Cisco Preferred Architecture for Enterprise Collaboration 11.0

Cisco Validated Design (CVD) Guide

Revised: November 20, 2015

Cisco Systems, Inc.
www.cisco.com

Cisco has more than 200 offices worldwide.
Addresses, phone numbers, and fax numbers
are listed on the Cisco website at
www.cisco.com/go/offices.

First Published: October 28, 2014

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

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

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

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

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

CONTENTS

Preface

Revised: November 20, 2015

Cisco Validated Designs (CVDs) explain important design and deployment decisions based on common use cases and current system releases. They incorporate a broad set of technologies, features, and applications to address customer needs. Cisco engineers have comprehensively tested and documented the guidelines within the CVDs in order to provide faster, more reliable, and fully predictable deployment. CVDs provide a tested starting point for Cisco partners and customers to begin designing and deploying systems using their own setup and configuration.

Documentation for Enterprise Collaboration

Cisco Preferred Architecture (PA) Design Overview guides help customers and sales teams select the appropriate architecture based on an organization's business requirements; understand the products that are used within the architecture; and obtain general design best practices. These guides support sales processes.

Cisco Validated Design (CVD) guides provide detailed steps for deploying the Cisco Preferred Architectures. These guides support planning, design, and implementation of the Preferred Architectures.

Cisco Collaboration System 11.*x* Solution Reference Network Design (SRND) guide provides detailed design options for Cisco Collaboration. The SRND should be referenced when design requirements are outside the scope of Cisco Preferred Architectures.

About This Guide

This Cisco Validated Design guide for the Cisco Enterprise Collaboration Preferred Architecture is for:

- Sales teams that sell, design, and deploy collaboration solutions
- Customers and sales teams who want detailed design best practices and ordered steps for deploying Cisco Collaboration

Readers of this guide should have a general knowledge of Cisco voice, video, and collaboration products and a basic understanding of how to deploy these products. We recommend that readers review the Cisco Preferred Architecture for Enterprise Collaboration 11.*x* Design Overview before reading this CVD document.

The design decisions within this CVD are in line with the framework outlined in the Cisco Collaboration SRND. While the SRND offers many design and deployment options, in this document a single deployment recommendation is selected based on fundamental assumptions for the Preferred Architecture design. Different assumptions can certainly lead to different design decisions, which then should be validated against the SRND. For large deployments with unique needs and advanced customization, it is recommended to work with your Cisco Account Manager for guidance beyond that contained in this CVD or the SRND.

This guide simplifies the design and sales process by:

- Building upon the product and design recommendations of the Cisco Preferred Architecture for Enterprise Collaboration 11.*x* Design Overview
- Cisco Preferred Architecture Design Overview
- Detailing a collaboration architecture, identifying best practices, and explaining the reasoning behind those recommendations

This CVD guide is organized into the following discrete modules that integrate together to form the overall Collaboration solution:

- Call Control — Explains fundamental concepts of dial plan design, Computer Telephony Integration (CTI), Survivable Remote Site Telephony (SRST), IM and Presence, LDAP directory integration, SIP trunks, and other aspects of call control. This chapter also lists the best practices for deploying call control in the Enterprise Collaboration Preferred Architecture.
- Conferencing — Describes the types of conferences available in the Enterprise Collaboration Preferred Architecture and explains how to deploy conferencing capability.
- Collaboration Edge — Explains how to deploy Cisco Collaboration Edge components to provide remote registration services, external communications, and interoperability.
- Core Applications — Lists the various applications and deployment tools available in the Enterprise Collaboration Preferred Architecture, and focuses on two core applications for unified messaging and conference scheduling.
- Sizing — Provides simplified sizing examples to size the components of the Enterprise Collaboration Preferred Architecture to fit the requirements of your deployment.

Revision History

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

http://www.cisco.com/go/cvd/collaboration

Visit the above website periodically and check for documentation updates by comparing the revision date of your copy with the revision date of the online document.

Table 1 lists the revision history for this document.

Table 1 Revision History for This CVD Guide

Revision Date	Comments
November 20, 2015	This document was updated for Cisco Collaboration System Release (CSR) 11.0. For details, in each chapter see *What's New in This Chapter.*
October 28, 2014	Initial release of this document.

Obtaining Documentation and Submitting a Service Request

For information on obtaining documentation, using the Cisco Bug Search Tool (BST), submitting a service request, and gathering additional information, see *What's New in Cisco Product Documentation* at: http://www.cisco.com/c/en/us/td/docs/general/whatsnew/whatsnew.html.

Subscribe to *What's New in Cisco Product Documentation*, which lists all new and revised Cisco technical documentation, as an RSS feed and deliver content directly to your desktop using a reader application. The RSS feeds are a free service.

