	Y
Video call	Y	Y ⁶	N
Survivable Remote Site Telephony (SRST) Support	N	Y	Y
Unicast Music on Hold (MoH)	Y	Y	Y
Multicast Music on Hold (MoH)	Y	Y	Y
Tone on Hold	N	Y	N
Mute	Y	Y	Y
Multilevel Precedence and Preemption (MLPP)	N	Y	N
Barge	N	Y	Y
cBarge	N	Y	Y
Disable Gratuitous Address Resolution Protocol (GARP)	N	N	N
Signaling and Media Encryption	N	N	N
Signaling Integrity	N	N	N
Manufacturing-Installed Certificate (X.509v3)	N	N	N
Field-Installed Certificate	N	N	N

Table 20-11 **Software-Based Endpoint Features (continued)**

Feature	Unified Personal Communicator	IP Communicator with SCCP	IP Communicator with SIP
Third-Party XML Service	N	Y	Y
Signaling Packet ToS Value Marking	N	0x60	0x60
Media Packet ToS Value Marking	0xB8	0xB8	0xB8
Skinny Client Control Protocol (SCCP)	N	Y	N
Session Initiation Protocol (SIP)	Y	N	N
G.711	Y	Y	Y
G.722	N	N	N
G.723	N	N	N
G.726	N	N	N
G.729	Y	Y	Y ⁷
Wideband Audio	N	Y	Y
Wideband Video	N	N	N
H.261	N	N	N
H.263	Y	N	N
H.264	Y	N	N
Voice Activity Detection (VAD)	Y	Y	Y
Comfort Noise Generation (CNG)	Y	Y	Y
DTMF – H.245	N	N	N
DTMF – SCCP	N	Y	N
DTMF – RFC2833	Y	Y	Y
DTMF – KPML	Y	N	Y

1. Cisco Unified Personal Communicator does not have an explicit transfer feature. Calls can be transferred by the other party, or the Cisco Unified Personal Communicator user can merge two calls and then disconnect to obtain the effect of a transfer.
2. Cisco Unified Personal Communicator does not support the “consult, then merge” feature (equivalent to conference on IP phones), but it does support **merge** (equivalent to **join** on IP phones) to conference calls.
3. Cisco Unified Personal Communicator cannot create a meet-me conference, but the user can join one by dialing the correct number.
4. Last number redial.
5. Cisco Unified Personal Communicator does not support the Unified CM speed dials page, but it does support dialing from the Contacts (buddy) list in a similar fashion.
6. In combination with Cisco Unified Video Advantage, Cisco IP Communicator operating in SCCP mode supports video calls.
7. Cisco IP Communicator with SIP supports G.729b and G.729ab in receiving mode only.



CHAPTER 21

Cisco Unified Presence

Last revised on: September 27, 2007

Cisco Unified Presence consists of many components that enhance the value of a Cisco Unified Communications system. The main presence component is the Cisco Unified Presence Server, which collects information regarding a user's availability status and communications capabilities. The user's availability status indicates whether or not the user is actively using a particular communications device such as a phone. The user's communications capabilities indicate the types of communications that user is capable of using, such as video conferencing, web collaboration, or basic audio.

The aggregated user information captured by the Cisco Unified Presence Server enables Cisco Unified Personal Communicator and Cisco Unified Communications Manager applications to increase user productivity. These applications help connect colleagues more efficiently by determining the most effective form of communication.

This chapter explains the basic concepts of Presence within the Cisco Unified Communications System and provides guidelines for how best to deploy the various presence components. This chapter covers the following topics:

- [Presence, page 21-1](#)
- [Unified CM Presence, page 21-4](#)
- [Cisco Unified Presence Server, page 21-8](#)
- [Cisco IP Phone Messenger Application, page 21-15](#)
- [Cisco Unified Personal Communicator, page 21-19](#)
- [Third-Party Presence Server Integration, page 21-23](#)

Presence

Presence refers to the ability and willingness of a user to communicate across a set of devices. It involves the following phases or activities:

- Publish user status
User status changes can be published automatically by recognizing user keyboard activity, phone use, or device connectivity to the network.
- Collect this status
The published information is gathered from all the available sources, privacy policies are applied, and then current status is aggregated, synchronized, and stored for consumption.

- Consume the information

Desktop applications, calendar applications, and devices can use the user status information to provide real-time updates for the end users to make better communication decisions.

Status combines the capabilities of what the device or user can do (voice, video, web collaboration, and so forth) and the attributes showing the state of the device or user (available, busy, on a call, and so forth). Presence status can be derived from automatic events such as computer login and telephone off-hook, or it can be derived from explicit notification events for changing status such as the user selecting Do Not Disturb from a change-status pick list.

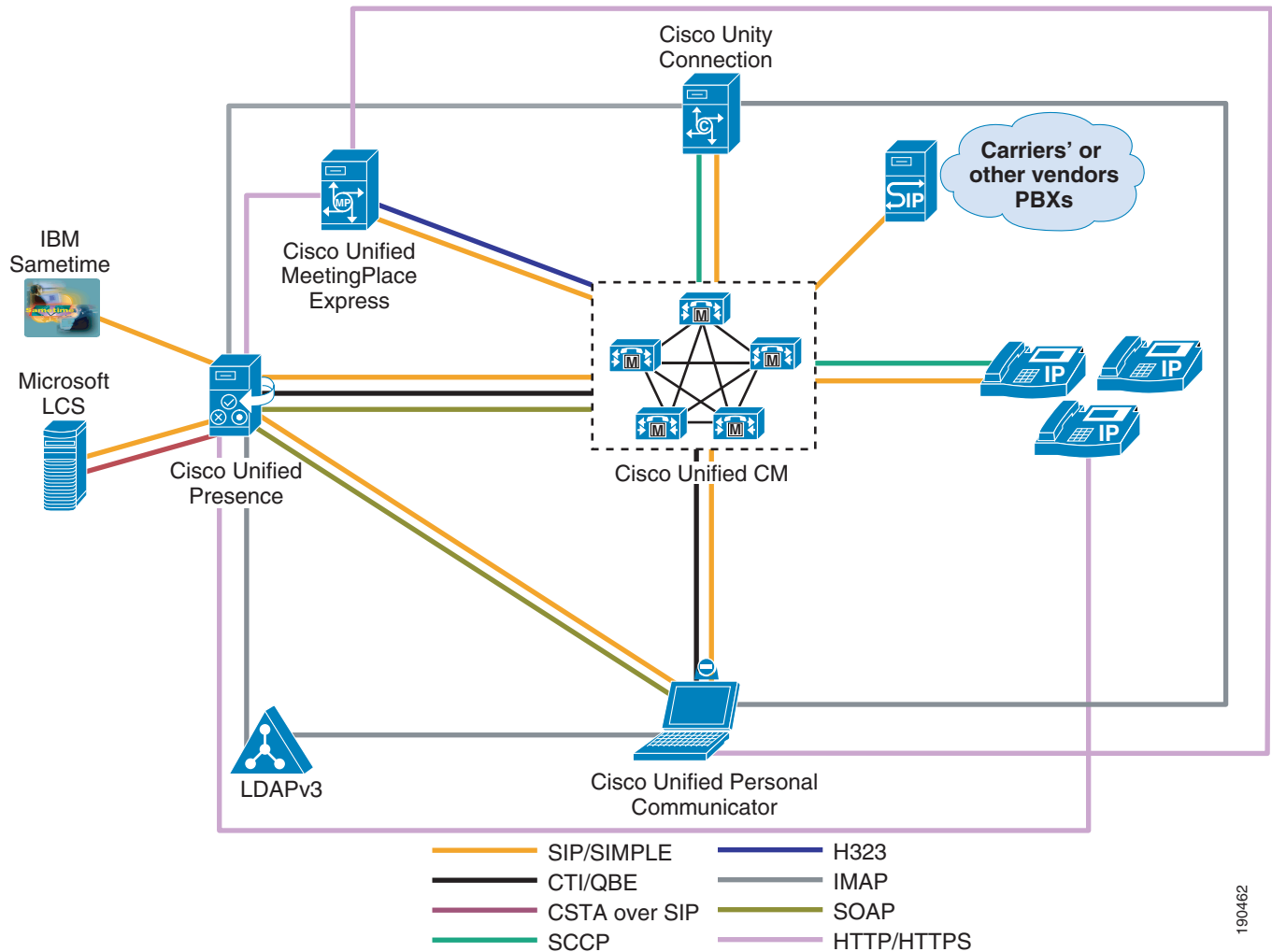
Terminology surrounding presence refers to a watcher, presence entity (*presentity*), and presence server. The presence entity publishes its current status via a PUBLISH or REGISTER message to the presence server, and it can be a directory number (DN) or a SIP uniform resource identifier (URI) that resides within or outside the communications cluster. A *watcher* (device or user) requests presence status about a presence entity by sending a SUBSCRIBE message to the presence server. The presence server responds to the watcher with a NOTIFY message containing the current status of the presence entity.

The reason for using SIP with presence is that it unifies all major communications services such as voice, video, web, email, presence, and instant messaging on a single infrastructure. SIP is an extensible protocol and allows for additional packages, such as presence events, to be utilized on an existing SIP network that already routes specified requests and responses to the appropriate locations. By aligning these services, SIP allows for tighter integration and reduces management complexity in delivering these combined communication services.

Cisco Unified Presence Components

Cisco Unified Presence encompasses the following components, illustrated in [Figure 21-1](#):

- Cisco Unified Presence Server
- Cisco Unified Communications Manager (Unified CM) Release 5.0(4) or higher
- Cisco Unified Personal Communicator
- Cisco Unified MeetingPlace Express
- Cisco Unity Connection
- Lightweight Directory Access Protocol (LDAP) Server v3.0
- Cisco Unified IP Phones
- Third-party presence server

Figure 21-1 Cisco Unified Presence Components

Cisco Unified Presence User

For presence, typically a user is described in terms of the user's presence status, the number of users on the system, or the user's presence capabilities.

As defined by Cisco Unified Presence, a user is specified in Unified CM as an End User and must be configured with a primary extension. The Primary Extension End User configuration is new to Cisco Unified CM as of Release 5.0. The user at this point is effectively tied to a directory number, and presence status is reflected for the user's primary extension rather than for the device to which the user is associated.

The concept of a *presence user* appears throughout this chapter; therefore, keep in mind the meaning of a user as defined for Cisco Unified Presence.

Unified CM Presence

All presence requests for users, whether inside or outside the cluster, are processed and handled by Cisco Unified CM Release 5.0(4) or higher.

A Unified CM watcher that sends a presence request will receive a direct response, including the presence status, if the watcher and presence entity are co-located within the Unified CM cluster.

If the presence entity exists outside the cluster, Unified CM will query the external presence entity through the SIP trunk. If the watcher has permission to monitor the external presence entity based on the SUBSCRIBE calling search space and presence group (both described in the section on [Unified CM Presence Policy](#), page 21-7), the SIP trunk will forward the presence request to the external presence entity, await the presence response from the external presence entity, and return the current presence status to the watcher.

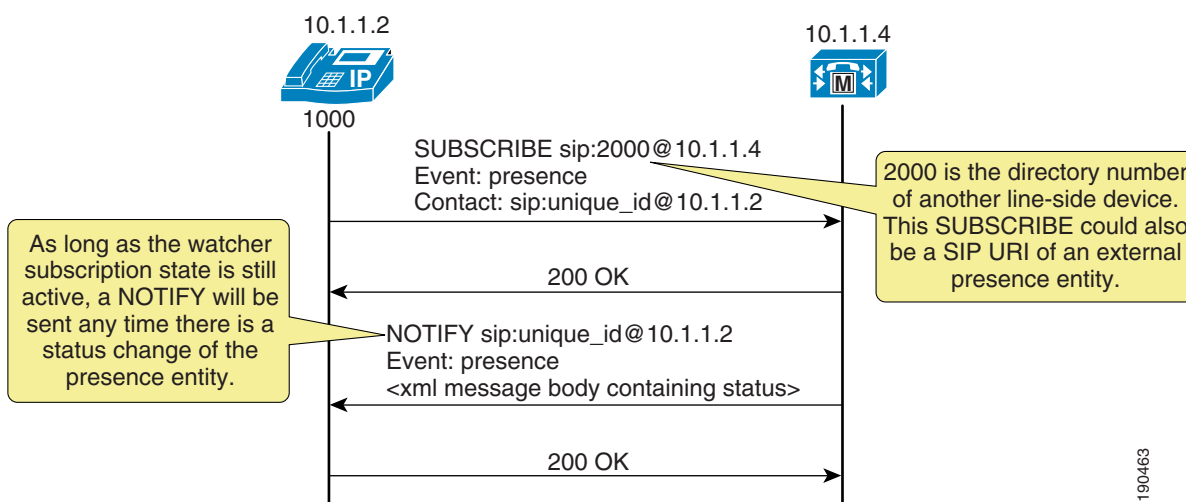
A watcher that is not in a Unified CM cluster can send a presence request to a SIP trunk. If Unified CM supports the presence entity, it will respond with the current presence status. If Unified CM does not support the presence entity, it will reject the presence request with a SIP error response.

Unified CM Presence with SIP

Unified CM uses the term *SIP line* to represent endpoints supporting SIP that are directly connected and registered to Unified CM and the term *SIP trunk* to represent trunks supporting SIP. SIP line-side endpoints acting as presence watchers can send a SIP SUBSCRIBE message to Unified CM requesting the presence status of the indicated presence entity.

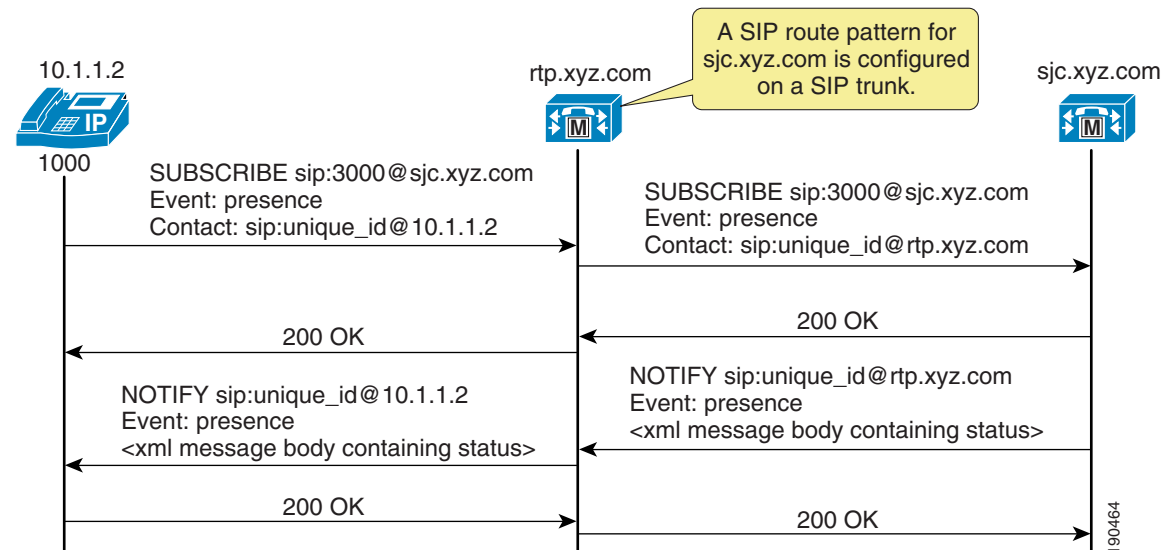
If the presence entity resides within the Unified CM cluster, Unified CM responds to a SIP line-side presence request by sending a SIP NOTIFY message to the presence watcher, indicating the current status of the presence entity. (See [Figure 21-2](#).)

Figure 21-2 SIP Line SUBSCRIBE/NOTIFY Exchange



If the presence entity resides outside the Unified CM cluster, Unified CM routes a SUBSCRIBE request out the appropriate SIP trunk, based on the SUBSCRIBE calling search space, presence group, and SIP route pattern. When Unified CM receives a SIP NOTIFY response on the trunk, indicating the presence entity status, it responds to the SIP line-side presence request by sending a SIP NOTIFY message to the presence watcher, indicating the current status of the presence entity. (See [Figure 21-3](#).)

Figure 21-3 SIP Trunk SUBSCRIBE/NOTIFY Exchange



SUBSCRIBE messages for any directory number or SIP URI residing outside the Unified CM cluster are sent or received on a SIP trunk in Unified CM. The SIP trunk could be an interface to another Unified CM or it could be an interface to the Cisco Unified Presence Server.

Unified CM Presence with SCCP

Unified CM supports Skinny Client Control Protocol (SCCP) line-side endpoints acting as presence watchers. There are no SCCP trunks. SCCP endpoints can request presence status of the indicated presence entity by sending SCCP messages to Unified CM.

If the presence entity resides within the Unified CM cluster, Unified CM responds to the SCCP line-side presence request by sending SCCP messages to the presence watcher, indicating the current status of the presence entity.

If the presence entity resides outside the Unified CM cluster, Unified CM routes a SUBSCRIBE request out the appropriate SIP trunk, based on the SUBSCRIBE calling search space, presence group, and SIP route pattern. When Unified CM receives a SIP NOTIFY response on the trunk, indicating the presence entity status, it responds to the SCCP line-side presence request by sending SCCP messages to the presence watcher, indicating the current status of the presence entity.

Unified CM Speed Dial Presence

Unified CM supports the ability for a speed dial to have presence capabilities via a busy lamp field (BLF) speed dial. BLF speed dials work as both a speed dial and a presence indicator. However, only the system administrator can configure a BLF speed dial; a system user is not allowed to configure a BLF speed dial.

The administrator must configure the BLF speed dial with a target directory number that is resolvable to a directory number within the Unified CM cluster or a SIP trunk destination. BLF SIP line-side endpoints can also be configured with a SIP URI for the BLF speed dial, but SCCP line-side endpoints cannot be configured with a SIP URI for BLF speed dial. The BLF speed dial indication is a line-level indication and not a device-level indication.

The following Cisco Unified IP Phones support BLF speed dial for SCCP:

- Cisco Unified IP Phone 7914G
- Cisco Unified IP Phone 7940G
- Cisco Unified IP Phone 7960G
- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

The following Cisco Unified IP Phones support BLF speed dial for SIP:

- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

The Cisco Unified IP Phones 7905, 7906, 7911, and 7912 do not support BLF speed dial.

Figure 21-4 lists the various types of BLF speed dial indications from the phones.

Figure 21-4 Indicators for Speed Dial Presence

State	Icon	LED
Idle		
Busy		
Unknown		

180485

Unified CM Call History Presence

Unified CM supports presence capabilities for call history lists (the Directories button on the phone). Call history list presence capabilities are controlled via the **BLF for Call Lists** Enterprise Parameter within Unified CM Administration. The **BLF for Call Lists** Enterprise Parameter impacts all pages using the phone Directories button (Missed, Received, and Placed Calls, Personal Directory, or Corporate Directory), and it is set on a global basis.

Presence capabilities for call history lists are supported for both SCCP and SIP on the following Cisco Unified IP Phones:

- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE

- Cisco Unified IP Phones 7970G and 7971G-GE

The Cisco Unified IP Phones 7905G, 7906G, 7911G, 7912G, 7940G, and 7960G do not support presence capabilities for call history lists.

The presence indicators for call history lists are the same as those listed in the Icon column in [Figure 21-4](#); however, no LED indications are available.

Unified CM Presence Policy

Unified CM provides the capability to set policy for users who request presence status. You can set this policy by configuring a calling search space specifically to route SIP SUBSCRIBE messages for presence status and by configuring presence groups with which users can be associated to specify rules for viewing the presence status of users associated with another group.

Unified CM Subscribe Calling Search Space

The first aspect of presence policy for Unified CM is the new SUBSCRIBE calling search space. Unified CM uses the SUBSCRIBE calling search space to determine how to route presence requests (SUBSCRIBE messages with the Event field set to Presence) that come from the watcher, which could be a phone or a trunk. The SUBSCRIBE calling search space is associated with the watcher and lists the partitions that the watcher is allowed to "see." This mechanism provides an additional level of granularity for the presence SUBSCRIBE requests to be routed independently from the normal call-processing calling search space.

The SUBSCRIBE calling search space can be assigned on a device basis or on a user basis. The user setting applies for originating subscriptions when the user is logged in to the device via Extension Mobility or when the user is administratively assigned to the device.

With the SUBSCRIBE calling search space set to <None>, BLF speed dial and call history list presence status does not work and the subscription messages is rejected as "user unknown." When a valid SUBSCRIBE calling search space is specified, the indicators work and the SUBSCRIBE messages are accepted and routed properly.



Note

Cisco strongly recommends that you do not leave any calling search space defined as <None>. Leaving a calling search space set to <None> can introduce presence status or dialing plan behavior that is difficult to predict.

Unified CM Presence Groups

The second aspect of the presence policy for Unified CM is presence groups. Devices, directory numbers, and users can be assigned to a presence group, and by default all users are assigned to the Standard Presence Group. A presence group controls the destinations that a watcher can monitor, based on the user's association with their defined presence group (for example, Contractors watching Executives is disallowed, but Executives watching Contractors is allowed). The presence group user setting applies for originating subscriptions when the user is logged in to the device via Extension Mobility or when the user is administratively assigned to the device.

When multiple presence groups are defined, the Inter-Presence Group Subscribe Policy service parameter is used. If one group has a relationship to another group via the Use System Default setting rather than being allowed or disallowed, this service parameter's value will take effect. If the Inter-Presence Group Subscribe Policy service parameter is set to **Disallowed**, Unified CM will block

the request even if the SUBSCRIBE calling search space allows it. The Inter-Presence Group Subscribe Policy service parameter applies only for presence status with call history lists and is not used for BLF speed dials.

Presence groups can list all associated directory numbers, users, and devices if you enable dependency records. Dependency records allow the administrator to find specific information about group-level settings. However, use caution when enabling the Dependency Record Enterprise parameter because it could lead to high CPU usage.

Unified CM Presence Guidelines

Unified CM enables the system administrator to configure and control user presence capabilities from within Unified CM Administration. Observe the following guidelines when configuring presence within Unified CM:

- Select the appropriate model of Cisco Unified IP Phones that have the ability to display user presence status.
- Define a presence policy for presence users.
 - Use SUBSCRIBE calling search spaces to control the routing of a watcher presence-based SIP SUBSCRIBE message to the correct destinations.
 - Use presence groups to define sets of similar users and to define whether presence status updates of other user groups are allowed or disallowed.
- Call history list presence capabilities are enabled on a global basis; however, user status can be secured by using a presence policy.
- BLF speed dials are administratively controlled and are not impacted by the presence policy configuration.

Cisco Unified Presence Server

The Cisco Unified Presence Server uses standards-based SIP and SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE) to provide a common demarcation and point for integrating all SIP or SIMPLE applications into the Cisco Unified Communications System. The Cisco Unified Presence Server collects, aggregates, and distributes user capabilities and attributes using this standards-based SIP and SIMPLE interface.

The core components of the Cisco Unified Presence Server consist of the SIP Presence Engine, which collects information regarding a user's availability status and communications capabilities, and the SIP Proxy/Registrar for the routing of presence-related and general SIP messaging. Applications (either Cisco or third-party) can integrate presence and provide services that improve the end user experience and efficiency. By default, the Cisco Unified Presence Server contains the IP Phone Messenger application to allow for instant messaging and presence status using Cisco Unified IP Phones.

The Cisco Unified Presence Server also contains support for interoperability with Microsoft Live Communications Server 2005 and the Microsoft Office Communicator client for any Cisco Unified IP Phone connected to a Unified CM. The Microsoft Office Communicator client interoperability includes click-to-dial functionality, phone control capability, and presence status of Cisco Unified IP Phones.

Cisco Unified Presence Server Cluster

The Cisco Unified Presence Server uses the same underlying appliance model and hardware as used by Unified CM, including a similar administration interface. For details on the supported platforms, refer to the *Cisco Unified Presence Server Administration Guide*, available at

http://www.cisco.com/en/US/products/ps6837/prod_maintenance_guides_list.html

The Cisco Unified Presence Server consists of two servers, a publisher and a subscriber, based on the same architecture as the Unified CM publisher and subscriber. However, a Cisco Unified Presence Server publisher and subscriber form their own cluster and are not formally integrated as part of the Unified CM cluster. The Cisco Unified Presence Server publisher utilizes and builds upon the database used by the Unified CM publisher by sharing the user and device information. A Cisco Unified Presence Server cluster supports only a single Unified CM cluster; therefore, all users of the Cisco Unified Presence Server must be defined within the same Unified CM cluster with Release 5.0(4) or higher.

Intracuster traffic participates at a very low level between the Cisco Unified Presence Server and Unified CM and between the Cisco Unified Presence Server publisher and subscriber. Both clusters share a common hosts file and have a strong trust relationship using IPTables. At the level of the database and services, the clusters are separate and distinct, and each Cisco Unified Presence Server and Unified CM cluster requires separate administration. There is currently no Transport Layer Security (TLS) or IPsec utilization for intracuster traffic.

The Cisco Unified Presence Server interface with external systems sends SIP traffic over UDP, TCP, or TLS. TLS mutual authentication requires the import and export of certificates between Cisco Unified Presence Server and the external system. TLS server authentication (Cisco Unified Presence Server presenting its TLS certificate to the client device for verification) validates the end user via HTTP digest authentication.

The Cisco Unified Presence Server publisher communicates directly with the Unified CM publisher via the AVVID XML Layer Application Program Interface (AXL API) using the Simple Object Access Protocol (SOAP) interface. When first configured, the Cisco Unified Presence Server publisher performs an initial synchronization of the entire Unified CM user and device database. All Cisco Unified Presence users are configured in the Unified CM End User configuration. During the synchronization, the Cisco Unified Presence Server populates these users in its database from the Unified CM database and does not provide end-user configuration from its administration interface.

The initial Cisco Unified Presence Server database synchronization from Unified CM might take a while, depending on the amount of information in the database as well as the load that is currently on the system. Subsequent database synchronizations from Unified CM to Cisco Unified Presence Server are performed in real time when any new user or device information is added to Unified CM. For planning purposes, use the values in [Table 21-1](#) as guidelines when executing the initial database synchronization with Unified CM using a single Cisco Unified Presence Server publisher.

Table 21-1 Synchronization Times for a Cisco Unified Presence Server Publisher

Server Platform	Number of Users	Synchronization Time
Cisco MCS 7825	1000	30 minutes
Cisco MCS 7835	1000	20 minutes
	2500	75 minutes
Cisco MCS 7845	1000	15 minutes
	2500	60 minutes
	30000	13 hours

For planning purposes, use the values in [Table 21-2](#) as guidelines when executing the initial database synchronization with Unified CM using a Cisco Unified Presence Server publisher and subscriber:

Table 21-2 Synchronization Times for a Cisco Unified Presence Server Publisher and Subscriber

Server Platform	Number of Users	Synchronization Time
Cisco MCS 7825	1000	60 minutes
Cisco MCS 7835	1000	40 minutes
	2500	150 minutes
Cisco MCS 7845	1000	30 minutes
	2500	120 minutes
	30000	26 hours



Note

When the Cisco Unified Presence Server is performing the initial database synchronization from Unified CM, do not perform any administrative activities on Unified CM while the synchronization agent is active.

If the database entries are not updating, you can check for broken connections with the synchronization agent by using the Real-Time Monitoring Tool (RTMT) to monitor the Critical Alarm **Cisco Unified Presence ServerSyncAgentAXLConnectionFailed**.

Cisco Unified Presence Server Redundancy

Unified CM provides a choice of the following optional 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.

For more information on Unified CM redundancy, see the chapter on [Call Processing, page 8-1](#).

The Cisco Unified Presence Server 1.0 cluster consists of only the Cisco Unified Presence Server publisher and subscriber, with no redundancy for either server. What this means is that, if one Cisco Unified Presence Server fails, the users associated with that failed server will not automatically fail-over to the other Cisco Unified Presence Server.

A Cisco Unified Presence Server cluster has two servers primarily for load balancing, which allows for the processing power to be scaled beyond the capacity of a single server to support a larger number of users. A Cisco Unified Presence Server cluster can initially be viewed as each server containing the same user information and services, but it currently does not allow for true redundancy.

Cisco Unified Presence Server Deployment Models

Unified CM provides a choice of the following deployment models:

- Single site
- Multisite WAN with centralized call processing
- Multisite WAN with distributed call processing

- Clustering over the WAN

Cisco Unified Presence Server Release 1.0 is supported with centralized call processing in the single-site and multisite deployment models; however, both the Cisco Unified Presence Server publisher and subscriber must be co-located with the Unified CM publisher. The Cisco Unified Presence Server 1.0 cluster is *not* supported in multisite Unified CM deployments with distributed call processing or in deployments with clustering over the WAN.

For more information on Unified CM deployment models, see the chapter on [Deployment Models](#), page 2-1.

Cisco Unified Presence Server Performance

Cisco Unified Presence Server clusters support publisher-only or publisher-and-subscriber configurations. However, if a subscriber is used, it must be on the same type of server platform as the publisher.

[Table 21-3](#) lists the hardware platform requirements for the Cisco Unified Presence Server as well as the maximum number of users supported per platform. (For example, if a Cisco Unified Presence Server publisher and subscriber are deployed on MCS 7845 platforms, a total of 10,000 users would be supported.)

Table 21-3 *Cisco Unified Presence Server Platforms and Number of Users Supported*

Server Platform	CPU	Physical Memory	Hard Disk Capacity	Users Supported Per Platform
Cisco MCS 7825	One 2.8 GHz P4	2 GB	Dual 80 GB	1000
Cisco MCS 7835	One 3.4 GHz Xeon	2 GB	Dual 72 GB	2000
Cisco MCS 7845	Two 3.4 GHz Xeon	4 GB	Dual 72 GB	5000

Cisco Unified Presence Server Licensing

User presence capabilities are assigned from Unified CM Administration under the licensing capabilities assignment. Because users are licensed for presence capabilities from Unified CM, the Cisco Unified Presence Server requires integration with Cisco Unified CM Release 5.0(4) or higher.

There are two checkboxes, one for Unified Presence Server and one for Unified Personal Communicator. To enable the user to send and receive presence SIP messaging updates, the Unified Presence Server checkbox must be enabled for that user. If the user is not enabled for Unified Presence Server, no presence messaging or status updates will be allowed for that user. To enable the use of Cisco Unified Personal Communicator, the Unified Personal Communicator checkbox must be enabled for the user.

Each checkbox (Enable UPS and Enable UPC) for the user will consume a Device License Unit. You can use the License Unit Calculator on Unified CM to view a report of the number of Device License Units consumed (that is, the number of users enabled for presence). For complete information regarding licensing on Unified CM, refer to the *Cisco Unified Communications Manager Administration Guide*, available at

<http://www.cisco.com>

Cisco Unified Presence Server Deployment

Figure 21-5 represents the interfaces between Cisco Unified Presence Server components and the interactions between those components for basic functionality. For complete information on Cisco Unified Presence Server administration and configuration, refer to the Cisco Unified Presence Server installation, administration, and configuration guides, available at

http://www.cisco.com/en/US/products/ps6837/tsd_products_support_series_home.html

Figure 21-5 Interactions Between Cisco Unified Presence Server Components

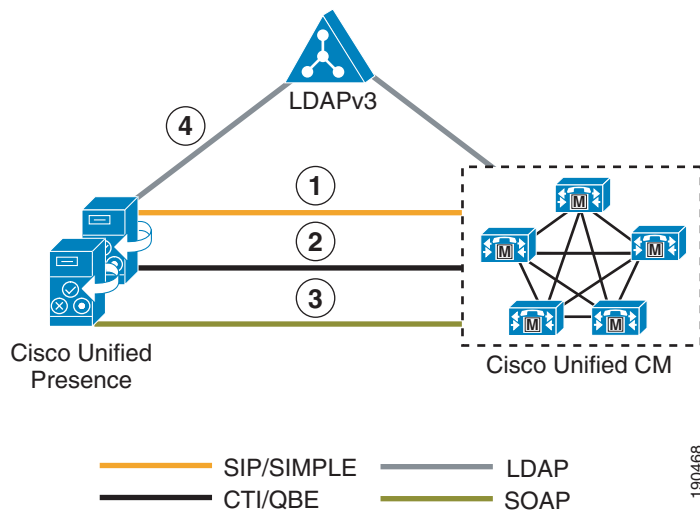


Figure 21-5 depicts the following interactions between Cisco Unified Presence Server components:

1. The SIP connection between the Cisco Unified Presence Server and Unified CM handles all the presence information exchange.
 - a. Unified CM configuration requires the Cisco Unified Presence Server publisher and subscriber to be added as application servers on Unified CM and also requires a SIP trunk pointing to the Cisco Unified Presence Server. The address configured on the SIP trunk could be a Domain Name System (DNS) server (SRV) fully qualified domain name (FQDN) that resolves to the Cisco Unified Presence Server publisher and subscriber, or it could simply be an IP address of the Cisco Unified Presence Server publisher or subscriber.
 - b. Configuration of the Cisco Unified Presence Server 1.0(1) occurs through the Backend Gateway for presence information exchange with Unified CM. The following information is configured:
Pattern token: *ccm.domain.com*



Note The *domain.com* must match the Proxy Domain service parameter.

Gateway: *server_fqdn:5070;transport=tcp*



Note The *server_fqdn* could be the FQDN of the Unified CM publisher, a DNS SRV FQDN that resolves to the Unified CM subscriber servers, or an IP address.

Configuration of the Cisco Unified Presence Server 1.0(2) or higher occurs through the Unified CM Presence Gateway for presence information exchange with Unified CM. The following information is configured:

CallManager Presence Gateway: *server_fqdn:5070*



Note The *server_fqdn* could be the FQDN of the Unified CM publisher, a DNS SRV FQDN that resolves to the Unified CM subscriber servers, or an IP address.

If DNS is highly available within your network and DNS SRV is an option, configure the SIP trunk on Unified CM with a DNS SRV FQDN of the Cisco Unified Presence Server publisher and subscriber. Also configure the Presence Gateway on the Cisco Unified Presence Server with a DNS SRV FQDN of the Unified CM subscribers, equally weighted. This configuration will allow for presence messaging to be shared equally among all the servers used for presence information exchange.

If DNS is not highly available or not a viable option within your network, use IP addressing. When using an IP address, presence messaging traffic cannot be equally shared across multiple subscribers because it points to a single subscriber.

2. The Computer Telephony Integration Quick Buffer Encoding (CTI-QBE) connection between Cisco Unified Presence Server and Unified CM handles all the CTI communication for users on the Cisco Unified Presence Server to control phones on Unified CM. This CTI communication occurs when Cisco Unified Personal Communicator is using Desk Phone mode to do Click to Call or when Microsoft Office Communicator is doing Click to Call through Microsoft Live Communications Server 2005.
 - a. Unified CM configuration requires the user to be associated with a CTI Enabled Group, and the primary extension assigned to that user must be enabled for CTI control (checkbox on the Directory Number page). The CTI Manager Service must also be activated on each of the Unified CM subscribers used for communication with the Cisco Unified Presence Server publisher and subscriber. Integration with Microsoft Live Communications Server 2005 requires that you configure an Application User, with CTI Enabled Group and Role, on Unified CM.
 - b. Cisco Unified Presence Server CTI configuration (CTI Server and Profile) for use with Cisco Unified Personal Communicator is automatically created during the database synchronization with Unified CM. All Cisco Unified Personal Communicator CTI communication occurs directly with Unified CM and not through the Cisco Unified Presence Server.

Cisco Unified Presence Server CTI configuration (CTI Gateway) for use with Microsoft Live Communications Server 2005 requires you to set the CTI Gateway for a primary address (Primary Unified CM CTI Manager), a failover address (Backup Unified CM CTI Manager), and a provider which is the application user configured previously in Unified CM. Only IP addresses can be used for CTI gateway configuration in the Cisco Unified Presence Server.
3. The AXL/SOAP interface handles the database synchronization from Unified CM to populate the Cisco Unified Presence Server database.
 - a. No additional configuration is required on Unified CM.
 - b. Cisco Unified Presence Server security configuration requires you to set a user and password for the Unified CM AXL account in the AXL configuration.
4. The LDAP interface is used for LDAP authentication of Cisco Unified Personal Communicator users during login. For more information regarding LDAP synchronization and authentication, see the chapter on [LDAP Directory Integration, page 17-1](#).

Unified CM is responsible for all synchronization of user information directly from LDAP, then the Cisco Unified Presence Server synchronizes all the user information from Unified CM. If a Cisco Unified Personal Communicator user logs into the Cisco Unified Presence Server and LDAP authentication is enabled on Unified CM, the Cisco Unified Presence Server will go directly to LDAP for the Cisco Unified Personal Communicator user authentication using the Bind operation. Once Cisco Unified Personal Communicator is authenticated, the Cisco Unified Presence Server forwards the information to Cisco Unified Personal Communicator to continue login.

Table 21-4 lists the ports used by the various protocol interfaces for Cisco Unified Presence Server.

Table 21-4 Cisco Unified Presence Server Port Usage

Protocol Interface	Port
SIP	5060
SIMPLE	5070
LDAP	389
LDAPS	636
CTI-QBE	2748
HTTP	80
HTTPS	443 or 8443
TLS (Server Authentication)	5061
TLS (Peer Authentication)	5062
Presence Engine Livebus Messaging	50000
Presence Engine Exchange Notification	50020
Presence Engine Management (oamagent)	60020
IPPM HTTP Listener	8081
DB Change Notification	8001
RISDC Server	2555
RISDC Client	2556
RISDC System Access (ServM)	8888 and 8889
SNMP Agent	161, 6161, 7161, or 8161
NTP	123
SSH/SFTP	22
DNS	53

Cisco Unified Presence Server Guidelines

- If LDAP integration is possible, LDAP synchronization with Unified CM should be used to pull all user information (number, ID, and so forth) from a single source. However, if there is both an LDAP server and Unified CM that do not have LDAP synchronization enabled, then the administrator should utilize consistent configuration across Unified CM and LDAP when configuring user directory number associations.

- Traffic marking has not been implemented fully with Cisco Unified Presence Server Release 1.0. Therefore, follow the traffic marking guidelines in the *Enterprise QoS Solution Reference Network Design (SRND)*, available at <http://www.cisco.com/go/designzone>
- Presence Policy for the Cisco Unified Presence Server currently cannot be configured.
- The Cisco Unified Presence Server publisher and subscriber must be co-located with the Unified CM publisher.

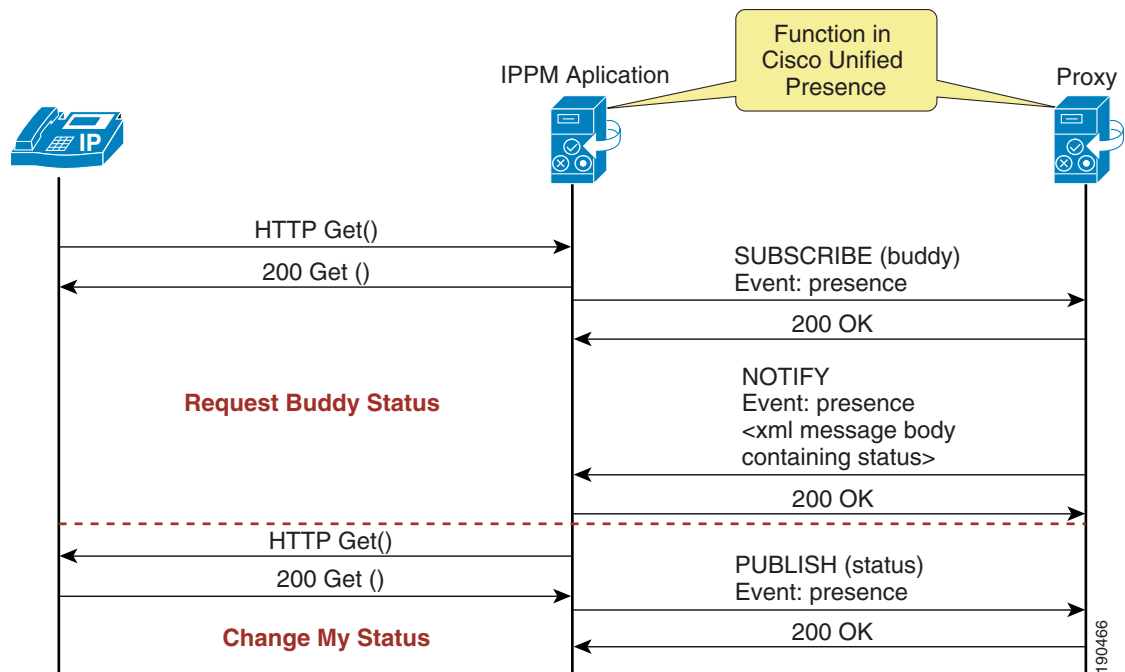
Cisco IP Phone Messenger Application

Cisco IP Phone Messenger is a Cisco Unified IP Phone Service that provides users with the ability to create a buddy list, watch their buddies' aggregated presence information, and exchange instant messages with their buddies' Cisco Unified IP Phone or compliant SIP or SIMPLE client or gateway.

The Cisco IP Phone Messenger (IPPM) application, which is a component of the Cisco Unified Presence Server, serves as a protocol translator between HTTP and SIP messaging. The IPPM application communicates with the Cisco Unified IP Phones using XML over HTTP (<http://www.cisco.com/go/apps>), and it communicates with the SIP Proxy/Registrar Server using SIP. The IPPM application can distinguish between two devices with the same directory number in different partitions and can also function when the user is logged in via Extension Mobility. However, it does rely on the availability of the Cisco Unified Presence Server publisher for new user logins.

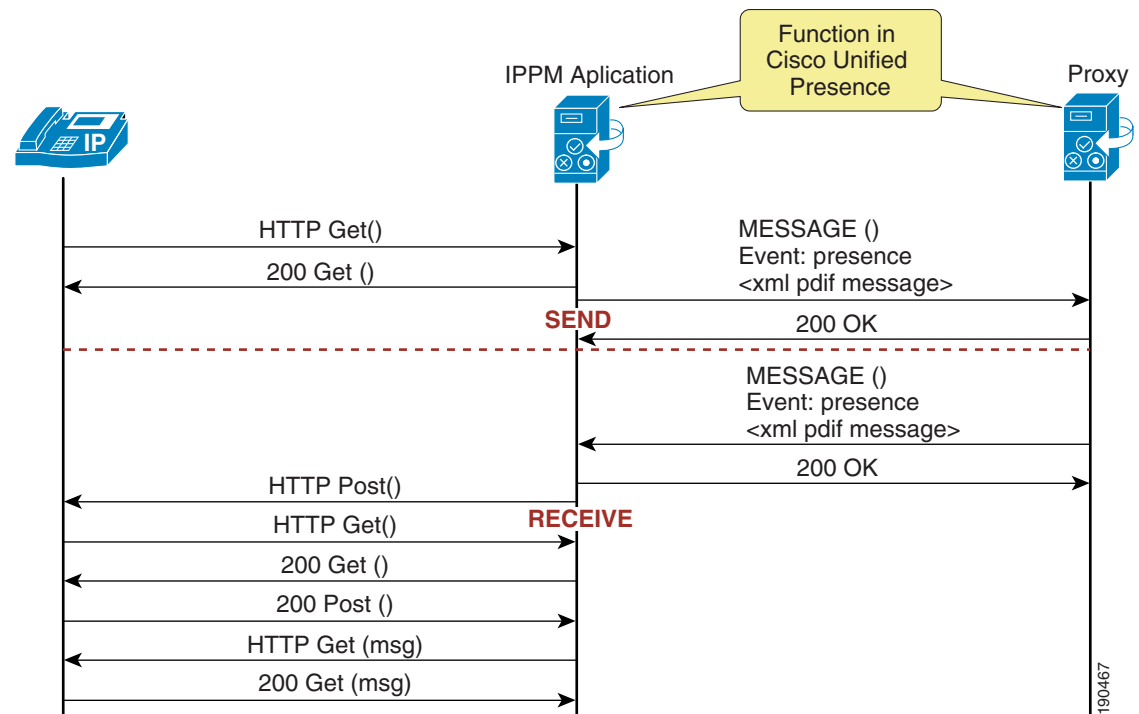
The IPPM application provides the following presence functionality (see [Figure 21-6](#)):

- Shows aggregated presence status of a buddy.
- Supports overriding the presence status manually (Available, Busy, Do Not Disturb).
- Upon phone login, SUBSCRIBE to all phone buddies' presence statuses. Upon phone logout, SUBSCRIBE with Expires=0 (terminate subscription).
- Update buddy presence status in the IPPM application upon receipt of a NOTIFY message from the presence engine.
- Manage the contact list from both the phone (Phone Messenger Service) and the web user interface (http://presence_server_address/ccmuser).

Figure 21-6 *IPPM Protocol Translation Presence*

The IPPM application provides the following instant messaging (IM) functionality (see [Figure 21-7](#)):

- Translates HTTP instant messages from the phone to outbound SIP MESSAGE messages.
- Translates inbound SIP MESSAGE messages to HTTP instant messages to the phone.
- Provides the ability to dial-back a buddy from either the buddy information screen or the IM screen.
- Manages the message history from the phone (Phone Messenger Service).
- Gives you the ability to configure canned system-wide and personal IM responses as well as the ability to compose messages.

Figure 21-7 IPPM Protocol Translation Instant Message

The following Cisco Unified IP Phones support Cisco IP Phone Messenger with SCCP:

- Cisco Unified IP Phone 7905G
- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phone 7912G
- Cisco Unified IP Phone 7940G
- Cisco Unified IP Phone 7960G
- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE
- Cisco Unified IP Phone 7920
- Cisco IP Communicator

The following Cisco Unified IP Phones support Cisco IP Phone Messenger with SIP:

- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

IP Phone Messenger is not supported with SIP on the Cisco IP Phones 7905, 7912, 7940, and 7960.

Current IP Phone Services in the Cisco Unified Communications System are configured with either an IP address or DNS A record entry for the HTTP Service URL, which can result in a single point of failure if no IP Phone Service redundancy is configured.

Without IP Phone Services redundancy, the IP Phone Messenger deployment should be load-balanced via configuration across both the Cisco Unified Presence Server publisher and subscriber.

On Unified CM, configure two phone services for IP Phone Messenger, one that uses the Cisco Unified Presence Server publisher and one that uses the Cisco Unified Presence Server subscriber, as illustrated in the following example:

- PhoneMessenger1:
`http://publisher.cups.com:8081/ippm/default?name=#DEVICENAME#`
- PhoneMessenger2:
`http://subscriber.cups.com:8081/ippm/default?name=#DEVICENAME#`

With Cisco IP Phone Messenger, you can deploy the Cisco Unified IP Phones in either of the following ways:

- Single phone service

With the single phone service, you configure half of the Cisco Unified IP Phones to point to the Cisco Unified Presence Server publisher (PhoneMessenger1 in the example above), while the other half is configured to point to the Cisco Unified Presence Server subscriber (PhoneMessenger2 in the example above).

Advantage — The administrator load-balances the IP Phone Messenger users via configuration.

Disadvantage — If the Cisco Unified Presence Server node running that phone service fails, the IP Phone Messenger service is unavailable to the users.

- Dual phone service

With the dual phone service, you configure all Cisco Unified IP Phones to have two IP Phone Messenger services (both PhoneMessenger1 and PhoneMessenger2 in the above example).

Advantage — If the Cisco Unified Presence Server node running the phone service fails, the user can try using the IP Phone Messenger service running on the second node.

Disadvantage — This method relies on the phone user to pick the IP Phone Messenger service to use from the Services menu. This method potentially leads to one Cisco Unified Presence Server node being selected more than the other, thus resulting in a disproportionate number of users on one particular Cisco Unified Presence Server node.

With IP Phone Services redundancy (see [IP Phone Services Redundancy, page 22-6](#)), IP Phone Messenger can be configured on Unified CM as a single phone service using the server load balancer (SLB) IP address, as illustrated in the following example:

- PhoneMessenger:
`http://slb_ip_address:8081/ippm/default?name=#DEVICENAME#`

Cisco IP Phone Messenger Bandwidth Considerations

The user message history and contact list are both stored in the Cisco Unified Presence Server database and have the potential to contain a lot of data. Every user login to the IP Phone Messenger application will download the message history and contact list. Therefore, if bandwidth might be a concern, you can

limit the message history size and contact list size via Cisco Unified Presence Server administration by setting the **Max Instant Message History Size** and **Max Contact List Size** under the IP Phone Messenger settings.

The user has the ability to set a Session Timer parameter for controlling the amount of time the user is logged into the current session and a Refresh Interval parameter for controlling the rate at which presence status updates occur. The administrator currently has no control over these parameters; therefore, the default settings (Session Timer = 480 minutes and Refresh Interval = 30 seconds) are most likely to be used.

Cisco Unified Personal Communicator

Cisco Unified Personal Communicator integrates the most frequently used communications applications and services into a single desktop software application. Cisco Unified Personal Communicator leverages Unified CM for IP Telephony, Cisco Unity Connection for voice messaging, Cisco Unified MeetingPlace Express for web conferencing, and the Cisco Unified Presence Server for services and presence information.

Cisco Unified Personal Communicator operates in one of two modes, Desk Phone (CTI control of the user's desk phone for Click to Call) and Soft Phone (software client operation). Cisco Unified Personal Communicator call control features are similar to the features on Cisco Unified IP Phones. For a complete list of features supported by Cisco Unified Personal Communicator, see the chapter on [IP Telephony Endpoints](#), page 20-1.

Cisco Unified Personal Communicator Deployment

[Figure 21-8](#) shows the interfaces and various components of Cisco Unified Personal Communicator, and lists the ports used by Cisco Unified Personal Communicator for the protocol exchanges. Cisco Unified Personal Communicator is supported only with centralized call processing deployments of Unified CM. For more information on Cisco Unified Personal Communicator, refer to the Cisco Unified Personal Communicator installation, operation, and maintenance guides, available at

http://www.cisco.com/en/US/products/ps6844/tsd_products_support_series_home.html

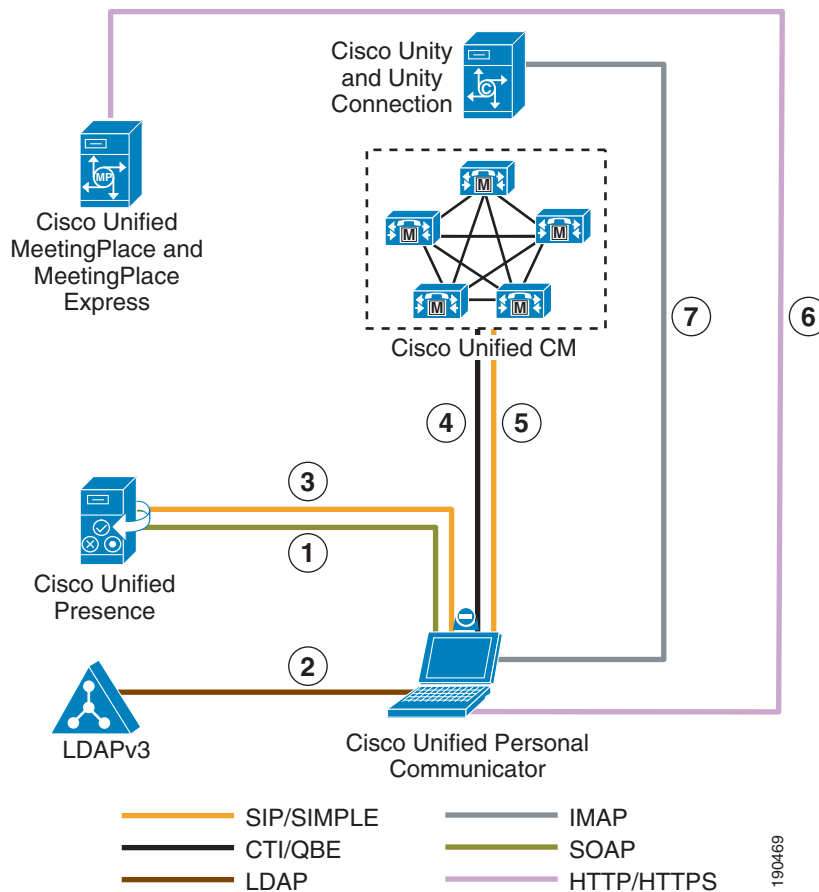
Figure 21-8 Cisco Unified Personal Communicator Components

Figure 21-8 depicts the following interactions between the components of Cisco Unified Personal Communicator:

1. Through the SOAP interface, a Cisco Unified Personal Communicator user logs in using TLS and retrieves the configuration profiles for the various interface components. Cisco Unified Presence Server is not an authenticated server with Cisco Unified Personal Communicator because no certificate validation is performed. However, this TLS connection protects all configuration information exchanged between Cisco Unified Personal Communicator and Cisco Unified Presence Server. If the Cisco Unified Personal Communicator user is authenticated via LDAP, Cisco Unified Presence Server does the bind to LDAP for the user during this login period before proceeding.
2. Cisco Unified Personal Communicator binds to the LDAP server for all the user-specific information (telephone number, contact, and so forth).
3. Cisco Unified Personal Communicator sends a SIP REGISTER for the user and a SIP SUBSCRIBE for all the users in the contact list, to the Cisco Unified Presence Server. A SIP NOTIFY is received from Cisco Unified Presence Server when a user status in the contact list needs to be updated. A SIP PUBLISH is sent when the user status is changed either manually or automatically.
4. If Cisco Unified Personal Communicator is configured for Desk Phone mode, a connection is established with the CTI Manager of Unified CM for phone control.

5. Cisco Unified Personal Communicator sends a SIP REGISTER to Unified CM to allow for media exchanges (voice and video), and it receives a SIP NOTIFY for Message Waiting Indication (MWI) status.
6. Calls between two Cisco Unified Personal Communicator users can be escalated, from within the conversation, to Cisco Unified MeetingPlace Express for a share-only web collaboration session using HTTP. Each connection will utilize one MeetingPlace User License, and the user initiating the escalation must have a MeetingPlace Express Profile.
7. Calls initiated to voicemail are communicated using IMAP with Cisco Unity Connection.

All SIP traffic, presence, and call setup between Cisco Unified Personal Communicator and Cisco Unified Presence Server, as well as Cisco Unified Personal Communicator and Unified CM, is not encrypted and is done via TCP or UDP. Cisco Unified Personal Communicator uses a standard SIMPLE interface for presence information exchange.

The Cisco Unified Personal Communicator deployment should be load-balanced via configuration across both the Cisco Unified Presence Server publisher and subscriber.

On the Cisco Unified Presence Server, configure two Cisco Unified Personal Communicator proxy profiles, one that uses the Cisco Unified Presence Server publisher as the primary proxy server and the Cisco Unified Presence Server subscriber as the backup proxy server, while the other uses the Cisco Unified Presence Server subscriber as the primary proxy server and the Cisco Unified Presence Server publisher as the backup proxy server, as illustrated in the following example:

- ProxyProfile1:
 - Primary Proxy Server: Cisco Unified Presence Server Publisher
 - Backup Proxy Server: Cisco Unified Presence Server Subscriber
- ProxyProfile2:
 - Primary Proxy Server: Cisco Unified Presence Server Subscriber
 - Backup Proxy Server: Cisco Unified Presence Server Publisher

Configure half of the Cisco Unified Personal Communicator users to point to ProxyProfile1 and configure the other half to point to ProxyProfile2.

Advantage — The administrator load-balances the Cisco Unified Personal Communicator users via configuration.

Disadvantage — If the Cisco Unified Presence Server node running that service fails, Cisco Unified Personal Communicator users can all try to use the working Cisco Unified Presence Server node.

Table 21-5 lists the Cisco Unified Presence Server ports used by the various protocol interfaces for Cisco Unified Personal Communicator.

Table 21-5 Cisco Unified Presence Server Port Usage for Cisco Unified Personal Communicator

Protocol Interface	Port
SIP or SIMPLE (outbound)	5060
SIP or SIMPLE (inbound)	50000 to 50063
LDAP	389
CTI-QBE	2748
HTTP or HTTPS	80
SOAP	443
TFTP	69

Table 21-5 Cisco Unified Presence Server Port Usage for Cisco Unified Personal Communicator

Protocol Interface	Port
IMAP	143
RTP	16384 to 16424

Design Considerations for Cisco Unified Personal Communicator

Cisco Unified Personal Communicator marks Layer 3 IP packets via Differentiated Services Code Point (DSCP). The Cisco Unified Personal Communicator client marks call signaling traffic with a DSCP value of 26, or a per-hop behavior (PHB) value of AF31, and it marks RTP media traffic with a DSCP value of 46 (PHB value of EF). However, personal computer traffic is typically untrusted, and the network will strip DSCP markings made by an application from the PC. Therefore, access routers and switches must be configured to allow these DSCP markings for the port ranges that Cisco Unified Personal Communicator utilizes. For details on traffic marking, refer to the *Enterprise QoS Solution Reference Network Design (SRND)*, available at

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

Cisco Unified Personal Communicator interfaces with Unified CM. Therefore, the following guidelines for the current functionality of Unified CM apply when Cisco Unified Personal Communicator voice or video calls are initiated:

- CTI scalability

In Desk Phone mode, calls from Cisco Unified Personal Communicator use the CTI interface on Unified CM. Therefore, observe the CTI limits as defined in the chapter on [Call Processing](#), [page 8-1](#). You must include these CTI devices when sizing Unified CM clusters.

- Call admission control

Cisco Unified Personal Communicator applies call admission control for voice and video calls via Unified CM locations or RSVP.

- Codec selection

Cisco Unified Personal Communicator voice and video calls utilize codec selection through the Unified CM regions configurations.

- Dial rules

Cisco Unified Personal Communicator does not download or store any local dial rules. Therefore, you might have to configure application dial rules on Unified CM to maintain the dial plans. (For example, if a user dials 18005551212 but you need only five digits, Unified CM can apply the Application Dial Rule: "Eleven-digit number beginning with 1 = retain only the last five digits.")

The following bandwidth considerations also apply to Cisco Unified Personal Communicator:

- All Cisco Unified Personal Communicator configuration and contacts are stored in the Cisco Unified Presence Server database and have the potential to contain a lot of data. The current communications list is limited to 50 users, while the contact list size has no limit. Therefore, bandwidth utilization for presence data exchange must be taken into consideration.
- For Cisco Unified MeetingPlace Express web collaboration sessions, see [Cisco Unified MeetingPlace Express](#), [page 15-1](#).
- For video calls, see the chapter on [IP Video Telephony](#), [page 16-1](#).
- For Cisco Unity Connection, see the section on [Managing Bandwidth](#), [page 13-6](#), in the chapter on [Cisco Unity](#), [page 13-1](#).

Third-Party Presence Server Integration

Cisco Unified Presence Server provides an interface based on SIP and SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE) for integrating applications into the Cisco Unified Communications System. You can configure and integrate a third-party presence server or application with this SIP/SIMPLE interface to provide presence aggregation and federation.

Microsoft Live Communications Server 2005

For setup, configuration, and deployment of Microsoft Live Communications Server 2005 and Microsoft Office Communicator, refer to the documentation at

<http://www.microsoft.com/>

Cisco does not recommend configuration, deployment, or best practice procedures for Microsoft Communications products, but Cisco does provide the following guidelines for integrating Cisco Unified Presence Server with Microsoft Live Communications Server 2005.

Guidelines for Integrating Cisco Unified Presence Server with Microsoft Live Communications Server 2005

- Communications between Cisco Unified Presence Server and Microsoft Live Communications Server 2005 uses the SIP/SIMPLE interface. However, Microsoft Live Communications Server 2005 tunnels Computer-Supported Telecommunications Applications (CSTA) traffic over SIP. Therefore, the CTI gateway on the Cisco Unified Presence Server must be configured to handle the CSTA-to-CTI conversion for Click to Call phone control.
- You must configure the same end-user ID in LDAP, Unified CM, and Microsoft Live Communications Server 2005. This practice avoids any conflicts between Microsoft Live Communications Server 2005 authentication with Active Directory (AD) and end-user configuration on Unified CM, as well as conflicts with user phone control on Unified CM.
For Active Directory, Cisco recommends that the user properties of General, Account, and Live Communications to all have the same ID.
- You must configure Microsoft Live Communications Server 2005 Host Authentication to contain the Cisco Unified Presence Server publisher and subscriber.
- You can configure routing of the SIP messages to the Cisco Unified Presence Server by means of Static Routes in the Microsoft Live Communications Server 2005 properties.
- You must configure an incoming access control list (ACL) on the Cisco Unified Presence Server to allow for communications with Microsoft Live Communications Server 2005.
- You must enable each user for Microsoft Office Communicator (MOC) in the Cisco Unified Presence Server CTI Gateway configuration.

IBM Sametime 7.5

For setup, configuration, and deployment of IBM Sametime Server 7.5, refer to the documentation at

<http://www.ibm.com/services/>



CHAPTER 22

Cisco Unified CM Applications

Last revised on: September 27, 2007

Cisco Unified Communications Manager (Unified CM) applications provide numerous operational and functional enhancements to basic IP telephony. External eXtensible Markup Language (XML) productivity applications or IP Phone Services can be run on the web server and/or client on most Cisco Unified IP Phones. For example, the IP phone on a user's desk can be used to get stock quotes, weather information, flight information, and other types of web-based information. In addition, custom IP phone service applications can be written that allow users to track inventory, bill customers for time, or control conference room environments (lights, video screen, temperature, and so forth). Unified CM also has a number of integrated applications that provide additional functionality, including:

- **Cisco Extension Mobility (EM)**
The Extension Mobility feature enables mobile users to configure a Cisco Unified IP Phone as their own, on a temporary basis, by logging in to that phone.
- **Cisco Unified Communications Manager Assistant (Unified CM Assistant)**
Unified CM Assistant is a Unified CM integrated application that enables assistants to handle one or more managers' incoming phone calls.
- **Cisco Unified Communications Manager Attendant Console (AC)**
The Attendant Console enables one or more receptionists to answer and transfer (or dispatch) calls within an organization.
- **Cisco WebDialer**
WebDialer is a click-to-dial application for Unified CM that enables users to place calls easily from their PCs using any supported phone device.

In some cases these integrated applications also invoke IP Phone Services to provide additional functionality.

This chapter examines the following Unified CM applications:

- [IP Phone Services, page 22-2](#)
- [Extension Mobility \(EM\), page 22-7](#)
- [Unified CM Assistant, page 22-13](#)
- [Attendant Console, page 22-28](#)
- [WebDialer, page 22-39](#)

IP Phone Services

Cisco Unified IP Phone Services are applications that utilize the web client and/or server and XML capabilities of the Cisco Unified IP Phone. The Cisco Unified IP Phone firmware contains a micro-browser that enables limited web browsing capability. These phone service applications provide the potential for value-added services and productivity enhancement by running directly on the user's desktop phone. For purposes of this chapter, the term *phone service* refers to an application that transmits and receives content to and from the Cisco Unified IP Phone.

IP Phone Services Phone Support

The following phones support IP Phone Services:

- Cisco Unified Wireless IP Phone 7921G
- Cisco Unified IP Phones 7940G, 7941G and 7941G-GE
- Cisco Unified IP Phones 7960G, 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G, and 7971G-GE

IP Phone Services can also run on the following IP phones, however these phone models support only text-based XML applications:

- Cisco Unified IP Phone 7905G
- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phones 7912G and 7912G-A
- Cisco Unified Wireless IP Phone 7920

All of the IP Phones listed above can process a limited set of Cisco-defined XML objects for enabling the user interface (UI) between the phone and the web server that contains the running phone service.

Note that the phones listed above support IP Phone Services for both the Skinny Client Control Protocol (SCCP) and Session Initiation Protocol (SIP). Beginning with Cisco Unified Communications Manager release 5.1.1(b), the Cisco Unified IP Phones 7942G, 7945G, 7962G, 7965G, and 7975G also support IP Phone Services with both SCCP and SIP loads.

Unified CM Services and IP Phone Service Enterprise Service Parameters

To enable IP Phone Services, the system administrator must ensure that the Cisco Unified IP Phone Service network service is enabled under the Unified CM Serviceability interface. In addition, the administrator can use service parameters within Unified CM to customize and configure certain aspects of Unified CM and application behavior. There are a number of enterprise service parameters that provide configuration and customization options for IP Phone Services, as described in the following sections.

Unified CM Services for IP Phone Services

IP Phone Services rely on the Cisco CallManager Cisco IP Phone Services network service on Unified CM in order to function. This feature is installed and activated by default when Unified CM is installed on a server.

IP Phone Service Enterprise Service Parameters

There are a number of pertinent Enterprise Service Parameters that relate to IP Phone Services. The following items represent a partial list of configuration parameters under the Phone URL Parameters section of the Unified CM Enterprise Service Parameters configuration page that relate to IP Phone Services and XML operation of IP Phones:

- URL Authentication (Default value = `http://<CM_IP_address>:8080/ccmcip/authenticate.jsp`)
This URL points to the `authenticate.jsp` service on Unified CM, which provides an authentication proxy service between Cisco Unified IP Phones and Unified CM. The URL is used to validate "push" requests made directly to the phone by the phone services. It is automatically configured at installation time. If no value is specified for this parameter, phone services will not be able to push content to the phone.
- URL Directories (Default value = `http://<CM_IP_address>:8080/ccmcip/xmldirectory.jsp`)
This URL points to the `xmldirectory.jsp` service on Unified CM, which generates and returns the directory menu presented when the user pushes the Directories (or book icon) button on the phone. The URL is automatically configured at installation time. If no value is specified for this parameter, the directory menu will not be available when the user pushes the Directories button.
- URL Idle (Default value = `<blank>`)
This URL, if specified, points to a service that provides text or images to be displayed on the phone screen when the phone is idle. This parameter is closely coupled with the URL Idle Time parameter, which indicates how long the phone must be idle before initiating the service. By default this parameter is left blank (not configured) at installation time.
- URL Idle Time (Default value = 0)
This parameter setting indicates the time in seconds that a phone will wait before initiating the URL Idle service. By default the parameter is set to 0 (zero) at installation time, indicating that the phone will never become idle.
- URL Information (Default value = `http://<CM_IP_address>:8080/ccmcip/GetTelecasterHelpText.jsp`)
This URL points to the `GetTelecasterHelpText.jsp` service on Unified CM, which generates and returns on-screen phone help for the phone keys and call statistics when the user presses the Help ("i" or "?") button (located to the right of the keypad). The URL is automatically configured at installation time. If no value is specified for this parameter, no help information will be displayed when the user pushes the Help button.
- URL Services (Default value = `http://<CM_IP_address>:8080/ccmcip/getservicesmenu.jsp`)
This URL points to the `getservicesmenu.jsp` service on Unified CM, which provides a list of user-subscribed phone services for the phone when the user presses the Services (or globe icon) button. It is automatically configured at installation time. If no value is specified for this parameter, a list of subscribed services will not be provided when the user pushes the Services button.

IP Phone Services Architecture

An IP Phone service can be initiated in several ways:

- User-initiated (pull)

An IP Phone user presses the Services button, which sends an HTTP GET message to Unified CM for displaying a list of user-subscribed phone services. [Figure 22-1](#) illustrates this functionality.

- Phone-initiated (pull)

An idle time value can be set within the IP Phone firmware, as indicated by the URL Idle Time parameter. When this timeout value is exceeded, the IP Phone firmware itself initiates an HTTP GET to the idle URL location specified by the URL Idle parameter.

- Phone service-initiated (push)

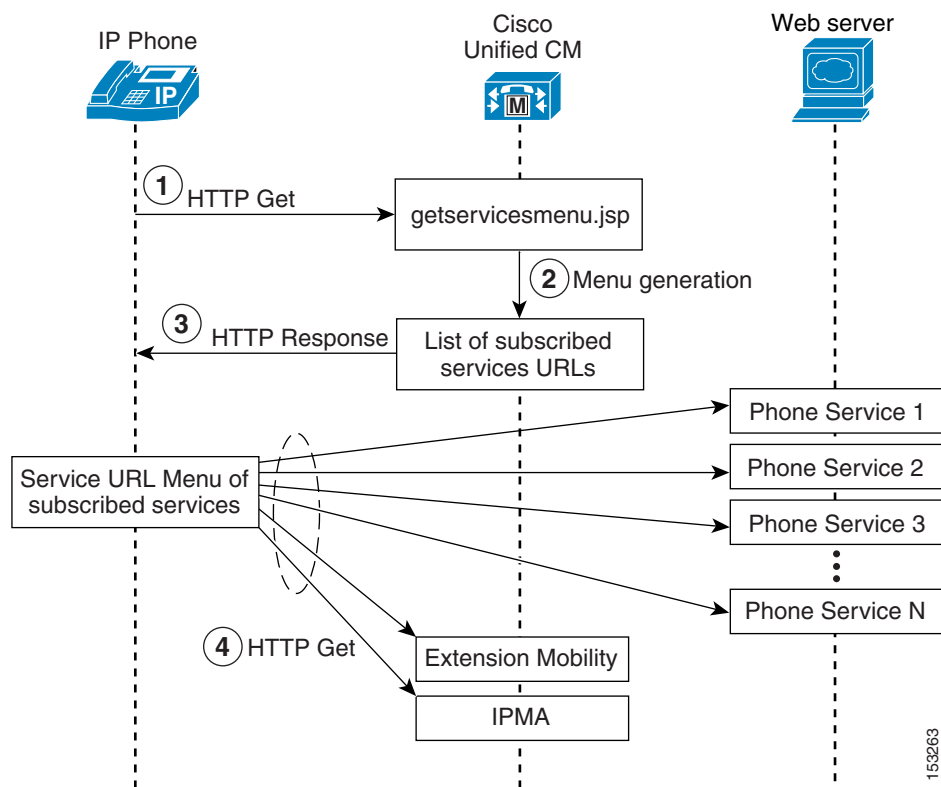
A phone service application can push content to the IP Phone by sending an HTTP POST message to the phone.

**Note**

Unlike with the user-initiated and phone-initiated pull functionality, whereby the phone's web client is used to invoke phone services, the phone service-initiated push functionality invokes action on the phone by posting content (via an HTTP POST) to the phone's web server (not to its client).

Figure 22-1 shows a detailed illustration of the user-initiated IP Phone service operation. When a user presses the Services button, an HTTP GET message is sent from the IP Phone to the Unified CM `getservicesmenu.jsp` script by default (step 1). You can specify a different script by changing the URL Services parameter (see [IP Phone Service Enterprise Service Parameters](#), page 22-3). The `getservicesmenu.jsp` script returns the list of phone service URL locations to which the individual user has subscribed (step 2). The HTTP response returns this list to the IP Phone (step 3). Any further phone service menu options chosen by the user continues the HTTP messaging between the user and the web server containing the selected phone service application (step 4).

Figure 22-1 User-Initiated IP Phone Service Architecture

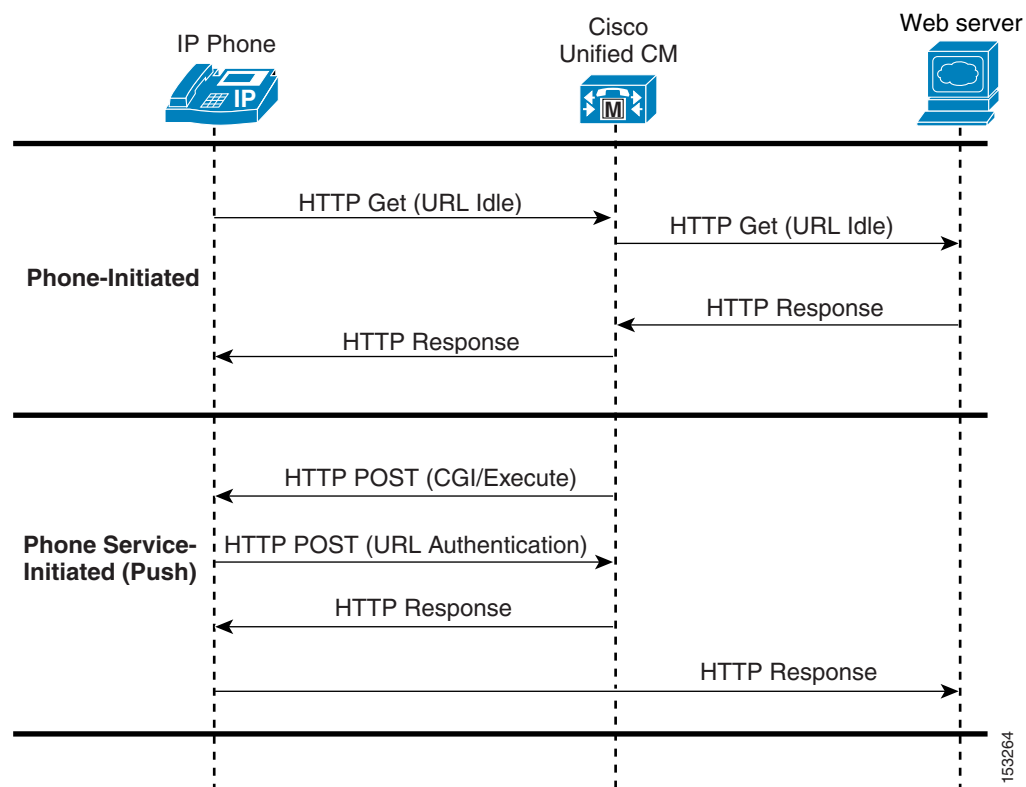


153263

Figure 22-2 shows examples of both phone-initiated and phone service-initiated push functionality. In the phone-initiated example, the phone automatically sends an HTTP GET to the location specified under the URL Idle parameter (see [IP Phone Service Enterprise Service Parameters, page 22-3](#)) when the URL Idle Time is reached. The HTTP GET is forwarded via Unified CM to the external web server. The web server sends back an HTTP Response, which is relayed by Unified CM back to the phone, and the phone displays the text and/or image on the screen.

In the phone service-initiated push example, the phone service on the external web server sends an HTTP POST with a Common Gateway Interface (CGI) or Execute call to the phone's web server. Before performing the CGI or Execute call, the phone authenticates the request using the proxy authentication service specified by the URL Authentication parameter (see [IP Phone Service Enterprise Service Parameters, page 22-3](#)). This proxy authentication service provides an interface between the phone and the Unified CM directory in order to validate requests made directly to the phone. If the request is authenticated, Unified CM forwards an HTTP Response to the phone. The phone's web server then performs the requested action, and the phone returns an HTTP response back to the external web server. If authentication fails, Unified CM forwards a negative HTTP Response, and the phone does not perform the requested CGI or Execute action but in turn forwards a negative HTTP Response to the external web server.

Figure 22-2 Phone-Initiated and Phone Service-Initiated IP Phone Service Architecture



IP Phone Services Redundancy

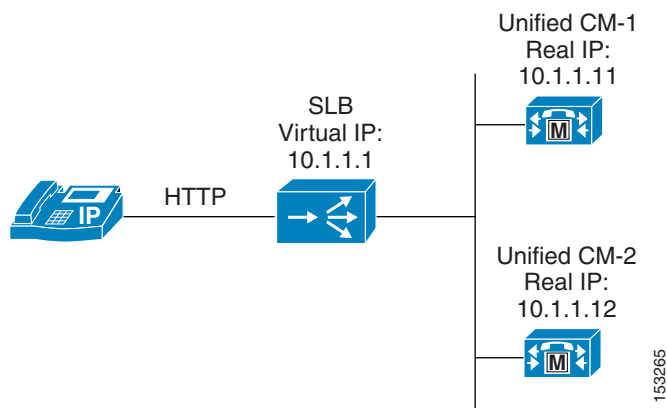
To ensure reliable services for phone users, you must maintain a high level of system availability, with a seamless transition to redundant systems during a system failure. While most of the back-end processing of a phone service occurs on a web server, the phones still depend upon Unified CM to redirect them to the phone service. And in the case of Extension Mobility and Unified CM Assistant phone services, the service actually runs on the Unified CM server(s).

Given the architecture of IP phone service functionality and the message flows shown in [Figure 22-1](#) and [Figure 22-2](#), there are two main failure scenarios to be considered:

Failure Scenario 1: Server with Unified CM Cisco Unified IP Phone Services Fails

Redundancy in this case depends upon some type of server load balancing (SLB), as illustrated in [Figure 22-3](#) where a virtual IP address is used to point to one or more Unified CM servers. This virtual IP address is used when configuring the URL Services parameter. Thus a Unified CM server failure does not prevent the IP Phone Services subscription list from being returned to the phone when the phone's Services button is pushed. In addition, phone services such as Extension Mobility and Unified CM Assistant that run on a Unified CM server are also potentially made redundant via this method.

Figure 22-3 Method for Providing Redundancy for Phone Services



Failure Scenario 2: External Web Server Hosting a Particular IP Phone Service Fails

In this scenario, the connection to the Unified CM server is preserved, but the link fails to the web server hosting the user-subscribed phone service. This is an easier scenario to provision for redundancy because the IP Phone is still able to access the Unified CM server when the Services button is pressed. In this case, the IP Phone is similar to any other HTTP client accessing a web server. As a result, you can again use some type of SLB functionality (similar to the one indicated in [Figure 22-3](#)) to redirect the HTTP request from the phone to one or more redundant web servers hosting the user-subscribe phone service.

IP Phone Services Scalability

Cisco Unified IP Phone Services act, for the most part, as an HTTP client. In most cases it uses Unified CM only as a redirect server to the location of the subscribed service. Because Unified CM acts as a redirect server to the phone service, there is minimal performance impact on Unified CM when a user initiates a phone service request by pressing the Services key.

**Note**

In the case of Extension Mobility and Unified CM Assistant phone service, Unified CM acts as more than a redirect server, and performance impacts should be considered. See the sections on [Extension Mobility \(EM\)](#), [page 22-7](#), and [Unified CM Assistant](#), [page 22-13](#), for specific performance and scalability considerations for these applications.

Because the IP Phone is either an HTTP client or server, estimating the required bandwidth used by an IP Phone service is similar to estimating the bandwidth of an HTTP browser accessing the same text as HTTP content residing on a web hosting server.

Guidelines and Restrictions for IP Phone Services

With the exception of the integrated Extension Mobility and Unified CM Assistant application's Phone Services, IP Phone services must reside on a separate web server. Running phone services other than Extension Mobility and IP Manager Application on the Unified CM server is not supported.

Extension Mobility (EM)

The Cisco Extension Mobility (EM) feature enables users to configure a Cisco Unified IP Phone as their own, on a temporary basis, by logging in to that phone. After a user logs in, the phone adopts the user's individual device profile information, including line numbers, speed dials, services links, and other user-specific properties of a phone. For example, when user X occupies a desk and logs in to the phone, that user's directory number(s), speed dials, and other properties appear on that phone; but when user Y uses the same desk at a different time, user Y's information appears. The EM feature dynamically configures a phone according to the authenticated user's device profile. The benefit of this application is that it allows users to be reached at their own extension on any phone within the Unified CM cluster, regardless of physical location, provided the phone supports EM.

EM Phone Support

The following Skinny Client Control Protocol (SCCP) phones support EM:

- Cisco Unified IP Phone 7905G
- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phones 7912G and 7912G-A
- Cisco Unified Wireless IP Phone 7920
- Cisco Unified Wireless IP Phone 7921G
- Cisco Unified IP Phones 7940G, 7941G, and 7941G-GE
- Cisco Unified IP Phones 7960G, 7961G, and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

The following Session Initiation Protocol (SIP) phones support EM:

- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G

- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

**Note**

EM is not supported on Cisco Unified IP Phones 7905G, 7912G, 7940G, or 7960G running SIP loads.

Beginning with Cisco Unified Communications Manager release 5.1.1(b), the Cisco Unified IP Phones 7942G, 7945G, 7962G, 7965G, and 7975G also support EM.

Unified CM Services and EM Service Parameters

To enable the EM application, the system administrator must activate and start a number of Unified CM services from the Unified CM Serviceability interface. In addition, EM service parameters provide configuration and customization options for determining how the EM application behaves.

Unified CM Services for EM

The EM application relies on the Cisco Extension Mobility feature service, which you must activate manually from the Serviceability page.

EM also relies on the following network services, which are activated automatically on Unified CM during installation:

- Cisco Extension Mobility Application
- Cisco CallManager Cisco IP Phone Services

The Cisco Extension Mobility Application service provides an interface between the EM user phone and the Cisco Extension Mobility service. In addition, the Cisco Extension Mobility Application service subscribes to the change notification indications within the cluster and maintains a list of nodes in the cluster that have an active Cisco Extension Mobility service. By subscribing to change notification within the cluster, the Cisco Tomcat network service and the Cisco Extension Mobility feature service do not have to be restarted after changes are made to the EM service parameters.

The Unified CM Cisco Unified IP Phone Services service is needed to provide access to the EM phone service. The URL used to define the EM phone service is:

`http://<Publisher_IP-Address>/emapp/EMAppServlet?device=#DEVICENAME#`

For example:

`http://10.1.1.1/emapp/EMAppServlet?device=#DEVICENAME#`

EM Service Parameters

The following items represent a partial list of Cisco EM Service Parameters related to Extension Mobility functionality:

- Enforce Maximum Login Time (Default value = False)

This parameter indicates whether EM users will be logged out automatically when the Maximum Login Time is reached. By default the value is set to False, meaning EM users are not automatically logged out.

- Maximum Login Time (Default value = 8:00)

This parameter indicates the number of hours and/or minutes (*hh:mm*) an EM user can stay logged in before being logged out automatically. Automatic logout at the time specified occurs only if the Enforce Maximum Login Time parameter is set to True.

- Multiple Login Behavior (Default value = Multiple Logins Not Allowed)

This parameter indicates whether an EM user is allowed to log in to more than one device at one time. By default multiple logins by a single user are not allowed, and attempts by a user to log in to another device while logged on to one device results in the following message:

```
Login Unsuccessful
[25]User logged in elsewhere.
```

- Remember the Last User Logged In (Default value = False)

This parameter indicates whether the last userID used to log in to a device will be remembered during subsequent attempts to log in to that same device. If this value is set to True, the last logged-in userID information is stored in a table in the Unified CM database for efficient retrieval. Upon a subsequent login attempt, the UserID field in the login screen on phone is pre-populated with the stored userID value.

- Clear the call log (Default value = False)

This parameter indicates whether the call logs specified for the Directories button menu are cleared at EM login and logout. This parameter affects the following logs: Missed Calls, Received Calls, and Placed Calls. If the value is set to True, then these logs are cleared during login and manual logout.

One exception is that these logs are not cleared when the user is logged out automatically. Therefore, logs are not cleared when the Maximum Login Time is reached (assuming Enforce Maximum Login Time is set to True) and the user is automatically logged out of the phone. Likewise, if a Unified CM Administrator clicks on the Log Out button under the Extension section of the phone/device configuration screen, the logs are not cleared.

For a complete list of Extension Mobility service parameters, consult the *Cisco Extension Mobility* chapter of the *Cisco Unified Communications Manager Features and Services Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

EM Architecture

Figure 22-4 depicts the message flows and architecture of the EM application. When a phone user wants to access the EM application, the following sequence of events occurs:

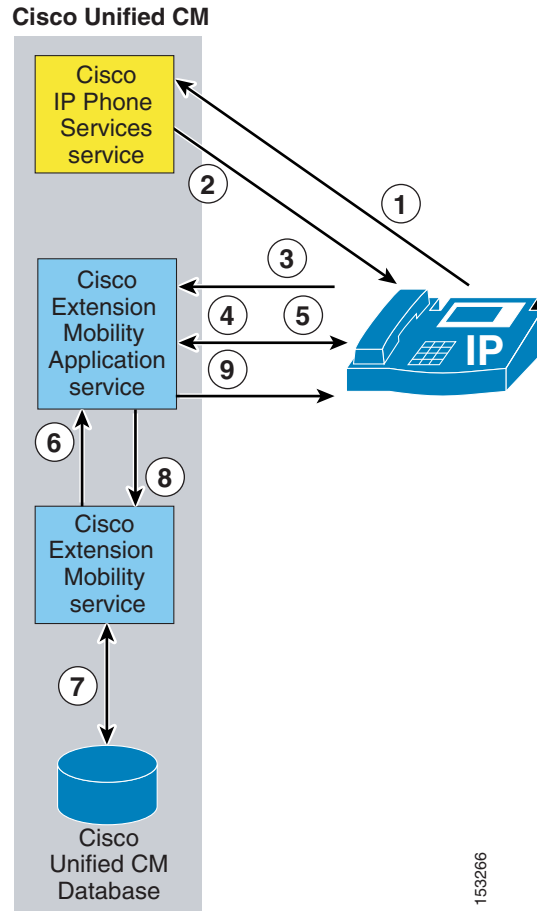
1. When the user presses the Services button on the phone, this action generates a call to the URL specified under the URL Services parameter on the Enterprise Parameter configuration page (see [IP Phone Service Enterprise Service Parameters](#), page 22-3) (see also step 1 in [Figure 22-4](#)).
2. An HTTP/XML call is generated to the IP Phone Services, which returns a list of all services to which the user's phone is subscribed (see step 2 in [Figure 22-4](#)).



Note

A Service URL button can be configured for EM on a user's phone so that the user can press a line or speed-dial button to generate a direct call to the Extension Mobility Application service. In this case, the interactions between the phone and the IP Phone Services (steps 1 and 2) are bypassed.

3. Next the user selects the Extension Mobility phone service listing. This selection in turn generates an HTTP call to the Extension Mobility Application service, which serves as the interface between the phone and the Cisco Extension Mobility service (see step 3 in [Figure 22-4](#)).
4. The Extension Mobility Application service then forwards an XML response back to the phone requesting user login credentials (userID and PIN) or, if the user is already logged in, a response asking if the user wants to log off the phone (see step 4 in [Figure 22-4](#)).
5. Assuming the user is attempting to log in, the user must use the phone's keypad to enter a valid userID and PIN. After the user presses the Submit softkey, a response containing the userID and PIN just entered is forwarded back to the Extension Mobility Application service (see step 5 in [Figure 22-4](#)).
6. The Extension Mobility Application next forwards this login information to the Extension Mobility service, which interacts with the Unified CM database to verify the user's credentials (see step 6 in [Figure 22-4](#)).
7. Upon successful verification of the user's credentials, the Extension Mobility service also interacts with the Unified CM database to read and select the appropriate user device profile and to write needed changes to the phone configuration based on this device profile (see step 7 in [Figure 22-4](#)).
8. Once these changes have been made, the Extension Mobility service sends back a successful response to the Extension Mobility Application service (see step 8 in [Figure 22-4](#)).
9. The Extension Mobility Application service, in turn, sends a reset message to the phone, and the phone resets and accepts the new phone configuration (see step 9 in [Figure 22-4](#)).

Figure 22-4 EM Application Architecture and Message Flow

153266

EM Redundancy

According to the EM architecture illustrated in [Figure 22-4](#), reads and writes to the Unified CM database are required. Based on current Unified CM cluster architecture, the publisher server is the only node capable of writing to the Unified CM database. Therefore, EM functionality is dependent upon the publisher server. If the publisher is unavailable, then EM logins and logouts are not possible. Because there is no mechanism for publisher redundancy, the publisher is a single point of failure when it comes to EM operation. For this reason, Cisco recommends that the three required services for EM (IP Phone Services, Extension Mobility, and Extension Mobility Application services) all run on the publisher. Running these services on other nodes in the cluster is pointless because they will provide no functionality if the publisher is unavailable.



Note

The Extension Mobility Application service is always activated and running on all nodes in the cluster. However, when you configure the EM phone service, always configure the Service URL to point to the publisher node.

As mentioned previously, because the Extension Mobility Application service subscribes to cluster change notification, it maintains a list of all nodes in the cluster with the Extension Mobility service activated. Therefore, the Extension Mobility service could be run on multiple servers within the cluster, and the Extension Mobility Application service could provide automatic failover to other nodes running the Extension Mobility service. However, this automatic failover is only for the Extension Mobility service. It does not provide failover for actual Extension Mobility functionality except in the very rare situation where the Extension Mobility service fails on the publisher but all other services, the Unified CM database, and the node itself remain operational. Typically, failure scenarios involve not just a single service but an entire node, node uplink switch, or switch port. Therefore, with regard to EM redundancy, running the required Extension Mobility service or IP Phone Services on any nodes other than the publisher is unnecessary. However, administrators might want to provide Extension Mobility service redundancy for the rare failure scenario described previously and can do so by running this service on additional nodes in the cluster. Additional redundancy could also be provided for the IP Phone Services and the Extension Mobility Application services through the use of SLB and virtual-IP-to-real-IP address mapping as depicted in [Figure 22-3](#). But again, this method provides redundancy only in rare situations where all these services fail on the publisher but the publisher and the Unified CM database continue to operate and handle EM login and logout requests.

Guidelines and Restrictions for EM

The following guidelines and restrictions apply with regard to the deployment and operation of EM within the Unified CM telephony environment:

- EM is supported only within a single Unified CM cluster.

EM is not currently supported between clusters. EM users of one Unified CM cluster can not log on to a phone in a second cluster unless a separate device profile and userID for that user has been created in the second cluster.

- EM users should not move between locations or sites within a cluster when Automated Alternate Routing (AAR) and/or the Voice over PSTN (VoPSTN) deployment model are in use.

EM functionality relies on the use of the IP network for routing calls. Call routing via the PSTN is more problematic because E.164 PSTN numbers are static and the PSTN is unable to account for movement of EM user directory numbers (DNs) from their home sites. AAR relies on the PSTN for call routing, as does the VoPSTN deployment model. In both cases, EM user movement between locations and sites is supported only if all sites the user is traversing are in the same AAR group. For additional information, see [Extension Mobility, page 10-30](#).

- EM functionality is completely dependent upon the Unified CM publisher server.

If the Unified CM publisher is down, EM users cannot log on or log off of a phone.

- Restarting the Extension Mobility Service or the node on which the service is running will affect auto-logout settings.

If Cisco Extension Mobility is stopped or restarted, the system does not auto-logout users who are already logged in after the expiration of the maximum login interval. These phones have to be logged out manually.

EM Performance and Capacity

The Cisco EM application relies on the publisher server for login and logout functionality. EM supports the following publisher capacities:

- Cisco MCS-7845 servers can support 50 sequential logins and/or logouts per minute.
- Cisco MCS-7835 servers can support 30 sequential logins and/or logouts per minute.

For additional information on EM capacity, refer to the Unified CM data sheets, documentation, and release notes available at

<http://www.cisco.com>

EM Interactions: Unified CM Assistant, AC, and WebDialer

EM can be used by both Unified CM Assistant Managers and Attendant Console users to log into their phones. For the details and guidelines governing the use EM with these other applications, see [Unified CM Assistant Interactions with EM, page 22-28](#) and [AC Interactions with EM, page 22-39](#).

WebDialer users can also use EM to log on to their phones. See [WebDialer Interactions with EM, page 22-49](#), for more information.

Unified CM Assistant

Cisco Unified Communications Manager Assistant (Unified CM Assistant) is a Unified CM integrated application that enables assistants to handle incoming calls on behalf of one or more managers. With the use of the Unified CM Assistant Console desktop application or the Unified CM Assistant Console phone service on the assistant phone, assistants can quickly determine a manager's status and determine what to do with a call. Assistants can manipulate calls using their phone's softkeys and service menus or via the PC interface with either keyboard shortcuts, drop-down menus, or by dragging and dropping calls to the managers' proxy lines.

Unified CM Assistant Phone Support

The following SCCP phones support Unified CM Assistant:

- Cisco Unified IP Phones 7940G, 7941G, and 7941G-GE
- Cisco Unified IP Phones 7960G, 7961G, and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE



Note

The Cisco Unified IP Phone Expansion Module 7914 is supported with any of the following phones: Cisco Unified IP Phone 7960G, 7961G, 7961G-GE, 7970G, or 7971G-GE. Up to two Cisco 7914 Modules are supported per phone.