Conventions

This document uses the following conventions:

bold font	Commands and keywords and user-entered text appear in **bold** font.
italic font	Document titles, new or emphasized terms, and arguments for which you supply values are in *italic* font.
[ ]	Elements in square brackets are optional.
{x | y | z }	Required alternative keywords are grouped in braces and separated by vertical bars.
[x | y | z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.
`courier` font	Terminal sessions and information the system displays appear in `courier` font.
< >	Nonprinting characters such as passwords are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

Note Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the manual.

Tip Means *the following information will help you solve a problem*. The tips information might not be troubleshooting or even an action, but could be useful information, similar to a Timesaver.

Caution Means *reader be careful*. In this situation, you might perform an action that could result in equipment damage or loss of data.

Timesaver Means *the described action saves time*. You can save time by performing the action described in the paragraph.

Warning **IMPORTANT SAFETY INSTRUCTIONS**

This warning symbol means danger. You are in a situation that could cause bodily injury. Before you work on any equipment, be aware of the hazards involved with electrical circuitry and be familiar with standard practices for preventing accidents. Use the statement number provided at the end of each warning to locate its translation in the translated safety warnings that accompanied this device.

SAVE THESE INSTRUCTIONS

Warning **Statements using this symbol are provided for additional information and to comply with regulatory and customer requirements.**

Introduction

Revised: November 20, 2015

In recent years, many new collaborative tools have been introduced to the market, enabling businesses to enhance communications and extend collaboration outside the walls of their businesses. Organizations realize the added value that collaboration applications bring to their businesses through increased employee productivity and enhanced customer relationships. Significant advances have been made in the collaboration space to simplify deployment, improve interoperability, and enhance the overall user experience.

Today's collaboration solutions offer organizations the ability to integrate video, audio, and web participants into a single, unified meeting experience. The guidelines within this Cisco Validated Design (CVD) guide are written with the overall collaboration architecture in mind. Subsystems are used for better organization of the content, and the recommendations within them are tested to ensure they align with recommendations in related subsystems.

What's New in This Chapter

Table 1-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

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

New or Revised Topic	Described in:	Revision Date
Added bandwidth management information	Table 1-2	November 20, 2015
Removed Cisco EX Series endpoints, and replaced Cisco IP Phone 7821 with 7811 model	Table 1-3	November 20, 2015

Architectural Overview

This CVD for the Enterprise Collaboration Preferred Architecture incorporates a subset of products from the total Cisco Collaboration portfolio that is best suited for the enterprise market segment. This Preferred Architecture deployment model is prescriptive, out-of-the-box, and built to scale with an organization as its business needs change. This prescriptive approach simplifies the integration of multiple system-level components while also enabling an organization to select the features, services, and capacities that best address its business needs.

This CVD for the Enterprise Collaboration Preferred Architecture provides end-to-end collaboration targeted for deployments larger than 1,000 users. For smaller deployments, consult the Preferred Architecture Design Overview and CVDs for Midmarket Collaboration.

This CVD for the Enterprise Collaboration Preferred Architecture provides high availability for critical applications. The architecture supports an advanced set of collaboration services that extend to mobile workers, partners, and customers through the following key services:

- Voice communications
- Instant messaging and presence
- High definition video and content sharing
- Rich media conferencing
- Enablement of mobile and remote workers
- Business-to-business voice and video communications
- Unified voice messaging

Because of the adaptable nature of Cisco endpoints and their support for IP networks, this architecture enables an organization to use its current data network to support both voice and video calls. The preferred architecture employs a holistic approach to bandwidth management that incorporates an end-to-end QoS architecture, call admission control, and video rate adaptation and resiliency mechanisms to provide the best possible user experience for deploying pervasive video over managed and unmanaged networks.

The Cisco Preferred Architecture for Enterprise Collaboration, shown in Figure 1-1, provides highly available and secure centralized services. These services extend easily to remote offices and mobile workers, providing availability of critical services even if communication with headquarters is lost. This should be viewed as a fundamental architecture from which to design a new deployment or to evolve an existing one. As the Preferred Architecture progresses, this architecture will be expanded upon with additional products and solutions.

Figure 1-1 Cisco Preferred Architecture for Enterprise Collaboration

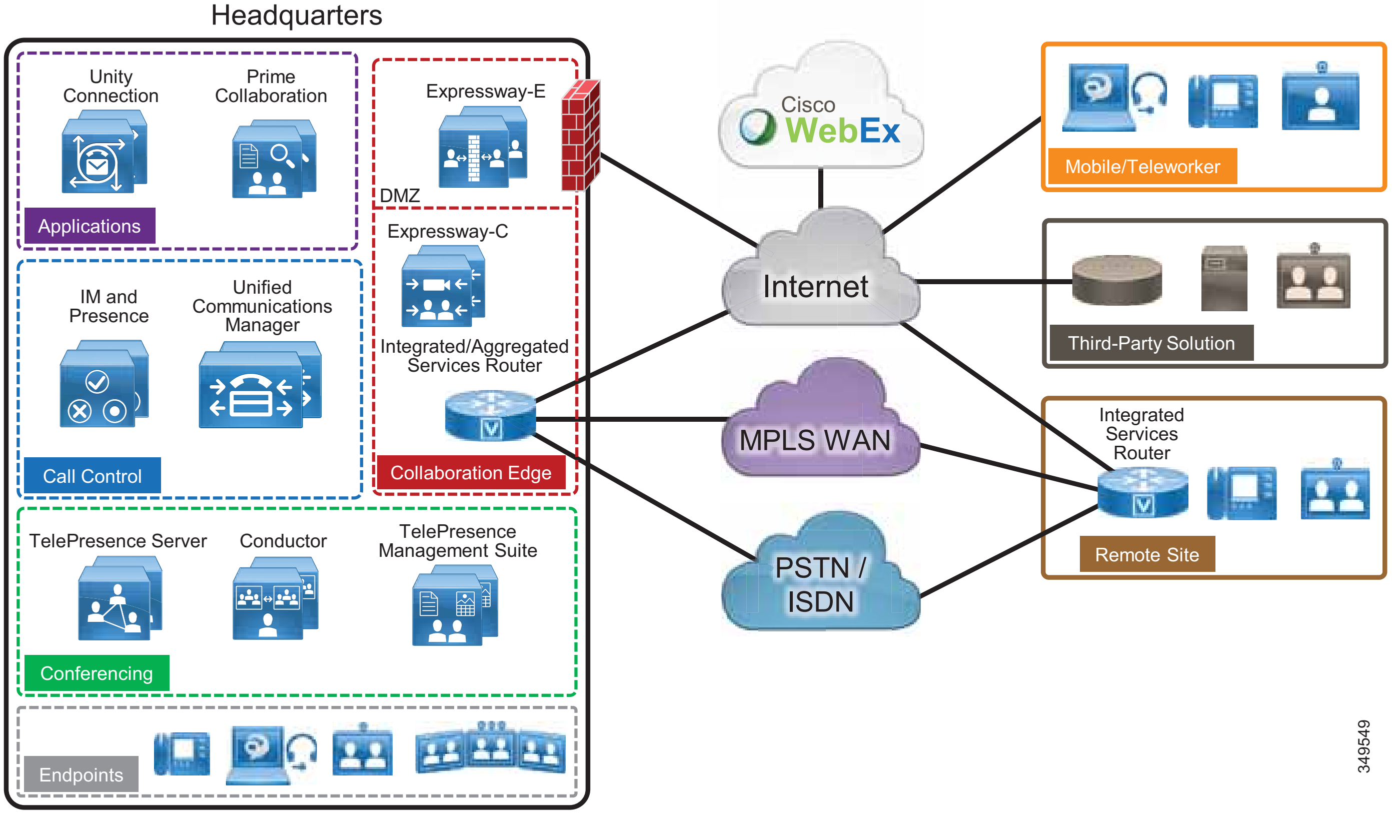


Table 1-2 lists the products in this architecture. For simplicity, products are grouped into modules to help categorize and define their roles. The content of this CVD is organized into the same modules.

Table 1-2 Components of the Cisco Preferred Architecture for Enterprise Collaboration

Module	Component(s)	Purpose
Call Control	Cisco Unified Communications Manager (Unified CM) Cisco Unified Communications Manager IM and Presence Service Cisco Integrated Services Router (ISR) G2/G3	Call control provides registration, call processing, resource management and instant messaging and presence for users and endpoints. It also encompasses remote site survivability for remote offices.
Conferencing	Cisco TelePresence Conductor Cisco TelePresence Server Cisco TelePresence Management Suite (TMS) Cisco WebEx Software as a Service (Cloud) Cisco WebEx Meetings Server (On-premises)	Conferencing allows three or more parties to communicate via voice, video, and content sharing in real time. Resources can be either on-premises or hosted in the cloud, or both.
Collaboration Edge	Cisco Expressway-C Cisco Expressway-E Cisco Integrated Services Router (ISR) G2/G3 Cisco Aggregation Services Routers (ASR)	Collaboration Edge provides remote registration services, external communications, and interoperability.
Core Applications	Cisco Unity Connection Cisco Prime Collaboration Deployment Cisco Prime License Manager	Applications cover a wide range of services including voice messaging and management.
Bandwidth Management	Network infrastructure and products from all chapters of this document	Bandwidth management incorporates an end-to-end QoS architecture, call admission control, and video rate adaptation and resiliency mechanisms to provide the best possible user experience for deploying pervasive video over managed and unmanaged networks.
Sizing	Products from all chapters of this document Virtual Machine Placement Tool (VMPT)	Sizing for all modules that are covered in this document, as well as a virtual machine placement example.