Beginning with Cisco Unified Communications Manager release 5.1.1(b), the Cisco Unified IP Phones 7942G, 7945G, 7962G, 7965G, and 7975G support Unified CM Assistant. The Cisco Unified IP Phones 7962G, 7965G, and 7975G also support up to two Cisco 7914 Modules.

Unified CM Assistant is not supported on SIP phones.

Unified CM Services and Unified CM Assistant Service Parameters

To enable the Unified CM Assistant application, the system administrator must activate and start several Unified CM feature services from the Unified CM Serviceability interface. In addition, Unified CM Assistant service parameters provide configuration and customization options for determining how the Unified CM Assistant application and services behave.

Unified CM Services for Unified CM Assistant

The Unified CM Assistant application relies on the following feature services, which must to be activated manually from the Serviceability page:

- Cisco IP Manager Assistant
- Cisco CTIManager

The Unified CM Assistant application also relies on the Cisco CallManager Cisco IP Phone Services network service, which is automatically activated on Unified CM during installation.

The Cisco CallManager IP Manager Assistant service provides an interface for the Unified CM Assistant Console and Manager Configuration applications as well as interacting with the Cisco CTIManager service and Unified CM database. The Cisco CTIManager service interfaces and interacts with the Cisco CallManager Service and the Cisco CallManager IP Manager Assistant service for phone and call control.

The Cisco Unified IP Phone Services are needed to provide access to the Unified CM Assistant phone service from the manager and assistant phones. The URL used to define the Unified CM Assistant phone service is:

```
http://<Server_IP-Address>:8080/ma/servlet/MAService?cmd=doPhoneService&Name=#DEVICENAME#
```

(where <Server_IP-Address> is the IP address of any node in the cluster)

**Note**

Cisco recommends that you configure two instances of the Unified CM Assistant phone service. One Unified CM Assistant phone service instance should point to the primary Unified CM Assistant server and the other should point to the backup Unified CM Assistant server. By configuring a primary and a secondary phone service in this manner, you can provide redundancy for the assistant phone console. See [Device and Reachability Redundancy, page 22-26](#), for additional information.

Unified CM Assistant Service Parameters

The following items represent a partial list of Cisco CallManager IP Manager Assistant (IPMA) Service Parameters related to Unified CM Assistant functionality:

- CTIManager Connection Security Flag (Default value = Non Secure)

This parameter determines whether a secure Transport Layer Security (TLS) connection is used between the Cisco CallManager IP Manager Assistant service and the CTIManager. If enabled, a secure connection is configured using the Certificate Authority Proxy Function (CAPF) profile configured for the instance ID of the application user Unified CM AssistantSecureSysUser. The instance ID must be specified under the service parameter CAPF Profile Instance ID for Secure Connection to CTIManager.

**Note**

The application user Unified CM AssistantSecureSysUser is a system account created automatically at installation. It cannot be deleted.

- CAPF Profile Instance ID for Secure Connection to CTIManager (Default value = <None>)
The CAPF Profile Instance ID is a unique string of numbers and/or letters used to identify the TLS connection or instance that is made between the Unified CM Assistant server and CTIManager for the Unified CM AssistantSecureSysUser application user. If the CTI Manager Connection Security Flag parameter is set to True, then this parameter must be configured with a value.
- CTIManager (Primary) IP Address (Default value = <blank>)
This parameter specifies the IP address of the primary CTIManager that the Cisco Unified CM Assistant server uses to process calls. A primary CTIManager can be configured on each Unified CM Assistant server.
- CTIManager (Backup) IP Address (Default value = <blank>)
This parameter specifies the IP address of the backup CTIManager that this Cisco Unified CM Assistant server uses to process calls when the primary CTIManager is down. A backup CTIManager can be configured on each Unified CM Assistant server.
- Cisco IPMA Server (Primary) IP Address (Default value = <blank>)
This parameter specifies the IP address of the primary Cisco IPMA server. This is a cluster-wide parameter and only two IPMA servers, a primary and a backup, may be configured.
- Cisco IPMA Server (Backup) IP Address (Default value = <blank>)
This parameter specifies the IP address of the backup Cisco IPMA server. The backup server provides Unified CM Assistant service when the primary IPMA server fails. This is a cluster-wide parameter.
- Cisco IPMA Console Heartbeat Interval (Default value = 30)
This parameter specifies how often, in seconds, the IPMA server will send keep-alive messages (commonly referred to as heartbeats) to each Unified CM Assistant Console desktop application. The Unified CM Assistant Console desktop applications will initiate a failover to the backup IPMA server when they fail to receive a keep-alive message from the primary server before the specified time expires.
- Cisco IPMA Console Request Timeout (Default value = 30)
This parameter specifies the time, in seconds, that Unified CM Assistant Console desktop applications will wait to receive a response from the active or primary IPMA server.
- Cisco IPMA RNA Forward Calls (Default value = False)
When set to True, this parameter enables calls to an assistant's phone to be ring-no-answer (RNA) forwarded to the manager's next available assistant when the RNA value specified by the Cisco IPMA RNA Timeout parameter expires. If this parameter is set to False, then the call will ring the first assistant indefinitely or, if a voicemail profile is configured, the call will be forwarded to voicemail.
- Cisco IPMA RNA Timeout (Default value = 10)
This parameter specifies the time, in seconds, that the Cisco IPMA server waits before RNA forwarding an unanswered call to the next available assistant. RNA forwarding will occur only if the Cisco IPMA RNA Forward Calls parameter is set to True. If a voicemail profile is configured on the line and no other assistant is available, the call will be forwarded to voicemail when the timeout expires.

For a complete list of Unified CM Assistant service parameters, refer to the Unified CM Assistant information in the *Cisco Unified Communications Manager Features and Services Guide*, available at http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

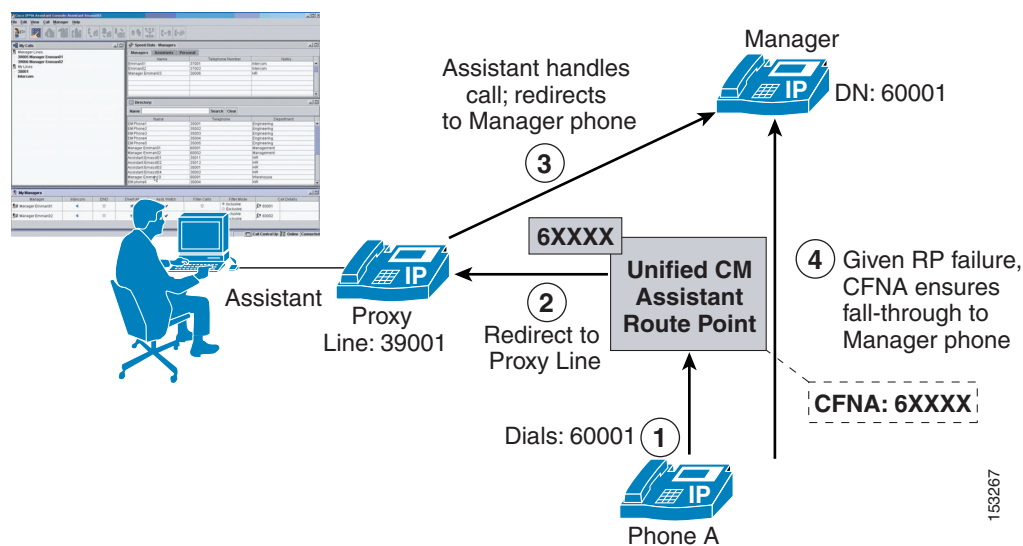
Unified CM Assistant Functionality and Architecture

The Unified CM Assistant application can operate in two modes: proxy line mode and shared line mode. The operation and functionality of each mode is different, and each has specific advantages and disadvantages. Both modes can be configured within a single cluster. However, mixing modes on the same assistant is not allowed. A single assistant providing support for one or more managers can support those managers in either shared line mode or proxy line mode.

Unified CM Assistant Proxy Line Mode

Figure 22-5 illustrates a simple call flow with Unified CM Assistant in proxy line mode. In this example, Phone A calls the Manager phone with directory number (DN) 60001 (step 1). The CTI/Unified CM Assistant Route Point (RP) intercepts this call based on a configured DN of 6XXXX. Next, based on the Manager DN, the call is redirected by the route point to the Manager's proxy line (DN: 39001) on the Assistant's phone (step 2). The Assistant can then answer or handle the call and, if appropriate, redirect the call to the Manager's phone (step 3). In the event of Unified CM Assistant application failure or if the Unified CM Assistant RP fails, a fall-through mechanism exists via the Call Forward No Answer (CFNA) 6XXXX configuration of the RP, so that calls to the Manager's DN will fall-through directly to the Manager's phone (step 4).

Figure 22-5 Unified CM Assistant Proxy Line Mode



Note

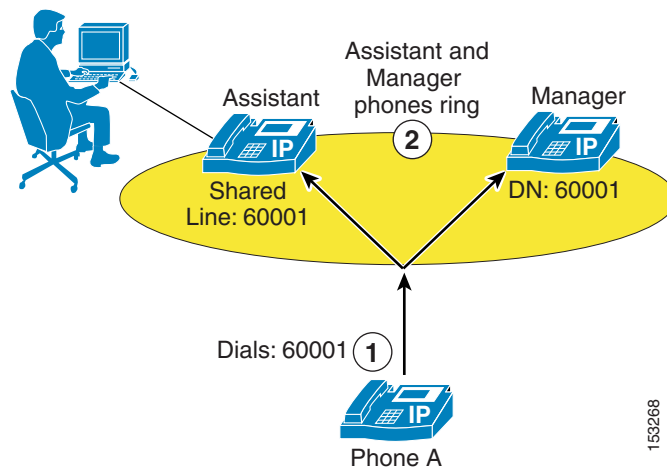
The CFNA fall-through mechanism illustrated in Figure 22-5 requires configuration of the same summarized digit-string as the Unified CM Assistant RP directory number in both the Forward No Answer Internal and Forward No Answer External fields under the Unified CM Assistant RP directory number configuration page. In addition, the calling search space (CSS) field for each of these call

forward parameters should be configured with the calling search space containing the partition with which the Manager phone DNs are configured, so that the Manager phone DNs can be reached if the Unified CM Assistant RP or Unified CM Assistant application fails.

Unified CM Assistant Share Lined Mode

Figure 22-6 illustrates a simple call flow with Unified CM Assistant in shared line mode. In this example, Phone A calls the Manager phone with directory number (DN) 60001, which is a shared line on the Assistant phone (step 1). The call will ring at both the Assistant and Manager phones unless the Manager has invoked the Do Not Disturb (DND) feature, in which case the Assistant's phone will be the only phone that rings audibly (step 2).

Figure 22-6 Unified CM Assistant Shared Line Mode



In Unified CM Assistant shared line mode, the Unified CM Assistant RP is not needed or required for intercepting calls to the Manager phone. However, the Do Not Disturb (DND) feature on the Manager phone and the Unified CM Assistant Console desktop application still depend on the Unified CM Assistant and Cisco CTIManager services. Furthermore, in Unified CM Assistant shared line mode, features such as call filtering, call intercept, assistant selection, and Assistant Watch are not available.

Unified CM Assistant Architecture

The architecture of the Unified CM Assistant application is as important to understand as its functionality. Figure 22-7 depicts the message flows and architecture of Unified CM Assistant. When Unified CM Assistant has been configured for Unified CM Assistant Manager and Assistant users, the following sequence of interactions and events can occur:

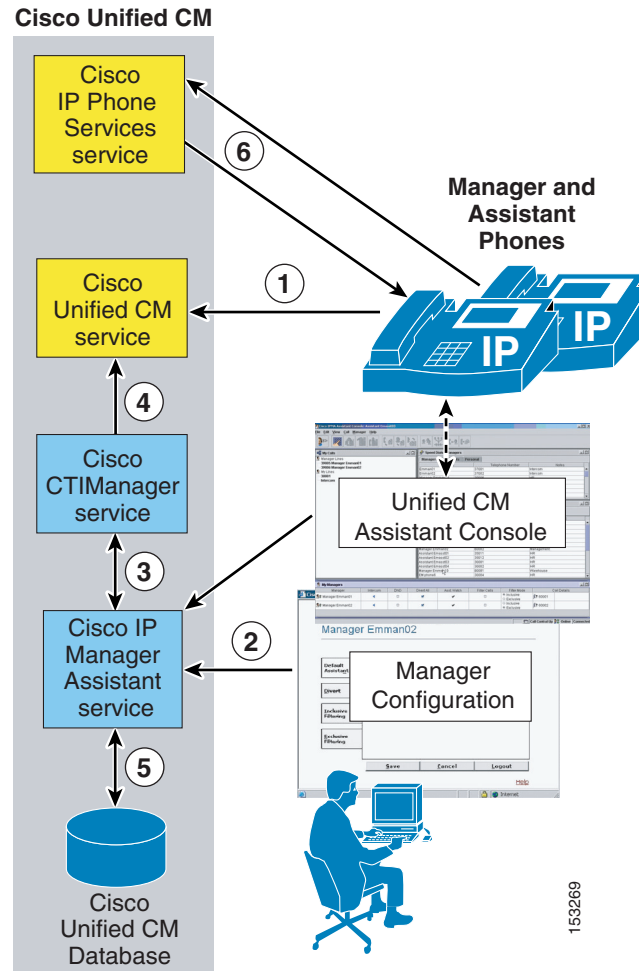
1. Manager and Assistant phones register with the Cisco CallManager Service, and the phone's keypad and softkeys are used to handle call flows (see step 1 in Figure 22-7).
2. Both the Unified CM Assistant Console desktop application and the Manager Configuration web-based application communicate and interface with the Unified CM Assistant service (see step 2 in Figure 22-7).
3. The Unified CM Assistant service in turn interacts with the CTIManager service for exchanging line monitoring and phone control information (see step 3 in Figure 22-7).

4. The CTIManager service passes Unified CM Assistant phone control information to the Cisco CallManager Service and also controls the Unified CM Assistant RP (see step 4 in [Figure 22-7](#)).
5. In parallel, the Unified CM Assistant service reads and writes Unified CM Assistant application information to and from the Unified CM database (see step 5 in [Figure 22-7](#)).
6. The Manager may choose to invoke the Unified CM Assistant phone service by pushing the Services button, thus generating a call to the IP Phone Service service that will return a list of all services (including the Unified CM Assistant phone service) to which the phone is subscribed (see step 6 in [Figure 22-7](#)).

The Unified CM Assistant phone service is controlled by the Unified CM Assistant service, and configuration changes made by the Manager using the phone are handled and propagated via the Unified CM Assistant service.

**Note**

A Service URL button can be configured for the Unified CM Assistant phone service on a Manager's phone so that the user can press a line or speed-dial button to generate a direct call to the Unified CM Assistant service, thus bypassing the IP Phone Service.

Figure 22-7 Unified CM Assistant Architecture**Note**

While [Figure 22-7](#) shows the IP Phone Services, Unified CM, CTIManager, and Unified CM Assistant services all running on the same node, this configuration is not a requirement. These services can be distributed between multiple nodes in the cluster but have been shown on the same node here for ease of explanation.

Unified CM Assistant Dial Plan Considerations

Dial plan configuration is extremely important for Unified CM Assistant configured in proxy line mode. To ensure that calls to Manager DNs are intercepted by the Unified CM Assistant RP and redirected to the Assistant phone, calling search spaces and partitions must be configured in such a way that Manager DNs are unreachable from all devices except the Unified CM Assistant RP and the Manager's proxy line on the Assistant phone.

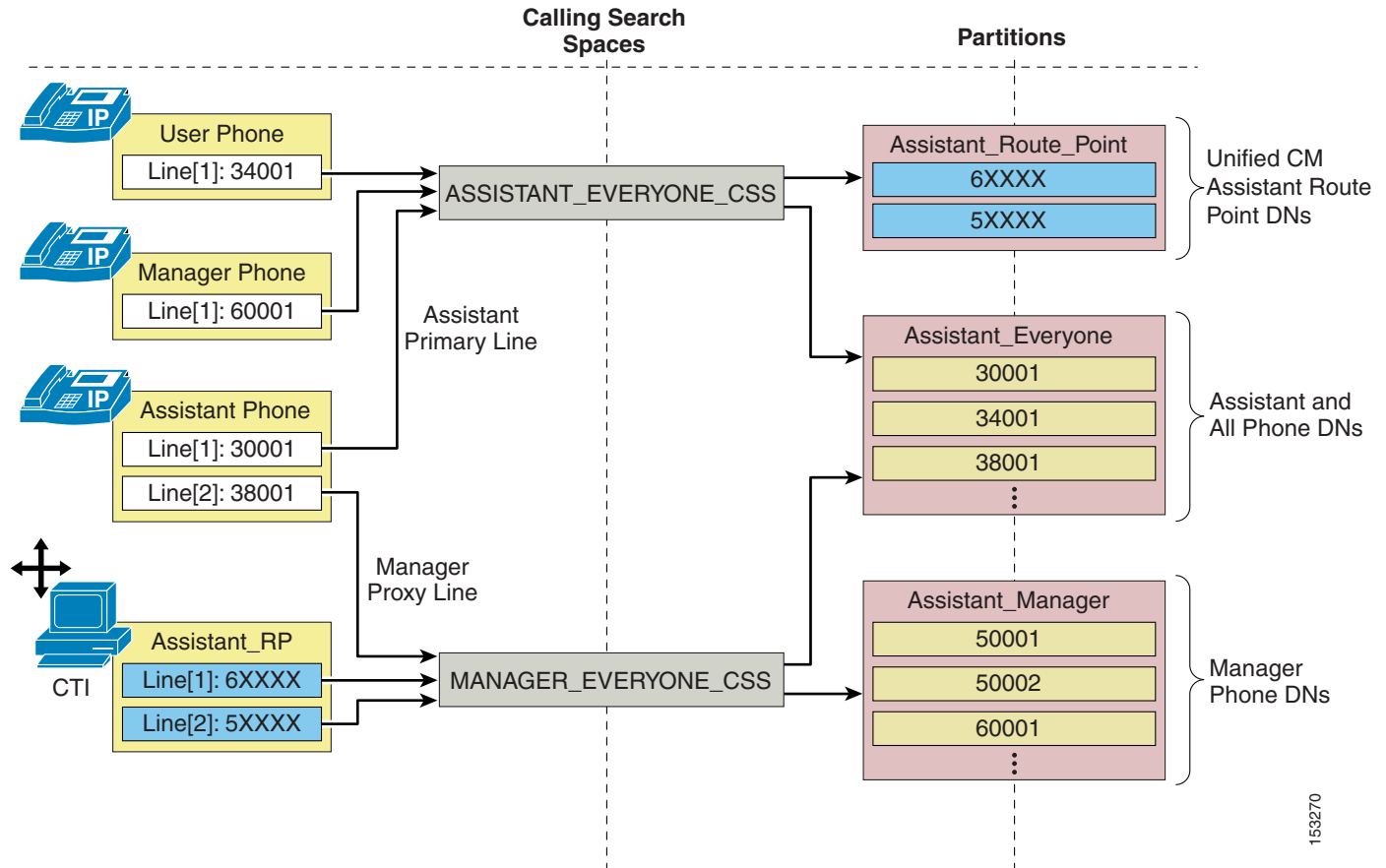
Figure 22-8 shows an example of a proxy line mode Unified CM Assistant dial plan with the minimum requirements for calling search spaces, partitions, and the configuration of various types of devices within these dial plan components. Three partitions are required for proxy line mode, and for the example in Figure 22-8 they are as follows:

- Assistant_Route_Point partition, containing all the Unified CM Assistant RP DNs
- Assistant_Everyone partition, containing all the Assistant and other user phone DNs
- Assistant_Manager partition, containing all the Manager phone DNs

In addition, two calling search spaces are required, and for the example in Figure 22-8 they are as follows:

- ASSISTANT_EVERYONE_CSS calling search space, containing both the Assistant_Route_Point and Assistant_Everyone partitions.
- MANAGER_EVERYONE_CSS calling search space, containing both the Assistant_Manager and Assistant_Everyone partitions.

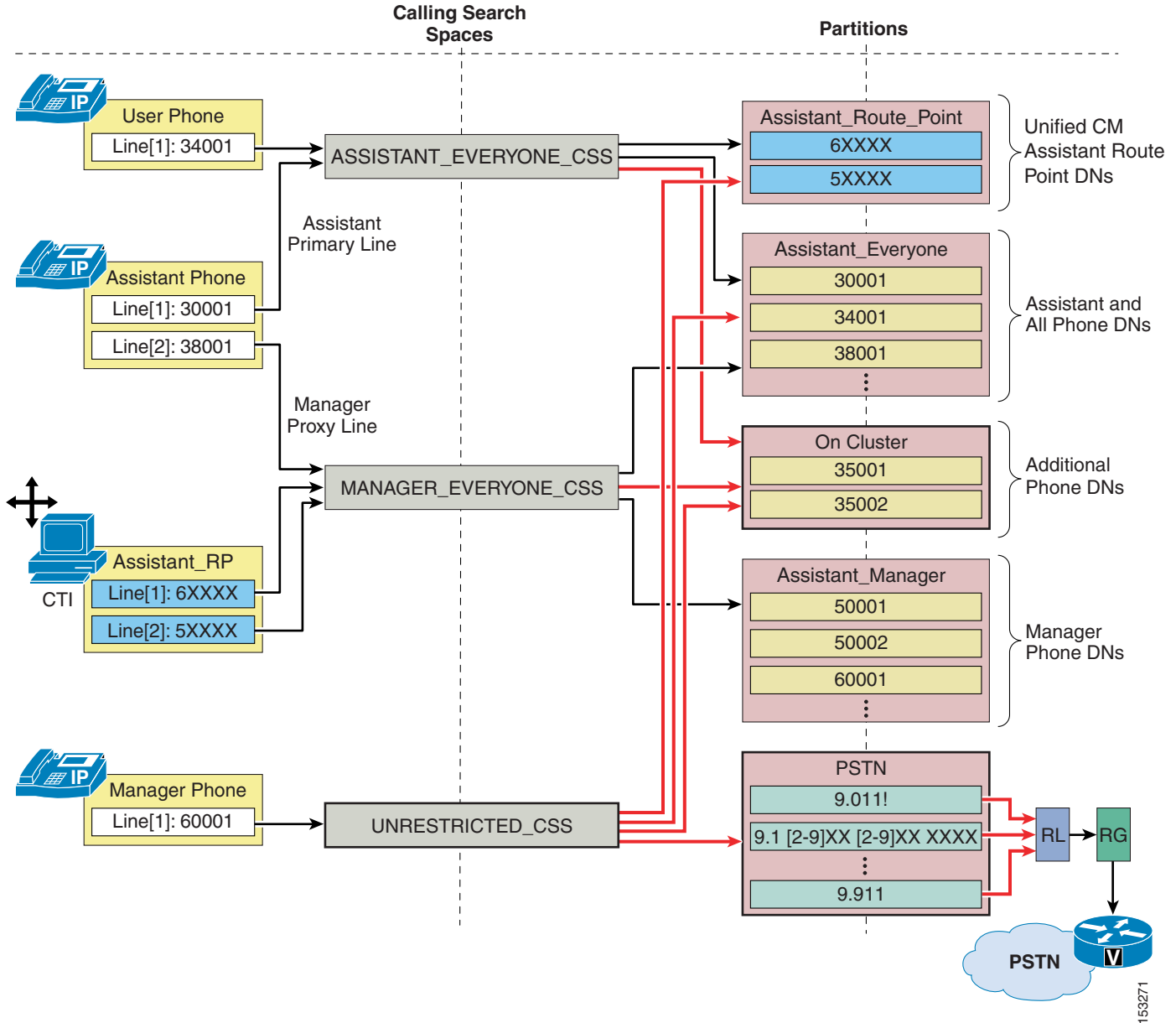
That is the extent of the dial plan for this example. However, it is also important to properly configure the various phone and Unified CM Assistant RP DNs or lines with the appropriate calling search spaces so that call routing works as required. In this case all user, Assistant primary (or personal), and Manager phone lines would be configured with the ASSISTANT_EVERYONE_CSS calling search space so that all of these lines can reach all the DNs in the Assistant_Everyone and Assistant_Route_Point partitions. Intercom lines and any other lines configured on devices within the telephony network would be configured with this same calling search space. All Manager proxy lines and all Assistant_RP lines are configured with the MANAGER_EVERYONE_CSS calling search space so that all of these lines can reach the Manager DNs in the Assistant_Manager partition as well as all the DNs belonging to the Assistant_Everyone partition. In this way, the dial plan ensures that only the Assistant_RP lines and the Manager proxy lines on the Assistant phones are capable of reaching the Manager phone DNs directly.

Figure 22-8 Unified CM Assistant Proxy Line Mode Dial Plan Example

153270

The example in Figure 22-8 shows the minimum dial plan requirements for Unified CM Assistant in proxy line mode. However, most real-world telephony networks will have additional or existing dial plan requirements that must be integrated with the Unified CM Assistant calling search spaces and partitions. Figure 22-9 illustrates such an integration dial plan. In this example, the previously discussed dial plan must now handle two additional partitions and an additional calling search space. The On Cluster partition has been added in Figure 22-9, and it contains some additional phone DNs. The On Cluster partition has been added to both of the existing Unified CM Assistant calling search spaces (ASSISTANT_EVERYONE_CSS and MANAGER_EVERYONE_CSS) so that existing devices can reach these added DNs. The UNRESTRICTED_CSS calling search space has also been added to the existing dial plan. This calling search space is configured with the Assistant_Route_Point, Assistant_Everyone, and the recently added On Cluster partitions. In addition, a second new partition called PSTN has been added, and it contains a set of route patterns used for routing calls to the PSTN via the common route list (RL), route group (RG), and voice gateway mechanism. This PSTN partition is configured as part of the UNRESTRICTED_CSS calling search space.

Phone and device line calling search space configurations may be adjusted to incorporate the newly added partitions and calling search spaces, provided the Assistant_RP and Assistant phone Manager proxy lines remain assigned to the MANAGER_EVERYONE_CSS calling search space. In this example, the Manager phone line has been moved from the originally configured ASSISTANT_EVERYONE_CSS calling search space to the new UNRESTRICTED_CSS because it is likely that a Manager would be given unrestricted access to the PSTN.

Figure 22-9 Unified CM Assistant Proxy Line Mode Dial Plan Integration Example

As Figure 22-9 illustrates, integrating additional partitions and calling search spaces into a new or existing Unified CM Assistant dial plan is feasible, but care must be taken to ensure that the underlying proxy line mode mechanism remains intact.

For Unified CM Assistant shared line mode, no special dial plan provisioning is required. Manager and Assistant phones can be configured with calling search spaces and partitions like any other phones in the network because there are no Unified CM Assistant RPs or proxy lines to be concerned about. The only requirement with regard to shared line mode is that the Manager and Assistant DN's must be in the same partition so that shared line functionality is possible.

Unified CM Assistant Console

The Unified CM Assistant Console desktop application or the Unified CM Assistant Console phone service is required in order for assistants to handle calls on a manager's behalf. The desktop application provides assistants with a graphical interface for handling calls, while the phone service provides a menu-driven interface for handling calls. Both the desktop application and the IP phone service allow the assistant to configure the Manager phone and environment and monitor line status and availability. In addition, the desktop application provides other functions such as click-to-dial speed dialing and directory entries, which can also be performed on the assistant phone using the traditional softkey and menu approach.

Unified CM Assistant Console Installation

The Unified CM Assistant Console desktop application can be installed from the following URL:

`https://<Server_IP-Address>:8443/ma/Install/IPMAConsoleInstall.jsp`

(where <Server_IP-Address> is the IP address of any node in the cluster)

The Unified CM Assistant Console phone service does not require any installation. To enable the Assistant's phone as a console, subscribe the phone to the Unified CM Assistant phone service. (This is the same service to which Manager phones must also be subscribed.)

Unified CM Assistant Desktop Console QoS

After installation, and in order to handle calls on a Manager's behalf, the Assistant must log on to the application by providing userID and password (as configured in the End-user directory on Unified CM) and will have to toggle status to "online" by clicking the Go Online icon or menu item. Once the user is logged in and online, the desktop application communicates with the Unified CM Assistant server at TCP port 2912. The application chooses an ephemeral TCP port when sourcing traffic. Because the Unified CM Assistant server on Unified CM interfaces with the desktop application for call control (generation and handling of call flows), traffic sourced from Unified CM on TCP port 2912 is QoS-marked by Unified CM as Differentiated Services Code Point (DSCP) of 24 or Per Hop Behavior (PHB) of CS3. In this way, Unified CM Assistant phone control traffic can be queued throughout the network like all other call signaling traffic.

In order to ensure symmetrical marking and queuing, the Unified CM Assistant Console application traffic destined for Unified CM TCP port 2912 should also be marked as DSCP 24 (PHB CS3) to ensure this traffic is placed in the appropriate call signaling queues along the network path toward Unified CM and the Unified CM Assistant server. The Unified CM Assistant Console application marks all traffic as best-effort. This means that you will have to apply an access control list (ACL) at the switch port level (or somewhere along the network path, preferably as close to the console PC as possible) to remark traffic sent by the application PC destined for Unified CM on TCP port 2912 from DSCP 0 (PHB Best Effort) to DSCP 24 (PHB CS3).

Unified CM Assistant Console Directory Window

The directory window within the Assistant Console desktop application enables an assistant to search for end-users in the Unified CM Directory. Search strings entered into the Name field of the directory window are sent to the Unified CM Assistant server, and searches are generated directly against the Unified CM database. Responses to search queries are then sent back to the desktop application by the Unified CM Assistant server.

While the additional traffic generated by directory searches within the desktop application is nominal, this traffic can be problematic in centralized call processing deployments when one or more Unified CM Assistant console applications are running at remote sites. A directory search resulting in a single entry generates approximately one (1) kilobit of traffic from the Unified CM Assistant server to the desktop application. Fortunately, a maximum of 25 entries can be retrieved per search, meaning that a maximum of approximately 25 kilobits of traffic can be generated for each search made by the desktop application. However, if directory searches are made by multiple Unified CM Assistant Console desktop applications across low-speed WAN links from the Unified CM Assistant server, the potential for congestion, delay, and queuing is increased. In addition, directory retrieval traffic is sourced from Unified CM on TCP port 2912, like all other Unified CM Assistant traffic to the desktop. This means that directory retrieval traffic is also marked with DSCP 24 (PHB CS3) and therefore is queued like call signaling traffic. As a result, directory retrieval could potentially congest, overrun, or delay call control traffic.

**Note**

If a directory search generates more than 25 entries, the assistant is warned via a dialog box with the message: “Your search returned more than 25 entries. Please refine your search.”

Given the potential for network congestion, Cisco recommends that administrators encourage Unified CM Assistant Console users to do the following:

- Limit their use of the directory window search function.
- To reduce the number of entries returned, enter as much information as possible in the Name field and avoid wild-card or blank searches when using the feature.

These recommendations are especially important if either of the following conditions is true:

- There are many Unified CM Assistant Assistants within the cluster.
- There are many assistants separated from the Unified CM and/or Unified CM Assistant servers by low-speed WAN links.

Unified CM Assistant Phone Console QoS

In order to handle calls on a Manager's behalf using the Unified CM Assistant Phone Console phone service, the Assistant must log on to the service by providing a userID and PIN (as configured in the End-user directory on Unified CM). Once the user is logged in, the phone console service communicates with Unified CM using HTTPS and SCCP. Call control traffic for Unified CM Assistant call generation and call handling is sent between the phone and Unified CM using SCCP. By default this traffic is marked as Differentiated Services Code Point (DSCP) of 24 or Per Hop Behavior (PHB) of CS3, thus ensuring it is queued throughout the network as call signaling traffic, therefore no additional QoS configuration or marking is required.

Unified CM Assistant Redundancy

Unified CM Assistant application redundancy can be provided at two levels:

- Redundancy at the component and service level

At this level, redundancy must be considered with regard to Unified CM Assistant service or server redundancy and CTIManager service redundancy. Likewise, the lack of publisher redundancy and the impact of this component failing should also be considered.

- Redundancy at the device and reachability level

At this level, redundancy should be considered as it relates to Assistant and Manager phones, the Unified CM Assistant route point, and the Unified CM Assistant Console desktop application and phone service, as well as redundancy in terms of Assistant and Manager reachability.

Service and Component Redundancy

As shown in [Figure 22-7](#), Unified CM Assistant functionality is primarily dependent on the Cisco CallManager IP Manager Assistant service and the Cisco CTIManager service. In both cases, redundancy is automatically built-in using a primary and backup mechanism. Two Unified CM Assistant servers (nodes running the Cisco Unified CM Assistant service) can be defined within a single cluster using the Cisco Unified CM Assistant Server (Primary) IP Address and Cisco Unified CM Assistant Server (Backup) IP Address service parameters (see [Unified CM Assistant Service Parameters](#), [page 22-14](#)). With the configuration of these parameters, the required Unified CM Assistant service is made redundant. Given a failure of the primary Unified CM Assistant server, the backup or standby Unified CM Assistant server is able to handle Unified CM Assistant service requests. Only one Unified CM Assistant server can be active and handling request at a given time, while the other Unified CM Assistant server will be in a standby state and will not handle requests unless the active server fails.

In addition, two CTIManager servers or services can be defined for each Unified CM Assistant server using the CTIManager (Primary) IP Address and CTIManager (Backup) IP Address service parameters (see [Unified CM Assistant Service Parameters](#), [page 22-14](#)). This results in a total of up to four CTIManagers per cluster for use by the Unified CM Assistant application. By configuring these parameters, you can make the CTIManager service redundant. Thus, given a failure of a primary CTIManager, CTIManager services can still be provided by the backup CTIManager. If all Unified CM Assistant and CTIManager services on cluster nodes fail, the Unified CM Assistant route point, Unified CM Assistant Console desktop application and phone service, and in turn the Unified CM Assistant application as a whole will fail. However as noted previously, given a failure of the Unified CM Assistant application, the CFNA fall-through mechanism will continue to work, allowing calls to a Manager to be routed directly to the Manager's phone.



Note

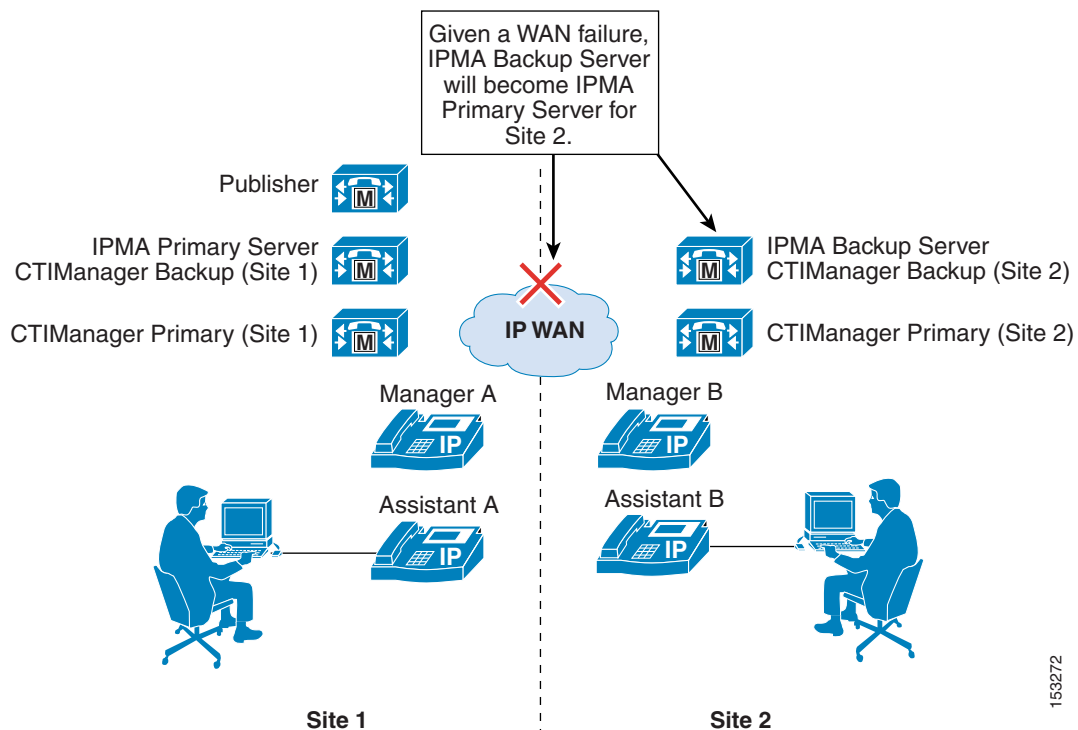
If configured in Unified CM Assistant shared-line mode, a complete failure of Unified CM Assistant and CTIManager service will not keep the Assistant from continuing to handle calls on behalf of the Manager because the phones will continue to share a line. However, the Unified CM Assistant Console desktop application and phone service and the DND feature will not be available.

[Figure 22-10](#) shows an example redundancy configuration for Unified CM Assistant and CTIManager primary and backup servers in a two-site deployment with clustering over the WAN. In order to provide maximum redundancy, a node at Site 1 is configured as the primary Unified CM Assistant server and a node at Site 2 is configured as the backup Unified CM Assistant server. In the event of a WAN failure, the backup Unified CM Assistant server at Site 2 will become a primary Unified CM Assistant server because the existing primary Unified CM Assistant server will be unreachable from Site 2. In this way, Unified CM Assistant servers can be made redundant in the clustering-over-the-WAN environment given a WAN failure. Furthermore, with a primary and backup CTIManager configured at both Site 1 and Site 2, CTIManager is made redundant given a WAN failure, and additional redundancy is provided for a CTIManager failure at each site.

**Note**

The redundancy scenario depicted in [Figure 22-10](#) shows a special circumstance. During normal operation it is not possible to have two active or primary Unified CM Assistant servers within the same cluster. If the two Unified CM Assistant servers can communicate over the network, then one server will be in backup mode and can not handle requests.

Figure 22-10 Unified CM Assistant Redundancy with Two-Site Clustering over the WAN



153272

As previously mentioned, the publisher is a single point of failure when it comes to writing to the Unified CM Database. The effect of a publisher failure on the Unified CM Assistant application is less dramatic than the impact such a failure has on Extension Mobility. Given a publisher failure, all aspects of the Unified CM Assistant application will continue to work; however, no changes to the Unified CM Assistant application configuration can be made. Configuration changes via the Unified CM Assistant Console desktop application, the Manager configuration web-based application, the phone softkeys, or the Unified CM Assistant phone service, will not be possible until the publisher is restored. This condition includes enabling or disabling features such as Do Not Disturb, DivertAll, Assistant Watch, and call filtering, as well as changing call filter and assistant selection configuration.

Device and Reachability Redundancy

Redundancy for Unified CM Assistant at the devices level relies on a number of mechanisms. First and foremost, manager and assistant phones as well as the Unified CM Assistant RP rely on the built-in redundancy provided by a combination of the device pool and Unified CM group configuration for device registration.

In addition, some devices rely on component services for additional redundancy and functionality. For example, the Unified CM Assistant RP also relies on CTIManager for call control functionality and therefore must rely on the primary and back CTIManager mechanism described in the previous section. The Unified CM Assistant Console desktop application also relies on the component services for redundancy and functionality. The Assistant Console desktop application supports automatic failover from the primary to the backup Unified CM Assistant server (and vice versa) in order to continue to handle incoming calls for managers. The amount of time this automatic failover will take can be controlled using the Cisco Unified CM Assistant Console Heartbeat Interval and the Cisco Unified CM Assistant Console Request Timeout service parameters (see [Unified CM Assistant Service Parameters, page 22-14](#)). Although the heartbeat or keep-alive frequency can be configured so that failures of the Unified CM Assistant server are detected by the desktop application more quickly, be careful not to affect the network adversely by sending keep-alives too frequently. This consideration is especially important if there are a large number of Assistant Console desktop applications in use.

The Unified CM Assistant Console phone service, unlike the Unified CM Assistant Console desktop application, requires manual intervention for redundancy given the failure of the primary Unified CM Assistant server. If the Primary Unified CM Assistant server goes down, assistants using the phone console will not see an indication of this condition. However, the assistant phone will receive a "Host not found Exception" message upon trying to use a softkey. In order to continue using the phone console with the backup Unified CM Assistant server, the user must manually select the secondary Unified CM Assistant phone service from the IP Services menu and log in again.

There are several other failover mechanisms which ensure that Manager and Assistant reachability are redundant. First, calls sent to a Manager's Assistant via the Unified CM Assistant application (in proxy line mode) can be forwarded to the Manager's next available Assistant if the call is not answered after a configured amount of time. If the next Assistant does not answer the call after the configured amount of time, the call can again be forwarded to the Manager's next available Assistant, and so on. The mechanism is configured using the Cisco Unified CM Assistant RNA Forward Calls and Cisco Unified CM Assistant RNA Timeout service parameters (see [Unified CM Assistant Service Parameters, page 22-14](#)). Second, as mentioned previously, if all Unified CM Assistant and CTI services on cluster nodes fail, the Unified CM Assistant RP will become unavailable. However, based on the CFNA configuration of the Unified CM Assistant RP, calls to all Manager DNs will fall-through directly to the Manager phones so that Manager reachability is sufficiently redundant.

Guidelines and Restrictions for Unified CM Assistant

Unified CM Assistant has the following limitations with regard to overlapping and shared extensions, which you should keep in mind when planning directory number provisioning:

- With Unified CM Assistant in proxy line mode, the proxy line number(s) on the assistant phone should be unique, even across different partitions.
- With Unified CM Assistant in proxy line mode, two Managers cannot have the same Unified CM Assistant controlled line number (DN), even across different partitions.

Additionally, Unified CM Assistant assistants should be made aware that, when upgrading from Cisco Unified CM release 4.x to Unified CM 5.x, the Unified CM Assistant Assistant Console desktop application will not automatically upgrade to the newest version. Instead, the previous version of the Unified CM Assistant Console desktop application will have to be un-installed and the new one will have to be installed.

Unified CM Assistant Performance and Capacity

The Cisco Unified CM Assistant application supports the following capacities:

- Maximum of 10 Assistants can be configured per Manager.
- Maximum of 33 Managers can be configured for a single Assistant (if each Manager has one Unified CM Assistant-controlled line).
- Maximum of 1250 Assistants and 1250 Managers can be configured per cluster with an MCS-7845 server.
- Maximum of two Unified CM Assistant servers can be deployed per cluster (primary and backup).
- Maximum of four CTIManagers can be configured for Unified CM Assistant per cluster. (A primary and backup CTIManager can be specified for each of two Unified CM Assistant servers.)

The Cisco Unified CM Assistant application interacts with the CTIManager for line monitoring and phone control. Each line on a Unified CM Assistant or Manager phone generates a connection to the CTIManager. In addition, each Unified CM Assistant route point generates a connection to the CTIManager. When you configure Unified CM Assistant, the number of required CTI connections must be considered with regard to the overall cluster limit for CTI connections (10,000 CTI connections per cluster with the MCS-7845 platform or 3200 CTI connections per cluster with the MCS-7835 and MCS-7825 servers). If additional CTI connections are required for other applications, they can limit the capacity of Unified CM Assistant.

For additional information on Unified CM Assistant capacity, refer to the Unified CM and Unified CM Assistant data sheets, documentation, and release notes available at

<http://www.cisco.com>

Unified CM Assistant Interactions with EM

Unified CM Assistant Managers can use EM to log in to their phones in both proxy-line and shared-lined modes. However, the Manager must be configured as a Mobile Manager under the Cisco Unified CM Assistant Manager configuration page of the End-user Directory. When using EM in conjunction with Unified CM Assistant, users should not be able to log in to more than one phone using EM. This behavior can be enabled/disabled via the EM service parameter Multiple Login Behavior (see [EM Service Parameters, page 22-8](#)). If multiple EM logins by the same user are required within the cluster, Unified CM Assistant Managers who use EM should be instructed not to log in to multiple phones. Allowing a manager to log in to two different phones with EM violates the previously stated restriction that, in proxy line mode, two Managers cannot have the same Unified CM Assistant controlled line number (DN), even across different partitions.



Note

Unified CM Assistant Assistants cannot use EM to log in to their phones because there is no concept of a Mobile Assistant.

Attendant Console

The Cisco Unified Communications Manager Attendant Console (AC) application enables a receptionist to answer and transfer or dispatch calls within an organization. The attendant can install the attendant console, a client/server Java application, on a PC that runs Windows 2000 or Windows XP. The attendant

console connects to the Cisco CallManager Attendant Console Server (Cisco Unified CallManager AC Server) for login services, line state, and directory services. Multiple attendant consoles can connect to a single AC server.

AC Phone Support

The following SCCP phones support AC functionality:

- Cisco Unified IP Phone 7905G
- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phones 7912G and 7912G-A
- Cisco Unified IP Phones 7940G, 7941G, and 7941G-GE
- Cisco Unified IP Phones 7960G, 7961G, and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

Beginning with Cisco Unified Communications Manager release 5.1.1(b), the Cisco Unified IP Phones 7942G, 7945G, 7962G, 7965G, and 7975G also support AC functionality.

AC is not supported on SIP phones.

Unified CM Services and AC Service Parameters

To enable the AC application, the system administrator must activate and start a number of Unified CM feature services from the Unified CM Serviceability interface. In addition, AC service parameters provide configuration and customization options for determining how the AC application behaves.

Unified CM Services for AC

The AC application relies on the following feature services, which must to be activated manually from the Serviceability page:

- Cisco CallManager Attendant Console Server
- Cisco CTIManager

The Cisco CallManager Attendant Console Server service provides an interface for the AC Desktop application and interacts with the Cisco CTIManager service and Unified CM database. The Cisco CTIManager service interfaces and interacts with the Cisco CallManager Service and the Cisco CallManager Attendant Console Server service for phone and call control. It also interfaces with the AC Desktop application.