Network Services

The Preferred Architecture for Enterprise Collaboration requires 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). A detailed description of how these basic network services are utilized by Cisco applications and endpoints can be found in the Network Services section of the Cisco Collaboration SRND.

Virtualization

Virtualizing multiple applications and consolidating them on physical servers lowers cost, minimizes rack space, lowers power requirements, and simplifies deployment and management. Virtualization also accommodates redeploying hardware and scaling software applications as organizational needs change.

Cisco Unified Communications on the Cisco Unified Computing System (UCS)

Cisco UCS servers are thoroughly tested with unified communications (UC) core applications to provide reliable and consistent performance in a virtualized environment. There are two options for deploying UC applications on UCS servers:

- UC on UCS Tested Reference Configurations (TRCs)

  UCS TRCs are specific hardware configurations of UCS server components. These components include CPU, memory, hard disks (in the case of local storage), RAID controllers, and power supplies. Specific TRCs are documented at the UC Virtualization Supported Hardware website.

- UC on UCS Spec-Based

  Specifications-based UCS hardware configurations are not explicitly validated with UC applications. Therefore, no prediction or assurance of UC application virtual machine performance is made when the applications are installed on UCS specs-based hardware. In those cases Cisco provides guidance only, and ownership of assuring that the pre-sales hardware design provides the performance required by UC applications is the responsibility of the customer.

Both options are fully supported by the Cisco Technical Assistance Center (TAC), provided all rules for Unified Communications in a Virtualized Environment are followed.

Cisco Business Edition 7000 (BE7000)

The Cisco BE7000 is built on a virtualized UCS that ships ready-for-use with a pre-installed virtualization hypervisor and application installation files. The BE7000 is a UCS TRC in that UC applications have been explicitly tested on its specific UCS configuration. The Cisco BE7000 solution offers premium voice, video, messaging, instant messaging and presence, and contact center features on a single, integrated platform. For more information about the Cisco BE7000, see the Cisco Business Edition 7000 Data Sheet.

Core Applications

In the Preferred Architecture for Enterprise Collaboration, the following virtualized applications are deployed on multiple Cisco UCS servers to provide hardware and software redundancy:

- Cisco Unified Communications Manager
- Cisco Unified Communications Manager IM and Presence Service
- Cisco Unity Connection
- Cisco Expressway, consisting of Expressway-C and Expressway-E
- Cisco TelePresence Conductor
- Cisco TelePresence Server
- Cisco TelePresence Management Suite

We recommend always deploying redundant configurations to provide the highest availability for critical business applications.

Collaboration Endpoints

The recommendations within this CVD guide assume a deployment of Cisco voice and video endpoints, including soft clients such as Cisco Jabber. These endpoints use SIP to register to Cisco Unified Communications Manager (Unified CM). Table 1-3 lists the preferred endpoints for optimal features, functionality, and user experience.

Table 1-3 Cisco Collaboration Endpoints

Product	Description
Mobile: • Jabber for Android • Jabber for iPhone and iPad **Desktop:** • Jabber for Mac • Jabber for Windows	Soft client with integrated voice, video, voicemail, instant messaging, and presence functionality as well as secure edge traversal for mobile devices and personal computers
Cisco IP Phone 7811	Public space, single-line phone
Cisco IP Phone 8800 Series	General office use, multiple-line audio and video phones
Cisco IP Phone 8831	IP conference phone
Cisco DX Series	Personal TelePresence endpoint for the desktop
Cisco MX Series	TelePresence multipurpose room endpoint
Cisco SX Series	Integrator series TelePresence endpoint
Cisco IX Series	Immersive TelePresence room system

CHAPTER 2

Call Control

Revised: November 20, 2015

This chapter describes the call control function for the Cisco Preferred Architecture (PA) for Enterprise Collaboration.

Certain requirements might put your deployment outside the PA design guidelines and recommendations, in which case you might have to use other documentation such as the *Cisco Collaboration SRND* and related product documentation.

The first part of this chapter provides an architectural overview and introduces some fundamental design concepts, while the second part explains more detailed deployment considerations. The Architecture section discusses topics such as redundancy concepts, high availability, Computer Telephony Integration (CTI), and IM and presence architecture, and it introduces a hypothetical customer topology used in the examples throughout this document. The focus of this chapter is the Deployment Overview section. The deployment examples in that section will help you to understand the background of certain design decisions more clearly than an abstract discussion of concepts can. Topics covered in the Deployment Overview section include DNS requirements, cluster provisioning, certificate management, dial plan configuration, user provisioning using LDAP, media resources, SIP trunking considerations, endpoint provisioning, and multi-cluster considerations. The order of the topics in the Deployment Overview section follows the recommended configuration order.