AC Service Parameters

The following items represent a partial list of Cisco CallManager Attendant Console Server Service Parameters related to AC functionality:

- Directory Sync Period (Default value = 3)

This parameter specifies the frequency interval, in hours, for the synchronization of the AC server AutoGenerated.txt file with the Unified CM End-user Directory. Changes to the End-user Directory will not be reflected in the AutoGenerated.txt file until this interval is reached.

- JTAPI Username (Default value = ac)
This parameter specifies the application user name that the AC server uses to log into and communicate with the CTIManager.
- Device Authentication Application Username (Default value = ACDeviceAuthenticationUser)
This parameter specifies the application user name that the AC server uses for authentication of the attendant phone.

For a complete list of Attendant Console service parameters, refer to the Unified CM Attendant Console information in the *Cisco Unified Communications Manager Features and Services Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

AC and AC Device Authentication Application Users

For proper AC operation, an application user named **ac** must be configured in Unified CM. The ac application user is required in order for the AC server to interact with the CTIManager. Without this application user configuration, attendants will not be able to receive calls.



Note

Application users are different than end users within the Cisco Unified CM 5.x database, and application users are stored separately from the end users in the directory. Therefore, directory searches will not return application user entries. For more details on Application Users and End Users in Cisco Unified CM 5.x, see [LDAP Directory Integration, page 17-1](#).

The ac user must be configured with the following group permissions under the application user configuration page:

- Standard CTI Allow Control of All Devices
- Standard CTI Allow Call Park Monitoring
- Standard CTI Enabled

By configuring the ac application user with the "Standard CTI Allow Control of All Devices" group permissions, you enable the CTI Super Provider feature. The Super Provider feature allows the AC application to control and monitor (via CTI) any device or line, which eliminates the need for associating attendant phones and AC pilot points to the ac application user, thus greatly simplifying configuration.

An administrator might want to change this application user name to something other than ac. If a user name other than ac is configured, the JTAPI Username service parameter (see [AC Service Parameters, page 22-29](#)) must be set to the new user name.

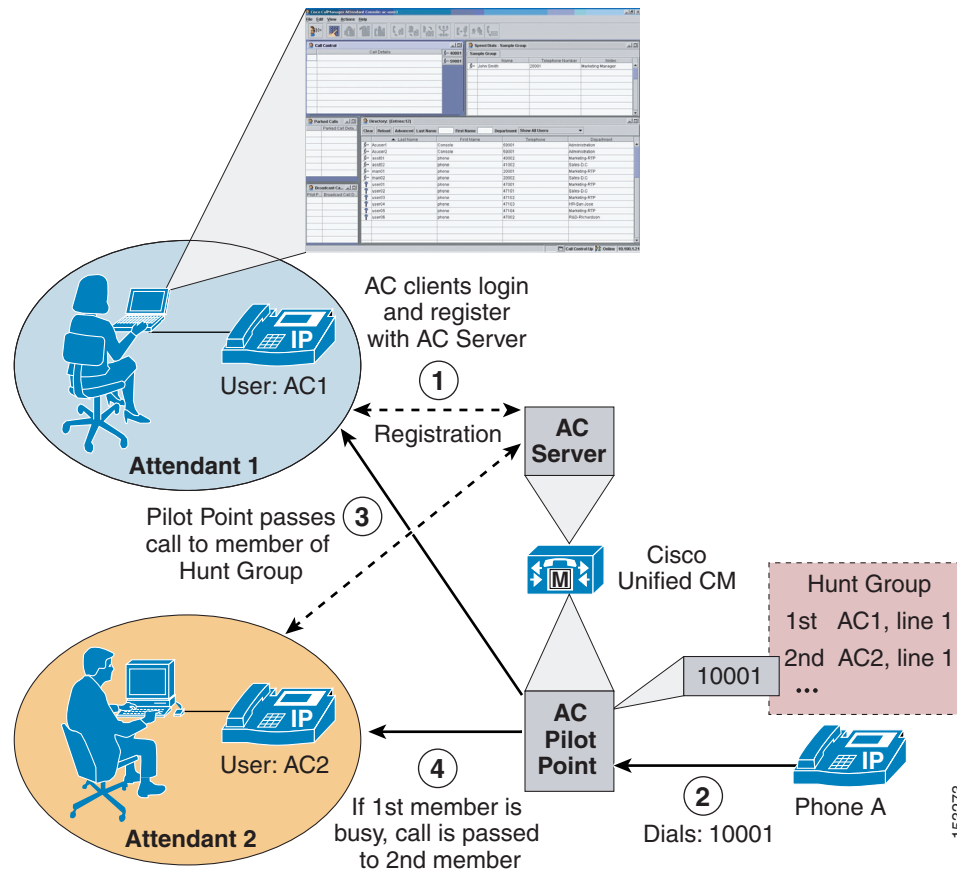
In addition to adding the ac user and assigning the group permissions listed above, you should create a device authentication application user named ACDeviceAuthenticationUser. Once configured, all attendant phones must be associated with this application user. This application user is used by the Cisco Unified CM Attendant Console server to authenticate attendant phones. No group permissions need to be configured for this application user account.

If an administrator chooses to use a different application user account name instead of ACDeviceAuthenticationUser, the Device Authentication Application Username service parameter (see [AC Service Parameters, page 22-29](#)) must be changed to match the configured application user name.

AC Functionality and Architecture

Figure 22-11 illustrates a basic example of AC functionality and operation. First, the AC clients log into and register with the AC server on Unified CM (step 1). Next, Phone A calls directory number (DN) 10001 configured for the AC pilot point on Unified CM (step 2). The AC pilot point intercepts this call and, based on the hunt group configuration, directs the call to one of its available members. In this case, the call is sent to line 2 of the Attendant user AC1's phone (step 3). If a second call comes into the pilot point number 10001 while user AC1 is still on the first call, the second call is routed to another available member of the hunt group, and in this case the call is forwarded to line 1 of Attendant user AC2's phone (step 4).

Figure 22-11 Basic AC Operation



For purposes of routing calls, pilot points determine the next available member of a hunt group based on one of the following routing algorithms (configured in the "Route Calls to" field under the pilot point):

- **First available**
With this algorithm, incoming calls are routed to the first member of the group who is available.
- **Longest idle**
With this algorithm, incoming calls are routed to the member who has been idle (not handled a call) the longest.
- **Circular hunting**
With this algorithm, incoming calls are routed to the available members in a round-robin fashion.

- Broadcast hunting

With this algorithm, incoming calls are queued, and indication is sent to the AC desktop application of all available members simultaneously.

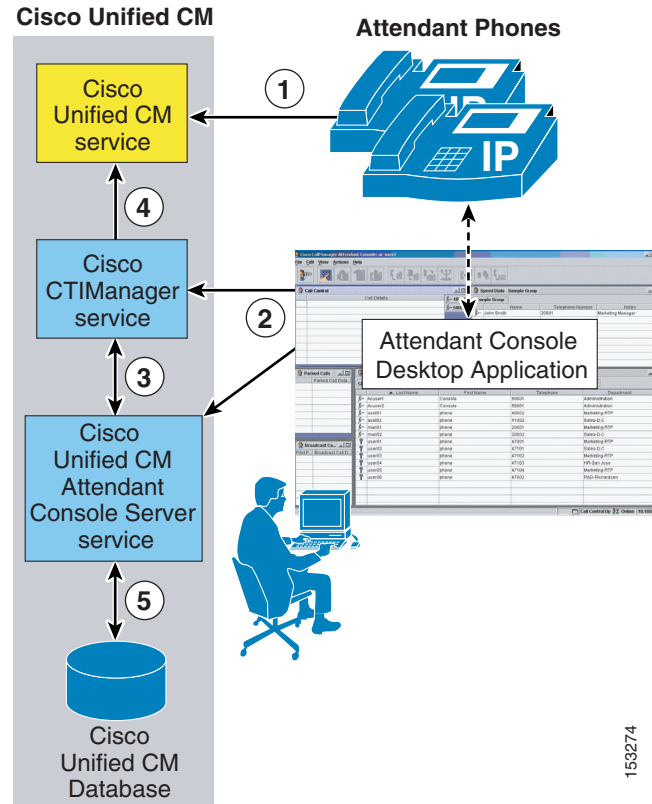
With the example shown in [Figure 22-11](#), the First Available algorithm is used. The hunt group routing algorithm as well as the queuing settings for the Broadcast routing algorithm are all configured under the Pilot Point configuration page on Unified CM.

AC Architecture

The architecture of the AC application is as important to understand as its functionality. [Figure 22-12](#) depicts the message flows and architecture of AC. When AC has been configured for AC users, the following sequence of interactions and events can occur:

1. Attendant phones register with the Cisco CallManager Service, and the phone's keypad and softkeys are used to handle call flows (see step 1 in [Figure 22-12](#)).
2. The Attendant Console desktop application communicates and interfaces with the CTIManager service, using JTAPI for phone and line control. In addition, the desktop application interfaces with the AC service and server via Remote Method Invocation (RMI) for AC functionality (see step 2 in [Figure 22-12](#)).
3. The AC server in turn interacts with the CTIManager service for exchanging line monitoring and phone control information (see step 3 in [Figure 22-12](#)).
4. Likewise, the CTIManager service passes AC phone control information to the Cisco CallManager Service and also controls the AC pilot point (see step 4 in [Figure 22-12](#)).
5. In parallel, the AC server reads and writes AC application information to and from the Unified CM Database (see step 5 in [Figure 22-12](#)).

Figure 22-12 AC Architecture

**Note**

Although Figure 22-12 shows the Unified CM, CTIManager, and Attendant Console Server services all running on the same node, this configuration is not a requirement. These services can be distributed among multiple nodes in the cluster, but they are shown on the same node here for ease of explanation.

Attendant Console Desktop Application

The Attendant Console desktop application is used by attendants to handle calls via a graphical virtual console. Besides call handling, the application provides additional functions such as click-to-dial speed dialing and directory entries, environment configuration, and line status and availability indication for other users within the directory and speed dial windows.

Attendant Console Installation

The Attendant Console desktop application can be downloaded from the following URL:

https://<Server_IP-Address>:8443/plugins/CiscoAttendantConsoleClient.exe

(where <Server_IP-Address> is the IP address of any node in the cluster)

Once the CiscoAttendantConsoleClient.exe file has been downloaded to an attendant's PC, it must also be installed.

Attendant Console QoS

After installing the Attendant Console desktop application, an attendant logs on to the console application by providing an AC userID and password (as configured under the Cisco Unified CM Attendant Console User page on Unified CM).



Note

AC userIDs are required for logging into the AC desktop application and are separate from the users configured in the end-user directory as well as the application users on Unified CM. Because these users are stored separately from the end user-directory, directory searches will not return AC user entries.

Once the AC user is logged on, the AC desktop application communicates with Unified CM using Remote Method Invocation (RMI) and Java Telephony Application Programming Interface (JTAPI) protocols predominantly. RMI is used for communication between the desktop client and the AC server including registration, keep-alives, and information exchange. RMI traffic is sourced from Unified CM on TCP ports 1101 to 1129 and sourced from the desktop application on one or more ephemeral TCP ports. All RMI traffic is marked as best-effort.

JTAPI traffic carries device and line control information and call control traffic between the CTIManager on Unified CM and the AC desktop application. JTAPI Traffic is sourced from Unified CM on TCP port 2748 and sourced from the desktop application on an ephemeral TCP port.

Because the JTAPI traffic between the CTIManager and AC client is used for call control (generation and handling of call flows), this traffic is QoS-marked by Unified CM with DSCP of 24 (PHB of CS3). In this way, AC phone control traffic can be queued throughout the network like all other call signaling traffic. In order to ensure symmetrical marking and queuing, the Attendant Console desktop application traffic destined for Unified CM TCP port 2748 should also be marked as DSCP 24 (PHB CS3) to ensure this traffic is placed in the appropriate call signaling queues along the network path toward Unified CM and the CTIManager. However, because the AC client application marks all traffic as best-effort, an access control list (ACL) must be configured to re-mark this traffic properly.

The AC server and desktop client marking can be summarized as follows:

- Unified CM appropriately marks all JTAPI traffic sourced from TCP port 2748 with DSCP of 24 (PHB of CS3).
- The Attendant Console desktop application marks JTAPI traffic destined for Unified CM TCP port 2748 as best-effort. This means that an ACL should be applied at the switch port level to re-mark the JTAPI traffic sent by the application to the Unified CM and AC server from a DSCP of 0 to a DSCP of 24 (PHB of CS3).

Attendant Console Directory Window

The directory window within the Attendant Console desktop application enables an attendant to search for end-users within the Unified CM telephony environment. Typically directory listings are obtained by searching against a directory file rather than searching against the Unified CM directory itself. One of the following directory files will be searched when an AC application user enters a search in the directory window:

- User list

This directory file is stored on the local PC or on a local drive path. In order for this file to be searched, its name and location must be configured in the Path Name of Local Directory File field under the Advanced tab of the Attendant Settings dialog box. If a file name and location has not been configured in this field, this option is skipped and a directory search will be conducted against one of the other directory files.

- AutoGenerated.txt

This directory file is automatically generated from the Unified CM database end-user table by the AC server and is stored on the Unified CM server. If the local directory user list file is not configured, then the AC desktop application will download this file from Unified CM automatically. The AutoGenerated.txt file is periodically regenerated or synchronized by the AC server against the end-user directory to ensure that the information in the file is accurate. The frequency of this synchronization is determined by the Directory Sync Period AC service parameter (see [AC Service Parameters, page 22-29](#)), which by default is set to three hours, meaning that the AutoGenerated.txt file will be updated every three hours.

- CorporateDirectory.txt

This file is available only if manually imported to Unified CM by an Administrator using the Cisco Unified CM Attendant Console User File Upload tool (under **Application > Cisco Unified CM Attendant Console**). If uploaded, this file will replace the AutoGenerated.txt file on the Unified CM server. Therefore, the AC desktop application will download this file rather than the AutoGenerated.txt file, provided that the local user list file has not been configured.

One of the directory files listed above is downloaded (in the case of the AutoGenerated or Corporate Directory.txt files) and loaded every time the AC desktop application is launched. The directory file is then downloaded and/or reloaded periodically, as long as the application is up, based on the Directory Reload Interval setting on the Advanced tab of the Attendant Settings dialog box. All the directory files conform to a comma-delimited format with one user entry per line.

While the additional traffic generated by downloading directory files for directory window searches within the desktop application is typically nominal, this traffic can be problematic for a number of reasons. First, if the Unified CM directory size is large, the directory file downloaded by the console application containing the entire directory can generate large amounts of traffic on the network. When this factor is coupled with large numbers of AC desktop applications within the network, short download intervals, a centralized call processing deployment, and console applications running at remote sites over low-speed WAN links, the potential for network congestion, delay, and queuing is highly probable.

Although the use of local user list files on the desktop application PCs could likely eliminate many of the concerns about network bandwidth and congestion, the Advanced search feature within the directory window of the AC desktop is more problematic. Whereas all other directory searches within the AC desktop application directory window are conducted against either the local user list file or one of the downloaded files, there is an exception for searches done using the Advanced search window triggered by the Advanced button in the directory window. Searches made using the Advanced search window bypass the directory file search convention and instead are generated directly against the Unified CM end-user directory in real time. This means additional traffic on the network above and beyond the periodic downloading of directory files. Furthermore, there is no limit on the number of entries that can be downloaded using the Advanced search feature. Not only will these real-time searches and retrievals generate additional network load, but because there is no limiting of returned entries, this additional load can be quite large. Fortunately, the traffic generated by both the directory file download and the Advanced directory search use the RMI protocol, which is marked as best-effort, so there is no risk of congesting priority voice media and provisioned call signaling queues on the network path. However, the possibility still exists that the AC desktop application directory traffic could cause congestion of the best-effort queues and lead to dropping of both directory traffic and other best-effort network data traffic.

Given the potential for network congestion with AC desktop application directory file downloads and directory searches, Cisco recommends the following practices:

- Administrators should encourage all Attendant Console users to limit their use of the Advanced directory search feature. Furthermore, if users are going to use the feature, they should be urged to enter as much information as possible in the Advanced search parameter fields in order to reduce the number of entries returned.

- In centralized call processing deployment scenarios, user list files on the AC client PCs or on a network shares should be utilized for remote site AC users to eliminate periodic downloading of a directory file from Unified CM across low-speed WAN links. One method for achieving this goal with minimal management overhead is to make a user list file available on a local network share at each remote site. This file can be scheduled to synchronize with the Unified CM database and then be loaded automatically onto the remote network shares during off-peak or overnight hours, thereby eliminating the potential for network congestion during peak business hours. Then each morning, when AC users launch the desktop console, the application can download the most up-to-date directory user list.

These recommendations are especially critical if one or more of the following conditions are true:

- There are many AC users within a Unified CM cluster.
- There are significant numbers of AC users separated from Unified CM servers by low-speed WAN links.
- The end-user directory is very large.

AC Redundancy

AC application redundancy can be provided at two levels:

- Redundancy at the component and service level

At this level, redundancy must be considered with regard to AC service or server redundancy and CTIManager service redundancy. Likewise, the lack of publisher redundancy and the impact of this component failing should also be considered.

- Redundancy at the device and reachability level

At this level, redundancy should be considered as it relates to attendant phones, the AC pilot point, and the Attendant Console desktop application, as well as redundancy in terms of attendant and pilot point reachability.

Service and Component Redundancy

As shown in [Figure 22-12](#), AC functionality is primarily dependent on the Cisco CallManager Attendant Console Server service and the Cisco CTIManager service. In both cases, redundancy is built into the Unified CM cluster architecture. Redundancy for both the AC Server service and the CTIManager service is determined by the number of nodes within the cluster where each service is running. Redundancy is determined by the number of server failures that can occur while still providing the required service, which can be expressed using the formula $(N - 1)$, where N is the number of servers running the service. For example, if three servers in the cluster are running the AC Server service, then $N = 3$. Redundancy for this service can then be calculated as $(3 - 1)$, or 2. Thus, redundancy is ensured for up to two server failures. CTIManager redundancy can be calculated using the same formula. In order to provide maximum redundancy for these services, Cisco recommends running both the AC Server and CTIManager services on all call processing nodes within the cluster. However, for minimum redundancy, each of these services should run on at least two call processing nodes within the cluster.

The publisher is a single point of failure when it comes to writing to the Unified CM Database. The affect of a publisher failure on the AC application is minimal. Given a publisher failure, all aspects of the AC application will continue to work; however, no changes to AC application configuration can be made. Configuration changes to AC pilot points, hunt groups, and attendant phones will not be possible until the publisher is restored.

Device and Reachability Redundancy

Redundancy for AC at the device level relies on a number of mechanisms. First and foremost, attendant phones as well as the AC pilot point rely on the built-in redundancy provided by a combination of the device pool and Unified CM group configuration for device registration.

In addition, some devices rely on component services for additional redundancy and functionality. For example, the AC pilot point also relies on CTIManager for call control functionality and must therefore rely on CTIManager redundancy as described in the previous section.

The Attendant Console desktop application also relies on the component services for redundancy and functionality. The AC desktop application supports automatic failover between redundant AC Servers and CTIManager services in order to continue to handle incoming calls. Redundancy of these services from perspective of the AC desktop application is determined by the Unified CM group mechanism as described below. First, when the AC desktop application is launched and the attendant logs in, the application downloads a list of Unified CMs based on the device pool and Unified CM group configuration of the attendant phone. This list is stored in the GlobalSettings.xml file on the local PC and determines CTIManager service redundancy for the desktop.

**Note**

Cisco recommends that the IP address of the primary Unified CM server configured in the Unified CM group for the attendant's phone be entered in the Attendant Server Host Name or IP Address field under the Basic tab of the Attendant Settings dialog boxes. This entry ensures that, if a failure occurs, both the attendant phone and the AC desktop application will simultaneously fail-over to the next server in the phone's configured Unified CM group.

Next, the desktop application relies on the device pool and Unified CM group of the AC pilot point (of which the attendant phone is a member) for AC server redundancy. In both cases, up to tertiary redundancy is provided because up to three servers can be configured in a Unified CM group.

Further redundancy for these services can be provided for the desktop application via the Call Processing Server Host Names or IP Addresses field under the Advanced tab of the Attendant Settings dialog box. By configuring a comma-separated list of Unified CM servers in this field, additional redundancy beyond the Unified CM group mechanism can be provided. However, because this additional redundancy is useful only if the group mechanism has been exhausted, it might be unnecessary. This additional redundancy would really be used only in the event that the first three servers providing registration services for the attendant phone and AC pilot point are unavailable, which means that the attendant phone and AC pilot point are also unavailable. If the phone and pilot point are unavailable, then the desktop application is of no use.

Finally, besides the reachability redundancy built into the hunt group mechanism under the AC pilot point, which provides incoming callers with attendant redundancy, additional redundancy can be provided for an AC pilot point failure. If an AC pilot point fails, incoming callers dialing the pilot point number will receive a busy tone. In order to provide a failover mechanism for pilot points, another AC pilot point number can be configured in the Call Forward No Answer (CFNA) fields of the pilot point line configuration screen. This CFNA mechanism ensures that callers to a failed pilot point will be forwarded to another pilot point for call handling and routing.

Guidelines and Restrictions for AC

AC is aware of partitions at the JTAPI level, therefore the Attendant Console desktop application is aware of partitions in terms of line control. However, other AC components are not aware of partitions, or they have some limitations with regard to overlapping and shared extensions. Keep the following guidelines in mind when planning directory number provisioning:

- Hunt groups
 - Shared lines must not be used by any hunt group members.
 - Overlapping extensions must not be used by any hunt group members.
 - Hunt group member directory numbers must not be added to Unified CM line groups.
- Pilot points
 - Shared lines must not be used as pilot point directory numbers.
 - Pilot point directory numbers must not be added to Unified CM line groups.
- Console directory and speed-dial windows

Line status displays within the console directory and speed-dial windows are unable to account for shared lines or overlapping extensions. As a result, only the status of the most recently changed line instance is shown when encountering shared or overlapping line appearances.

Also ensure that all AC pilot points are configured with a calling search space that includes all partition in which hunt group member directory numbers reside. Failure to do so will result in one or more members being unreachable.

AC Performance and Capacity

The Cisco AC application supports the following capacities:

- Maximum of 500 attendants per cluster.
- Maximum of 500 pilot points per cluster.
- Maximum of 125 attendant and pilot point pairs per Cisco MCS-7845 server.
- Maximum of 100 attendant and pilot point pairs per Cisco MCS-7835 server.
- Maximum of 75 attendant and pilot point pairs per Cisco MCS-7825 server.
- The Cisco MCS-7845 server supports a maximum of 1250 AC devices.
- The Cisco MCS-7835 server supports a maximum of 1000 AC devices.
- The Cisco MCS-7825 server supports a maximum of 750 AC devices.



Note

The AC device capacity numbers can be divided among hunt pilots and hunt pilot members. For example, with an MCS-7845 server, the maximum number of AC devices is 1250. This capacity can be allocated in a number of ways, such as 125 hunt pilots with 10 members in each hunt pilot, or 10 hunt pilots with 125 members in each hunt pilot.

In order to support the maximum of 500 attendants and 500 pilot points, attendants and pilot points must be distributed across multiple servers in groups of no more than 125 pairs per MCS-7845 server, no more than 100 pairs per MCS-7835 server, and no more than 75 pairs per MCS-7825 server.

The Cisco AC application interacts with the CTIManager for line monitoring and phone control. Each line on an Attendant phone generates a connection to the CTIManager. In addition, each AC pilot point generates a connection to the CTIManager. When you configure the AC application, the number of required CTI connections must be considered with regard to the overall cluster limit for CTI connections (10,000 CTI connections per cluster with the MCS-7845 platform or 3200 CTI connections per cluster with the MCS-7835 and MCS-7825 servers). If additional CTI connections are required for other applications, they can limit the capacity of the AC application.

AC Interactions with EM

AC Console users can log in to their phones using EM. However, the AC desktop application requires AC users to log out and back into the application whenever there is a configuration change on the attendant phone. Because an EM login (or logout) results in a configuration change on the phone, when using AC and EM together, users should first log in to their phone using EM and then log in to the AC desktop application, thus eliminating the need to log out and back on to the desktop application.

In addition, EM and AC user DN's should be added to AC pilot point hunt groups as User Members rather than Device Members. This practice ensures that incoming calls will not be routed to an AC user who is unavailable due to the fact that they are not logged into their phone using EM. User Members of a hunt group are configured with both a user name and line number, while Device Members are simply configured with a directory number. The pilot point routes calls to Device Members after checking to see only that the directory number is not busy. The pilot point routes calls to User Members after verifying that the line number of the attendant phone is available and that the AC user is logged on and online. Thus, by adding EM AC users to a hunt group as User Members, you can ensure that a call will be sent to that user phone only if the EM AC user is logged in.

WebDialer

WebDialer is a click-to-dial application for Unified CM that enables users to place calls easily from their PCs using any supported phone device. There is no requirement for administrators to manage CTI links or build JTAPI or TAPI applications because Cisco WebDialer provides a simplified web application and HTTP or Simple Objects Access Protocol (SOAP) interface for those who want to provide their own user interface and authentication mechanisms. Either way, the solution scales to support the entire Unified CM cluster with full redundancy.

WebDialer Phone Support

The following SCCP phones support WebDialer:

- Cisco Unified IP Phone 7902G
- Cisco Unified IP Phone 7905G
- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phones 7912G and 7912G-A
- Cisco Unified Wireless IP Phone 7920
- Cisco Unified Wireless IP Phone 7921G
- Cisco Unified IP Phones 7940G, 7941G, and 7941G-GE

- Cisco Unified IP Phones 7960G, 7961G, and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

The following SIP phones also support WebDialer:

- Cisco Unified IP Phones 7941G and 7941G-GE
- Cisco Unified IP Phones 7961G and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

Beginning with Cisco Unified Communications Manager release 5.1.1(b), the Cisco Unified IP Phones 7942G, 7945G, 7962G, 7965G, and 7975G running SCCP or SIP loads also support WebDialer.


Note

WebDialer is not supported on Cisco Unified IP Phones 7905G, 7906G, 7911G, 7912G, 7912G-A, 7940G, or 7960G running SIP loads.

Unified CM Services and WebDialer Service Parameters

To enable the WebDialer application, the system administrator must activate and start a number of Unified CM feature services from the Unified CM Serviceability interface. In addition, WebDialer service parameters provide configuration and customization options for determining how the WebDialer application and services behave.

Unified CM Services for WebDialer

The WebDialer application relies on the following feature services, which must be activated manually from the Serviceability page:

- Cisco WebDialer Web Service
- Cisco CTIManager

The Cisco WebDialer Service is the interface point between click-to-dial web-based and desktop-based applications and Unified CM. The Cisco CTIManager Service processes requests received from the WebDialer service by interacting with the Unified CM call processing and database layers. The net result exposes phone and call control capabilities to the applications.

WebDialer Service Parameters

The following items represent a partial list of Cisco WebDialer Web Service service parameters related to WebDialer functionality:

- CTIManager Connection Security Flag (Default value = Non Secure)

This parameter determines whether a secure Transport Layer Security (TLS) connection is used between the Cisco WebDialer Web service and the CTIManager. If enabled, a secure connection is configured using the Certificate Authority Proxy Function (CAPF) profile configured for the instance ID of the application user WDSecureSysUser. The instance ID must be specified under the CAPF Profile Instance ID for Secure Connection to CTIManager service parameter.


Note

The application user WDSecureSysUser is a system account created automatically at installation. It cannot be deleted.

- CAPF Profile Instance ID for Secure Connection to CTIManager (Default value = <None>)
The CAPF Profile Instance ID is a unique string of numbers and/or letters used to identify the TLS connection or instance that is made between the Cisco WebDialer Web service and CTIManager for the WDSecureSysUser application user. If the CTI Manager Connection Security Flag parameter is set to True, then this parameter must be configured with a value.
- Primary Cisco CTIManager (Default value = 127.0.0.1)
This parameter specifies the IP address of the Unified CM subscriber running the CTIManager service that WebDialer should use when processing requests. This is a cluster-wide parameter.
- Backup Cisco CTIManager (Default value = <blank>)
This parameter specifies the IP address of the Unified CM subscriber running the backup instance of the CTIManager service that WebDialer should use when processing requests. This is a cluster-wide parameter.
- List of WebDialers (Default value = <blank>)
This parameter specifies the IP address and port numbers of all WebDialers in the enterprise. Use a space to separate multiple entries. This parameter must be populated if Redirector functionality is needed.
- User Session Expiry (Default value = 0)
This parameter specifies the time in hours after which the user session or browser cookie will expire. A value of 0 indicates that the session or cookie will never expire.

For a complete list of WebDialer service parameters, refer to the Cisco WebDialer information in the *Cisco Unified Communications Manager Features and Services Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

WebDialer Functionality and Architecture

The WebDialer application contains two servlets: the WebDialer servlet and the Redirector servlet. The operation and functionality of each servlet is similar, and they can be configured to run simultaneously.

WebDialer Servlet

Figure 22-13 illustrates a simple WebDialer example. In this example, user John Smith launches WebDialer via a web-based or desktop application (step 1). WebDialer responds with a request for login credentials. The user must respond with a valid userID and password as configured in the Unified CM end-user directory. In this case, John Smith submits userID = jsmith and password = cisco (step 2). Next, based on this login, WebDialer responds with the Cisco WebDialer Preferences configuration page, and the user must indicate either “User permanent device” or “Use Extension Mobility” (assuming the user has an EM device profile). In this case, user John Smith selects “User permanent device” and selects the appropriate MAC address (SEP00036BC7B973) and directory number (10001) for his phone from drop-down menus on the configuration page (step 3). Finally, the user is presented with a screen requesting the phone number to be called (this value may already be indicated) and must click Dial. In this case, John Smith enters 10002 and, after clicking Dial, a call is automatically generated from his phone to Phone B at number 10002 (step 4).

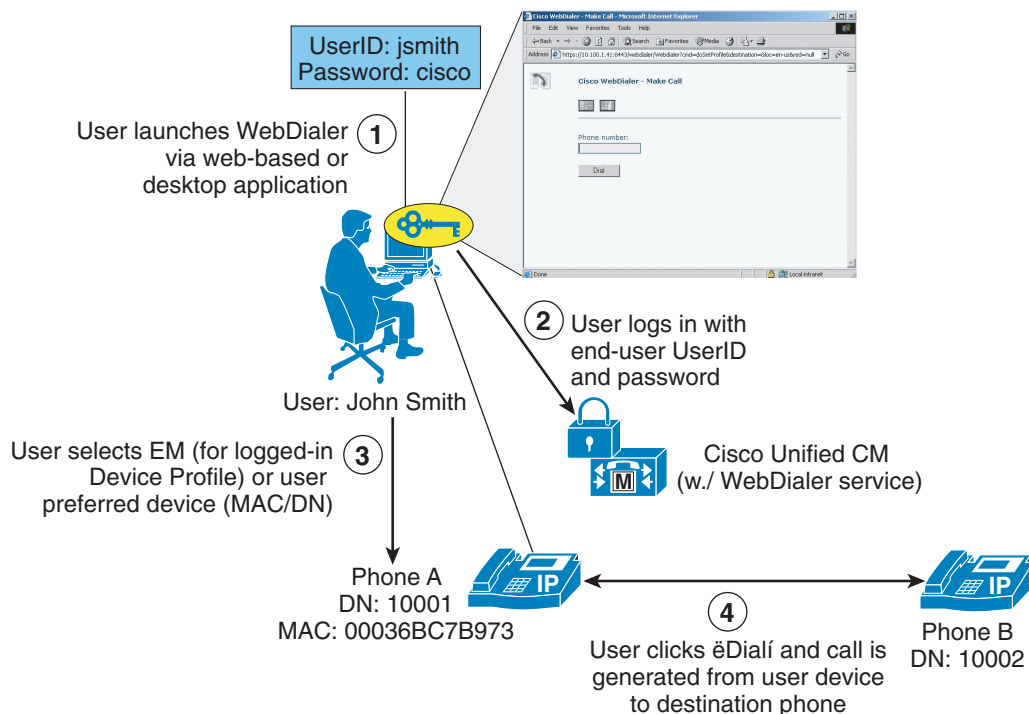


Note

If the user has previously logged in to the WebDialer application and a web browser and server cookie are still active, the user will not be prompted to log in again during subsequent requests. The user will be prompted to log in again when the cookie has been cleared at the browser or by a restart of the

WebDialer server. Alternatively, the user web browser cookie can be set to expire automatically after a certain number of hours as configured by the User Session Expiry WebDialer service parameter (see [WebDialer Service Parameters, page 22-40](#)).

Figure 22-13 WebDialer Servlet Operation



153275

Redirector Servlet

The Redirector servlet provides WebDialer functionality in a multi-cluster or distributed call processing environment. This functionality allows the use of a single enterprise-wide web-based WebDialer application between all Unified CM clusters. [Figure 22-14](#) illustrates the basic operation of the Redirector servlet as part of the WebDialer application. In this example, the enterprise has three Unified CM clusters: New York, Chicago, and San Francisco. All three clusters have been configured with a single WebDialer application. The San Francisco cluster has been designated as the Redirector. In order to designate the San Francisco WebDialer as the enterprise-wide Redirector, each cluster WebDialer server must have the service parameter List of WebDialers (see [WebDialer Service Parameters, page 22-40](#)) configured with its own IP address and the San Francisco WebDialer IP address. The San Francisco WebDialer server will be configured with its own IP address and the addresses of all other WebDialer servers in the enterprise. Based on this example, the List of WebDialers service parameter field for each WebDialer server would be configured as follows:

- New York WebDialer — List of WebDialers: 10.1.1.10:8443 10.3.1.0:8443
- Chicago WebDialer — List of WebDialers: 10.1.1.10:8443 10.2.1.0:8443
- San Francisco WebDialer — List of WebDialers: 10.1.1.10:8443 10.2.1.0:8443 10.3.1.0:8443

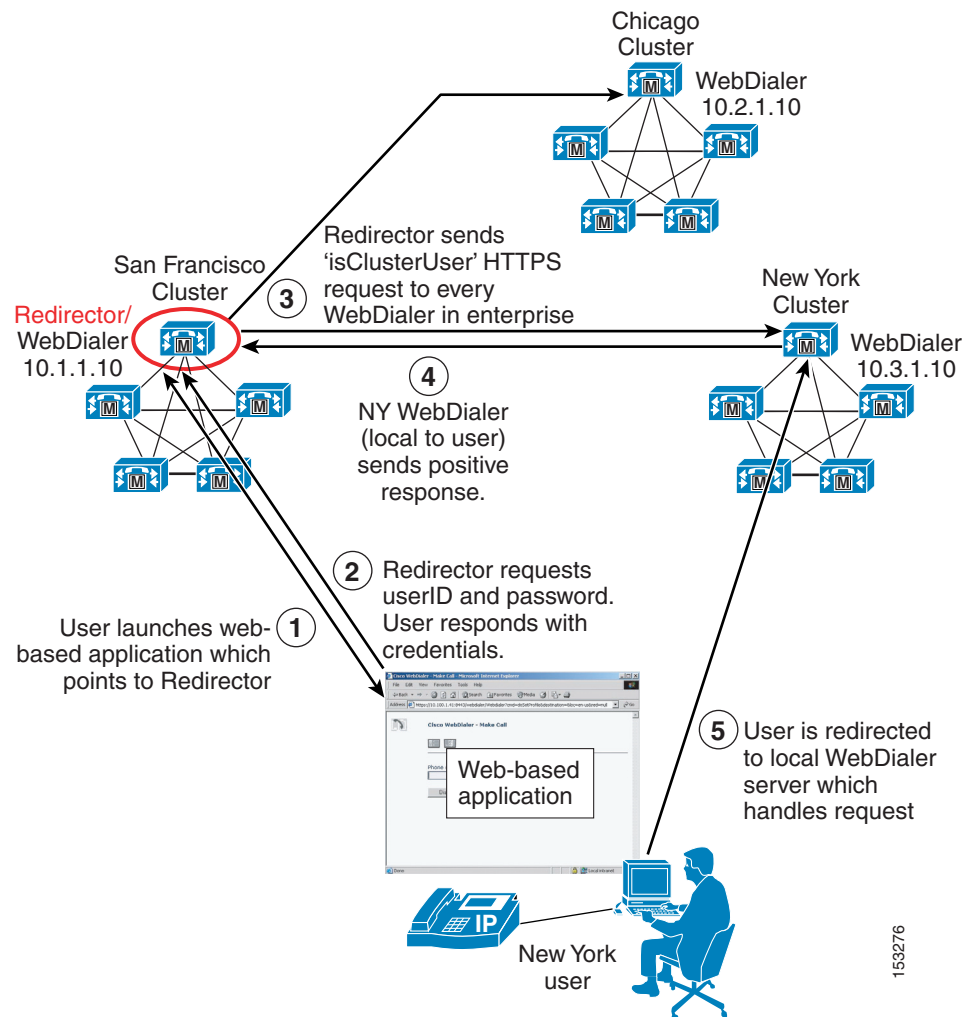
The enterprise-wide web-based application points to the San Francisco Redirector and is launched by the New York user (see step 1 in [Figure 22-14](#)). Next the Redirector requests user login, and the New York user responds back with their userID and password (see step 2 in [Figure 22-14](#)).

**Note**

If the user has previously logged in to the WebDialer application and a web browser and server cookie are still active, the user will not be prompted to log in again during subsequent requests. Alternatively, the user web browser cookie can be set to expire automatically after a certain number of hours as configured by the User Session Expiry WebDialer service parameter (see [WebDialer Service Parameters](#), page 22-40).

The Redirector then broadcasts an `isClusterUser` HTTPS request to every WebDialer in the enterprise simultaneously (as configured in the List of WebDialers service parameter). In this example, the requests go to the Chicago and New York WebDialer servers (see step 3 in [Figure 22-14](#)). Because the New York user is local to the New York cluster, the New York WebDialer responds with a positive response (see step 4 in [Figure 22-14](#)). Finally, the New York user is redirected to their local WebDialer server, which will handle the application request (see step 5 in [Figure 22-14](#)). The user is not notified of the redirect; however, the URL in the browser address bar will be changed as the user is redirected from the Redirector to the local WebDialer server).

Figure 22-14 Redirector Servlet Operation



**Note**

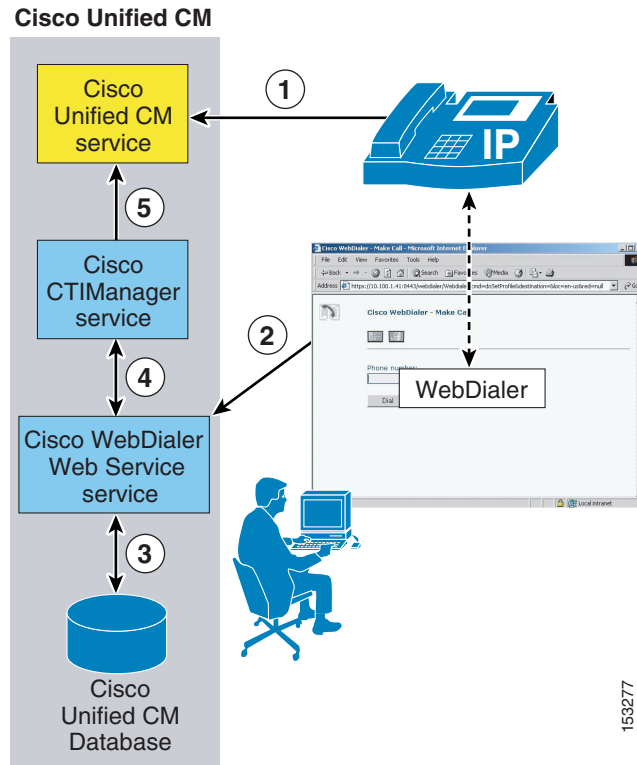
Because the Redirector application is an enterprise-wide application that requires user authentication against the Unified CM Database, Cisco highly recommends that all end-user userIDs be unique across all Unified CM clusters. If they are not, then it is possible that more than one positive response to the `isClusterUser` request could be received by the Redirector application. If this happens, the user will be asked by the Redirector application to select their local WebDialer server manually. The user will then have to know which server is their local server. If the wrong server is chosen, the WebDialer request will fail.

WebDialer Architecture

The architecture of the WebDialer application is as important to understand as its functionality.

[Figure 22-15](#) depicts the message flows and architecture of WebDialer. The following sequence of interactions and events can occur:

1. WebDialer user phones register and make and receive calls via the Cisco CallManager Service (see step 1 in [Figure 22-15](#)).
2. The WebDialer application on the user's PC communicates with the Cisco WebDialer Web Service (see step 2 in [Figure 22-15](#)) via one of the following interfaces:
 - HTML over HTTPS
This interface is used by web-based applications based on the HTTPS protocol. This is the only interface that provides access to the WebDialer and Redirector servlets.
 - Simple Object Access Protocol (SOAP) over HTTP
This interface is used by desktop applications based on the SOAP interface.
3. The WebDialer Web service reads user and phone information from the Unified CM Database (see step 3 in [Figure 22-15](#)).
4. The WebDialer Web service in turn interacts with the CTIManager service for exchanging line and phone control information (see step 4 in [Figure 22-15](#)).
5. The CTIManager service passes WebDialer phone control information to the Cisco CallManager Service (see step 5 in [Figure 22-15](#)).

Figure 22-15 WebDialer Architecture**Note**

Although Figure 22-15 shows the Unified CM, CTIManager, and WebDialer Web Service services all running on the same node, this configuration is not a requirement. These services can be distributed among multiple nodes in the cluster, but they are shown on the same node here for ease of explanation.

WebDialer URLs

The WebDialer application can be accessed from web-based applications via the HTML-over-HTTPS interface using the following URLs:

- WebDialer servlet

`https://<Server-IP_Addr>:8443/webdialer/Webdialer?destination=<Number_to_dial>`

(where <Server_IP-Address> is the IP address of any node in the cluster running the Cisco WebDialer Web Service service, and where <Number_to_dial> is the number that the WebDialer user wishes to dial)

- Redirector servlet

`https://<Server-IP_Addr>:8443/webdialer/Redirector?destination=<Number_to_dial>`

(where <Server_IP-Address> is the IP address of any node in the enterprise running the Cisco WebDialer Web Service service, and where <Number_to_dial> is the number that the WebDialer user wishes to dial)

Figure 22-16 gives an example of HTML source code used in a click-to-dial web-based application calling the Cisco WebDialer application. In this example, the URL `https://10.1.1.1:8443/webdialer/Webdialer?destination=30271` in the HTML source view corresponds to the "Phone: 30721" link for user Steve Smith within the web browser view. A user clicking on this link would launch the WebDialer application and, after logging in and clicking Dial, would generate a call from the user's phone to Steve Smith's phone. The same code could be used for a click-to-dial application using the Redirector function by changing the URL to `https://10.1.1.1:8443/webdialer/Redirector?destination=30271`.

Figure 22-16 WebDialer URL HTML Example

HTML source view:

```
<html>
<center><h3>WebDialer click-to-dial HTML sample</h3></center>
<b>Username:</b> Adams, Sally<br>
<b>Email:</b> <a href="mailto:sadams@cisco.com">a</a><br>
<b>Phone:</b> <a href="https://10.1.1.1:8443/webdialer/Webdialer?destination=23923">23923</a><br>
<b>Department:</b> Human Resources<br>
<br>
<b>Username:</b> Smith, Steve<br>
<b>Email:</b> <a href="mailto:ssmith@cisco.com">:ssmith</a><br>
<b>Phone:</b> <a href="https://10.1.1.1:8443/webdialer/Webdialer?destination=30271">30271</a><br>
<b>Department:</b> Human Resources
<hr>
</html>
```

Web browser view:

WebDialer click-to-dial HTML sample

Username: Adams, Sally
Email: sadams
Phone: [23923](https://10.1.1.1:8443/webdialer/Webdialer?destination=23923)
Department: Human Resources

Username: Smith, Steve
Email: ssmith
Phone: [30271](https://10.1.1.1:8443/webdialer/Webdialer?destination=30271)
Department: Human Resources

153278

For information and examples of SOAP-over-HTTP source code to be used in click-to-dial desktop applications, refer to the WebDialer API Programming information in the *Cisco Unified Communications Manager Developers Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/products_programming_reference_guides_list.html

WebDialer Redundancy

WebDialer application redundancy can be provided at two levels:

- Redundancy at the component and service level

At this level, redundancy must be considered with regard to WebDialer and CTIManager service redundancy. Likewise, the lack of publisher redundancy and the impact of this component failing should also be considered.

- Redundancy at the device and reachability level

At this level, redundancy should be considered as it relates to user phones and the WebDialer user interface.

Service and Component Redundancy

As shown in [Figure 22-15](#), WebDialer functionality is primarily dependent on the Cisco WebDialer Web Service and the Cisco CTIManager services. In the case of the WebDialer service, redundancy is provided by listing multiple WebDialer server IP addresses in the List of WebDialers service parameter (see [WebDialer Service Parameters, page 22-40](#)) and enabling the service on multiple nodes within the cluster. In the case of CTIManager, redundancy is automatically built-in using a primary and backup mechanism. Two CTIManager servers or services can be defined within the cluster using the Primary Cisco CTIManager and the Backup Cisco CTIManager service parameters (see [WebDialer Service Parameters, page 22-40](#)). By configuring these parameters, you can make the CTIManager service redundant. Thus, if the primary CTIManager fails, CTIManager services can still be provided by the backup CTIManager. If the WebDialer server to which the web-based (or desktop) application is pointing fails and the primary and backup CTIManager services on cluster nodes also fail, the WebDialer application will fail. The WebDialer service is not dependant upon the Unified CM publisher

Device and Reachability Redundancy

Redundancy for WebDialer at the device level relies on a number of mechanisms. First and foremost, user phones rely on the built-in redundancy provided by a combination of the device pool and Unified CM group configuration for device registration.

The WebDialer service can run on multiple Unified CM subscribers to provide redundancy, however many applications might not be equipped to handle more than one IP address. Cisco recommends using the Server Load Balancing (SLB) feature available in Cisco IOS to mask the presence of multiple WebDialer servers in the enterprise. The SLB feature provides a virtual IP address or DNS-resolvable hostname that front-ends the real IP addresses of the WebDialer and/or Redirector servers. The SLB feature can be configured to monitor the status of multiple WebDialer servers, automatically redirecting requests during failure events. The SLB feature can also be configured to load-balance WebDialer requests when additional click-to-dial capacity is required. DNS Service (SRV) records can also be used to provide redundancy.

In enterprise deployments, link cost might also be an important consideration. The Cisco Global Site Selector (GSS) feature in Cisco IOS builds upon the capabilities of the SLB feature by adding link cost and availability to the load-balancing algorithm.

Guidelines and Restrictions for WebDialer

The following guidelines and restrictions apply with regard to deployment and operation of WebDialer within the Unified CM telephony environment:

- The administrator should ensure that all WebDialer users are associated with a phone or device profile in the Unified CM end-user directory.
 - If the user selects "Use permanent device" under the Cisco WebDialer Preferences screen with no phone association, then the following message is received when the Dial button is pressed:
"No supported device configured for user"
 - If the user selects Use Extension Mobility under the Cisco WebDialer Preferences screen with no device profile association (or the user is not logged in using a profile), then the following message is received when the Dial button is pressed:
"Call to <dialled_ number> failed: User not logged in on any device"

**Note**

WebDialer and EM applications may be used together. For information about WebDialer interaction with EM, see [WebDialer Interactions with EM, page 22-49](#).

- When configuring the List of WebDialers service parameter (see [WebDialer Service Parameters, page 22-40](#)), port number 8443 must be specified along with WebDialer IP address.
- If using Client Matter Codes (CMC) or Forced Authorization Codes (FMC), WebDialer users must enter the proper code at the tone by using the phone's keypad. Failure to enter the appropriate code at the tone will result in call failure signaled by a reorder tone.

WebDialer Performance and Scalability

The WebDialer and Redirector services can run on one or more subscriber nodes within a Unified CM cluster, and they support the following capacities:

- Each WebDialer service can make up to 2 calls per second (7,200 calls per hour) per node.
- Each Redirector service can handle up to 10 calls or requests per second. Because each WebDialer service can handle 2 calls per second, a single Redirector service can support a maximum of five WebDialer services.

Use the following formula to determine the number of WebDialer servers required to handle the user load:

$$(\text{Number of WebDialer users}) * (\text{Average BHCA}) / 3600 = \text{Number of calls per second}$$

**Note**

The typical user makes an average of 6 calls during the busy hour (6 BHCA), but the average varies from organization to organization.

For example, suppose an organization wishes to enable 3000 users for click-to-dial using WebDialer. Each user is assumed to make 6 calls per hour. Using the formula above, this organization needs to deploy enough WebDialers to handle 5 calls per second:

$$(3000 \text{ WebDialer users}) * (6 \text{ BHCA}) / 3600 = \text{Requirement of 5 calls per second}$$

Because each WebDialer service can make up to 2 calls per second, a total of 3 nodes should be configured to run the WebDialer service in this example.

In order to maintain WebDialer capacity during a server failure, additional backup WebDialer servers should be deployed to provide redundancy.

The Cisco WebDialer application interacts with the CTIManager for phone control. When enabled, each WebDialer service opens a single persistent CTI connection to the CTIManager. In addition, each WebDialer MakeCall request generates a temporary CTI connection. The number of CTI connections required to handle WebDialer call rates must be considered with regard to the overall cluster limit for CTI connections (10,000 CTI connections per cluster with the MCS-7845 platform or 3200 CTI connections per cluster with the MCS-7835 and MCS-7825 servers). If additional CTI connections are required for other applications, they can limit the capacity of the WebDialer application.

WebDialer Interactions with EM

WebDialer users can log in to their phones using EM. EM users wishing to use WebDialer simply need to select the Use Extension Mobility setting under the Cisco WebDialer Preferences page.



CHAPTER 23

Cisco Mobility Applications

Last revised on: September 27, 2007

Cisco mobility applications provide the ability to deliver features and functionality of the enterprise IP Communications environment to mobile workers wherever they might be. Mobility functionality delivered within the Cisco Unified Communications solution is provided via the Cisco Unified Mobility application server. The Mobility server communicates with Cisco Unified Communications Manager (Unified CM) via the Java Telephony Application Programming Interface (JTAPI) and AVVID XML Layer (AXL) and provides the following mobility application functionality:

- **Mobile Connect**

Mobile Connect or Single Number Reach provides Cisco Unified Communications users with the ability to be reached via a single enterprise phone number that will ring on both their IP desk phone and their cellular phone simultaneously. Mobile Connect users can pick up an incoming call on either their desk or cellular phones and at any point can move the in-progress call from one of these phones to the other without interruption.

- **Mobile Voice Access**

Mobile Voice Access or Enterprise Dial tone provides mobile users with the ability to make calls from their mobile phone as if they were calling from their enterprise IP desk phone. The feature provides a cost savings in terms of toll charges for long distance or international calls, and it provides the enterprise with an easy way to track phone calls made by employees via a uniform and centrally located set of call detail records.

This chapter examines the following mobility applications as provided by the Cisco Unified Mobility application server:

- [Mobile Connect, page 23-1](#)
- [Mobile Voice Access, page 23-15](#)

This chapter also examines general guidelines and restrictions as well as performance and capacity of the Cisco Unified Mobility application server:

- [Unified Mobility, page 23-21](#)

Mobile Connect

Mobile Connect is an application that allows an incoming call to an enterprise user to be offered to the user's IP desk phone as well as up to four configurable remote destinations. Typically a user's remote destination is their mobile or cellular telephone. Once the call is offered to both the desktop and remote

destination phone(s), the user can answer at any of those phones. Upon answering the call on one of the remote destination phones or on the IP desk phone, the user has the option to hand off or pick up the call on the other phone.

Mobile Connect Phone Support

Mobile Connect is supported on any of the following phone models:

- Cisco Unified IP Phone 7906G
- Cisco Unified IP Phone 7911G
- Cisco Unified IP Phones 7940G, 7941G, and 7941G-GE
- Cisco Unified IP Phones 7960G, 7961G, and 7961G-GE
- Cisco Unified IP Phones 7970G and 7971G-GE

All phones listed above support Mobile Connect remote destination pickup with both Skinny Client Control Protocol (SCCP) and Session Initiation Protocol (SIP).

Beginning with Cisco Unified Communications Manager release 5.1.1(b), the Cisco Unified IP Phones 7942G, 7945G, 7962G, 7965G, and 7975G are also supported with Mobile Connect for both SCCP and SIP.

The following phones support Mobile Connect with SCCP only:

- Cisco Unified IP Phone 7905G
- Cisco Unified IP Phones 7912G and 7912G-A
- Cisco Unified Wireless IP Phones 7920 and 7921G
- Cisco Unified IP Phone 7940G
- Cisco Unified IP Phone 7960G

Mobile Connect calls that are answered at a remote destination phone can be handed off to any of the IP phone models listed above, regardless of remote destination phone type, because pickup at the IP desk phone requires only that the user hang up the remote destination phone. All pickup functionality in these scenarios is handled by Unified CM, Unified Mobility, and the enterprise PSTN gateway that handled the outgoing call to the remote destination.



Note

The Cisco Unified IP Phone 7905G, 7912G, 7912G-A, 7940G, and 7960G devices running SIP can be used with Mobile Connect to allow incoming calls to ring at both the desk phone and remote destination phone. However, once the call is answered, desk phone pickup or remote destination phone pickup is not possible with these phones when running SIP.

Unified CM Services and Mobility System Parameters

For the Mobile Connect application to function properly, you must enable certain services on Unified CM and configure certain parameters on Cisco Unified Mobility. This section lists those required services and parameters.

Unified CM Services for Mobile Connect

The Mobile Connect application relies in part on the Cisco CTIManager and Cisco AXL Web Service feature services, which must be activated manually from the Serviceability page on one or more nodes. The Mobile Connect application also relies on the Cisco CallManager Cisco IP Phone Services network service, which is automatically activated on Unified CM nodes during installation.

The Cisco CTIManager service interfaces and interacts with the Cisco CallManager Service and the Mobility application server for Mobile Connect call control and monitoring. The Cisco AXL Web Service interfaces and interacts with the Unified CM database and the Mobility application server for authentication of Mobile Connect users.

The Cisco Unified IP Phone Services service is needed to provide access to the Mobile Connect phone service for remote phone pickup from the desktop phone. The URL used to define the Mobile Connect phone service is:

`http://<Mobility_Server_IP-Address>:8081/ip_phone/main.xml`

(where <Mobility_Server_IP-Address> is the IP address of the Mobility application server)

In addition, you must configure a parameter named **dn** (case-sensitive) for the Mobile Connect phone service. This is a required parameter. When a device is subscribed to the Mobile Connect phone service, the shared line directory number must be entered in this parameter field.

Unified Mobility System Parameters

The following items represent a partial list of Cisco Unified Mobility system parameters related to Mobile Connect functionality:

- **Enable Mobile Connect Feature** (Default value: yes)
This parameter determines whether the Mobile Connect feature is enabled. This is a system-wide parameter. (Note that this parameter can also be configured per remote destination per user within Unified Mobility.)
- **Maximum Wait Time For Desk Phone Pickup** (Default value: 10000)
This parameter specifies the maximum time, in seconds, that the system will allow the desk phone to pick up after the remote destination phone ends a Mobile Connect call. If this time is exceeded, the call will be terminated and desk phone pickup will not be possible. (Note that this parameter can also be configured per user within Unified Mobility.)
- **Enable Cellular Phone Pickup** (Default value: yes)
This parameter determines whether a Mobile Connect call answered by the desk phone can be picked up at the remote destination phone. This is a system-wide parameter. (Note that this parameter can also be configured per user under the line appearance configuration page within Unified Mobility.)
- **Minimum Cellular Phone Ring Timer** (Default value: 1500)
This parameter specifies the minimum time, in milliseconds, that must pass before the remote destination phone can be answered after the call is extended to this phone. If the call is answered before this time, the system assumes that the remote destination voicemail has picked up the call.

This feature is designed to prevent calls to enterprise phones from being forwarded to the remote destination voicemail system. The Enable Minimum Cellular Phone Ring Timer parameter must be set to **yes** in order for this parameter to be used. (Note that this parameter can also be configured per remote destination per user within Unified Mobility.)

- Maximum Cellular Phone Ring Timer (Default value: 19000)

This parameter specifies the maximum time, in milliseconds, that a call to the remote destination phone will be extended. If the call is not answered before this time, the remote destination call leg is disconnected. The Enable Maximum Cellular Phone Ring Timer parameter must be set to **yes** in order for this parameter to be used. (Note that this parameter can also be configured per remote destination per user within Unified Mobility.)

- Delay Before Ringing Cellular Phone (Default value: 4000)

This parameter specifies the minimum time, in milliseconds, that the system will wait before extending a Mobile Connect call to configured remote destinations. The Enable Delay Before Ringing Cellular Phone system parameter must be set to **yes** in order for this parameter to be used. (Note that this parameter can also be configured per user within Unified Mobility.)

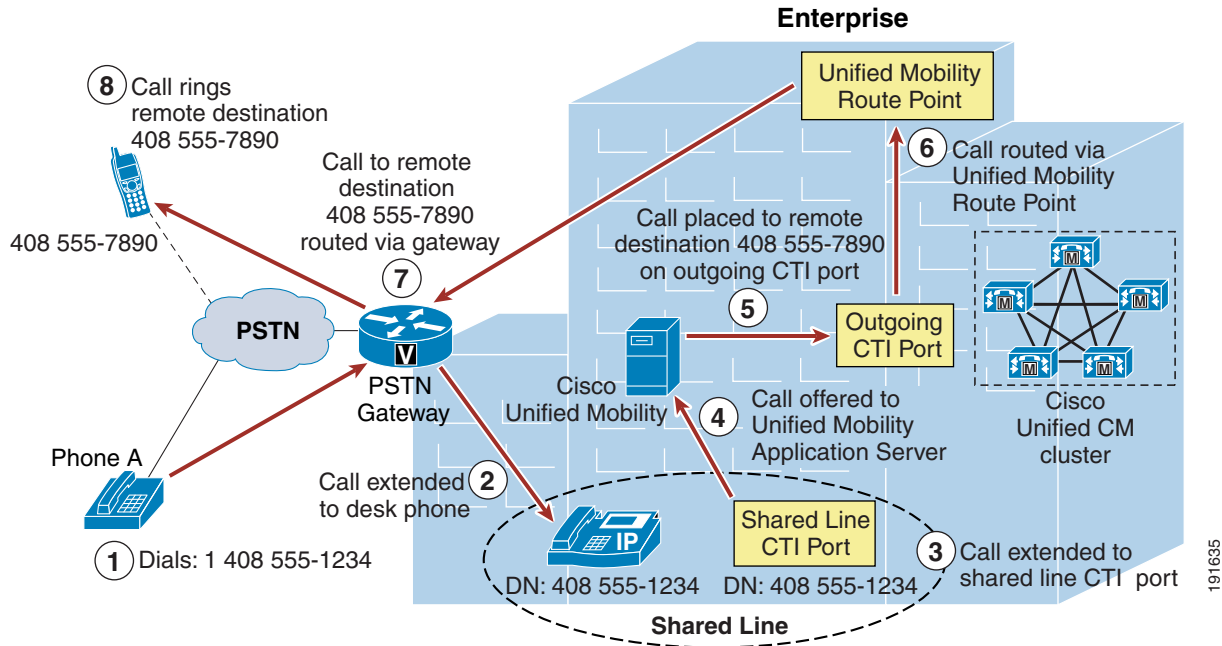
- Unified CM AXL Server Name or IP Address (Default value: *<blank>*)

This parameter specifies the Unified CM node that should be used for AXL interaction.

Mobile Connect Functionality

[Figure 23-1](#) illustrates a basic Mobile Connect call flow. In this example, Phone A on the PSTN calls a Mobile Connect user's enterprise directory number (DN) 408-555-1234 (step 1). The call comes into the enterprise PSTN gateway and is extended via Unified CM to the IP phone with DN 408-555-1234 (step 2), and this phone begins to ring. The call is also extended to the Unified CM shared line CTI port with the same DN (step 3). In turn, this call is offered to the Unified Mobility application server (step 4). Unified Mobility then places a call to the Mobile Connect user's configured remote destination (in this case 408-555-7890) via a Unified CM outgoing CTI port (step 5). The outgoing call is routed via the Mobility CTI route point in Unified CM (step 6) and then routed to the remote destination via the PSTN gateway (step 7). Finally the call rings at the remote destination PSTN phone with number 408 555-7890 (step 8). The call can then be answered at either phone.

Figure 23-1 Mobile Connect



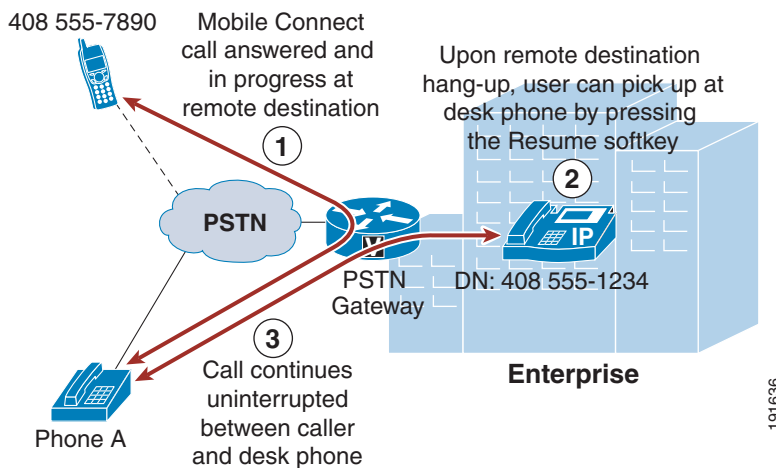
Typically a Mobile Connect user's configured remote destination is their mobile phone on a global system for mobile communications (GSM) or cellular network; however, any destination reachable via the PSTN can be configured as a user's remote destination. Furthermore, a Mobile Connect user can have up to four remote destinations configured, so an incoming call could potentially ring as many as four PSTN phones as well as the user's desk phone. If the incoming call is answered at the remote destination, the voice media path will be hairpinned within the enterprise PSTN gateway utilizing two DS0s or analog ports. This utilization must be considered when deploying the Mobile Connect feature.

**Note**

In order for Mobile Connect to work as in [Figure 23-1](#), ensure that the system-wide Enable Mobile Connect system parameter is set to **yes** (see [Unified Mobility System Parameters, page 23-3](#)) and that the Enable Mobile Connect parameter has been set to **yes** on the Mobile Connect user's configured remote destination(s).

Desk Phone Pickup

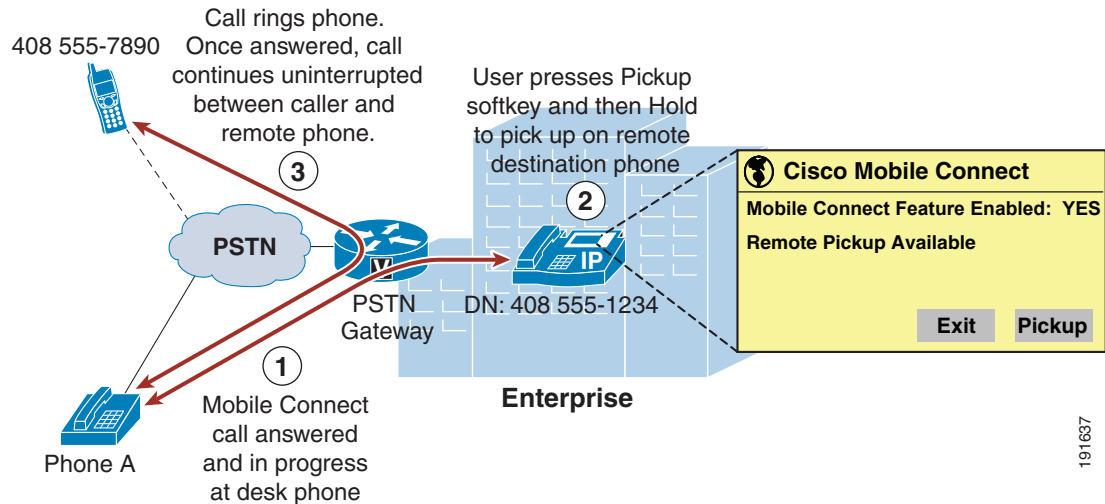
As illustrated in [Figure 23-2](#), once a user answers a Mobile Connect call at the remote destination device (step 1: in this case, 408 555-7890), at any point the user can hang up the call at the remote destination and pick it up again at their desk phone by simply pressing the Resume softkey on the desk phone (step 2: at DN 408 555-1234 in this case). The call continues uninterrupted between the original caller at Phone A and the desk phone (step 3).

Figure 23-2 Desk Phone Pickup

Desk phone pickup can occur whenever a call is in progress at a configured remote destination phone and that phone is hung up. The option to pick up or resume the call at the desk phone is always available for a certain amount of time. For this reason, it is good practice for the Mobile Connect user to ensure that the calling phone hangs up before the remote destination phone is hung up. This ensures that the call cannot be resumed at the desk phone by someone else. By default, the call remains available for pickup at the desk phone for 10 seconds after the remote destination phone hangs up; however, you can configure this setting by changing the system-wide parameter *Maximum Wait Time For Desk Phone Pickup* or the specific user line appearance parameter (see [Unified Mobility System Parameters](#), page 23-3).

Remote Destination Phone Pickup

Figure 23-3 illustrates Mobile Connect remote destination phone pickup functionality. Assuming Phone A calls Mobile Connect user's enterprise DN 408 555-1234 and the call is answered at the user's desk phone and is in progress (step 1), the user must push the *Services* button on the phone and select the Mobile Connect phone service. Assuming the Mobile Connect feature is enabled for this phone and that remote pickup is available, the user presses the *Pickup* softkey and then the *Hold* softkey (step 2). A call is generated to the user's remote destination phone (in this case, 408 555-7890), and the remote phone begins to ring. Once the call is answered at the remote phone, the call continues uninterrupted between Phone A and the Mobile Connect user's remote phone with number 408 555-7890 (step 3).

Figure 23-3 Remote Destination Phone Pickup

After pushing the Pickup softkey, the Mobile Connect user can also wait to push the Hold softkey until the remote destination phone has begun to ring or even until the remote destination phone has been answered. However, in order for the voice media path to be extended to the remote destination phone, the Hold softkey must be pushed. When a Mobile Connect user has multiple remote destinations configured, each remote destination will ring when the Pickup softkey is pressed, and the user can answer the desired phone.

**Note**

In order for remote destination phone pickup to work as in [Figure 23-3](#), ensure that the system-wide parameter Enable Cellular Phone Pickup is set to **yes** (see [Unified Mobility System Parameters, page 23-3](#)) and that the Enable Cellular Phone Pickup parameter has been set to either **<System Default>** or **Enable** on each Mobile Connect user's line appearance configuration page.

Single Enterprise Voicemail Box

An additional feature that Unified Mobility provides in conjunction with Mobile Connect is the ability to have a single voicemail box for all enterprise business calls. This prevents a user from having to check multiple mailboxes (enterprise, cellular, home, and so forth) for calls to their enterprise phone number. To help implement this feature, Unified Mobility provides a set of timers that can be used in conjunction with call-forward timers on Unified CM. The purpose of these timers is to ensure that, when and if a call is forwarded to a voicemail box on ring-no-answer, the call is forwarded to the enterprise voicemail box rather than any remote destination voicemail box. This behavior can be achieved in one of two ways:

- Ensure the forward-no-answer time is shorter at the desk phone than at the remote destination phones.

To do so, ensure that the global Forward No Answer Timer field in Unified CM or the No Answer Ring Duration field under the individual phone line is configured with a value that is less than the amount of time a remote destination phone will ring before forwarding to the remote destination voicemail box. In addition, the Unified Mobility system parameter Delay Before Ringing Cellular Phone (see [Unified Mobility System Parameters, page 23-3](#)) can be used to delay the ringing of the remote destination phone in order to further lengthen the amount of time that must pass before a remote destination phone will forward to its own voicemail box.

- Ensure that the remote destination phone stops ringing before it is forwarded to its own voicemail box.

You can accomplish this by setting the Unified Mobility system parameter Maximum Cellular Phone Ring Timer (see [Unified Mobility System Parameters, page 23-3](#)) to a value that is less than the amount of time that a remote destination phone will forward to its voicemail box. This ensures that the remote destination phone stops ringing before the call can be forwarded to its own voicemail box.

To ensure that the call is still forwarded to the enterprise voicemail box in cases where a user's phone is busy and call waiting is not available or in cases where a user's cellular phone is out of range, you can use the Unified Mobility system parameter Minimum Cellular Phone Ring Timer (see [Unified Mobility System Parameters, page 23-3](#)). If the call is forwarded and answered right away by the remote destination voicemail, this parameter ensures that the call leg extended to the remote destination phone will be disconnected, thus allowing additional time for the desk phone to be answered or for the enterprise voicemail system to handle the call.

Mobile Connect Architecture

The architecture of the Mobile Connect feature is as important to understand as its functionality. [Figure 23-4](#) depicts the message flows and architecture required for Mobile Connect. The following sequence of interactions and events can occur between the Unified Mobility applications server, Unified CM, and the Mobile Connect user:

1. Unified Mobility communicates with Unified CM CTIManager for CTI monitoring and call control via JTAPI (see step 1 in [Figure 23-4](#)). This communication is maintained via the Unified CM Links on the Unified Mobility application servers. Specifically, these are the Outgoing Port User Link and the Shared Line User Link.
2. Unified Mobility interacts with Unified CM via the AXL interface for user configuration authentication (required for using of the Cisco Unified Mobility user web interface) and CTI user link authentication (see step 2 in [Figure 23-4](#)).
3. The Mobile Connect phone user who wishes to pick up an in-progress call on their remote destination phone pushes the Services button on their desk phone to get a list of services (see step 3 in [Figure 23-4](#)).
4. The Unified CallManager Cisco IP Phone Services service returns a list of phone services to which the user's phone is subscribed (see step 4 in [Figure 23-4](#)).



Note

A Service URL button can be configured for the Mobile Connect phone service on a user's phone so that the user can press a line or speed-dial button to generate a direct call to the Mobile Connect service. In this case, the interactions between the phone and the Unified CallManager Cisco IP Phone Services service (steps 3 and 4) are bypassed.

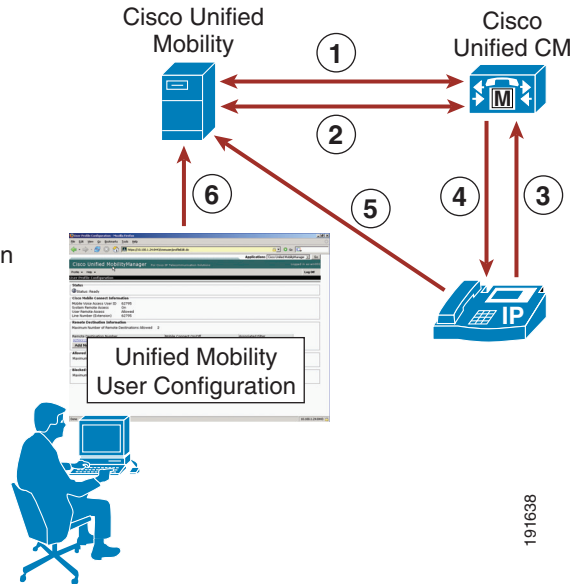
5. The Mobile Connect user then selects the Mobile Connect phone service for Remote Pickup and is redirected to the Unified Mobility application server that is hosting this phone service (see step 5 in [Figure 23-4](#)).
6. In addition, Mobile Connect users can use the user web interface to configure their own settings via the web-based configuration pages on the Mobility application server at

`http://<Mobility_Server_IP-Address>/cmmuser/`

where `<Mobility_Server_IP-Address>` is the IP address of the Unified Mobility application server (see step 6 in [Figure 23-4](#)).

Figure 23-4 Mobile Connect Architecture

- ① Links between Unified Mobility and Unified CM CTIManager for CTI monitoring and call control via JTAPI.
- ② Unified Mobility interacts with Unified CM via AXL for user configuration authentication and CTI user link authentication.
- ③ Mobile Connect phone user pushes Services button to get list of services.
- ④ Unified CM IP Phone Services Service returns subscribed phone services list.
- ⑤ User selects Mobile Connect phone service for Cell Pickup hosted on Unified Mobility server.
- ⑥ Web-based user configuration of Mobile Connect features.



Mobile Connect Dial Plan Considerations

Dial plan configuration is extremely important for Mobile Connect. To ensure that calls to a Mobile Connect user's enterprise DN not only ring the desk phone but also get extended to the user's configured remote destination or destinations, it is important to configure the shared line and outgoing CTI ports as well as the CTI route point with the appropriate dial plan class of control.

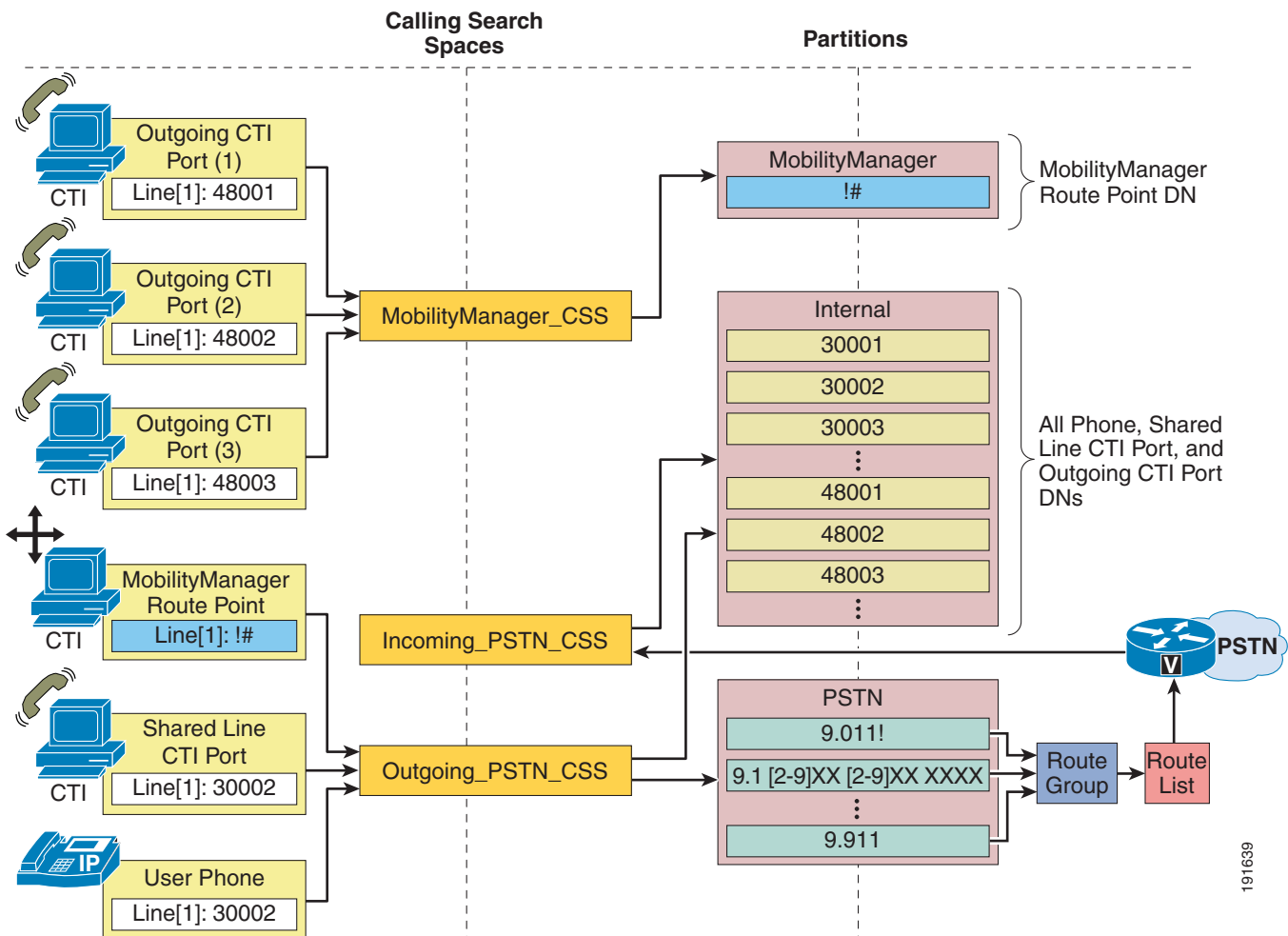
Figure 23-5 shows an example of a Mobile Connect dial plan with the requirements for calling search spaces and partitions as well as the configuration of the various types of devices within these dial plan components. One partition is required for Mobile Connect, and that partition is called Mobility for the example in Figure 23-5. It contains the Mobility CTI Route Point DN.

In addition, one calling search space is required for Mobile Connect, and in this example the calling search space is called MobilityManager_CSS. It contains the Mobility partition.

Besides showing the minimum dial plan requirements for Mobile Connect in the example in Figure 23-5, additional dial plan configuration has been depicted because most real-world telephony networks will have added or existing dial plan requirements that must be integrated with the Mobile Connect dial plan. First, two additional partitions, Internal and PSTN, are depicted. The Internal partition includes all the internal DNs, including the Mobile Connect user desk phone DNs as well as the Shared Line CTI port DNs and the Outgoing CTI port DNs. The PSTN partition includes a set of route patterns used for routing calls to the PSTN via the common route group (RG), route list (RL), and voice gateway mechanism. Next, two additional calling search spaces, Incoming_PSTN_CSS and Outgoing_PSTN_CSS, are shown. Incoming_PSTN_CSS serves as the calling search space to which incoming calls from the PSTN are sent. In order for these incoming calls to reach the appropriate enterprise phones, this calling search space contains the Internal partition so that calls can reach all the internal DNs and the Shared Line CTI ports. The Outgoing_PSTN_CSS calling search space contains both the Internal partition and the PSTN partition so that devices pointing to this calling search space can reach not only internal phones but also external PSTN phones.

This is the extent of the Mobile Connect dial plan in this example. However, it is also important to properly configure the various phone and CTI ports and route point DN or lines with the appropriate calling search spaces so that call routing works as required. In this case all Mobile Connect user phones, Shared Line CTI Ports, and the Mobility CTI Route Point lines would be configured with the Outgoing_PSTN_CSS calling search space so that all of these lines can reach the route patterns of the PSTN partitions and, therefore, the remote destination PSTN phone numbers configured for all Mobile Connect users. Likewise, these lines also need to be able to reach all internal phone, Shared Line CTI Port, and Outgoing CTI Port DNs for internal call routing. Next, all Outgoing CTI Port lines are configured with the MobilityManager_CSS calling search space so that all of these lines can reach the Mobility Route Point DN in the Mobility partition. In this way, the dial plan ensures that calls extended by the Unified Mobility application server to the Outgoing CTI Ports to Mobile Connect users' configured remote destinations will be routed via the Mobility Route Point.

Figure 23-5 Mobile Connect Dial Plan Example

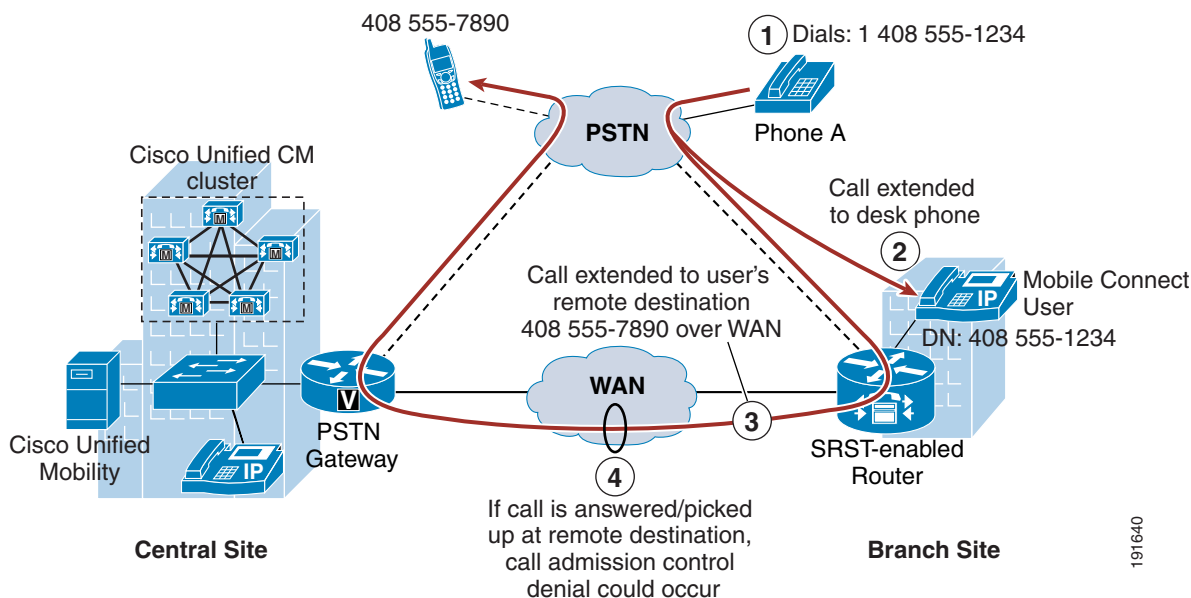


Centralized Gateway Call Routing

Unified Mobility call routing to the PSTN for Mobile Connect is handled by the Outgoing CTI port pool and the Mobility CTI Route Point configured within Unified CM. However, in order to integrate with Unified CM so that this routing mechanism can be used, Unified Mobility relies on a pair of CTI connections (Shared Line User Link and Outgoing Port User Links). While Unified Mobility has the ability to configure and maintain more than one set of CTI connections to Unified CM, it has no mechanism for controlling which outgoing calls are routed on which set of CTI connections. As a result, there is no mechanism for supporting more than one Outgoing CTI port pool or more than one Mobility CTI Route Point within Unified CM. What this means for routing Mobile Connect calls to the PSTN is that all Mobile Connect calls must use the same Mobility CTI route point. Therefore, there is no ability to easily route groups of Mobile Connect calls to the PSTN based on groups of users and their location. Instead, all Mobile Connect calls are typically routed out one or more centralized gateways for all users regardless of their location. For this reason, PSTN call routing is dependent on centralized gateways.

The use of centralized gateways can present problems for Mobile Connect calls from the perspective of a dial plan and call routing when Unified Mobility is deployed in a centralized call processing environment where there are multiple remote branch sites with local PSTN gateways and Mobile Connect users. As [Figure 23-6](#) illustrates, call admission control on the WAN between sites can prevent Mobile Connect calls from being completed if there is not sufficient bandwidth on the WAN. In this example, Phone A (a PSTN phone) dials a Mobile Connect user's enterprise DID, DN 408 555-1234 (step 1). The incoming call comes into the enterprise via the branch site PSTN gateway and is extended to the user's desk phone (step 2). The call is also extended to the user's configured remote destination (in this case, 408 555-7890) over the WAN to the centralized gateway in the central site (step 3). Because this call could potentially be routed over the WAN and out the centralized gateway, it is possible the call will fail due to a call admission control denial if there is not sufficient bandwidth on the WAN link (step 4).

Figure 23-6 Mobile Connect with Centralized Gateway



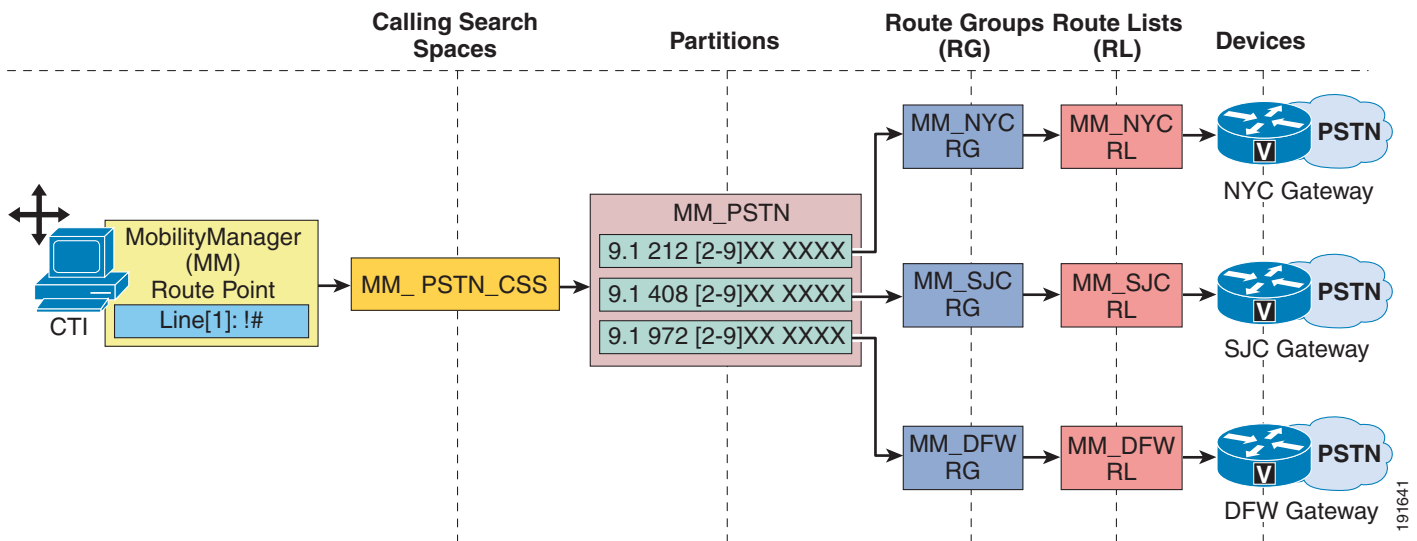
191640

Distributed Gateway Call Routing

As the call flow depicted in [Figure 23-6](#) illustrates, the outgoing PSTN call to the Mobile Connect user's remote destination egresses the enterprise on a different PSTN gateway than the original incoming PSTN call to the user's enterprise DN. In centralized call processing environments with low-speed WAN links between the central site and remote branch sites where Mobile Connect functionality is desired, it might be appropriate to modify the Mobile Connect dial plan so that Mobile Connect calls to a user's remote destination can be routed out the same PSTN gateway on which the call came into the enterprise. In many cases, this provides more efficient call routing and prevents the potential for call admission control denials on WAN links interconnecting sites.

[Figure 23-7](#) shows a dial plan method for achieving distributed gateway call routing in a Mobile Connect environment. In this example the Mobility CTI route point points to the MM_PSTN_CSS calling search space, which contains the MM_PSTN partition. The MM_PSTN partition in turn includes a series of site-specific route patterns in the form of 9 + 1 + area code or Number Plan Area (NPA), each pointing to a different route group (RG), route list (RL), and PSTN gateway mechanism, thus ensuring that outgoing PSTN calls from the Mobility CTI route point are routed out a particular site's gateway or gateways. In this case all outgoing Mobile Connect calls dialed with an NPA or area code of 212 will be forwarded out the NYC RG, RL, and PSTN gateway construct (after the appropriate number of digits have been discarded as dictated by site-specific requirements). Likewise, outgoing Mobile Connect calls dialed with the 408 and 972 NPAs will be forwarded out the SJC RG, RL, and PSTN gateway construct and the DFW RG, RL, and PSTN gateway construct, respectively. Of course, this method of distributed gateway call routing can be made as granular as necessary to facilitate multiple gateway sites within the same NPA by adding more specific route patterns in the MM_PSTN partition.

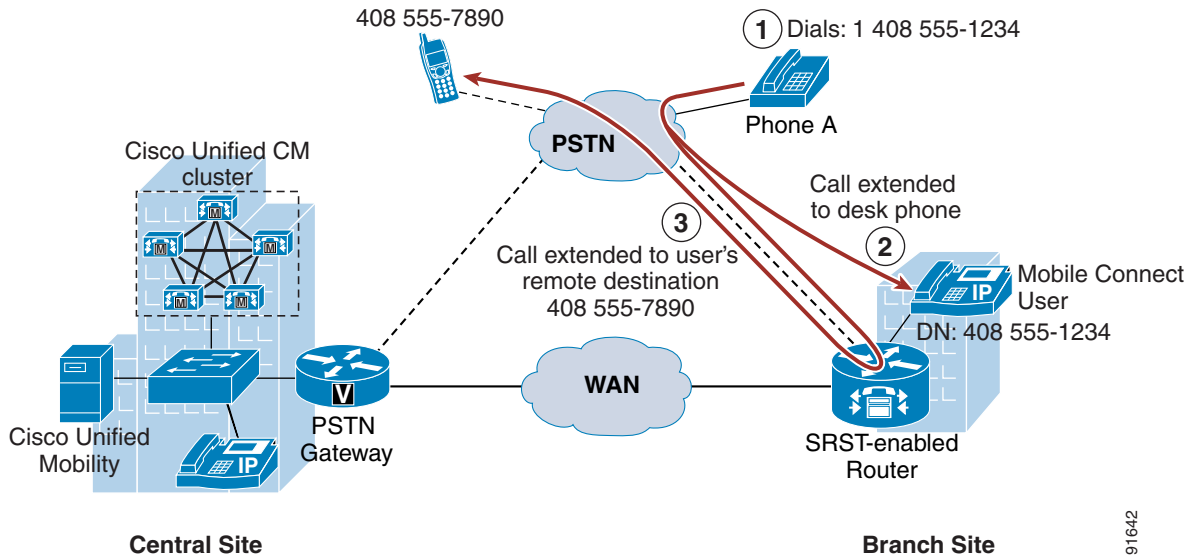
Figure 23-7 Mobile Connect Distributed Gateway Call Routing Dial Plan Example



With a distributed gateway call routing dial plan as depicted in [Figure 23-7](#), it is possible to avoid sending most outbound remote destination Mobile Connect calls across the WAN. [Figure 23-8](#) shows the previous call flow with this new dial plan in place. Now when the PSTN phone (Phone A) dials the Mobile Connect user's enterprise DID DN 408 555-1234 (step 1) just as before, the incoming call comes into the enterprise via the branch site PSTN gateway and is extended to the user's desk phone (step 2). However, now the call is extended to the user's configured remote destination (in this case, 408

555-7890) out the branch site PSTN gateway rather than across the WAN to the centralized gateway (step 3). By hairpinning the call at the local branch gateway, you can avoid the potential for call failure due to a call admission control denial if there is not sufficient bandwidth on the WAN link.

Figure 23-8 *Mobile Connect With Distributed Gateways*



One scenario that the dial plan example in [Figure 23-7](#) does not account for is one in which the NPA of a Mobile Connect user's remote destination phone is different than their enterprise DID NPA. As an example, referring again to [Figure 23-8](#), suppose the Mobile Connect user's remote destination phone number is 212 555-7890 rather than 408 555-7890. In that case, the outbound call to the user's remote destination phone would be extended across the WAN to the site with the gateway that handles the 212 NPA rather than routing out the local branch PSTN gateway. In order to handle situations like this one, route patterns with steering digits or site code prefixes can be used in the MM_PSTN partition so that calls are distributed to remote gateways, not based on the NPA of the called number, but based on the configured prefix of a user's remote destination. This adds complexity to the dial plan and increases the administrative burden for maintaining such a dial plan. It also puts the onus on the Mobile Connect user to understand and know the appropriate site code to prefix their remote destination numbers when configuring them via the Unified Mobility user web interface. However, this might be an acceptable trade-off if more efficient call routing is achieved.

Ultimately, each network administrator must decide whether the centralized gateway or distributed gateway (with or without site prefixes) method of Mobile Connect call routing is most appropriate for their enterprise, and they must weigh dial plan complexity against the potential for undesirable PSTN call routing.