What's New in This Chapter

Table 2-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 2-1 New or Changed Information Since the Previous Release of This Document

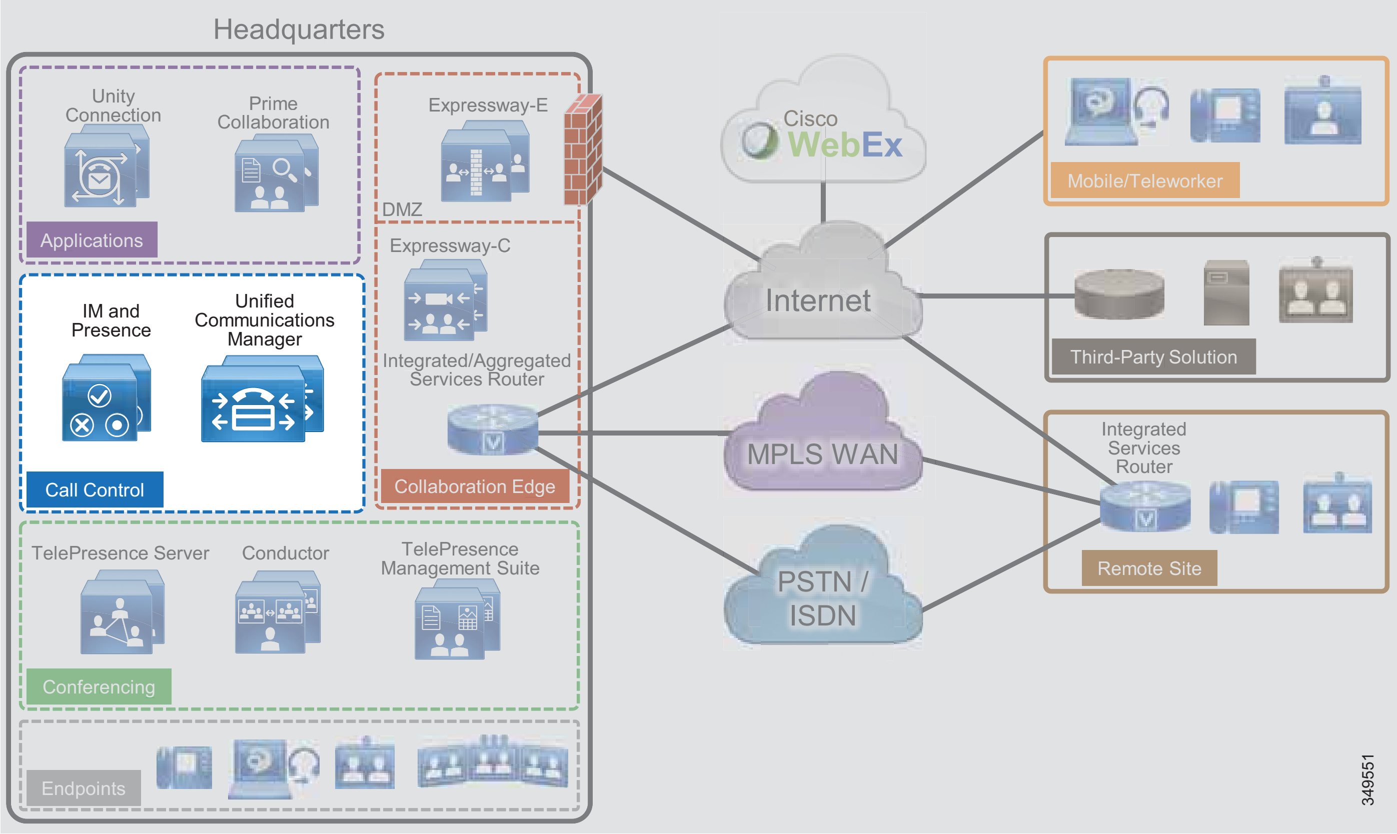
New or Revised Topic	Described in:	Revision Date
No new content has been added for this release, but some sections of this chapter have been reorganized.		November 20, 2015

Core Components

The core architecture contains these key elements (Figure 2-1):

- Cisco Unified Communications Manager
- Cisco Unified Communications Manager IM and Presence Service
- Cisco Integrated Services Router (ISR)

Figure 2-1 Preferred Architecture Overview



Key Benefits

- Call control is centralized at a single location that serves multiple remote sites.
- Management and administration are centralized.
- Common telephony features are available across voice and video endpoints.
- Single call control and a unified dial plan are provided for voice and video endpoints.
- Critical business applications are highly available and redundant.

Architecture

The handling and processing of voice and video calls is a critical function provided by enterprise communications systems. This functionality is handled by some type of call processing entity or agent. Given the critical nature of call processing operations, it is important to design unified communications deployments to ensure that call processing systems are scalable enough to handle the required number of users and devices and are resilient enough to handle various network and application outages or failures.