Mobile Connect Redundancy

The Mobile Connect feature relies on the following components:

- Cisco Unified Mobility application server
- Unified CM servers
- PSTN gateway

Each component must be redundant or resilient in order for Mobile Connect to continue functioning fully during various failure scenarios.

Unified Mobility Application Server Redundancy

The application server provides the basis for Mobile Connect functionality. If this server fails, Mobile Connect users associated to the server will not be able to use Mobile Connect. There is no backup or redundancy mechanism in place to handle a complete Unified Mobility server failure.

In addition to the physical server itself, if any of the subcomponents within the Unified Mobility application server fail, Mobile Connect functionality can be lost or limited.

In order for Mobile Connect to function, both the Outgoing CTI User Link and the Shared Line CTI User Links must be in service. These links rely on the availability of Unified CM CTIManager nodes. Redundancy for these links is provided by specifying both a primary and secondary (or backup) Unified CM CTIManager node under the link configuration page. If the primary CTIManager node fails or if communication to the primary node is lost, the secondary CTIManager node can continue to provide connectivity for these links, thus ensuring that Mobile Connect functionality continues.

Cisco recommends dividing Mobile Connect users among three Shared Line CTI User links. This requires creation of three Mobility Shared Line CTI port users on Unified CM for each Unified Mobility application server. Each Mobility Shared Line user on Unified CM should be associated with one-third of the Shared Line CTI ports. On each Unified Mobility application server, three Shared Line CTI User Links should be created, with each one configured to use a different Mobility Shared Line CTI port user on Unified CM and a different primary Unified CM CTIManager node. By ensuring that Mobile Connect users are evenly distributed across the three Shared Line CTI User Links and across three CTIManager nodes, you can distribute the CTI load so that any CTIManager node failure affects only one-third of the Mobile Connect users. Furthermore, with this distributed configuration, system initialization is faster and failover to the backup CTIManager is faster.



Note

The Mobility CTI route point should be associated to only one of the Mobility Shared Line CTI port users within Unified CM.

In order for the Mobile Connect user web interface to function so that Mobile Connect users can configure their applications settings (remote destinations, cellular timers, caller filters, and so forth), the Mobility application server relies on AXL connectivity to a Unified CM server in order to authenticate Mobile Connect users. AXL connectivity is determined by the Unified CM AXL Server Name or IP Address system parameter on the Unified Mobility server (see [Unified Mobility System Parameters, page 23-3](#)). This parameter specifies a single Unified CM server for AXL connectivity. However, if this server were to fail, Unified Mobility has no failover mechanism for AXL connectivity and must wait until the configured Unified CM AXL server is back in service. If this connectivity is not available, the user cannot be authenticated, and therefore the user cannot make changes; however, existing Mobile Connect configuration and functionality will continue.

Unified CM Server Redundancy

The Unified CM server is an integral part of the Mobile Connect feature. Unified CM server failures can be problematic for Mobile Connect functionality in certain scenarios. First, CTIManager service node failures can affect Unified Mobility Shared Line and Outgoing CTI Port links as previously discussed, but redundancy is provided in the Unified Mobility server for primary and secondary nodes. In addition, Mobile Connect functionality is very dependent on a number of Unified CM CTI components including the shared line CTI ports, the outgoing CTI port pool, and the Mobility CTI route point. However, both

Unified CM group constructs (via device pools) for providing registration redundancy and CTIManager node redundancy for CTI connectivity within the cluster ensure that these CTI components will survive Unified CM node failures.

PSTN Gateway Redundancy

Because the Mobile Connect feature relies on the ability to extend a second call leg to the PSTN to reach the Mobile Connect user's remote destination phone, PSTN gateway redundancy is important. Should a PSTN gateway fail or be out of capacity, the Mobile Connect call cannot complete. Typically, enterprise IP telephony dial plans provide redundancy for PSTN access by providing physical gateway redundancy and call re-routing capabilities as well as enough capacity to handle expected call activity. Assuming that Unified CM has been configured with sufficient capacity, multiple gateways, and route group and route list constructs for call routing resiliency, the Mobile Connect feature can rely on this redundancy for uninterrupted functionality.

Mobile Voice Access

Mobile Voice Access, also referred to as System Remote Access, is a feature built on top of the Mobile Connection application, and it enables a Mobile Connect user who is outside the enterprise to make a call as though they are directly connected to Unified CM. This functionality is commonly referred to as Direct Inward System Access (DISA) in traditional telephony environments. This feature benefits the enterprise by limiting toll charges and consolidating phone billing directly to the enterprise rather than billing to each mobile user. Mobile Voice Access is accessed by calling a system-configured DID number that is answered and handled by an H.323 VoiceXML gateway. The VoiceXML gateway plays interactive voice response (IVR) prompts to the Mobile Voice Access user, requesting user authentication and input of a number to be dialed via the user phone keypad. Once the call is connected, users can pick up the call on their desk phones just as with a Mobile Connect call.

Mobile Voice Access Phone Support

Mobile Voice Access is supported for all incoming PSTN devices capable of sending DTMF digits in response to the IVR prompts. Cisco Unified IP Phone support for Mobile Voice Access is the same as for Mobile Connect (see [Mobile Connect Phone Support](#), page 23-2). Although an IP phone is not required to use Mobile Voice Access, a user wishing to pick up an in-progress Mobile Voice Access call on a desk phone will need access to one of the Mobile Connect supported phone models.

Cisco Unified CM Services and Mobility System Parameters

For the Mobile Voice Access application to function properly, you must enable certain services on Unified CM and configure certain parameters on Cisco Unified Mobility. This section lists those required services and parameters.

Unified CM Services for Mobile Voice Access

The Mobile Voice Access application relies in part on the Cisco CTIManager feature services, which must be activated manually from the Serviceability page on one or more nodes. The Cisco CTIManager service interfaces and interacts with the Cisco CallManager Service and the Mobility application server for Mobile Voice Access call control and monitoring.

Unified Mobility Mobile Voice Access System Parameters

The following items represent a partial list of Cisco Unified Mobility system parameters related to Mobile Voice Access functionality:

- Mobile Voice Access Numbers (Default value: <blank>)

This parameter indicates the Mobile Voice Access DID number(s). Use commas to delimit multiple DID numbers. This field can have a maximum of 200 characters, so the maximum number of DIDs will vary depending on the number of digits in each DID. For example, if all DIDs are seven digits, then a maximum of 25 DIDs can be configured ($7 * 25 + 24$ comma delimiters = 199 characters). As another example, if all DIDs are 10 digits, then a maximum of 18 DIDs can be configured ($10 * 18 + 17$ comma delimiters = 197 characters).

- Enable System Remote Access (Default value: no)

This parameter specifies whether Mobile Voice Access is enabled or not. To enable Mobile Voice Access system-wide, set this parameter to **yes**. (Note that this parameter can also be configured per Mobile Connect user within Unified Mobility.)

In addition to configuring the system parameters above, you must enable Mobile Voice Access for each user individually by setting the Enable User Remote Access field to **yes** under the Mobile Connect User Configuration page.

Mobile Voice Access IVR VoiceXML Gateway URL

The Mobile Voice Access feature requires the Unified Mobility VoiceXML application to reside on the H.323 gateway. The URL used to load this application is:

`http://<Mobility_Server_IP-Address>:8080/cmmivr/pages/IVRMainpage.vxml`

where <Mobility_Server_IP-Address> is the IP address of the Unified Mobility application server.

Mobile Voice Access Functionality

Figure 23-9 illustrates a Mobile Voice Access call flow. In this example, the Mobile Voice Access user on PSTN phone 408 555-7890 dials the Mobile Voice Access enterprise DID DN 408-555-2345 (step 1).



Note

The Mobile Voice Access enterprise DID number or numbers must be reachable and routeable via Unified CM and must also be configured in the Mobile Voice Access Number system parameter of the Unified Mobility application server (see [Unified Mobility Mobile Voice Access System Parameters, page 23-16](#)).

The call comes into the enterprise PSTN H.323 gateway, which also serves as the VoiceXML gateway. The user is prompted via IVR to enter their numeric user ID (followed by the # sign), PIN number (followed by the # sign), and then a 1 to make a Mobile Voice Access call, followed by the phone number they wish to reach. In this case, the user enters 9 1 972 555 3456 as the number they wish to reach (followed by the # sign) (step 2).

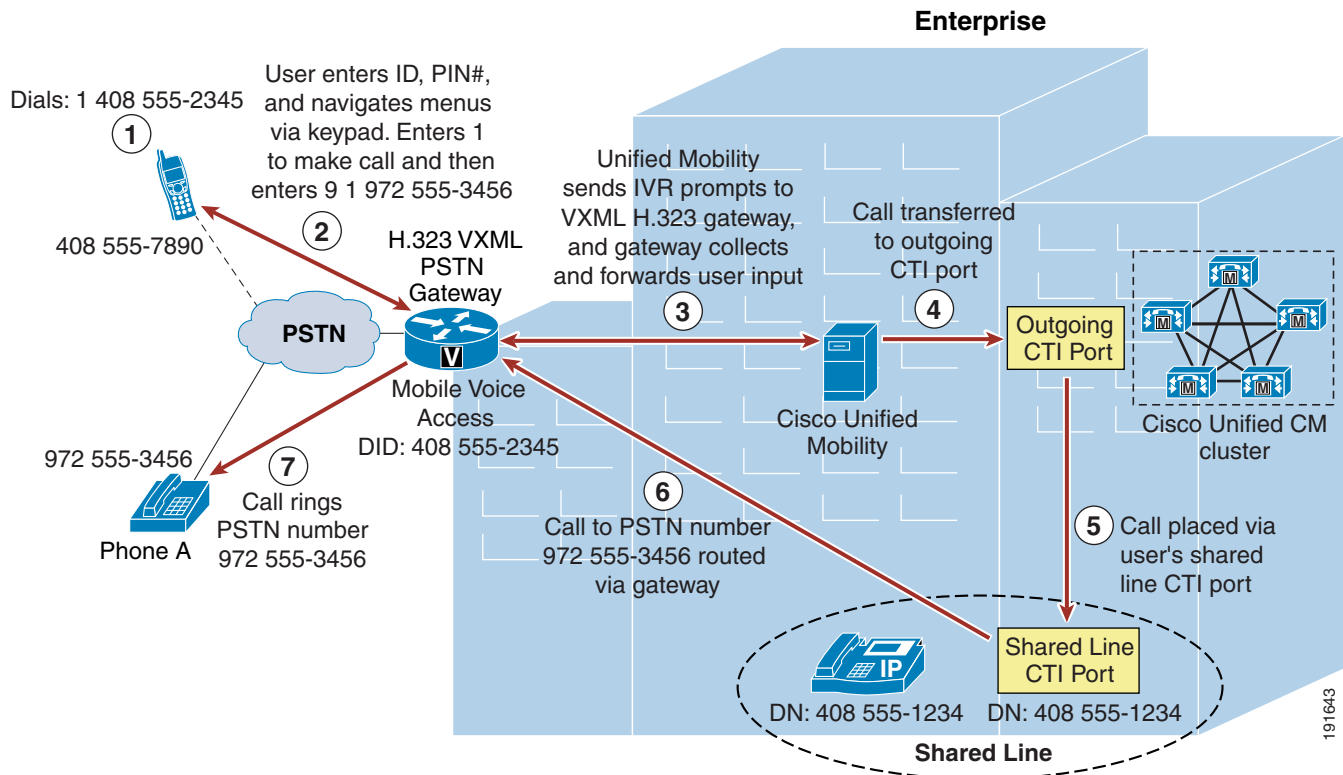


Note

If the PSTN phone from which the Mobile Voice Access user is calling is configured as a Mobile Connect remote destination for that user and the incoming caller ID can be matched against this remote destination by Unified Mobility, the user does not have to enter their numeric user ID. Instead they will be prompted to enter just the PIN number.

In the meantime, the Unified Mobility application server has forwarded IVR prompts to the VoiceXML gateway, the gateway has played these prompts to the user, and the gateway has collected user input including the numeric ID and PIN number of the user. This information is forwarded to the Unified Mobility server for authentication and to generate the call to 9 1 972 555 3456 (step 3). After authenticating the user and receiving the number to be dialed, Unified Mobility transfers the call to an outgoing CTI port (step 4). The call is then placed via the user's shared line CTI port (step 5). The outbound call to 972 555-3456 is routed via the PSTN gateway (step 6). Finally, the call rings at the PSTN destination phone with number 972 555-3456 (step 7). The voice media path of each Mobile Voice Access call is hairpinned within the enterprise PSTN gateway utilizing two DS0s or analog ports.

Figure 23-9 Mobile Voice Access



Note

In order for Mobile Voice Access to work as in [Figure 23-9](#), ensure that the system-wide Enable System Remote Access system parameter is set to **yes** (see [Unified Mobility Mobile Voice Access System Parameters, page 23-16](#)) and that the Enable System Remote Access parameter has also been set to **yes** on the Mobile Connect user's configuration page.

Mobile Voice Access Using Hairpinning

In deployments where the enterprise PSTN gateways are not using H.323 protocol, Mobile Voice Access functionality can still be provided using hairpinning on a separate gateway running H.323. Mobile Voice Access using hairpinning relies on off-loading the VoiceXML functionality to a separate H.323 gateway. [Figure 23-10](#) illustrates a Mobile Voice Access call flow using hairpinning. In this example, just as in the previous example, the Mobile Voice Access user on PSTN phone 408 555-7890 dials the Mobile Voice Access enterprise DID DN 408-555-2345 (step 1). The call comes into the enterprise PSTN

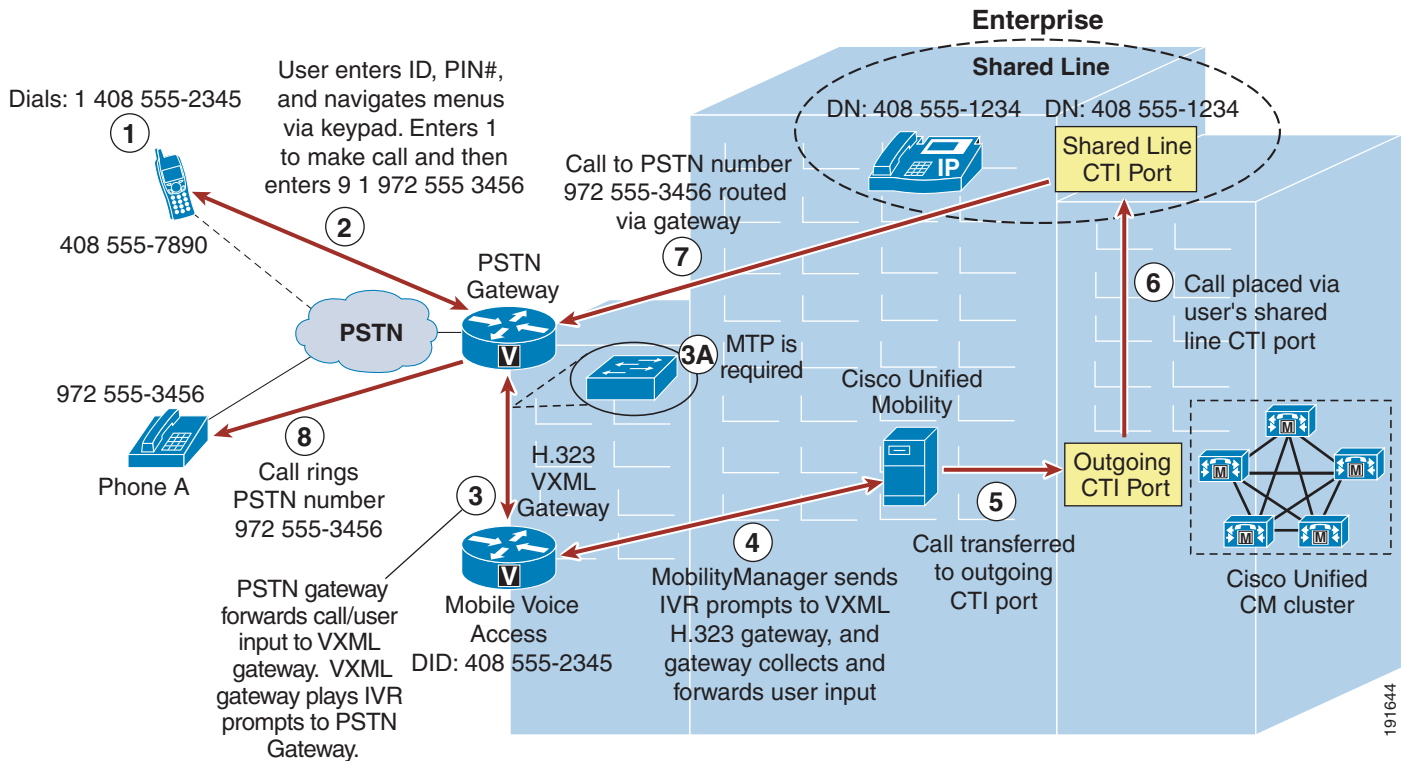
gateway and, just as before, the user is prompted via IVR to enter their numeric user ID (if required), PIN number, and then a 1 to make a Mobile Voice Access call, followed by the phone number they wish to reach. Again the user enters 9 1 972 555 3456 as the number they wish to reach (followed by the # sign) (step 2). Next the PSTN gateway collects and forwards the user input to the separate H.323 VoiceXML gateway with DID 408-555-2345, and the VoiceXML gateway plays IVR prompts to the PSTN gateway and the Mobile Voice Access user (step 3). Note that the media path between the PSTN gateway and the H.323 VoiceXML gateway requires a media termination point (MTP), therefore the media (including IVR prompts) must travel through this MTP (step 3A).

**Note**

In order to insert the MTP for Mobile Voice Access calls using hairpinning, the Media Termination Point Required setting must be checked (enabled) for the H.323 VoiceXML gateway configuration in Unified CM. The MTP will be allocated based on the configured Media Resource Group List (MRGL) of this gateway in Unified CM.

In the meantime, the Unified Mobility application server has forwarded IVR prompts to the H.323 VoiceXML gateway, and the VoiceXML gateway in turn has collected user input (including the numeric ID and PIN number of the user) from the PSTN gateway. This information is forwarded to the Unified Mobility server for authentication and to generate the call to 9 1 972 555 3456 (step 4). After authenticating the user and receiving the number to be dialed, Unified Mobility transfers the call to an outgoing CTI port (step 5). The call is then placed via the user's shared line CTI port (step 6). The outbound call to 972 555-3456 is routed via the PSTN gateway (step 7). Finally, the call rings at the PSTN destination phone with number 972 555-3456 (step 8). The voice media path in the case of each Mobile Voice Access call using hairpinning is not only hairpinned within the enterprise PSTN gateway utilizing two DS0s or analog ports, but it is also hairpinned at the H.323 VoiceXML gateway.

Figure 23-10 *Mobile Voice Access Using Hairpinning*



If you anticipate heavy or frequent use of Mobile Voice Access using hairpinning, Cisco recommends a hardware-based MTP or Cisco IOS software-based MTP. Whenever possible, the MTP resource should be co-resident on the H.323 VoiceXML gateway or at the very least in the same physical location as the H.323 VoiceXML gateway. This will minimize the chances of unnecessary hairpinning of media for Mobile Voice Access calls and will keep media hairpinning within the same H.323 VoiceXML gateway or the same gateway location.

Desk and Remote Destination Phone Pickup

Because Mobile Voice Access is tightly integrated with the Mobile Connect feature, once a Mobile Voice Access call has been established (with or without hairpinning), the user does have the option of using Mobile Connect functionality to pick up the in-progress call on their desk phone by simply hanging up the call on the originating phone and pushing the Resume softkey on their desk phone. The Mobile Voice Access call can then be picked up on the user's configured remote destination phone by selecting the Mobile Connect phone service and pushing the Pickup softkey.

Enabling and Disabling Mobile Connect

In addition to providing Mobile Voice Access users with the ability to make calls from the PSTN as though they are within the enterprise, the IVR functionality provided by Mobile Voice Access on the H.323 VoiceXML gateway also gives users the ability to remotely enable and disable their Mobile Connect functionality for each remote destination via their phone keypad. Rather than entering a 1 to make a call, Mobile Voice Access users enter a 2 to turn the Mobile Connect feature on and a 3 to turn

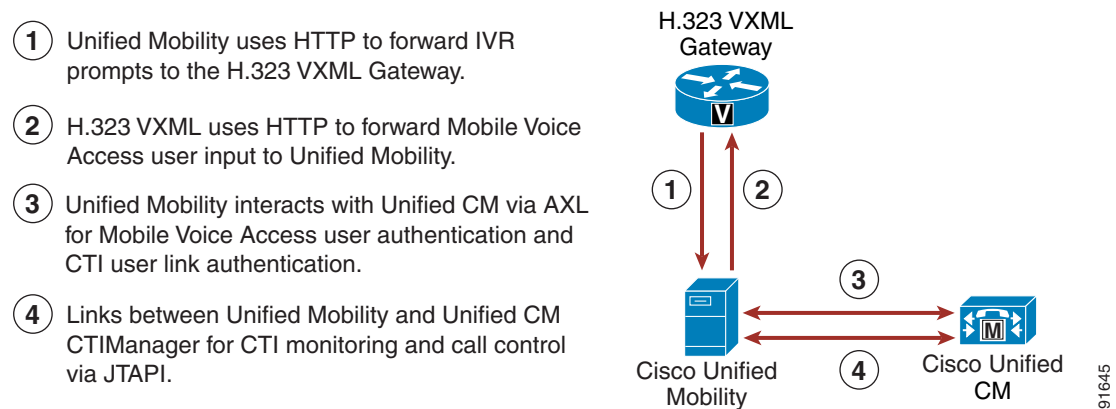
the Mobile Connect feature off. If a user has more than one remote destination configured, they are prompted to key in the remote destination phone number for which they wish to enable or disable the Mobile Connect feature.

Mobile Voice Access Architecture

The architecture of the Mobile Voice Access feature is as important to understand as its functionality. [Figure 23-11](#) depicts the message flows and architecture required for Mobile Voice Access. The following sequence of interactions and events can occur between the Unified Mobility applications server, Unified CM, and the H.323 VoiceXML (VXML) gateway:

1. The Unified Mobility application forwards IVR prompts and instructions to the H.323 VXML gateway via HTTP (see step 1 in [Figure 23-11](#)). This provides the VXML gateway with the ability to play these prompts for the inbound Mobile Voice Access callers.
2. The H.323 VXML gateway uses HTTP to forward Mobile Voice Access user input back to Unified Mobility (see step 2 in [Figure 23-11](#)).
3. Unified Mobility interacts with Unified CM via the AXL interface for Mobile Voice Access user authentication (see step 3 in [Figure 23-11](#)).
4. Unified Mobility communicates with Unified CM CTIManager for CTI monitoring and call control via JTAPI (see step 4 in [Figure 23-11](#)). This communication is maintained via the Outgoing Port User and Shared Line User Unified CM links on the Unified Mobility application servers.

Figure 23-11 Mobile Voice Access Architecture



Mobile Voice Access Dial Plan Considerations

Dial plan configuration for Mobile Voice Access requires the same dial plan class of control as required for Mobile Connect functionality. Mobile Voice Access relies on the same calling search spaces, partitions, route lists, route groups, and route patterns for proper operation. For more details, see [Mobile Connect Dial Plan Considerations](#), page 23-9.

One additional consideration for the Mobile Voice Access dial plan is the Mobile Voice Access DID. This DID provides access from the PSTN to the VoiceXML and IVR functionality of Mobile Voice Access. Given a centralized call processing environment with multiple branch sites, it is a good idea to provide local DIDs and VoiceXML or IVR H.323 gateway service at each branch site. This not only

provides toll savings by providing Mobile Voice Access users with a non-toll local access number, but by providing local H.323 VoiceXML gateway service in the case of Mobile Voice Access using hairpinning, it also prevents the call media from hairpinning over the WAN to a centralized H.323 VoiceXML gateway.

Mobile Voice Access Redundancy

The Mobile Voice Access feature relies on the same components and redundancy mechanisms as the Mobile Connect feature (see [Mobile Connect Redundancy, page 23-13](#)). However, the H.323 VoiceXML gateway and Mobile Voice Access DID requirements of Mobile Voice Access do present additional redundancy requirements. Unlike the Mobile Connect feature, Mobile Voice Access relies on a specific gateway protocol (H.323) and functionality (VoiceXML) as well as the need for specific feature access DIDs. Therefore, in addition to the components required for the Mobile Connect feature, the H.323 VoiceXML gateways required for Mobile Voice Access must also be redundant. As with regular PSTN gateways, H.323 VoiceXML gateways (whether for Mobile Voice Access or Mobile Voice Access using hairpinning) must be redundant so that there is sufficient capacity as well as multiple gateways and route group and route list constructs for call routing resiliency. The Mobile Voice Access feature can achieve resilient functionality only if multiple H.323 VoiceXML gateways and feature access DIDs are available.

Unified Mobility

The Cisco Unified Communications solution delivers mobility functionality via the Cisco Unified Mobility application server. The Unified Mobility server communicates with Unified CM via JTAPI and AXL, and it provides Mobile Connect and Mobile Voice Access functionality.

Guidelines and Restrictions for Unified Mobility

The following guidelines and restrictions apply with regard to deployment and operation of Mobile Connect within the Unified CM telephony environment:

- Ensure that the Shared Line CTI port user names and the Outgoing CTI port user names configured in Unified CM are configured with all capital letters.
- Mobile Connect does not work with Multilevel Precedence and Preemption (MLPP). If a call is preempted with MLPP, Mobile Connect features are disabled for that call.
- Mobile Connect services do not extend to video calls. A video call received at the desktop phone cannot be picked up on the cellular phone.
- The Unified CM Forced Authorization Code and Client Matter Code (FAC/CMC) feature does not work with Mobile Voice Access.

For additional guidelines and restrictions, refer to the latest version of the *Cisco Unified Mobility Administration Guide*, available at

<http://www.cisco.com>

Unified Mobility Performance and Capacity

The Unified Mobility application server (version 1.2 or higher) supports the following capacities:

- The Cisco MCS-7845, MCS-7835, and MCS-7825 servers support a maximum of 1500 users.
- The Cisco MCS-7815 server supports a maximum of 1000 users.
- Unified CM supports up to three Unified Mobility application servers per cluster.

The Unified Mobility application server interacts with the Unified CM CTIManager for CTI line monitoring and call control. CTI connections within Unified CM are required in order to interact with Unified Mobility. CTI connection quantities needed for Unified Mobility interaction are as follows:

- Each shared line CTI port generates a single connection to the Unified CM CTIManager.

A shared line is required for each Mobile Connect or Mobile Voice Access user. Therefore, for each Mobility application server, a maximum of 1500 shared line CTI ports would be required, which equates to 1500 CTI connections per server. This would mean a maximum of 4500 CTI connections per cluster, given a maximum of three Unified Mobility application servers (MCS-7825, 7835, or 7845) per cluster.

- Each outgoing CTI port generates a connection to the Unified CM CTIManager.

An outgoing CTI port is required for each Mobile Connect or Mobile Voice Access call. Cisco recommends one outgoing CTI port for every five shared-line CTI ports (a ratio of 1:5). Therefore, for each Mobility application server, a maximum of 300 outgoing CTI ports would be required, which equates to 300 CTI connections per server. This would mean a maximum of 900 CTI connections per cluster, given a maximum of three Unified Mobility application servers (MCS-7825, 7835, or 7845) per cluster. Note that, if the ratio of outgoing CTI ports to shared-line CTI ports is higher, then the server-wide and cluster-wide numbers would be higher also.

- Each Mobility CTI route point generates a connection to the Unified CM CTIManager.

A single CTI route point is required for each Unified Mobility application server. This would translate to a maximum of 3 CTI connections per cluster, given a maximum of three Unified Mobility application servers per cluster.

A total of 1801 CTI connections per Unified Mobility application server is possible, given an MCS-7825, 7835, or 7845 server. Each Unified Mobility application server should be configured to use a different primary and backup Unified CM CTIManager node in order to distribute the CTI load across multiple nodes in the cluster, if more than one server is deployed within a cluster.

Assuming a Mobility deployment at full capacity with MCS-7825, 7835, or 7845 servers (three Unified Mobility application servers with 1500 Mobile Connect or Mobile Voice Access users each), a total of 5403 CTI connections would be required within the Unified CM cluster. The number of required CTI connections for Unified Mobility within Unified CM must be considered with regard to the overall cluster limit for Unified CM CTI connections (10,000 CTI connections per cluster with the MCS-7845 platform or 3200 CTI connections per cluster with the MCS-7835 and MCS-7825 servers). If additional CTI connections are required for other applications, they can limit the capacity of the Unified Mobility application server.



APPENDIX **A**

Recommended Hardware and Software Combinations

Last revised on: September 27, 2007

For the most recent information on recommended hardware platforms, software releases, and firmware versions for Cisco Unified Communications System deployments based on Cisco Unified Communications Manager (Unified CM), frequently refer to the latest *Cisco Unified Communications System Release Notes for IP Telephony* available at the following location:

http://www.cisco.com/en/US/products/ps6884/tsd_products_integrated_systems_documentation09186a0080621410.html



Note

The platforms and software versions recommended in the *System Release Notes for IP Telephony* 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 Unified Communications deployments, contact your Cisco account representative.

For additional information on the Cisco Unified Communications System, refer to the documentation available at:

<http://www.cisco.com/go/unified-techinfo>



GLOSSARY

Last revised on: September 27, 2007

A

AA	Automated attendant
AAD	Alerts and Activities Display
AAR	Automated Alternate Routing
AC	Cisco Attendant Console
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
AES	Advanced Encryption Standards
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 Program Interface
ARJ	Admission Reject
ARP	Address Resolution Protocol

ARQ	Admission Request
ASA	Adaptive Security Appliance
ASP	Active server page
ASR	Automatic speech recognition
ATA	Cisco Analog Telephone Adapter
ATM	Asynchronous Transfer Mode
AXL	AVVID XML Layer

B

BAT	Cisco Bulk Administration Tool
BBWC	Battery-backed write cache
BGP	Border Gateway Protocol
BHCA	Busy hour call attempts
BHCC	Busy hour call completions
BLF	Busy lamp field
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
CAM	Content-addressable memory
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
CCA	Clear channel assessment
CCS	Common channel signaling
CDP	Cisco Discovery Protocol
CDR	Call detail record
CGI	Common Gateway Interface
CIR	Committed information rate
CKM	Cisco Centralized Key Management
CLEC	Competitive local exchange carrier
CLID	Calling line identifier
CM	Cisco Unified Communications Manager (Unified CM)
CMC	Client Matter Code
CME	Cisco Unified Communications Manager Express (Unified CME)
CMI	Cisco Messaging Interface
CMM	Cisco Communication Media Module
CNG	Comfort noise generation
CO	Central office
Co-located	Two or more devices in the same physical location, with no WAN or MAN connection between them
COM	Component Object Model
COR	Class of restriction
Co-resident	Two or more services or applications running on the same server
CoS	Class of service
CPA	Cisco Unity Personal Assistant
CPI	Cisco Product Identification tool
CPN	Calling party number
CRS	Cisco Customer Response Solution
cRTP	Compressed Real-Time Transport Protocol
CSTA	Computer-Supported Telecommunications Applications
CSUF	Cross-Stack UplinkFast

CTI Computer telephony integration

CUE Cisco Unity Express

D

DC Domain controller

DDNS Dynamic Domain Name Server

DFS Dynamic Frequency Selection

DHCP Dynamic Host Configuration Protocol

DID Direct inward dial

DIT Directory Information Tree

DMZ Demilitarized zone

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

DTIM Delivery Traffic Indicator Message

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
FQDN	Fully qualified domain name
FR	Frame Relay
FWSM	Firewall Services Module
FXO	Foreign Exchange Office
FXS	Foreign Exchange Station

G

GARP	Gratuitous Address Resolution 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
ICMP	Internet Control Message Protocol
ICS	IBM Cabling System
ICT	Intercluster trunk
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IIS	Microsoft Internet Information Server
IM	Instant messaging
IntServ	Integrated Services
IntServ/DiffServ	Integrated Services/Differentiated Services
IP	Internet Protocol
IPCC	Cisco IP Contact Center
IPIP GW	IP-to-IP gateway
IPPM	Cisco IP Phone Messenger
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

KPML Key Press Markup Language

L

LAN Local area network

LBR Low bit-rate

LCD Liquid crystal display

LCF Location Confirm

LDAP Lightweight Directory Access Protocol

LDAPS LDAP over SSL

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
MCM	Multimedia Conference Manager
MCS	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
MITM	Man-in-the-middle
MLA	Cisco Multi-Level Administration
MLP	Multilink Point-to-Point Protocol
MLPP	Multilevel Precedence and Preemption
MLTS	Multi-line telephone system
MOC	Microsoft Office Communicator
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	Named Service Event
NSF	Network Specific Facilities
NTE	Named Telephony Event
NTP	Network Time Protocol

O

OSPF	Open Shortest Path First
OU	Organizational unit

P

PAC	Protected Access Credential
PBX	Private branch exchange
PC	Personal computer
PCI	Peripheral Component Interconnect
PCM	Pulse code modulation
PD	Powered device
PHB	Per-hop behavior
PIN	Personal identification number
PINX	Private integrated services network exchange
PIX	Private Internet 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
PSK	Pre-Shared Key
PSTN	Public switched telephone network
PVC	Permanent virtual circuit

Q

QBE	Quick Buffer Encoding
QBSS	QoS Basic Service Set
QoS	Quality of Service
QSIG	Q signaling

R

RADIUS	Remote Authentication Dial-In User Service
RAS	Registration Admission Status
RCP	Remote Copy Protocol
RDNIS	Redirected Dialed Number Information Service
RF	Radio frequency
RFC	Request for Comments
RIP	Routing Information Protocol
RSNA	Reservationless Single Number Access
RSP	Route/Switch Processor
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 Description Protocol
SE	Cisco Systems Engineer
SF	Super Frame
SI	Standard Image
SIMPLE	SIP for Instant Messaging and Presence Leveraging Extensions
SIP	Session Initiation Protocol
SIW	Service Inter-Working
SLB	Server load balancing
SLDAP	Secure LDAP
SMDI	Simplified Message Desk Interface
SNMP	Simple Network Management Protocol
SOAP	Simple Object Access Protocol
SQL	Structured Query Language
SRND	Solution Reference Network Design
SRST	Survivable Remote Site Telephony
SRTP	Secure Real-Time Transport Protocol

SRV	Server
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
TKIP	Temporal Key Integrity Protocol
TLS	Transport layer security
ToD	Time of day
ToS	Type of service
TPC	Transmit Power Control
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
UN	Unsolicited SIP Notify
UNC	Universal Naming Convention
UPS	Uninterrupted power supply
URI	Uniform resource identifier
USB	Universal Serial Bus
UTIM	Cisco Unity Telephony 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
VoWLAN	Voice over Wireless LAN (WLAN)
VPN	Virtual private network
VWIC	Voice/WAN interface card

W

WAN	Wide area network
WEP	Wired Equivalent Privacy
WFQ	Weighted fair queuing
WINS	Windows Internet Naming Service
WLAN	Wireless local area network
WLSM	Cisco Wireless LAN Services Module
WMM	Wi-Fi Multimedia
WMM TSPEC	Wi-Fi Multimedia Traffic Specification
WPA	Wi-Fi Protected Access

X

XML	Extensible Markup Language



INDEX

Symbols

- ! in route patterns [10-18](#)
- <None> calling search space [21-7](#)
- @ in route patterns [10-17](#)

Numerics

- 1700 Series Routers [6-10, 6-12](#)
- 1A and 2A cabling [3-25](#)
- 2800 Series Routers [6-9, 6-12, 6-17, 6-21](#)
- 3500 Series Video Gateways [4-36](#)
- 3511 MCU [16-20](#)
- 3515 MCU [16-20](#)
- 3540 MCU [16-20](#)
- 3545 MCU [16-20](#)
- 3800 Series Routers [6-9, 6-12, 6-17, 6-21](#)
- 4ESS [4-20, 4-21](#)
- 508 conformance [2-27](#)
- 5ESS [4-20, 4-21](#)
- 7902G IP Phone [20-6](#)
- 7905_7912 dial rules [10-13, 10-74](#)
- 7905G IP Phone [20-6](#)
- 7906G IP Phone [20-6](#)
- 7911G IP Phone [20-7](#)
- 7912G IP Phone [20-7](#)
- 7914 Expansion Module [20-10](#)
- 7920 Wireless IP Phone [16-48, 20-13](#)
- 7921G Wireless IP Phone [20-13](#)
- 7935 IP Conference Station [20-19](#)
- 7936 IP Conference Station [20-19](#)
- 7940_7960_OTHER dial rules [10-13, 10-74](#)
- 7940G IP Phone [20-7](#)

- 7941G-GE IP Phone [20-8](#)
- 7941G IP Phone [20-8](#)
- 7960G IP Phone [20-8](#)
- 7961G-GE IP Phone [20-9](#)
- 7961G IP Phone [20-8](#)
- 7970G IP Phone [20-9](#)
- 7971G-GE IP Phone [20-10](#)
- 7985G IP Video Phone [20-23, 20-35](#)
- 802.1s [3-4](#)
- 802.1w [3-4, 3-6](#)
- 802.3af PoE [3-23](#)
- 9.@ route pattern [10-17, 10-18](#)
- 911 calls [10-60, 11-1](#)

A

- AAR
 - dial plan considerations [10-27](#)
 - for video calls [4-40, 16-8](#)
 - for Voice over PSTN [2-10, 2-12](#)
 - with Cisco Unity [13-7](#)
 - with hunt pilot [10-92](#)
 - with wireless IP phones [20-19](#)
- abbreviated dialing [10-4](#)
- abbreviations [GL-1](#)
- AC [1-7, 8-14, 22-28](#)
- ac application user name [22-30](#)
- access codes [10-7, 10-28](#)
- access control list (ACL) [19-24, 19-26, 20-35](#)
- Access Control Server (ACS) [3-69, 20-16](#)
- accessibility of IP Telephony features [2-27](#)
- Access Layer [3-4](#)
- access point (AP) [3-63, 3-67, 20-13](#)

- ACF [10-43](#)
- ACL [19-24, 19-26, 20-35](#)
- acronyms [GL-1](#)
- ACS [3-69, 20-16](#)
- Active Directory (AD) [3-69, 17-11, 17-15, 17-17, 17-22](#)
- AD [3-69, 17-11, 17-15, 17-17, 17-22](#)
- Adaptive Security Appliance (ASA) [19-29, 19-31](#)
- addresses
 - Admission Request (ARQ) [10-43](#)
 - MAC [19-13](#)
 - resolution [10-43, 10-44](#)
 - security [19-5](#)
- Address Resolution Protocol (ARP) [3-68, 19-20](#)
- Add Traffic Stream (ADDTS) [20-18](#)
- ADDTS [20-18](#)
- Admission Confirm (ACF) [10-43](#)
- Admission Reject (ARJ) [10-43](#)
- Admission Request (ARQ) [10-43](#)
- Advanced Encryption Standards (AES) [3-69](#)
- advanced formulas for bandwidth calculations [3-60](#)
- AES [3-69](#)
- AFT [11-19](#)
- agents for call processing [1-4, 2-17](#)
- Aironet [16-48](#)
- ALI [11-4, 11-19](#)
- ALI Formatting Tool (AFT) [11-19](#)
- all trunks busy [11-11](#)
- alternate
 - endpoints [5-14](#)
 - gatekeeper [5-14, 8-24](#)
 - TFTP file locations [3-22](#)
- analog
 - gateways [4-7, 4-18, 4-27, 20-2](#)
 - interface modules [20-2, 20-4](#)
- Analog Telephone Adaptor (ATA) [20-6, 20-26](#)
- ANI [4-17, 11-4, 11-6, 11-7](#)
- Annex M1 [5-14](#)
- annunciator [6-18](#)
- answer supervision [11-12](#)
- antivirus [19-40](#)
- AP [3-63, 3-67, 20-13](#)
- Application ID for RSVP [3-47, 3-56, 9-27, 16-4](#)
- applications
 - Attendant Console [22-28](#)
 - described [22-1](#)
 - Extension Mobility [22-7, 22-28, 22-39, 22-49](#)
 - for mobile users [23-1](#)
 - for video telephony [16-45](#)
 - general [1-6](#)
 - IP Manager Assistant [22-13](#)
 - IP Phone Services [22-2](#)
 - security [19-39](#)
 - sizing and scalability [8-13](#)
 - third-party [1-2](#)
 - Unified Communications Manager Assistant [22-13](#)
 - WebDialer [22-39](#)
- application users [17-8](#)
- architecture
 - for Attendant Console [22-31](#)
 - for Cisco Unified Communications Manager Assistant [22-16, 22-17](#)
 - for directories [17-8](#)
 - for Extension Mobility [22-9](#)
 - for IP Phone Services [22-3](#)
 - for IP Telephony [1-2](#)
 - for Mobile Connect [23-8](#)
 - for Mobile Voice Access [23-20](#)
 - for WebDialer [22-41, 22-44](#)
- area code [10-28](#)
- ARJ [10-43](#)
- ARP [3-68, 19-20](#)
- ARQ [10-43](#)
- ASA [19-29, 19-31](#)
- Assistant Console [22-23](#)
- asynchronous H.323 client [16-26, 16-31](#)
- Asynchronous Transfer Mode (ATM) [2-6, 2-15, 3-33](#)
- ATA [20-6, 20-26](#)
- ATM [2-6, 2-15, 3-33](#)

Attendant Console (AC) [1-7, 8-14, 16-46, 22-28](#)

attributes for directory data [17-13](#)

audio-only calls [16-8](#)

Audio Server [14-13](#)

audio sources [7-3, 7-9](#)

authentication

of phones [3-68, 19-11, 20-14](#)

of users [17-11, 17-19](#)

open [20-15](#)

shared key [20-15](#)

auto-detection [8-30](#)

AutoGenerated.txt directory file [22-35](#)

automated alternate routing (AAR)

dial plan considerations [10-27](#)

for video calls [4-40, 16-8](#)

for Voice over PSTN [2-10, 2-12](#)

with Cisco Unity [13-7](#)

with hunt pilot [10-92](#)

with wireless IP phones [20-19](#)

Automatic Location Identification (ALI) [11-4, 11-19](#)

Automatic Number Identification (ANI) [4-17, 11-4, 11-6, 11-7](#)

AUTO negotiate [3-24](#)

AVVID XML Layer (AXL) [23-1](#)

AXL [23-1](#)

B

BackboneFast [3-6](#)

bandwidth

advanced formulas [3-60](#)

best-effort [3-32](#)

call control traffic [3-58, 3-60, 3-63](#)

consumption [3-49, 3-52](#)

for Cisco Unified MeetingPlace [14-8](#)

for Cisco Unified MeetingPlace Express [15-7](#)

for Cisco Unity [13-6](#)

for RSVP [3-54, 3-61](#)

for screen sharing [15-7](#)

for shared line appearances [3-61](#)

for virtual tie lines [3-63](#)

for web applications [15-7](#)

for wireless networks [3-71](#)

general rule [2-19](#)

guaranteed [3-31](#)

management of [9-14](#)

provisioning [3-29, 3-31, 3-49](#)

request for [5-14](#)

requirements for call admission control [9-13](#)

requirements for gatekeepers [9-14](#)

voice class requirements [3-36](#)

basic IP phones [20-6](#)

B-Channel [4-42](#)

beacons [3-68](#)

Bearer Capabilities Information Element (bearer-caps) [4-45](#)

bearer-caps command [4-45](#)

bearer traffic [3-51, 3-54](#)

benefits of

distributed call processing [2-16](#)

single-site deployment [2-3](#)

best-effort bandwidth [3-32](#)

best practices for

call admission control [9-2](#)

centralized call processing [2-7](#)

Cisco Unified Communications Manager Express (Unified CME) [8-32](#)

distributed call processing [2-16](#)

fax support [4-25](#)

IP-to-IP gateway [8-34, 9-30](#)

LDAP synchronization [17-17](#)

line/device approach to building classes of service [10-84](#)

modem support [4-27](#)

music on hold [7-8](#)

RSVP [3-49](#)

single-site deployment [2-4](#)

WAN design [3-30](#)

BHCA [2-24, 4-2, 8-18, 8-19, 10-94](#)

BHCC [10-94](#)
 bill-to number (BTN) [11-5](#)
 binding of channels [4-42](#)
 BLF [21-5](#)
 border element [5-4](#)
 BPDU [3-6](#)
 branch office router [7-16](#)
 bridge protocol data unit (BPDU) [3-6](#)
 BTN [11-5](#)
 built-in conferencing [6-11](#)
 bump in the wire [19-34](#)
 bursting [3-38](#)
 bursty traffic [4-2](#)
 business IP phones [20-7](#)
 busy hour call attempts (BHCA) [2-24, 4-2, 8-18, 8-19, 10-94](#)
 busy hour call completions (BHCC) [10-94](#)
 busy lamp field (BLF) [21-5](#)
 busy-out channels [4-42](#)

C

C5421 chipset [6-6](#)
 C542 chipset [6-7](#)
 C549 chipset [6-6](#)
 C5510 chipset [6-4](#)
 cabling

- Category 3 [3-24](#)
- IBM type 1A and 2A [3-25](#)

 CAC (*see* call admission control)
 calculations for server capacities [8-16](#)
 call admission control

- bandwidth management [9-14](#)
- bandwidth requirements [9-13](#)
- best practices [9-2](#)
- centralized call processing [9-37, 9-41, 9-46, 9-51](#)
- components [9-11](#)
- described [9-1](#)
- design considerations [9-36](#)
- distributed call processing [9-38, 9-43, 9-48, 9-54](#)

elements [9-11](#)
 for Cisco IP Communicator [20-13](#)
 for Cisco Unified MeetingPlace [14-8](#)
 for Cisco Unified MeetingPlace Express [15-7](#)
 for Cisco Unified MeetingPlace Video [14-19](#)
 for music on hold [7-15](#)
 for wireless access points [20-18](#)
 gatekeeper [8-20, 9-14, 10-41](#)
 locations [16-8](#)
 migration from static locations to RSVP [9-25](#)
 moving devices to a new location [11-13](#)
 MPLS [9-11](#)
 regions [16-5](#)
 RSVP [3-48](#)
 RSVP-enabled locations [9-16](#)
 static locations [9-12](#)
 topologies [9-36](#)
 topology-aware [9-7](#)
 topology-unaware [9-3](#)
 callback

- for emergency services [11-8, 11-14](#)
- from the PSAP [11-8, 11-14](#)

 call control traffic [3-58, 3-63](#)
 call detail record (CDR) [2-20](#)
 call flows

- Cisco Unified MeetingPlace Video [14-27](#)
- media hold [7-27](#)
- music on hold [7-5, 7-20, 7-23](#)

 calling line ID (CLID) [4-17, 10-19](#)
 calling party number (CPN) [11-5](#)
 calling privileges [10-21, 10-50](#)
 calling restrictions [10-21, 10-50](#)
 calling search spaces [10-21, 10-23, 10-61, 10-79, 10-83, 21-7](#)
 CallManager (*see* Unified CM)
 call processing

- agents [1-4, 2-17](#)
- centralized [2-4, 9-37, 9-41, 9-46, 9-51, 13-13, 13-14](#)
- distributed [2-14, 9-38, 9-48, 9-54](#)
- distributed deployments [9-43](#)

- guidelines [8-1](#)
- hardware platforms [8-2](#)
- redundancy [4-7, 8-8](#)
- subscriber server [8-7](#)
- with gatekeeper [8-20](#)
- call-related traffic [3-63](#)
- call routing for emergency calls [11-18](#)
- calls
 - 911 [11-1](#)
 - audio-only [16-8](#)
 - classification of [10-20](#)
 - coverage of [10-91](#)
 - emergency [10-60](#)
 - flow between clusters [16-10](#)
 - flows for Cisco Unified MeetingPlace Video [14-27](#)
 - forwarding [10-25, 10-87](#)
 - H.323 [5-13](#)
 - hairpinning [13-14](#)
 - history [21-6](#)
 - hold [7-6](#)
 - inbound [4-37, 4-43, 10-60, 10-65, 10-72](#)
 - load balancing [5-9](#)
 - media hold [7-27](#)
 - music on hold [7-1](#)
 - number of calls per DSP resource [6-4, 6-6, 6-7](#)
 - outbound [4-39, 4-43, 5-9, 10-60, 10-64, 10-69](#)
 - per second (cps) [4-2](#)
 - pickup at desk phone [23-5, 23-19](#)
 - pickup at remote destination phone [23-6, 23-19](#)
 - preservation of [4-15](#)
 - privileges [10-21](#)
 - restrictions [10-50](#)
 - routing [4-37, 4-39, 10-15, 10-38, 10-41, 11-18, 23-11, 23-12](#)
 - scenarios [16-9](#)
 - signaling [4-44, 4-45](#)
 - simultaneous [4-2](#)
 - speed of [3-53](#)
 - tromboning [13-14](#)
 - types supported [16-3](#)
 - within a cluster [10-60, 10-63, 10-68](#)
- CAM [19-13](#)
- CAMA [11-5](#)
- campus
 - access switch [3-3](#)
 - infrastructure requirements [3-1](#)
- cancellation of echo [4-26](#)
- CanMapAlias [5-14](#)
- capability exchange for T.38 fax relay [4-33](#)
- capacity planning for
 - CTI route points and ports [8-20](#)
 - gateways [8-18](#)
 - music on hold [7-12, 7-13](#)
 - phones [8-17](#)
 - third-party controlled lines [8-20](#)
 - trunks [8-19](#)
 - Unified CM servers [8-15, 8-16, 8-17](#)
 - wireless networks [20-15](#)
- Capacity Tool [8-15, 8-16, 8-17](#)
- Category 3 cabling [3-24](#)
- CCA [3-68](#)
- CDP [19-12, 20-20](#)
- CDR [2-20](#)
- Centralized Automatic Message Accounting (CAMA) [11-5](#)
- centralized call processing
 - call admission control [9-37, 9-41, 9-46, 9-51](#)
 - call coverage [10-92](#)
 - centralized messaging [13-13](#)
 - deployment model [2-4](#)
 - distributed messaging [13-14](#)
 - hunt lists [10-92](#)
 - Voice over the PSTN [2-10](#)
- centralized gatekeeper deployment [10-45](#)
- centralized messaging [13-3, 13-13, 13-17, 13-21](#)
- centralized TFTP services [3-19, 3-20](#)
- channels
 - binding [4-42](#)
 - for video calls [4-42](#)

- for wireless devices [3-65](#)
- rollover [4-42](#)
- chipsets
 - C542 [6-7](#)
 - C5421 [6-6](#)
 - C549 [6-6](#)
 - C5510 [6-4](#)
- CIF [20-25](#)
- CIR [3-38](#)
- Cisco Aironet [16-48](#)
- Cisco Centralized Key Management (Cisco CKM) [20-15](#), [20-16](#)
- Cisco Discovery Protocol (CDP) [19-12](#), [20-20](#)
- Cisco Emergency Responder (ER) [11-9](#), [11-13](#), [16-45](#), [20-19](#)
- Cisco IOS
 - calling privileges [10-50](#)
 - call routing [10-38](#), [10-41](#)
 - classes of service [10-88](#)
 - digit manipulation [10-52](#)
 - DSP resources supported [6-4](#), [6-6](#), [6-7](#)
 - gatekeeper [16-22](#)
 - gateways [4-30](#), [4-32](#)
 - minimum release required [20-4](#)
 - software MTP [6-17](#)
- Cisco IP Communicator [20-31](#), [20-37](#)
- Cisco IP Conference Station [20-19](#), [20-26](#)
- Cisco IP Phone Messenger (IPPM) [21-15](#)
- Cisco IP SoftPhone [11-14](#), [16-47](#), [20-37](#)
- Cisco IP Voice Media Streaming Application [6-18](#)
- Cisco LEAP [3-69](#), [20-14](#), [20-15](#)
- Cisco Messaging Interface (CMI) [12-2](#)
- Cisco Multimedia Conference Manager (MCM) [5-14](#), [16-36](#)
- Cisco Security Agent [19-39](#)
- Cisco Unified Border Element [5-4](#)
- Cisco Unified Communications Manager (*see* Unified CM)
- Cisco Unified Communications Manager Assistant (Unified CM Assistant) [1-7](#), [8-14](#), [16-46](#)
- Cisco Unified Communications Manager Express (Unified CME) [2-17](#), [8-30](#)
- Cisco Unified Contact Center (Unified CC) [16-46](#)
- Cisco Unified IP IVR [16-21](#), [16-46](#)
- Cisco Unified MeetingPlace [14-1](#), [16-47](#)
- Cisco Unified MeetingPlace Express [15-1](#)
- Cisco Unified MeetingPlace Video [14-16](#)
- Cisco Unified Mobility (*see* Mobility)
- Cisco Unified Personal Communicator [20-10](#), [20-31](#), [21-19](#)
- Cisco Unified Presence [21-1](#), [21-8](#)
- Cisco Unified Video Advantage
 - classification of traffic [20-35](#)
 - described [16-1](#), [20-20](#)
 - QoS recommendations [20-31](#)
- Cisco Unified Wireless IP Phone 7920 [16-48](#)
- Cisco Unity [13-1](#)
- Cisco Unity Personal Assistant (CPCA) [13-3](#)
- Cisco Unity Telephony Integration Manager (UTIM) [13-10](#), [13-12](#)
- CKM [20-15](#), [20-16](#)
- classes of service for users [10-76](#), [10-80](#), [10-88](#)
- classification of
 - calls [10-20](#)
 - traffic [3-4](#), [3-27](#), [3-70](#), [14-8](#), [14-20](#), [15-8](#), [20-25](#), [20-34](#)
- class of restriction (COR) [10-50](#), [10-88](#)
- Class of Service (CoS) [3-4](#), [20-26](#)
- clear channel assessment (CCA) [3-68](#)
- CLEC [11-4](#)
- CLID [4-17](#), [10-19](#)
- Client Matter Code (CMC) [10-20](#)
- clients
 - H.323 [16-26](#)
 - zones [16-34](#)
- clipping [2-7](#)
- clocking source for fax/modem support [4-32](#)
- clustering over the WAN
 - Cisco Unity [13-17](#), [13-19](#)
 - described [2-17](#)
 - failover with Cisco Unity [13-20](#)

- local failover [2-22](#)
- MeetingPlace [14-6](#)
- MeetingPlace Express [15-5](#)
- music on hold [7-20](#)
- remote failover [2-26](#)
- troubleshooting [2-21](#)
- WAN considerations [2-18](#)
- clusters
 - co-located [9-52](#)
 - design guidelines [8-2](#)
 - Emergency Responder (ER) [11-18, 11-19](#)
 - for presence servers [21-9](#)
 - for Unified CM [8-2](#)
 - multiple, for Cisco Unity [13-5](#)
 - redundancy [8-9](#)
 - services [8-3](#)
- clusterwide parameters [9-23](#)
- CMC [10-20](#)
- CMI [12-2](#)
- CMM [7-3, 20-5](#)
- codecs
 - complexity modes [6-2](#)
 - flex mode [6-3](#)
 - for music on hold [7-8](#)
 - for video telephony [20-23](#)
 - low bit-rate (LBR) [6-24](#)
 - pass-through [9-22](#)
 - selection of [20-12](#)
 - supported by complexity mode [6-4](#)
 - supported by endpoint devices [16-6, 20-25](#)
 - types [7-3, 20-12](#)
- collaboration
 - capabilities [1-6](#)
 - solutions [16-47](#)
- co-located
 - DHCP server [3-13](#)
 - Unified CM clusters [9-52](#)
- COM [17-4](#)
- combined deployment models for messaging [13-16](#)
- Committed Information Rate (CIR) [3-38](#)
- Common Intermediate Format (CIF) [20-25](#)
- Communication Media Module (CMM) [7-3, 20-5](#)
- Communications Manager (*see* Unified CM)
- Communicator [20-10, 20-11, 20-31, 20-37, 21-19](#)
- competitive local exchange carrier (CLEC) [11-4](#)
- complexity modes for codecs [6-2](#)
- Component Object Model (COM) [17-4](#)
- components of
 - Cisco Unified MeetingPlace [14-2](#)
 - IP Video Telephony [16-1](#)
 - messaging system [13-4](#)
 - presence [21-2](#)
- compressed Real-Time Transport Protocol (cRTP) [3-33, 3-35](#)
- Computer Telephony Integration (CTI) [8-12, 16-2, 16-45](#)
- Conference Station [20-19, 20-26](#)
- conferencing
 - built-in resource [6-11](#)
 - capabilities [1-6](#)
 - described [6-8](#)
 - hardware resources [6-9, 6-10](#)
 - MeetingPlace [1-6](#)
 - resources [6-8, 16-13, 16-21](#)
 - rich media [1-1](#)
 - software resources [6-9](#)
 - video [14-29](#)
- configuration examples for
 - access control list (ACL) [19-25, 19-27](#)
 - ATA 188 and IP phones [20-26](#)
 - DHCP snooping [19-18](#)
 - Dynamic ARP Inspection [19-21](#)
 - endpoint gatekeeper [16-41](#)
 - fax/modem support [4-30, 4-32](#)
 - firewalls [19-35, 19-37](#)
 - gatekeeper [8-20](#)
 - IP Source Guard [19-23](#)
 - IP-to-IP gateways [9-32](#)
 - lobby phone security [19-41](#)

QoS [20-25](#)
 software-based endpoints [20-31](#)
 switch port security [19-15](#)
 Unified CME [8-30](#)
 VG224 gateways [20-25](#)
 VG248 gateways [20-25](#)
 via-zone gatekeepers [9-32](#)
 Wireless IP Phones [20-33](#)
 zones [16-34](#)
 conformance with Section 508 [2-27](#)
 connectivity options for the WAN [2-6, 2-15](#)
 console
 for attendants [16-46, 22-28](#)
 for Unified CM Assistant assistant [22-23](#)
 Contact Center [1-1, 16-46](#)
 contact center traffic patterns [4-3, 4-4](#)
 content-addressable memory (CAM) [19-13](#)
 continuous meeting server [14-13](#)
 continuous-presence conference view [16-15](#)
 control signaling [3-58, 3-63](#)
 COR [10-50, 10-88](#)
 Core Layer [3-9](#)
 co-resident
 DHCP [3-14](#)
 MoH [7-3](#)
 core switch [3-3](#)
 CorporateDirectory.txt directory file [22-35](#)
 CoS [3-4, 20-26](#)
 coverage of calls [10-91](#)
 CPCA [13-3](#)
 CPN [11-5](#)
 cps [4-2](#)
 CPU utilization in gateways [4-5](#)
 cRTP [3-33, 3-35](#)
 CTI [8-12, 16-2, 16-45, 23-11, 23-12](#)
 CTI Manager [8-3, 8-12](#)
 CTI route points [6-17, 8-20, 23-11, 23-12](#)
 customer contact [1-1](#)
 cutover [18-1, 18-2](#)

D

DAI [19-19, 19-20](#)
 database replication [8-4](#)
 database synchronization for Unified CM [17-24](#)
 data center [3-10, 19-39](#)
 delay
 of packets [2-18, 2-20, 4-25, 4-27](#)
 variation (jitter) [4-25, 4-27](#)
 Delayed Offer [5-16](#)
 Delivery Traffic Indicator Message (DTIM) [3-66](#)
 Demilitarized Zone (DMZ) [14-7, 15-5, 19-44](#)
 deployment models
 clustering over the WAN [2-17, 7-20, 14-6, 15-5](#)
 combined for messaging [13-16](#)
 described [2-1](#)
 DHCP [3-13](#)
 for Cisco Unity [13-2](#)
 for presence servers [21-10](#)
 for Unified CME [8-31](#)
 for Unified MeetingPlace [14-3](#)
 for Unified MeetingPlace Express [15-1](#)
 for Unified MeetingPlace Video [14-18](#)
 multisite dial plan [10-55](#)
 multisite WAN with centralized call processing [2-4, 6-24, 7-14, 10-92, 14-4, 15-3](#)
 multisite WAN with distributed call processing [2-14, 6-25, 7-19, 10-56, 10-93, 14-5, 15-4](#)
 music on hold [7-14](#)
 single site [2-2, 6-24, 7-14, 14-3, 15-2](#)
 voice over the PSTN [2-10](#)
 desk phone pickup [23-5, 23-19](#)
 desktop application for Attendant Console [22-33](#)
 desktop phones [20-6](#)
 destination of a call [10-27](#)
 devices
 hunt list [10-94](#)
 limits per server [8-16](#)
 line group [10-37](#)

- mobility [11-13, 20-18](#)
- pools [2-23, 2-26](#)
- route group [10-21](#)
- DFS [3-65](#)
- DHCP
 - binding information [19-19](#)
 - deployment options [3-13](#)
 - described [3-11](#)
 - lease times [3-12](#)
 - Option 150 [3-12](#)
 - servers [3-14](#)
 - Snooping [19-16, 19-19](#)
 - starvation attack [19-18](#)
- dialed pattern recognition [10-3, 10-74](#)
- dial-in conferences [16-21](#)
- dial peers [10-38, 10-50, 10-52](#)
- dial plan
 - 911 calls [11-1](#)
 - abbreviated dialing [10-4](#)
 - access codes [10-7](#)
 - approaches to [10-57](#)
 - calling privileges [10-21, 10-50](#)
 - calling search space [10-79, 10-83](#)
 - call routing [10-15](#)
 - classes of service [10-76, 10-80, 10-88](#)
 - design considerations [10-54](#)
 - dial peers [10-38, 10-50, 10-52](#)
 - distribution of digits [10-5](#)
 - elements [10-8](#)
 - emergency call string [11-10](#)
 - Extension Mobility [10-30, 10-79, 10-85](#)
 - for Cisco Unified MeetingPlace [14-12](#)
 - for Cisco Unified MeetingPlace Video [14-29](#)
 - for distributed call processing [10-56](#)
 - for Mobile Connect [23-9](#)
 - for Mobile Voice Access [23-20](#)
 - for multisite deployments [10-55](#)
 - for Unified CM Assistant [22-19](#)
 - for Voice over PSTN [2-13](#)
 - functions [10-1](#)
 - hunt lists [10-33, 10-36](#)
 - international calls [10-18](#)
 - line groups [10-33, 10-36](#)
 - number of digits [10-4](#)
 - on-net vs. off-net [10-3](#)
 - overlapping extensions [10-4, 10-61](#)
 - partitions [10-79, 10-83](#)
 - planning considerations [10-2, 10-8](#)
 - shared line appearance [11-14](#)
 - site codes [10-6](#)
 - string length [10-4](#)
 - uniform on-net dialing [10-5, 10-58](#)
 - variable length on-net dialing [10-6, 10-61, 10-66](#)
 - voicemail [10-60, 10-65, 10-72](#)
- dial rules [10-9, 10-11, 10-13, 10-74](#)
- DID [4-17, 11-5](#)
- differential threshold [20-16](#)
- differentiated services code point (DSCP) [3-4, 3-34](#)
- different versions of Unified CM in the same cluster [3-20](#)
- digital gateways [4-7, 4-19, 4-27](#)
- Digital PBX Adapter (DPA) [12-4, 12-6](#)
- digital set emulation (DSE) [12-4](#)
- digital signal processor (*see* DSP resources)
- digit manipulation [4-38, 10-19, 10-26, 10-52](#)
- Direct Inward Dial (DID) [4-17, 11-5](#)
- directories
 - access [17-4](#)
 - architecture [17-8](#)
 - attributes [17-13](#)
 - authentication of users [17-11, 17-19, 17-24](#)
 - filters for importing user data [17-24](#)
 - for Attendant Console [22-34](#)
 - for Unified CM Assistant [22-23](#)
 - integration with Cisco Unified MeetingPlace [14-9](#)
 - integration with Cisco Unified MeetingPlace Express [15-9](#)
 - integration with IP telephony system [17-1, 17-2](#)
 - integration with Unified CM 4.x [17-7](#)

- integration with Unified CM 5.0 [17-6](#)
- LDAP [17-1](#)
- schema [17-1](#)
- search base [17-13](#)
- security [17-16](#)
- sn attribute [17-11](#)
- synchronization [17-11, 17-24](#)
- UserID [17-11](#)
- directory gatekeeper [8-27, 10-48](#)
- directory number (DN) [10-94](#)
- disaster recovery [14-12, 14-13](#)
- distributed call processing [2-14, 9-38, 9-43, 9-48, 9-54, 10-93](#)
- distributed gatekeeper deployment [10-47](#)
- distributed messaging [13-3, 13-14, 13-19](#)
- Distribution Layer [3-7](#)
- distribution of digits in a dial plan [10-5](#)
- DMZ [14-7, 15-5, 19-44](#)
- DN [10-94](#)
- DNS [3-10](#)
- Domain Name System (DNS) [3-10](#)
- Domino Unified Communications Services (DUC) [13-3](#)
- DPA [12-4, 12-6](#)
- DS0 [8-19](#)
- DSCP [3-4, 3-34](#)
- DSE [12-4](#)
- DSP resources
 - C5421 chipset [6-6](#)
 - C542 chipset [6-7](#)
 - C549 chipset [6-6](#)
 - C5510 chipset [6-4](#)
 - calculations [6-22](#)
 - described [6-2](#)
 - for voice termination [6-4](#)
 - in multisite deployment model [2-5, 2-15](#)
 - in single-site deployment model [2-2](#)
 - number of calls [6-4, 6-6, 6-7](#)
 - PVDM [6-20](#)
- DTIM [3-66](#)
- DTMF [4-7, 4-11, 5-14, 5-15, 6-15, 6-16](#)

- DTPC [3-68](#)
- dual conference servers [14-13](#)
- dual-mode configuration [3-20](#)
- dual PBX integration [12-5, 12-6](#)
- dual tone multifrequency (DTMF) [4-7, 4-11, 5-14, 5-15, 6-15, 6-16](#)
- DUC [13-3](#)
- dynamic ANI interface [11-7](#)
- Dynamic ARP Inspection (DAI) [19-19, 19-20](#)
- Dynamic Frequency Selection (DFS) [3-65](#)
- Dynamic Host Configuration Protocol (DHCP) [3-11, 19-16, 19-18, 19-19](#)
- Dynamic Transmit Power Control (DTPC) [3-68](#)

E

- E.164 address [10-69, 10-70, 11-4, 11-5, 11-7](#)
- E911 [11-1, 11-3](#)
- EAP [3-68, 3-69](#)
- EAP-FAST [3-68, 20-14](#)
- Early Media [5-16](#)
- Early Offer [5-16](#)
- echo cancellation [4-26](#)
- ECM [4-26](#)
- ECS [16-2](#)
- efficiency of links [3-35](#)
- elements of a dial plan [10-8](#)
- ELIN [11-6, 11-7](#)
- EM (*see* Extension Mobility)
- emergency calls [10-60](#)
- emergency call string [11-10](#)
- emergency location identification number (ELIN) [11-6, 11-7](#)
- Emergency Responder (ER) [10-60, 11-9, 11-13, 16-45](#)
- emergency response location (ERL) [11-6, 11-7, 11-13](#)
- emergency services [11-1](#)
- EMP [16-13](#)
- Empty Capabilities Set (ECS) [16-2](#)
- encryption

- for phones [19-11](#)
 - for signaling [3-59, 3-61](#)
 - endpoints
 - alternate [5-14](#)
 - analog gateways [20-2](#)
 - codecs supported [16-6](#)
 - defined [1-5](#)
 - directory access [17-4](#)
 - features [20-37](#)
 - gatekeeper [16-23, 16-26](#)
 - gatekeeper output [8-27](#)
 - gatekeeper registration [8-27](#)
 - H.323 [20-36](#)
 - H.323 clients [16-26](#)
 - hiding IP addresses [6-26](#)
 - line group devices [10-37](#)
 - SCCP [20-20](#)
 - SIP [20-36](#)
 - software-based [1-5, 20-10, 20-31](#)
 - Sony [20-24](#)
 - supplementary services [6-13](#)
 - Tandberg [20-24, 20-35](#)
 - time to live [16-41](#)
 - types of [20-1](#)
 - video [1-5, 14-21, 16-1, 20-20, 20-34](#)
 - wireless [1-6, 20-13](#)
 - end users [17-8, 21-3](#)
 - Enhanced Media Processor (EMP) [16-13](#)
 - Enhanced Media Termination Points [16-4](#)
 - enhancements in Unified CM for IP Video Telephony [16-2](#)
 - Enterprise Dial Tone (*see* Mobile Voice Access)
 - Enterprise MCM [8-20](#)
 - equations for calculating
 - bandwidth [3-59, 3-60](#)
 - calling search spaces [10-79, 10-83](#)
 - music on hold server capacity [7-12](#)
 - partitions [10-79, 10-83](#)
 - ER [10-60, 11-13, 16-45, 20-19](#)
 - ERL [11-6, 11-7, 11-13](#)
 - Error Correction Mode (ECM) [4-26](#)
 - error rate [2-21](#)
 - ettercap virus [19-20](#)
 - example configurations [16-34, 16-41](#)
 - executive IP phones [20-9](#)
 - Expansion Module 7914 [20-10](#)
 - Extensible Authentication Protocol (EAP) [3-68, 3-69, 20-14](#)
 - Extensible Authentication Protocol-Flexible Authentication via Secure Tunneling (EAP-FAST) [3-68, 20-14](#)
 - Extension Mobility (EM)
 - described [1-7, 22-7](#)
 - dial plan [10-30, 10-79, 10-85](#)
 - interactions with Attendant Console [22-39](#)
 - interactions with Unified CM Assistant [22-28](#)
 - interactions with WebDialer [22-49](#)
 - profiles [8-19](#)
 - server capacity and performance [8-14, 8-19](#)
 - extensions, overlapping [10-4](#)
-
- ## F
- FAC [10-20](#)
 - failover
 - between subscriber servers [2-20](#)
 - Cisco Unity [13-4, 13-20](#)
 - clustering over the WAN [2-22, 2-26](#)
 - to PSTN [10-69, 10-70](#)
 - Fast Start [6-13](#)
 - FastStart [5-11](#)
 - fax
 - clocking source [4-32](#)
 - Error Correction Mode [4-26](#)
 - features supported [4-30](#)
 - gateway support for [4-7, 4-25](#)
 - interface modules [20-2, 20-3](#)
 - interoperability of features [4-29](#)
 - pass-through mode [4-25](#)

- protocols supported [4-28](#)
 - relay mode [4-25](#)
 - Super-Group 3 [4-27](#)
 - supported platforms and features [4-27](#)
 - T.38 [4-33](#)
 - features of endpoints [20-37](#)
 - filters for directory synchronization and authentication [17-24](#)
 - firewalls
 - around gateways [19-29](#)
 - bump in the road [19-34](#)
 - centralized deployment [19-44](#)
 - configuration example [19-35, 19-37](#)
 - described [19-31](#)
 - routed mode [19-34, 19-37](#)
 - stealth mode [19-34](#)
 - transparent mode [19-34, 19-37](#)
 - Firewall Services Module (FWSM) [19-29, 19-31, 19-37](#)
 - flash used for music on hold [7-16](#)
 - flat addressing [10-57, 10-66](#)
 - flex mode for codecs [6-3](#)
 - flows for
 - calls between clusters [16-10](#)
 - video calls with Cisco Unified MeetingPlace [14-27](#)
 - Forced Account Codes (FAC) [10-20](#)
 - Foreign Exchange Office (FXO) [11-6](#)
 - Foreign Exchange Station (FXS) [12-3](#)
 - forwarding calls [10-25, 10-87](#)
 - Frame Relay [2-6, 2-15, 3-33](#)
 - French national numbering plan [10-83](#)
 - full-duplex [3-24](#)
 - FWSM [19-29, 19-31, 19-37](#)
 - FXO [11-6](#)
 - FXS [12-3](#)
-
- G**
- GARP [19-7, 19-20](#)
 - gatekeeper
 - alternate [5-14, 8-24](#)
 - call admission control [2-16, 9-14](#)
 - call routing [10-41](#)
 - centralized deployment [10-45](#)
 - clustering [8-24](#)
 - configuration examples [8-20](#)
 - described [16-22](#)
 - design considerations [8-20](#)
 - directory [8-27, 10-48](#)
 - distributed deployment [10-47](#)
 - for Cisco Unified MeetingPlace [14-11](#)
 - for Cisco Unified MeetingPlace Express [15-11](#)
 - for Cisco Unified MeetingPlace Video [14-20](#)
 - for endpoints [8-27, 16-23, 16-26](#)
 - geographical resiliency [16-24](#)
 - H.225 trunks [5-6, 5-13](#)
 - incompatibilities [16-24](#)
 - intercluster trunks [5-5](#)
 - IOS [16-22](#)
 - legacy [9-33](#)
 - output example [8-27](#)
 - proxy [16-36, 16-38, 16-39](#)
 - redundancy [8-21, 8-27](#)
 - roles [16-23](#)
 - scalability [16-24](#)
 - summary [16-41](#)
 - supported platforms [16-25](#)
 - trunk redundancy [5-6](#)
 - via-zone [9-29, 9-34, 10-44, 14-21](#)
 - zones [9-14, 16-34](#)
 - gatekeeper-controlled
 - H.225 trunks [5-6, 5-13](#)
 - H.323 client [16-26, 16-30](#)
 - intercluster trunks [5-5](#)
 - Gatekeeper Transaction Message Protocol (GKTMP) [5-14](#)
 - Gatekeeper Update Protocol (GUP) [5-6, 8-24](#)
 - gateways
 - 911 services [11-11](#)

- all trunks busy [11-11](#)
 - analog [4-7, 4-18, 4-27, 20-2](#)
 - automated alternative routing [4-40](#)
 - blocking [11-11](#)
 - capabilities [4-45](#)
 - capacity calculations [8-18](#)
 - centralized [23-11](#)
 - Cisco IOS [4-30, 4-32](#)
 - Cisco Unified Videoconferencing 3500 Series Video Gateways [4-36](#)
 - configuration examples for fax/modem support [4-30](#)
 - configuration in Unified CM [4-44](#)
 - controlled with Named Service Event (NSE) [4-33](#)
 - core feature requirements [4-7](#)
 - CPU utilization [4-5](#)
 - digital [4-7, 4-19, 4-27](#)
 - digit manipulation [4-38](#)
 - distributed [23-12](#)
 - DS0 [8-19](#)
 - fax support [4-25](#)
 - features [20-37](#)
 - firewalls [19-29](#)
 - for contact centers [4-4](#)
 - for local failover [2-25](#)
 - for Mobile Connect [23-11, 23-12](#)
 - for music on hold [7-3](#)
 - for video telephony [4-36](#)
 - H.320 [16-33, 16-38](#)
 - H.323/SIP [14-14](#)
 - Intel NetStructure PBX-IP Media Gateway (PIMG) [13-22](#)
 - IP [14-14](#)
 - IP-to-IP [9-28, 9-34, 9-58, 10-44](#)
 - modem support [4-26](#)
 - performance overload [4-5](#)
 - performance tuning [4-5](#)
 - PIMG [13-22](#)
 - placement [11-11](#)
 - protocols [4-8](#)
 - QoS configuration examples [20-25](#)
 - QSIG support [4-24](#)
 - redundancy [4-15](#)
 - security [19-28](#)
 - selection of [4-7](#)
 - service prefixes [4-39](#)
 - SIP [4-12, 4-16](#)
 - site-specific requirements [4-16](#)
 - sizing for contact center traffic [4-4](#)
 - sizing for traffic [4-2](#)
 - V.34 modem support [4-27](#)
 - V.90 modem support [4-27](#)
 - VG224 [4-18, 12-2, 20-5](#)
 - VG248 [4-31, 12-3, 20-5](#)
 - voice applications [4-1, 20-2, 20-5](#)
 - VoiceXML [23-16, 23-17](#)
 - WS-X6624 [12-2, 12-3](#)
 - zone prefixes [16-40](#)
 - general security [19-2](#)
 - generic topologies [9-50](#)
 - geographical resiliency [16-24](#)
 - GKTMP [5-14](#)
 - glossary [GL-1](#)
 - Gratuitous Address Resolution Protocol (GARP) [19-7, 19-20](#)
 - groups for
 - call routing [10-21](#)
 - Emergency Responder (ER) [11-15, 11-16](#)
 - line numbers (hunting) [10-33](#)
 - media resources [6-1](#)
 - ports [13-5](#)
 - redundancy [14-24](#)
 - Unified CM redundancy [5-4, 8-8](#)
 - guaranteed bandwidth [3-31](#)
 - GUP [5-6, 8-24](#)
-
- ## H
- H.225 trunks [5-6, 5-13](#)

- H.245 [4-33](#)
 - H.320 [16-33, 16-38](#)
 - H.323
 - analog gateways [4-18](#)
 - Annex M1 [5-14](#)
 - call hairpinning [8-30](#)
 - call preservation enhancements [4-15](#)
 - calls [5-13](#)
 - classes of service [10-88](#)
 - clients [16-26, 16-36](#)
 - dial peers for call routing [10-38](#)
 - digital gateways [4-19, 4-20, 4-21, 4-22](#)
 - Fast Start [6-13](#)
 - FastStart [5-11](#)
 - fax and modem support [4-28](#)
 - gateways [4-8, 14-10, 15-9](#)
 - in single-site deployment model [2-4](#)
 - in Unified CM [5-12](#)
 - MCU resources [16-18](#)
 - SIP IP Gateway [14-14](#)
 - supplementary services [6-13](#)
 - T.38 fax relay [4-35](#)
 - trunks [5-1, 5-3, 5-4, 5-11](#)
 - video endpoints [16-2, 20-36](#)
 - zones prefixes [16-36](#)
 - hairpinning [8-30, 13-14, 23-17](#)
 - half-duplex [3-24](#)
 - hardware
 - analog interface modules [20-4](#)
 - audio conferencing bridge [6-9, 6-10](#)
 - DSP resources [6-4, 6-6, 6-7](#)
 - gatekeepers [8-21](#)
 - media resource capacities [6-20](#)
 - MTP resources [6-17](#)
 - music on hold [7-12](#)
 - recommendations [2-2, A-1](#)
 - transcoder [6-12, 6-13](#)
 - types of platforms [8-2](#)
 - headers for voice packets [3-51](#)
 - hiding IP addresses of endpoints [6-26](#)
 - high availability of
 - network services [3-4](#)
 - voice services [2-8](#)
 - high-availability servers [8-2](#)
 - high-density analog interface modules [20-3](#)
 - high-performance servers [8-2](#)
 - history of
 - calls [21-6](#)
 - hold [7-1, 7-6](#)
 - holdee [7-5](#)
 - holder [7-5](#)
 - hold time [4-2](#)
 - Hot Standby Router Protocol (HSRP) [2-16, 3-7, 8-20, 8-21, 14-31](#)
 - HSRP [2-16, 3-7, 8-20, 8-21, 14-31](#)
 - hub-and-spoke topology [3-3, 3-30, 9-14, 9-36, 10-41](#)
 - hunt
 - groups [10-33](#)
 - lists [10-33, 10-36, 10-94](#)
 - pilot [10-33, 10-35, 10-94](#)
-
- IBM Cabling System (ICS) [3-25](#)
 - IBM Sametime 7.5 [21-23](#)
 - IButton [10-14](#)
 - ICCS [2-19, 2-24, 8-4](#)
 - ICMP [4-16](#)
 - ICS [3-25](#)
 - iDivert [10-32](#)
 - IDS [2-19, 19-29](#)
 - Immediate Divert (iDivert) [10-32](#)
 - impairments without QoS [3-29](#)
 - inbound calls [4-37, 4-43, 10-60, 10-65, 10-72](#)
 - incompatibilities [16-24](#)
 - Informix Dynamic Server (IDS) [2-19](#)
 - infrastructure (*see* network infrastructure)
 - infrastructure gatekeeper [16-23](#)

- inline power [3-23](#)
 - Integrated Services (IntServ) model [3-44, 3-49](#)
 - Integrated Services/Differentiated Services (IntServ/DiffServ) model [3-46, 3-49](#)
 - integrating MeetingPlace with IP Telephony [14-1, 15-1](#)
 - integrations with Cisco Unity [13-5](#)
 - Intel NetStructure PBX-IP Media Gateway (PIMG) [13-22](#)
 - interactive voice response (IVR) [2-4, 16-21, 16-46](#)
 - intercluster trunks
 - gatekeeper controlled [5-5](#)
 - non-gatekeeper controlled [5-5](#)
 - using SIP [5-18](#)
 - interface modules [20-2](#)
 - interface types for 911 calls [11-4](#)
 - interference to wireless communications [3-66](#)
 - international calls [10-18](#)
 - Internet Control Message Protocol (ICMP) [4-16](#)
 - interoperability of fax and modem features [4-29](#)
 - Intra-Cluster Communication Signaling (ICCS) [2-19, 2-24, 8-4](#)
 - Intrusion Detection System (IDS) [19-29](#)
 - IntServ/DiffServ model [3-46, 3-49](#)
 - IntServ model [3-44, 3-49](#)
 - invia [9-29, 10-44, 16-34](#)
 - IOS
 - calling privileges [10-50](#)
 - call routing [10-38, 10-41](#)
 - classes of service [10-88](#)
 - digit manipulation [10-52](#)
 - DSP resources supported [6-4, 6-6, 6-7](#)
 - Gatekeeper [16-22](#)
 - minimum release required [20-4](#)
 - software MTP [6-17](#)
 - IP/H323 feature set [8-20](#)
 - IP/VC 3500 Series Video Gateways [4-36](#)
 - IP addresses
 - hiding [6-26](#)
 - security [19-5](#)
 - IP Communicator [1-5, 20-11, 20-31, 20-37](#)
 - IP Conference Station [20-19, 20-26](#)
 - IP Gateway [14-14](#)
 - IP-IP gateway (IPIPGW) [9-28, 9-34, 9-58, 10-44](#)
 - IPIPGW [9-28, 9-34, 9-58, 10-44](#)
 - IP IVR [16-46](#)
 - iPlanet Directory Server [17-11, 17-16](#)
 - IPMA [22-13](#)
 - IP Manager Assistant (IPMA) [22-13](#)
 - IP Phone Messenger (IPPM) [21-15](#)
 - IP phones [20-6](#)
 - IP Phone Services [1-6, 22-2](#)
 - IPPM [21-15](#)
 - IP Precedence [3-4, 3-34](#)
 - IP PSTN [6-26](#)
 - IPSec [2-6, 2-15](#)
 - IP Security Protocol (IPSec) [2-6, 2-15](#)
 - IPSG [19-22](#)
 - IP SoftPhone [16-47](#)
 - IP Source Guard (IPSG) [19-22](#)
 - IP Telephony [1-1, 1-2, 14-1](#)
 - IP-to-IP gateway (IPIPGW) [9-28, 9-34, 9-58, 10-44](#)
 - IP Video Telephony
 - components [16-1](#)
 - described [1-1, 1-6, 16-1](#)
 - enhancements in Unified CM [16-2](#)
 - MTP [16-4](#)
 - security [19-9](#)
 - transcoding [16-4](#)
 - IP VOICE feature set [8-30](#)
 - IP Voice Media Streaming Application [6-9, 6-17, 6-18, 6-19, 8-13](#)
 - ISDN [2-8, 2-10, 4-42](#)
 - IVR [2-4, 16-21, 16-46](#)
-
- ## J
- jitter [2-18, 4-25, 4-27](#)
 - JTAPI [8-12, 16-2](#)

K

Key Press Markup Language (KPML) [10-3, 10-9, 10-11](#)
 KPML [10-3, 10-9, 10-11](#)

L

LAN infrastructure [3-4](#)
 Layer 2 [2-16, 3-4](#)
 Layer 3 [3-4](#)
 layers of security [19-3](#)
 LBR [6-24](#)
 LCF [8-27, 10-44](#)
 LCR [4-41](#)
 LDAP [8-4, 17-1, 17-24](#)
 LDN [11-5](#)
 LEAP [3-68, 3-69, 20-14, 20-15](#)
 leased lines [2-6, 2-15, 3-33](#)
 lease times for DHCP [3-12](#)
 least-cost routing (LCR) [4-41](#)
 LEC [11-2, 11-11](#)
 legacy gatekeeper [9-33](#)
 LFI [3-33, 3-35, 3-36](#)
 licenses [21-11](#)
 Lightweight Directory Access Protocol (LDAP) [8-4, 17-1](#)
 Limit Client Power setting on access points [3-68](#)
 line/device approach to classes of service [10-80](#)
 line appearances [3-61, 8-18](#)
 line group devices [10-37](#)
 line groups [10-33, 10-36, 10-94](#)
 line speed mismatch [3-38](#)
 link efficiency [3-35](#)
 link fragmentation and interleaving (LFI) [3-33, 3-35, 3-36](#)
 listed directory number (LDN) [11-5](#)
 LLQ [3-33, 3-34](#)
 LMHOSTS file [3-10](#)
 load balancing [3-19, 5-6, 5-9, 8-11, 14-12](#)
 lobby phone security [19-41](#)
 local dialing area [10-29](#)

local exchange carrier (LEC) [11-2, 11-11](#)

local failover deployment model [2-22](#)

Location Confirm (LCF) [8-27, 10-44](#)

Location Reject (LRJ) [10-44](#)

Location Request (LRQ) [8-27, 10-44](#)

locations

 RSVP-enabled [9-16](#)

 static [9-12, 16-8](#)

 topology-aware [16-4](#)

loose gateway [4-33](#)

loss of packets [4-25, 4-27](#)

low bit-rate (LBR) codecs [6-24](#)

low-density analog interface modules [20-2](#)

low-latency queuing (LLQ) [3-33, 3-34](#)

LRJ [10-44](#)

LRQ [8-27, 10-44](#)

LRQ blast [8-27](#)

M

MAC address [19-13](#)

Manager Assistant [16-46](#)

manager IP phones [20-8](#)

manipulation of digits [10-26, 10-52](#)

marking traffic [14-8, 14-20, 15-8](#)

masking IP addresses of endpoints [6-26](#)

maximum sessions per RSVP Agent [9-21](#)

MC [16-13](#)

MCM [5-14, 8-20, 16-22, 16-36](#)

MCU

 capacity and sizing [16-20](#)

 configuration [16-31](#)

 for video telephony [16-1, 16-13](#)

 with H.323 or SIP [16-18](#)

 with Skinny Client Control Protocol (SCCP) [16-16](#)

 zone prefixes [16-38](#)

 zones [16-36](#)

media gateway [13-22](#)

Media Gateway Control Protocol (MGCP) [2-4, 4-8, 4-18, 4-23, 4-28, 16-2](#)

media hold [7-27](#)

media resource group (MRG) [6-22, 9-18, 16-17](#)

media resource group list (MRGL) [6-22, 9-18, 16-17](#)

media resources

described [6-1](#)

design guidelines [6-22](#)

for local failover [2-25](#)

hardware and software capacities [6-20](#)

PVDM [6-20](#)

security [19-28](#)

Media Streaming Application [6-9, 6-17, 6-18, 6-19, 8-13](#)

media termination point (MTP)

described [6-13](#)

for PSTN calls [6-26](#)

for video calls [16-4](#)

hiding IP addresses of endpoints [6-26](#)

in multisite deployment model [2-5, 2-15](#)

in single-site deployment model [2-2](#)

Named Telephony Events [6-14](#)

requirements for trunks [8-20](#)

with H.323 trunk [5-11](#)

with SIP trunk [5-14, 5-17](#)

meeting ID [14-29](#)

MeetingPlace

Audio Server [14-13](#)

components [14-2](#)

described [1-6](#)

dial plan [14-12](#)

direct integration with H.323 or SIP [14-10](#)

directory integration [14-9](#)

Express [15-1](#)

gatekeeper integration [14-11](#)

H.323/SIP IP Gateway [14-14](#)

integration with IP Telephony [14-1](#)

load balancing [14-12](#)

meeting ID [14-29](#)

RASAggregator trunk [14-24](#)

redundancy [14-12](#)

Video Administration component [14-17, 14-22, 14-23](#)

video call flows [14-27](#)

video conferences [16-47](#)

Video Integration component [14-16](#)

Web Server [14-15](#)

Message Waiting Indicator (MWI) [12-6](#)

messaging

bandwidth management [13-6](#)

capabilities [1-6](#)

centralized [13-3, 13-13, 13-17, 13-21](#)

Cisco Unity [13-1](#)

combined deployment models [13-16](#)

deployment models [13-2](#)

distributed [13-3, 13-14, 13-19](#)

failover [13-4, 13-20](#)

system components [13-4](#)

MGCP [2-4, 4-8, 4-18, 4-23, 4-28, 16-2](#)

Microsoft Active Directory (AD) [17-11, 17-15, 17-17, 17-22](#)

Microsoft Live Communications Server 2005 [21-23](#)

Microsoft ViewMail for Outlook (VMO) [13-3](#)

migration

from static locations to RSVP call admission control [9-25](#)

from Unified CM 4.0 [16-44](#)

parallel cutover [18-2](#)

phased method [18-1](#)

to IP Telephony [18-1](#)

MISTP [3-4](#)

mixed-mode operation [3-20](#)

MLP [3-33](#)

MLPP [6-18](#)

MLTS [11-1](#)

Mobile Connect

architecture [23-8](#)

centralized gateways [23-11](#)

described [1-7, 23-1](#)

desk phone pickup [23-5](#)

dial plan [23-9](#)

- distributed gateways [23-12](#)
- enabling and disabling [23-19](#)
- functionality [23-4](#)
- phones supported [23-2](#)
- redundancy [23-13](#)
- remote destination phone pickup [23-6](#)
- system parameters [23-3](#)
- Unified CM services [23-3](#)
- voicemail [23-7](#)
- mobile devices [20-18](#)
- Mobile Voice Access
 - architecture [23-20](#)
 - described [1-8, 23-1, 23-15](#)
 - desk phone pickup [23-19](#)
 - dial plan [23-20](#)
 - enabling and disabling Mobile Connect [23-19](#)
 - functionality [23-16](#)
 - hairpinning [23-17](#)
 - IVR VoiceXML gateway [23-16](#)
 - phones supported [23-15](#)
 - redundancy [23-21](#)
 - remote destination phone pickup [23-19](#)
 - system parameters [23-16](#)
 - Unified CM services [23-15](#)
- Mobility
 - applications [23-1](#)
 - described [23-1, 23-21](#)
 - guidelines for deploying [23-21](#)
 - performance [23-22](#)
 - system parameters [23-3, 23-16](#)
- models for deployments (*see* deployment models)
- modem
 - clocking source [4-32](#)
 - features supported [4-27, 4-30](#)
 - gateway support for [4-7, 4-26](#)
 - interoperability of features [4-29](#)
 - pass-through mode [4-26](#)
 - platforms supported [4-27](#)
 - protocols supported [4-28](#)
 - relay mode [4-26](#)
 - upspeed [4-26](#)
 - V.34 [4-27](#)
 - V.90 [4-27](#)
- MoH [2-25, 7-1](#)
- moves, adds, and changes [11-9](#)
- MP [16-13, 16-15](#)
- MPLS [2-6, 2-15, 3-30, 3-33, 9-11, 9-44](#)
- MRG [6-22, 9-18, 16-17](#)
- MRGL [6-22, 9-18, 16-17](#)
- MTP
 - audio conferencing bridge [6-18](#)
 - described [6-13](#)
 - for PSTN calls [6-26](#)
 - for video calls [16-4](#)
 - hardware resources [6-17, 6-18](#)
 - hiding IP addresses of endpoints [6-26](#)
 - in multisite deployment model [2-5, 2-15](#)
 - in single-site deployment model [2-2](#)
 - Named Telephony Events [6-14](#)
 - requirements for trunks [8-20](#)
 - software resources [6-17](#)
 - with H.323 trunk [5-11](#)
 - with SIP trunk [5-14, 5-17](#)
- multicast music on hold [7-2, 7-8, 7-10, 7-16, 7-20](#)
- multicast traffic on WLAN [3-66](#)
- Multilevel Precedence Preemption (MLPP) [6-18](#)
- multi-line telephone system (MLTS) [11-1](#)
- Multilink Point-to-Point Protocol (MLP) [3-33](#)
- multi-media collaboration [1-6](#)
- Multimedia Conference Manager (MCM) [5-14, 8-20, 16-22](#)
- multiple clusters for Cisco Unity [13-5](#)
- Multiple Instance Spanning Tree Protocol (MISTP) [3-4](#)
- multiple Unified CM servers [13-21](#)
- multipoint conferencing [16-13](#)
- Multipoint Controller (MC) [16-13](#)
- Multipoint Control Unit (MCU)
 - capacity and sizing [16-20](#)
 - configuration [16-31](#)