This chapter provides guidance for designing scalable and resilient call processing systems with Cisco Unified Communications Manager (Unified CM) and Survivable Remote Site Telephony (SRST). A centralized Unified CM cluster implements call processing services for all customer sites. Unified CM IM and Presences Service as part of the centralized Unified CM cluster implements instant messaging and presence services for the enterprise. Cisco Survivable Remote Site Telephony (SRST) is used to implement backup services for remotes sites when the corporate WAN reliability does not match the voice services availability requirements.

Cisco Unified CM provides call processing services for small to very large single-site deployments, multi-site centralized call processing deployments, and/or multi-site distributed call processing deployments. Unified CM is at the core of a Cisco Collaboration solution, and it serves as a foundation to deliver voice, video, TelePresence, IM and presence, messaging, mobility, web conferencing, and security.

Access to the enterprise collaboration network and to Unified CM from the Internet to enable remote access and business-to-business secure telepresence and video communications, is also available through various collaboration edge solutions such as VPN and Cisco Expressway.

Role of Unified CM

Cisco Unified CM is the central call control component in any Cisco collaboration deployment. Unified CM provides foundation services including call control, endpoint registration, endpoint configuration, call admission control, codec negotiation, trunk protocol translation, and CTI. Unified CM is the central point of administration and provisioning. All SIP trunks to other components – including conferencing media resources, gateways, and other components – are terminated on Unified CM so that Unified CM can orchestrate access to all of those components. Call routing is controlled by the dial plan configuration applied to Unified CM.

Role of IM and Presence Service

The Cisco Unified CM IM and Presence Service provides on-premises instant messaging and presence. It uses standards-based XMPP and also supports SIP for interoperability with SIP IM providers. Cisco Unified CM IM and Presence Service is an on-premise solution. The other Cisco instant messaging and presence service, Cisco WebEx Messenger, is a cloud-based service and is not covered in this document.

Role of SRST

When deploying Cisco desk phones in branch locations separated from a centralized call processing platform by a low-speed or unreliable WAN link, it is important to consider local call processing redundancy. By leveraging Survivable Remote Site Telephony (SRST) on a Cisco IOS router in each branch location, basic IP telephony services can be maintained for the desk phones if connectivity to the centralized call processing platform is lost. However, the set of available user-facing features is much smaller when a device is registered to SRST than when the phone is registered to Unified CM.

Unified CM Redundancy with Survivable Remote Site Telephony (SRST)

Cisco IOS SRST provides highly available call processing services for endpoints in locations remote from the Unified CM cluster. Unified CM clustering redundancy schemes provide a high level of redundancy for call processing and other application services within a LAN or MAN environment. However, for remote locations separated from the central Unified CM cluster by a WAN or other low-speed links, SRST can be used as a redundancy method to provide basic call processing services to these remote locations in the event of loss of network connectivity between the remote and central sites. We recommend deploying SRST-capable Cisco IOS routers at each remote site where call processing services are considered critical and need to be maintained in the event that connectivity to the Unified CM cluster is lost. Endpoints at these remote locations must be configured with an appropriate SRST reference within Unified CM so that the endpoint knows what address to use to connect to the SRST router for call processing services when connectivity to Unified CM subscribers is unavailable.

Unified CM and IM and Presence Service Clustering

Unified CM supports the concept of clustering. The Unified CM architecture enables a group of server nodes to work together as a single call processing entity. This grouping of server nodes is known as a cluster.

There are two types of Cisco Unified CM nodes: publisher and subscriber.

- Unified CM publisher

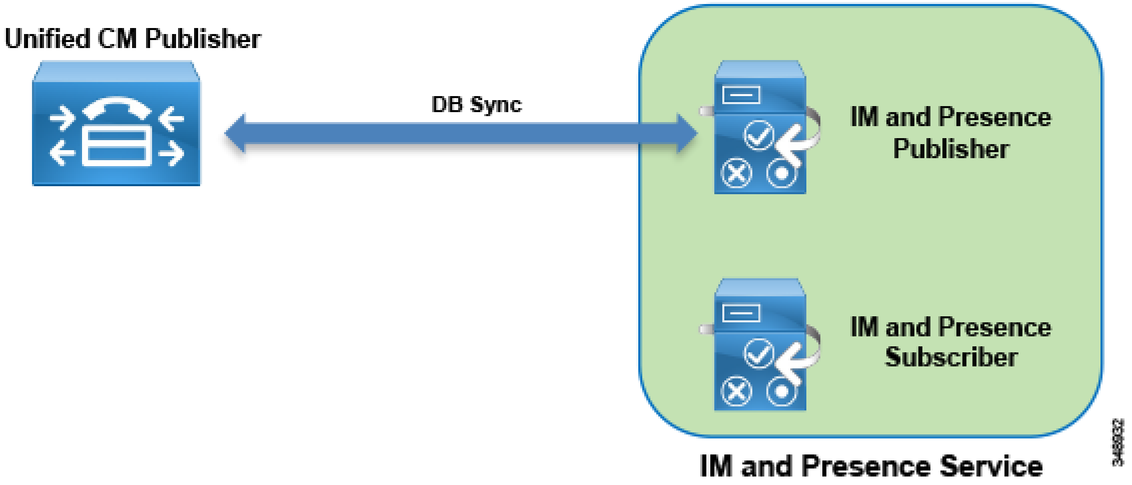
  The publisher is a required server node in all clusters. There can be only one publisher per cluster. This server node contains the cluster configuration, and it provides the database services to all other subscribers in the cluster. In this design, the Unified CM publisher is a dedicated node; it does not handle TFTP requests, endpoint registration, or call processing.

- Unified CM subscriber

  Subscriber nodes subscribe to the publisher to obtain a copy of the database information. Subscriber nodes include, for example, the Unified CM TFTP nodes and the Unified CM call processing subscriber nodes.

Cisco IM and Presence nodes have the same clustering concept. The first IM and Presence node is the IM and Presence publisher. The other IM and Presence nodes are the IM and Presence subscribers, and they obtain a copy of their database from the IM and Presence publisher. The IM and Presence publisher communicates with the Unified CM publisher and most of the IM and Presence configuration is actually done through the Unified CM publisher (for instance, the Unified CM users, the UC services available to presence users, and the service activation). Hence, all IM and Presence nodes, including the IM and Presence publisher, are considered subscribers of the larger Unified CM and IM and Presence Service cluster. Figure 2-2 shows the relationship between the Unified CM publisher and a two-node IM and Presence cluster.

Figure 2-2 Relationship Between Unified CM and a Two-Node IM and Presence Cluster



High Availability

Unified CM and IM and Presence nodes should be deployed in a highly available infrastructure. For example, the use of dual power supplies combined with the use of uninterruptible power supply (UPS) sources will provide maximum power availability. From a network perspective, the platform servers should be connected to multiple upstream switches.

Unified CM and IM and Presence systems also handle high availability at the application level.

With Unified CM in this design, two TFTP servers should be deployed for redundancy. The call processing nodes should be deployed with one-to-one (1:1) redundancy, where for every primary call processing subscriber there is a backup call processing subscriber. This 100%:0% redundancy design compared to a 50%:50% redundancy design has a number of advantages, including the reduction of Unified CM groups and device pools and simplified configuration and distribution of devices with fewer redundancy options.

Cisco IOS Survivable Remote Site Telephony (SRST) provides highly available call processing services for endpoints in locations remote from the Unified CM cluster when the WAN links are down.

Individual Cisco IM and Presence nodes are grouped in subclusters. A subcluster can have one or two nodes. Adding the second node in a subcluster provides high availability. High availability is recommended, and therefore in this design each subcluster consists of two nodes. A two-node subcluster allows for users associated with one server of the subcluster to use the other server in the subcluster automatically if a failover event occurs. We recommend balancing the user assignment between the two nodes in each pair. The IM and Presence publisher handles IM and Presence information from presence clients just like any other IM and Presence subscriber does, and it is deployed as one of the two nodes in an IM and Presence subcluster.

Computer Telephony Integration (CTI)

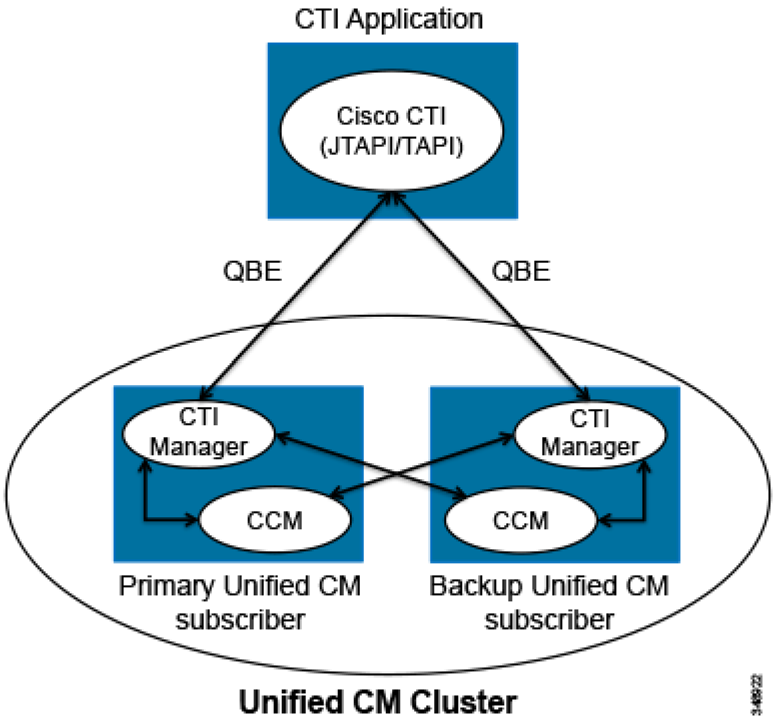
Cisco Computer Telephony Integration (CTI) extends the rich feature set available on Cisco Unified CM to third-party applications.