- for video telephony [16-1, 16-13](#)
- with H.323 or SIP [16-18](#)
- with Skinny Client Control Protocol (SCCP) [16-16](#)
- Multipoint Processor (MP) [16-13, 16-15](#)
- Multiprotocol Label Switching (MPLS) [2-6, 2-15, 3-30, 3-33, 9-11, 9-44](#)
- multisite dial plan [10-55](#)
- multisite WAN deployment model
 - with centralized call processing [2-4, 6-24, 7-14, 10-92, 14-4, 15-3](#)
 - with distributed call processing [2-14, 6-25, 7-19, 10-93, 14-5, 15-4](#)
- music on hold (MoH) [2-25, 7-1](#)
- MWI [12-6](#)

N

- Named Service Event (NSE) [4-28, 4-33](#)
- Named Telephony Event (NTE) [4-12, 6-14](#)
- National Emergency Number Association (NENA) [11-6, 11-19](#)
- native transcoding with Cisco Unity [13-8](#)
- NENA [11-6, 11-19](#)
- Netscape Directory Server [17-11, 17-16](#)
- NetStructure PBX-IP Media Gateways (PIMG) [13-22](#)
- 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 [1-3](#)
 - requirements [3-1](#)
 - roles [3-3](#)
 - security [19-4](#)
 - WAN [3-30](#)
 - WLAN [3-63](#)
- network modules [6-21](#)
- network services [3-10](#)

- Network Specific Facilities (NSF) [4-21](#)
- Network Time Protocol (NTP) [3-22](#)
- NFAS [2-4, 4-21](#)
- NIC teaming [8-3](#)
- NM-HD-1V/2V/2VE module [6-9, 6-12, 6-17](#)
- NM-HDV2 module [6-9, 6-12, 6-17](#)
- NM-HDV module [6-10, 6-12](#)
- nomadic phones [11-9](#)
- Non-Facility Associated Signaling (NFAS) [2-4, 4-21](#)
- non-gatekeeper controlled H.323 client [16-26, 16-27, 16-31](#)
- non-gatekeeper controlled intercluster trunks [5-5](#)
- non-IOS hardware platforms [6-7](#)
- NPA [10-28](#)
- NSE [4-28, 4-33](#)
- NSF [4-21](#)
- NTE [4-12, 6-14](#)
- NTP [3-22](#)
- Numbering Plan Area (NPA) [10-28](#)
- number of digits dialed [10-4](#)

O

- off-net dialing [10-3](#)
- on-net dialing [10-3, 10-5, 10-6, 10-58, 10-61, 10-66](#)
- open authentication [3-68, 20-14, 20-15](#)
- Open Shortest Path First (OSPF) [19-34](#)
- Option 150 [3-11, 3-12](#)
- OSPF [19-34](#)
- outbound calls [4-39, 4-43, 5-9, 10-60, 10-64, 10-69](#)
- outvia [9-29, 10-44, 16-34](#)
- overlapping
 - channels [3-65](#)
 - dial plans [10-61](#)
 - extensions [10-4](#)
- overlap receiving [10-18](#)
- overlap sending [10-18](#)
- oversubscription of a link [3-38](#)

P

PA [16-47](#)

PAC [3-68](#), [20-14](#)

packets

delay [2-18](#), [2-20](#), [4-27](#)

headers [3-51](#)

jitter [2-18](#)

loss of [2-18](#), [4-25](#)

parallel cutover [18-2](#)

parameters

clusterwide [9-23](#)

for Mobile Connect [23-3](#)

for Mobile Voice Access [23-16](#)

service parameters [22-2](#), [22-3](#), [22-8](#), [22-14](#), [22-29](#), [22-40](#)

partitioned addressing [10-57](#), [10-61](#)

partitions [10-21](#), [10-22](#), [10-61](#), [10-79](#), [10-83](#)

passive-interface command [3-9](#)

pass-through codec [9-22](#)

pattern recognition in dialing [10-3](#), [10-74](#)

PBX-IP Media Gateway (PIMG) [13-22](#)

PC

Access to Voice VLAN [20-20](#)

port on IP phone [19-6](#), [20-20](#)

performance

call rate [8-1](#)

of Attendant Console [22-38](#)

of Cisco Unified Mobility [23-22](#)

of Extension Mobility [22-13](#)

of presence servers [21-11](#)

of servers [8-15](#)

of Unified CM Assistant [22-28](#)

of WebDialer [22-48](#)

overload on gateways [4-5](#)

tuning of gateways [4-5](#)

Per-Port/Per-VLAN ACLs [20-35](#)

Personal Assistant (PA) [16-47](#)

Personal Communicator [1-5](#), [20-10](#), [20-31](#), [21-19](#)

phased migration [18-1](#)

phones

7902G [20-6](#)

7905G [20-6](#)

7906G [20-6](#)

7911G [20-7](#)

7912G [20-7](#)

7914 Expansion Module [20-10](#)

7940G [20-7](#)

7941G [20-8](#)

7941G-GE [20-8](#)

7960G [20-8](#)

7961G [20-8](#)

7961G-GE [20-9](#)

7970G [20-9](#)

7971G-GE [20-10](#)

7985G IP Video Phone [20-23](#), [20-35](#)

Attendant Console [22-28](#)

authentication and encryption [19-11](#)

basic models [20-6](#)

built-in conferencing [6-11](#)

business models [20-7](#)

call pickup at desk phone [23-5](#), [23-19](#)

capacity calculations [8-17](#)

configuration [20-16](#)

desktop IP models [20-6](#)

dialed pattern recognition [10-74](#)

executive models [20-9](#)

Extension Mobility [22-7](#)

features [20-37](#)

for Mobile Connect [23-2](#)

for Mobile Voice Access [23-15](#)

IP Phone Services [22-2](#)

line appearances [8-18](#)

location for 911 purposes [11-9](#)

manager models [20-8](#)

nomadic [11-9](#)

PC port [19-6](#)

QoS [20-26](#)

remote destination call pickup [23-6](#), [23-19](#)

- roaming [3-65, 20-16, 20-18](#)
- SCCP [10-9](#)
- security [19-5, 19-41](#)
- service parameters [22-2, 22-3, 22-8, 22-14](#)
- services [22-2](#)
- settings [19-10](#)
- SIP [10-9, 10-11, 20-25](#)
- software-based [1-5, 20-10, 20-31](#)
- Type-A [10-2, 10-9](#)
- Type-B [10-2, 10-11](#)
- Unified Communications Manager Assistant [22-13](#)
- user input [10-9, 10-11](#)
- video telephony [20-34](#)
- web access [19-9](#)
- WebDialer [22-39](#)
- wireless [1-6, 20-13, 20-33](#)
- Wireless IP Phone 7920 [20-13](#)
- Wireless IP Phone 7921G [20-13](#)
- with Cisco Unified Video Advantage [16-1](#)
- with Cisco Unified Video Advantage [20-20](#)
- physical security [19-4](#)
- pilot number for hunt lists [10-33, 10-35, 10-94](#)
- PIMG [13-22](#)
- ping utility [2-20](#)
- PINX [12-6](#)
- PIX [19-29, 19-31](#)
- plain old telephone service (POTS) [11-6](#)
- platforms [2-2, 8-2, 8-21, 16-25](#)
- PoE [3-23](#)
- policy
 - for network security [19-2](#)
 - for presence [21-7](#)
 - for RSVP [3-56, 9-23](#)
- PortFast [3-6](#)
- port groups for Cisco Unity Connection [13-5](#)
- ports
 - access [19-14](#)
 - CTI [8-20](#)
 - enable/disable [20-20](#)
 - for call signaling [4-44](#)
 - for Cisco Unified Video Advantage [20-35](#)
 - for integration of Cisco Unity with Unified CM [13-10, 13-12](#)
 - for presence server [21-14](#)
 - groups [13-5](#)
 - on the IP phone [19-6](#)
 - PC connection [20-20](#)
 - security [19-13](#)
- positive disconnect supervision [12-11](#)
- POTS [11-6](#)
- Power over Ethernet (PoE) [3-23](#)
- precedence settings for network traffic [3-4, 3-34](#)
- prefixes
 - for access code [10-28](#)
 - for gatekeeper zones [9-34](#)
 - gateway [16-34](#)
 - MCU [16-32](#)
 - service [4-39, 16-19](#)
 - zones [16-36, 16-38, 16-40](#)
- presence
 - call history [21-6](#)
 - Cisco IP Phone Messenger (IPPM) [21-15](#)
 - clusters [21-9](#)
 - components [21-2](#)
 - deployment models [21-10](#)
 - described [21-1](#)
 - end user [21-3](#)
 - groups [21-7](#)
 - guidelines [21-8](#)
 - IBM Sametime 7.5 [21-23](#)
 - integration with third-party applications [21-23](#)
 - interactions between components [21-12](#)
 - licensing of users [21-11](#)
 - Microsoft Live Communications Server 2005 [21-23](#)
 - policy [21-7](#)
 - port usage [21-14](#)
 - presentity [21-1](#)
 - SCCP [21-5](#)

- server guidelines [21-14](#)
- server performance [21-11](#)
- server redundancy [21-10](#)
- servers [21-8](#)
- SIP [21-4](#)
- speed dial [21-5](#)
- SUBSCRIBE calling search space [21-7](#)
- synchronization of servers [21-9](#)
- Unified CM [21-4](#)
- presentity [21-1](#)
- preservation of calls [4-15](#)
- Pre-Shared Key (PSK) [3-69](#)
- PRI [11-5](#)
- primary extension [21-3](#)
- Primary Rate Interface (PRI) [11-5](#)
- prioritization of traffic [3-34](#)
- priority, urgent [10-19](#)
- Priority Queue [3-56](#)
- Private Integrated Services Network Exchange (PINX) [12-6](#)
- Private Internet Exchange (PIX) [19-29](#), [19-31](#)
- privileges for making calls [10-21](#), [10-50](#)
- profiles for Extension Mobility [8-19](#)
- progress_ind alert enable 8** command [11-12](#)
- propagation of database [8-4](#)
- Protected Access Credential (PAC) [3-68](#), [20-14](#)
- Protocol Auto Detect [5-13](#)
- protocols
 - ARP [3-68](#), [19-20](#)
 - CDP [19-12](#), [20-20](#)
 - cRTP [3-33](#), [3-35](#)
 - DHCP [3-11](#), [19-16](#), [19-18](#), [19-19](#)
 - features supported [16-3](#)
 - GARP [19-7](#), [19-20](#)
 - GKTMP [5-14](#)
 - GUP [5-6](#), [8-24](#)
 - H.225 [5-6](#), [5-13](#)
 - H.245 [4-33](#)
 - H.320 [16-33](#), [16-38](#)
 - H.323 [2-4](#), [4-8](#), [4-18](#), [4-19](#), [4-20](#), [4-21](#), [4-22](#), [4-28](#), [4-35](#), [5-1](#), [5-3](#), [5-4](#), [5-11](#), [8-30](#), [10-38](#), [10-88](#), [14-10](#), [15-9](#), [16-2](#), [16-18](#), [16-26](#), [20-36](#)
 - HSRP [2-16](#), [3-7](#), [8-20](#), [8-21](#), [14-31](#)
 - IPSec [2-6](#), [2-15](#)
 - JTAPI [16-2](#)
 - LDAP [8-4](#), [17-1](#)
 - MGCP [2-4](#), [4-8](#), [4-18](#), [4-23](#), [4-28](#), [16-2](#)
 - MISTP [3-4](#)
 - MLP [3-33](#)
 - MPLS [9-11](#)
 - NTP [3-22](#)
 - RAS [10-41](#), [16-22](#)
 - RCP [19-20](#)
 - RIP [19-34](#)
 - routing [3-9](#)
 - RSTP [3-4](#), [3-6](#)
 - RSVP [3-30](#), [3-40](#), [9-7](#), [9-28](#), [16-4](#)
 - RTP [2-16](#), [16-2](#)
 - SCCP [4-8](#), [4-28](#), [7-20](#), [10-3](#), [10-9](#), [16-2](#), [16-16](#), [20-20](#), [20-24](#), [21-5](#)
 - SDP [4-33](#), [5-16](#), [6-14](#)
 - SIMPLE [21-8](#)
 - SIP [2-16](#), [4-12](#), [4-16](#), [4-18](#), [4-19](#), [4-20](#), [4-21](#), [4-22](#), [5-1](#), [5-3](#), [5-14](#), [6-18](#), [7-23](#), [10-3](#), [10-9](#), [10-11](#), [10-13](#), [10-74](#), [13-9](#), [14-10](#), [15-10](#), [16-2](#), [20-25](#), [20-36](#), [21-4](#)
 - SMDI [12-1](#)
 - SNMP [11-9](#)
 - SOAP [21-9](#)
 - SRTP [3-51](#)
 - STP [3-6](#)
 - TAPI [16-2](#)
 - TFTP [3-12](#), [3-15](#), [8-3](#), [8-12](#), [20-20](#)
 - UDP [2-16](#), [5-6](#)
- provisioning
 - H.320 gateways [16-33](#)
 - H.323 clients [16-26](#)
 - MCUs [16-31](#)
 - servers [8-15](#), [8-16](#), [8-17](#)
- proxy

- for gatekeeper [8-20, 16-36, 16-38, 16-39](#)
- line mode with Unified CM Assistant [22-16](#)
- PSAP [11-2, 11-8, 11-14, 20-19](#)
- PSK [3-69](#)
- PSTN [2-2, 2-6, 2-10, 2-15, 4-3, 5-4, 6-26, 10-27, 11-1](#)
- public safety answering point (PSAP) [11-2, 11-8, 11-14, 20-19](#)
- Public Switched Telephone Network (PSTN) [2-2, 2-6, 2-15, 4-3, 5-4, 10-27, 11-1](#)
- publisher server [2-20, 8-6](#)
- PVDM [6-20](#)

Q

- Q.SIG [5-14](#)
- QBSS [3-68, 3-72, 20-16, 20-17, 20-18](#)
- QBSS-Differential Threshold [20-16](#)
- QCIF [20-25](#)
- QoS
 - Attendant Console [22-34](#)
 - Cisco Unified MeetingPlace [14-8](#)
 - Cisco Unified MeetingPlace Express [15-8](#)
 - Cisco Unified MeetingPlace Video [14-20](#)
 - configuration examples [20-25](#)
 - for LAN [3-25](#)
 - for music on hold [7-11](#)
 - for security [19-24](#)
 - for WAN [3-30, 3-33](#)
 - for wireless LAN [3-70](#)
 - general [1-4](#)
 - RSVP [3-43](#)
 - Unified CM Assistant [22-23](#)
- QoS Basic Service Set (QBSS) [3-68, 3-72, 20-16, 20-17, 20-18](#)
- QSIG [4-19, 4-24, 12-6, 18-3](#)
- Quality of Service (QoS)
 - configuration examples [20-25](#)
 - for Attendant Console [22-34](#)
 - for Cisco Unified MeetingPlace [14-8](#)
 - for Cisco Unified MeetingPlace Express [15-8](#)

- for Cisco Unified MeetingPlace Video [14-20](#)
- for LAN [3-25](#)
- for music on hold [7-11](#)
- for security [19-24](#)
- for Unified CM Assistant [22-23](#)
- for WAN [3-30, 3-33](#)
- for wireless LAN [3-70](#)
- general [1-4](#)
- RSVP [3-43](#)
- quantity of
 - calls [8-19](#)
 - gateways [8-18](#)
 - phones [8-18](#)
 - trunks [8-19](#)
- Quarter Common Intermediate Format (QCIF) [20-25](#)
- queue depth [3-62](#)
- queuing of voice traffic [3-28, 3-71](#)
- quiescent traffic [3-63](#)

R

- radio frequency (RF) [20-14](#)
- RADIUS [3-69](#)
- Rapid Spanning Tree Protocol (RSTP) [3-4, 3-6](#)
- RAS [5-6, 9-14, 10-41, 16-22](#)
- RASAggregator trunk [14-24, 16-26, 16-30, 16-31](#)
- Rate Matching (RM) module [16-13, 16-16](#)
- rate of error [2-21](#)
- RBOC [11-2](#)
- RCF [16-41](#)
- RCP [19-20](#)
- RDNIS [13-7](#)
- Real Time Monitoring Tool (RTMT) [17-2](#)
- Real-Time Transport Protocol (RTP) [2-16, 16-2](#)
- recognition of dialed patterns [10-74](#)
- recommended hardware and software versions [2-2, A-1](#)
- Redirected Dialed Number Information Service (RDNIS) [13-7](#)
- Redirector servlet [22-42](#)

redundancy

- Attendant Console [22-36](#)
- call processing [8-8](#)
- cluster configurations [8-9](#)
- during software upgrades [8-8](#)
- Extension Mobility [22-11](#)
- for Cisco Unified MeetingPlace [14-12](#)
- for Cisco Unified MeetingPlace Express [15-12](#)
- for Mobile Connect [23-13](#)
- for Mobile Voice Access [23-21](#)
- for music on hold [7-11](#)
- for presence servers [21-10](#)
- for remote sites [2-8](#)
- for trunks [5-6](#)
- for Unified CM Assistant [22-24](#)
- for video [14-30](#)
- gatekeeper [8-21](#)
- gateway support for [4-7, 4-15](#)
- groups [14-24](#)
- IP Phone Services [22-6](#)
- IP-to-IP gateways [9-31](#)
- load balancing [8-11](#)
- TFTP services [3-18](#)
- WebDialer [22-47](#)

Regional Bell Operating Company (RBOC) [11-2](#)

regions [16-5, 16-7](#)

Registration Admission Status (RAS) [5-6, 9-14, 10-41, 16-22](#)

Registration Confirm (RCF) [16-41](#)

registration of RSVP Agent [9-21](#)

Registration Request (RRQ) [16-41](#)

Relative Signal Strength Indicator (RSSI) [20-16, 20-17](#)

Remote Authentication Dial-In User Service (RADIUS) [3-69](#)

Remote Copy Protocol (RCP) [19-20](#)

remote destination phone pickup [23-6, 23-19](#)

remote failover deployment model [2-26](#)

remote-mount servers [3-21](#)

remote RSVP Agent [9-55](#)

remote site survivability [2-8](#)

re-packetization of a stream [6-13](#)

replication of database [8-4](#)

request for

bandwidth [5-14](#)

Reservationless Single Number Access (RSNA) [14-12](#)

resilience [5-6, 8-1](#)

resolution of addresses [10-43, 10-44](#)

Resource Reservation Protocol (RSVP) [3-30, 3-40, 9-7, 9-28, 16-4](#)

restrictions for

Attendant Console [22-38](#)

Extension Mobility [22-12](#)

IP Phone Services [22-7](#)

Unified CM Assistant [22-27](#)

WebDialer [22-48](#)

Retry Video Call as Audio [16-8](#)

RF [20-14](#)

RFC 2833 [4-12, 6-14](#)

rich-media conferencing [1-1](#)

RIP [19-34](#)

RJ-45 [3-25](#)

RM [16-13, 16-16](#)

roaming [3-65, 11-9, 20-16, 20-18](#)

rogue

DHCP server [19-16](#)

network extensions [19-15](#)

roles

in the network infrastructure [3-3](#)

of a gatekeeper [16-23](#)

rollover of channels [4-42](#)

root guard [3-6](#)

round-trip time (RTT) [2-20, 2-24](#)

Route/Switch Processor (RSP) [4-26](#)

routed firewall

ASA or PIX [19-34](#)

FWSM [19-37](#)

routers

access control list (ACL) [19-26](#)

branch office [7-16](#)

- flash [7-16](#)
- roles and features [3-3](#)
- RSVP [3-43](#)
- selective for E911 [11-3](#)
- routes
 - filters [10-18](#)
 - group devices [10-21](#)
 - groups [10-19, 10-21](#)
 - lists [10-20](#)
 - patterns [10-15, 10-17](#)
 - selection of [10-28](#)
- routing
 - calling line ID [10-19](#)
 - calls [10-15, 10-38, 10-41, 23-11, 23-12](#)
 - digit manipulation [10-19](#)
 - inbound calls [4-37](#)
 - least-cost [4-41](#)
 - outbound calls [4-39](#)
 - protocols [3-9](#)
 - time-of-day (ToD) [10-37](#)
- Routing Information Protocol (RIP) [19-34](#)
- RRQ [16-41](#)
- RSNA [14-12](#)
- RSP [4-26](#)
- RSSI [20-16, 20-17](#)
- RSSI-Differential Threshold [20-16](#)
- RSTP [3-4, 3-6](#)
- RSVP
 - call admission control [9-7](#)
 - Cisco RSVP Agent [9-18, 9-21, 9-55](#)
 - described [3-40](#)
 - IP-to-IP Gateway [9-28](#)
 - locations enabled for RSVP [9-16, 16-4](#)
 - policy [9-23](#)
 - WAN infrastructure [3-30](#)
- RTMT [17-2](#)
- RTP [2-16, 16-2](#)
- RTT [2-20, 2-24](#)

S

- scalability of
 - gatekeepers [16-24](#)
 - IP Phone Services [22-6](#)
 - Unified CM [8-1](#)
- SCCP
 - dialed pattern recognition [10-3](#)
 - fax and modem support [4-8](#)
 - gateway support for [4-8](#)
 - MCU resources [16-16](#)
 - music on hold (MoH) [7-20](#)
 - phones [10-9](#)
 - presence [21-5](#)
 - user input on phones [10-9](#)
 - video endpoints [16-2, 20-20, 20-24](#)
- schema [17-1](#)
- screen sharing [15-7](#)
- SDK [17-4](#)
- SDP [4-33, 5-16, 6-14](#)
- search base for directories [17-13](#)
- Section 255 [2-27](#)
- Section 508 [2-27](#)
- security
 - access control list (ACL) [19-24, 19-26](#)
 - antivirus [19-40](#)
 - Cisco Security Agent [19-39](#)
 - configuration example [19-15, 19-18, 19-21, 19-23, 19-25, 19-27, 19-35, 19-37, 19-41](#)
 - data center [19-39](#)
 - DHCP Snooping [19-16](#)
 - DHCP starvation attack [19-18](#)
 - directories [17-16](#)
 - firewalls [19-31, 19-44](#)
 - gateways [19-28](#)
 - infrastructure [19-4](#)
 - in general [1-8, 19-1, 19-2](#)
 - intracluster communications [8-5](#)
 - layers [19-3](#)

- lobby phone example [19-41](#)
- MAC CAM flooding [19-13](#)
- media resources [19-28](#)
- MeetingPlace [14-7](#)
- MeetingPlace Express [15-5](#)
- PC port on the phone [19-6](#)
- phones [19-5](#)
- phone settings [19-10](#)
- physical access [19-4](#)
- policy [19-2](#)
- QoS [19-24](#)
- rogue network extensions [19-15](#)
- servers [19-39, 19-40](#)
- switch port [19-13](#)
- Video Capabilities [19-9](#)
- voice VLAN [19-8](#)
- web access [19-9](#)
- wireless network [3-68](#)
- selecting the proper route [10-28](#)
- selective router [11-3](#)
- separate integrations for Cisco Unity [13-5](#)
- Sequenced Routing Update Protocol (SRTP) [3-51](#)
- sequential LRQs [8-27](#)
- servers
 - capacity planning [8-15, 8-16, 8-17](#)
 - clusters [8-2, 21-9](#)
 - co-located [3-13](#)
 - co-resident DHCP [3-14](#)
 - co-resident MoH [7-3](#)
 - CTI Manager [8-12](#)
 - data center [3-10](#)
 - farm [3-10](#)
 - for DHCP [3-14](#)
 - for Extension Mobility (EM) [8-14](#)
 - for media resources [6-1](#)
 - for music on hold [7-3, 7-4, 7-12](#)
 - for presence [21-8](#)
 - for Unified CM [8-2](#)
 - high-availability [8-2](#)
 - high-performance [8-2](#)
 - maximum number of devices [8-16](#)
 - multiple Unified CM servers [13-21](#)
 - performance [8-15, 21-11](#)
 - publisher [2-20, 8-6](#)
 - redundancy [14-13, 21-10](#)
 - remote mount [3-21](#)
 - security [19-39, 19-40](#)
 - shadow server [14-13](#)
 - standalone [3-14, 7-3](#)
 - subscriber [2-20, 8-7](#)
 - synchronization [21-9](#)
 - TFTP [8-12](#)
 - types [8-2](#)
- Service Inter-Working (SIW) [2-6, 2-15, 3-33](#)
- service parameters
 - for Attendant console [22-29](#)
 - for Extension Mobility [22-8](#)
 - for IP Phone Services [22-2, 22-3](#)
 - for Unified CM Assistant [22-14](#)
 - for WebDialer [22-40](#)
- services
 - for IP phones [22-2](#)
 - prefix [4-39, 16-19, 16-32, 16-34](#)
 - supplementary [4-7](#)
 - template [16-19](#)
 - within a cluster [8-3](#)
- service set identifier (SSID) [3-64, 3-68](#)
- servlet for
 - Redirector [22-42](#)
 - WebDialer [22-41](#)
- Session Description Protocol (SDP) [4-33, 5-16, 6-14](#)
- Session Initiation Protocol (SIP)
 - analog gateways [4-18](#)
 - annunciator [6-18](#)
 - delayed offer [5-16](#)
 - dialed pattern recognition [10-3](#)
 - dial rules [10-13, 10-74](#)
 - digital gateways [4-19, 4-20, 4-21, 4-22](#)

- early offer [5-16](#)
- for distributed call processing [2-16](#)
- gateways [4-16](#)
- gateway support for [4-12](#)
- intercluster trunks [5-18](#)
- music on hold (MoH) [7-23](#)
- phones [10-9, 10-11, 20-25](#)
- presence [21-4](#)
- trunks [5-1, 5-3, 5-14, 13-9, 14-10, 15-10](#)
- Type-A phones [10-9](#)
- Type-B phones [10-11](#)
- video endpoints [16-2, 20-36](#)
- settings for IP phones [19-10](#)
- shadow server [14-13](#)
- shaping traffic [3-37](#)
- shared
 - key authentication [20-15](#)
 - line appearances [3-61, 11-14](#)
 - line mode with Unified CM Assistant [22-17](#)
 - T.120 applications [16-47](#)
- shielded twisted-pair (STP) [3-25](#)
- signaling encryption [3-59, 3-61](#)
- Signaling System 7 [2-4](#)
- SIMPLE [21-8](#)
- Simple Network Management Protocol (SNMP) [11-9](#)
- Simple Object Access Protocol (SOAP) [21-9](#)
- Simplified Message Desk Interface (SMDI) [12-1](#)
- simultaneous calls [4-2](#)
- Single Number Reach (*see* Mobile Connect)
- single site
 - deployment model [2-2, 6-24, 7-14, 14-3, 15-2](#)
 - messaging model [13-2](#)
- SIP
 - analog gateways [4-18](#)
 - annunciator [6-18](#)
 - delayed offer [5-16](#)
 - dialed pattern recognition [10-3](#)
 - dial rules [10-13, 10-74](#)
 - digital gateways [4-19, 4-20, 4-21, 4-22](#)
 - early offer [5-16](#)
 - for distributed call processing [2-16](#)
 - gateways [4-16](#)
 - gateway support for [4-12](#)
 - intercluster trunks [5-18](#)
 - music on hold (MoH) [7-23](#)
 - phones [10-9, 10-11, 20-25](#)
 - presence [21-4](#)
 - trunks [5-1, 5-3, 5-14, 13-9, 14-10, 15-10](#)
 - Type-A phones [10-9](#)
 - Type-B phones [10-11](#)
 - video endpoints [16-2, 20-36](#)
- SIP for Instant Messaging and Presence Leveraging Extensions (SIMPLE) [21-8](#)
- site
 - dialing codes [10-6, 10-72](#)
 - survey for wireless network [20-14](#)
- SIW [2-6, 2-15, 3-33](#)
- sizing
 - MCUs [16-20](#)
 - Unified CM servers [8-15, 8-16, 8-17](#)
- Skinny Client Control Protocol (SCCP)
 - dialed pattern recognition [10-3](#)
 - fax and modem support [4-28](#)
 - gateway support for [4-8](#)
 - MCU resources [16-16](#)
 - music on hold (MoH) [7-20](#)
 - phones [10-9](#)
 - presence [21-5](#)
 - user input on phones [10-9](#)
 - video endpoints [16-2, 20-20, 20-24](#)
- SMDI [12-1](#)
- sn attribute [17-11](#)
- SNMP [11-9](#)
- snooping [19-16](#)
- SOAP [21-9](#)
- soft clients [11-14](#)
- SoftPhone [11-14, 16-47, 20-37](#)
- software

- audio conferencing bridge [6-9](#)
 - endpoints [20-10](#)
 - media resource capacities [6-20](#)
 - MTP resources [6-17](#)
 - phones [20-37](#)
 - recommendations [2-2](#)
 - versions [2-2, 20-4, 20-5, A-1](#)
 - software-based endpoints [20-31](#)
 - Software Development Kit (SDK) [17-4](#)
 - Sony endpoints [20-24](#)
 - Source Guard [19-22](#)
 - Spanning Tree Protocol (STP) [3-6](#)
 - speed dial presence [21-5](#)
 - speed of calls [3-53](#)
 - SRST [2-6, 2-8, 7-16, 8-3, 10-93, 11-3](#)
 - S RTP [3-51](#)
 - SS7 [2-4](#)
 - SSID [3-64, 3-68](#)
 - standalone server [3-14, 7-3](#)
 - standard server [8-2](#)
 - standby preempt** command [3-7](#)
 - standby track** command [3-7](#)
 - star topology [9-36](#)
 - static ANI interface [11-7](#)
 - static locations [9-12](#)
 - Static Wired Equivalent Privacy [3-69](#)
 - Static Wire Equivalent Privacy (WEP) [3-68](#)
 - stealth firewall [19-34](#)
 - STP [3-6, 3-25](#)
 - string length [10-4](#)
 - subnets [16-40](#)
 - SUBSCRIBE calling search space [21-7](#)
 - subscriber server [2-20, 8-7](#)
 - summary of endpoint gatekeepers [16-41](#)
 - Sun ONE Directory Server [17-11, 17-16](#)
 - Super-Group 3 fax support [4-27](#)
 - supplementary services
 - for H.323 endpoints [6-13](#)
 - on gateways [4-7, 4-12](#)
 - supported
 - call types [16-3](#)
 - codecs [16-6, 20-25](#)
 - platforms for gatekeepers [16-25](#)
 - protocols [16-2, 16-3](#)
 - survey of wireless network [20-14](#)
 - Survivable Remote Site Telephony (SRST) [2-6, 2-8, 7-16, 8-3, 10-93, 11-3](#)
 - switchback [9-21](#)
 - switches
 - port security [19-13](#)
 - roles and features [3-3](#)
 - switchover [9-21](#)
 - synchronization of
 - directories [17-11, 17-24](#)
 - presence servers [21-9](#)
 - Unified CM database [17-24](#)
 - synchronous H.323 client [16-27](#)
-
- ## T
- T.120 application sharing [16-47](#)
 - T.38 fax relay [4-33](#)
 - Tail-End Hop-Off (TEHO) [10-55](#)
 - Tandberg endpoints
 - classification of traffic [20-35](#)
 - described [16-1, 20-24](#)
 - TAPI [8-12, 16-2](#)
 - TCP/UDP ports [20-35](#)
 - TCS [16-10](#)
 - TEHO [10-55](#)
 - Telecommunications Act [2-27](#)
 - telephone record and playback (TRaP) [13-3](#)
 - telephone user interface (TUI) [13-3](#)
 - templates to define service settings [16-19](#)
 - Temporal Key Integrity Protocol (TKIP) [3-69](#)
 - Terminal Capabilities Set (TCS) [16-10](#)
 - termination of calls [6-2](#)
 - test calls for 911 [11-14](#)

- TFTP [3-12, 3-15, 8-3, 8-12, 20-20](#)
- third-party
 - controlled lines [8-20](#)
 - SIP phones [20-25](#)
 - software applications [1-2](#)
 - video endpoints [20-24](#)
 - voicemail systems [12-1, 12-11](#)
- threshold, differential [20-16](#)
- time-of-day (ToD) routing [10-37](#)
- timers for call signaling [4-45](#)
- time synchronization [3-22, 3-23](#)
- Time to Live (TTL) [16-41](#)
- TKIP [3-69](#)
- ToD [10-37](#)
- Token Ring [3-25](#)
- topology
 - for call admission control [9-36](#)
 - generic [9-50](#)
 - hub-and-spoke [9-14, 9-36, 10-41](#)
 - MPLS-based [9-44](#)
 - star [9-36](#)
 - two-tier hub-and-spoke [9-40](#)
- topology-aware
 - call admission control [9-7](#)
 - locations [16-4](#)
- topology-unaware call admission control [9-3](#)
- TPC [3-65](#)
- tracking domain [11-18](#)
- traditional approach to classes of service [10-76](#)
- traffic
 - bearer traffic [3-51, 3-54](#)
 - bursty [4-2](#)
 - call control [3-58, 3-63](#)
 - call-related [3-63](#)
 - classification [3-4, 3-27, 3-70, 14-8, 14-20, 15-8, 20-25, 20-34](#)
 - contact center traffic patterns [4-3, 4-4](#)
 - gateway sizing [4-2](#)
 - normal business traffic [4-3](#)
 - prioritization [3-34](#)
 - provisioning for [3-51](#)
 - PSTN traffic patterns [4-3](#)
 - queuing [3-28, 3-71](#)
 - quiescent [3-63](#)
 - shaping [3-37](#)
 - traffic patterns [4-2](#)
 - video bearer traffic [3-53, 3-55](#)
 - voice bearer traffic [3-51, 3-55](#)
- Traffic Specification (TSPEC) [20-17, 20-18](#)
- transcoding
 - Cisco Unity [13-8](#)
 - described [6-11](#)
 - for video telephony [16-4](#)
 - hardware resources [6-12, 6-13](#)
 - IP PSTN [6-26](#)
 - resources [6-12](#)
- translation of digits
 - patterns [10-26](#)
 - voice translation profiles [10-52](#)
- Transmit Power Control (TPC) [3-65](#)
- transparent firewall
 - ASA or PIX [19-34](#)
 - FWSM [19-37](#)
- TRaP [13-3](#)
- Trivial File Transfer Protocol (TFTP) [3-12, 3-15, 8-3, 8-12, 20-20](#)
- tromboning [13-14](#)
- troubleshooting for clustering over the WAN [2-21](#)
- trunks
 - capacity calculations [8-19](#)
 - comparing H.323 and SIP [5-1](#)
 - described [5-1](#)
 - features supported [5-1](#)
 - H.225 [5-6, 5-13](#)
 - H.323 [5-3, 5-4, 5-11](#)
 - intercluster, gatekeeper controlled [5-5](#)
 - intercluster, non-gatekeeper controlled [5-5](#)
 - load balancing [5-6](#)
 - MTP requirements [8-20](#)

PSTN [5-4](#)
 RASAggregator [16-26, 16-30, 16-31](#)
 redundancy [5-6](#)
 SIP [5-3, 5-14, 6-18, 13-9](#)
 to service provider network [5-4](#)
 trust [20-25](#)
 TSPEC [20-17, 20-18](#)
 TTL [16-41](#)
 TUI [13-3](#)
 tuning gateway performance [4-5](#)
 Tunneled Q.SIG [5-14](#)
 two-tier hub-and-spoke topology [9-40](#)
 Type-A phones [10-2, 10-9](#)
 Type-B phones [10-2, 10-11](#)

U

UAC [20-6](#)
 UAS [20-6](#)
 UDC [3-25](#)
 UDLD [3-6](#)
 UDP [2-16, 3-35, 5-6](#)
 UN [4-12](#)
 unicast music on hold [7-2, 7-8, 7-10, 7-20](#)
 UniDirectional Link Detection (UDLD) [3-6](#)
 Unified CC [16-46](#)
 Unified CM
 Capacity Tool [8-15, 8-16, 8-17](#)
 co-located clusters [9-52](#)
 configuration of Cisco IOS gateways for fax/modem support [4-32](#)
 database synchronization [17-24](#)
 described [1-4](#)
 different versions in the same cluster [3-20](#)
 directory attributes [17-13](#)
 enhancements for IP Video Telephony [16-2](#)
 groups [2-23, 2-26](#)
 H.323 [5-12](#)
 integration with MeetingPlace [14-1](#)
 integration with MeetingPlace Express [15-1](#)
 load balancing [14-16](#)
 migration from Release 4.0 [16-44](#)
 mixed-mode operation [3-20](#)
 presence [21-4](#)
 redundancy [14-16](#)
 Release 3.3 [10-33](#)
 Release 4.0 [10-33](#)
 services [22-2, 22-8, 22-14, 22-29, 22-40](#)
 Unified CM Assistant [1-7, 8-14, 16-46, 22-13](#)
 Unified CMCT [8-15, 8-16, 8-17](#)
 Unified CME [2-17, 8-30](#)
 Unified CM Express (Unified CME) [2-17, 8-30](#)
 Unified Communications [1-1](#)
 Unified Communications Manager Assistant (Unified CM Assistant) [1-7, 8-14, 16-46, 22-13](#)
 Unified Communications Manager Capacity Tool (Unified CMCT) [8-15, 8-16, 8-17](#)
 Unified IP IVR [16-46](#)
 Unified MeetingPlace [14-1](#)
 unified messaging (*see also* messaging) [13-1](#)
 Unified Mobility (*see* Mobility)
 Unified Personal Communicator [1-5, 20-31](#)
 Unified Presence [21-1](#)
 Unified Video Advantage
 classification of traffic [20-35](#)
 described [16-1, 20-20](#)
 QoS recommendations [20-31](#)
 uniform on-net dial plan [10-5, 10-58](#)
 uninterrupted power supplies (UPS) [3-23](#)
 Unity [13-1](#)
 Unity Telephony Integration Manager (UTIM) [13-5, 13-10, 13-12](#)
 universal data connector (UDC) [3-25](#)
 Unsolicited SIP Notify (UN) [4-12](#)
 upgrading Unified CM releases [8-8](#)
 UplinkFast [3-6](#)
 UPS [3-23](#)
 upspeed [4-26](#)
 Urgent Priority [10-19](#)

URLs for WebDialer [22-45](#)
 user agent client (UAC) [20-6](#)
 user agent server (UAS) [20-6](#)
 User Datagram Protocol (UDP) [2-16, 3-35, 5-6](#)
 user hold [7-6](#)
 UserID [17-11](#)
 users
 application users [17-8](#)
 classes of service [10-76, 10-80, 10-88](#)
 directory search base [17-13](#)
 end users [17-8](#)
 input on phones [10-9, 10-11](#)
 User-to-User Information Element (UUIE) [5-13](#)
 utilization of
 DSOs [8-19](#)
 phones [8-18](#)
 trunks [8-20](#)
 UTIM [13-5, 13-10, 13-12](#)
 UUIE [5-13](#)

V

V.34 modems [4-27](#)
 V.90 modems [4-27](#)
 V3PN [2-6, 2-15](#)
 VAD [4-4, 4-26, 8-13, 16-14](#)
 VAF [3-37](#)
 variable length on-net dial plan [10-6, 10-61, 10-66](#)
 VATS [3-39](#)
 VG224 Voice Gateway [4-18, 12-2, 20-5, 20-25](#)
 VG248 Analog Phone Gateway [4-31, 12-3, 20-5, 20-25](#)
 via-zone gatekeeper [9-29, 9-34, 10-44, 14-21](#)
 VIC [20-2, 20-3](#)
 video
 bearer traffic [3-53, 3-55](#)
 call flows [14-27](#)
 capabilities [19-9](#)
 conferences [14-29](#)
 described [16-1](#)

 dial plan [14-29](#)
 enable/disable [20-20](#)
 endpoints [1-5, 14-21, 16-1, 20-20, 20-34](#)
 features [1-1, 1-6](#)
 gatekeeper integration [14-20](#)
 gateways [4-36](#)
 redundancy [14-30](#)
 traffic classification [3-28, 14-20, 20-34](#)
 VLAN [19-12](#)
 with Cisco Unified MeetingPlace [14-16](#)
 Video Administration for Cisco Unified MeetingPlace [14-17, 14-22, 14-23](#)
 Video Capabilities [19-9](#)
 Video Integration component of Cisco Unified MeetingPlace [14-16](#)
 video telephony (*see* IP Video Telephony)
 ViewMail for Outlook (VMO) [13-3](#)
 virtual LAN (VLAN) [3-4, 3-64, 20-25](#)
 Virtual Private Network (VPN) [2-6, 2-15](#)
 virtual tie lines [3-63](#)
 VLAN
 access control list (ACL) [19-24](#)
 number of devices per VLAN [3-4](#)
 separate VLANs for voice and data [3-64](#)
 video [19-12](#)
 VLAN ID [20-25](#)
 voice [19-8, 19-12](#)
 VMO [13-3](#)
 voice
 bandwidth requirements [3-36](#)
 bearer traffic [3-51, 3-55](#)
 gateways [4-1, 20-2, 20-5](#)
 port integration [13-10, 13-12](#)
 termination [6-2](#)
 translation profiles [10-52](#)
 VLAN [19-8, 19-12](#)
 voice/WAN interface card (VWIC) [20-2](#)
 voice-activated conference view [16-14](#)
 voice activity detection (VAD) [4-4, 4-26, 8-13, 16-14](#)

Voice-Adaptive Fragmentation (VAF) [3-37](#)
 Voice-Adaptive Traffic Shaping (VATS) [3-39](#)
 Voice and Video Enabled IPSec VPN (V3PN) [2-6, 2-15](#)
 voice interface card (VIC) [20-2, 20-3](#)
 voicemail
 centralized [12-6](#)
 Cisco Unity [13-1](#)
 dial plan [10-60, 10-65, 10-72](#)
 dual PBX integration [12-5, 12-6](#)
 for local failover [2-25](#)
 integration with IP telephony system [12-1](#)
 positive disconnect supervision [12-11](#)
 SIP trunks [13-9](#)
 third-party systems [12-1, 12-10, 12-11](#)
 unified messaging [13-1](#)
 with Mobile Connect [23-7](#)
 voice over IP (VoIP) [3-51](#)
 voice over the PSTN (VoPSTN) [2-10](#)
voice rtp send-recv command [11-12](#)
 VoiceXML (VXML) [23-16, 23-17](#)
 VoIP [3-51](#)
 VoPSTN [2-10](#)
 VPN [2-6, 2-15](#)
 VWIC [20-2](#)
 VXML [23-16, 23-17](#)

W

Wait for Far-End to Send TCS [16-10](#)
 WAN
 aggregation router [3-3](#)
 infrastructure [3-30](#)
 web
 access from IP phone [19-9](#)
 applications [15-7](#)
 server [14-15](#)
 WebDialer [1-7, 22-13, 22-39](#)
 weighted fair queuing [3-34](#)
 WEP [3-68, 3-69, 20-14](#)

Wi-Fi Multimedia (WMM) [3-71](#)
 Wi-Fi Multimedia Traffic Specification (WMM TSPEC) [3-72, 20-17, 20-18](#)
 Wi-Fi Protected Access (WPA) [3-69, 20-14](#)
 Wi-Fi Protected Access 2 (WPA2) [3-69, 20-14](#)
 Wi-Fi Protected Access 2 Pre-Shared Key (WPA2-PSK) [20-15](#)
 Wi-Fi Protected Access Pre-Shared Key (WPA-PSK) [20-15](#)
 wildcard route pattern [10-17, 10-18](#)
 Windows Internet Naming Service (WINS) [3-14](#)
 WINS [3-14](#)
 Wired Equivalent Privacy (WEP) [3-69, 20-14](#)
 wireless
 endpoints [20-13](#)
 IP Phone 7920 [16-48, 20-13](#)
 IP Phone 7921G [20-13](#)
 IP phones [1-6, 20-13, 20-33](#)
 LAN [3-63](#)
 networking solutions [16-48](#)
 wireless LAN (WLAN) [3-63](#)
 Wireless LAN Services Module (WLSM) [20-16](#)
 WLAN infrastructure [3-63](#)
 WLSM [20-16](#)
 WMM [3-71](#)
 WMM TSPEC [3-72, 20-17, 20-18](#)
 WPA [3-69, 20-14](#)
 WPA2 [3-69, 20-14](#)
 WPA2-PSK [20-15](#)
 WPA-PSK [20-15](#)
 WS-SVC-CMM-ACT module [6-10, 6-12, 6-17](#)
 WS-X6608-E1 module [6-10, 6-13, 6-18](#)
 WS-X6608-T1 module [6-10, 6-13, 6-18](#)
 WS-X6624-FXS analog interface module [20-5](#)
 WS-X6624 module [12-2, 12-3](#)

X

XML services [16-48](#)

Z

zones

- clients [16-34](#)
- configuration on gatekeeper [16-34](#)
- for gatekeepers [9-14](#)
- H.320 gateways [16-38](#)
- MCU [16-36](#)
- prefixes [9-34](#), [16-36](#), [16-38](#), [16-40](#)
- subnets [16-40](#)