CTI Architecture

Cisco CTI consists of the following components (Figure 2-3), which interact to enable applications to take advantage of the telephony feature set available in Cisco Unified CM:

- CTI application — Cisco or third-party application written to provide specific telephony features and/or functionality. It can use a JTAPI or TAPI interface. The protocol between the CTI application and Unified CM is Quick Buffer Encoding (QBE).
- Unified CM subscriber with the following services:
  - CCM — The Cisco CallManager Service, the telephony processing engine.
  - CTI Manager (CTIM) — A service that runs on one or more Unified CM subscribers operating in primary/secondary mode and that authenticates and authorizes telephony applications to control and/or monitor Cisco IP devices.

Figure 2-3 CTI Architecture



High Availability for CTI

High availability for CTI Manager relies on the CTI application being able to connect to the backup CTI Manager Service in case the primary CTI Manager fails. In case both the CTI Manager and CCM services on the primary Unified CM subscriber fail (for example, if the entire primary Unified CM subscriber fails), then both CCM and CTI Manager services running on the backup Unified CM subscriber will become active, and the CTI Manager service will monitor and control the devices that are registered to the CCM service located on the same backup Unified CM subscriber. If the primary CTI Manager Service fails but the primary CCM Service is still running (assuming you have 1:1 redundancy with a distribution of 100%/0% on the primary/backup Unified CM subscribers), then all the devices will stay registered to the CCM Service running on the primary Unified CM subscriber, and the CTI Manager running on the backup Unified CM subscriber will become active and will monitor and control the CTI devices even though they are registered to a CCM service running on a different node (the primary Unified CM subscriber in this case).

Capacity Planning for CTI

Ensure the capacity limits are not exceeded for the three types of CTI resources:

- The maximum number of CTI applications connecting to a given CTI Manager instance (Unified CM node running the CTI Manager service). This number is typically low with CTI server-based application, but with CTI client-based applications such as Jabber clients in deskphone mode where each Jabber client is considered a CTI application, it is important to ensure the limit is not exceeded when deploying a large number of Jabber clients.
- The maximum number of CTI-enabled endpoints registered to a given Unified CM call processing subscriber.
- The maximum number of CTI-enabled endpoints monitored and controlled by a CTI Manager instance. Ideally, the CTI Manager service running on a Unified CM node monitors only the endpoints registered to that Unified CM node. But it is possible that a CTI Manager service also monitors endpoints registered to other Unified CM nodes.

The CTI limits are the same for all three CTI resources described above. The CTI capacity limits vary with the type of OVA template. If the CTI limit is reached, deploy another pair of Unified CM call processing nodes running the CTI Manager service.

IM and Presence Architecture

The Cisco Unified CM IM and Presence Service provides on-premises instant messaging and presence. The main presence component of the solution is the IM and Presence Service, which incorporates the Extensible Communications Platform (XCP) and supports SIP/SIMPLE and Extensible Messaging and Presence Protocol (XMPP) for collecting 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.

Applications (either Cisco or third-party) can integrate presence and provide services that improve the end user experience and efficiency. In addition, Cisco Jabber is a supported client of the IM and Presence Service that also integrates instant messaging and presence status.

The IM and Presence Service uses the same underlying appliance model and hardware used by Unified CM on the Cisco Unified Computing System (UCS) platform.

The IM and Presence Service is deployed as an IM and Presence cluster. The IM and Presence cluster consists of up to six nodes, including one designated as a publisher and up to five subscriber nodes. As discussed in the sections on Unified CM and IM and Presence Service Clustering and High Availability, the IM and Presence nodes are grouped in subclusters and each subcluster consists of two nodes for high availability. As discussed in the sizing section, a single subcluster can be deployed in order to support up to 15,000 users. The IM and Presence publisher handles IM and presence requests, just like the IM and Presence subscribers do, so the first subcluster consists of the IM and Presence publisher and one IM and Presence subscriber.

As discussed in the section on Unified CM and IM and Presence Service Clustering, the IM and Presence nodes are considered part of the larger Unified CM and IM and Presence Service cluster.

Deployment of the Unified CM and IM and Presence Service Cluster

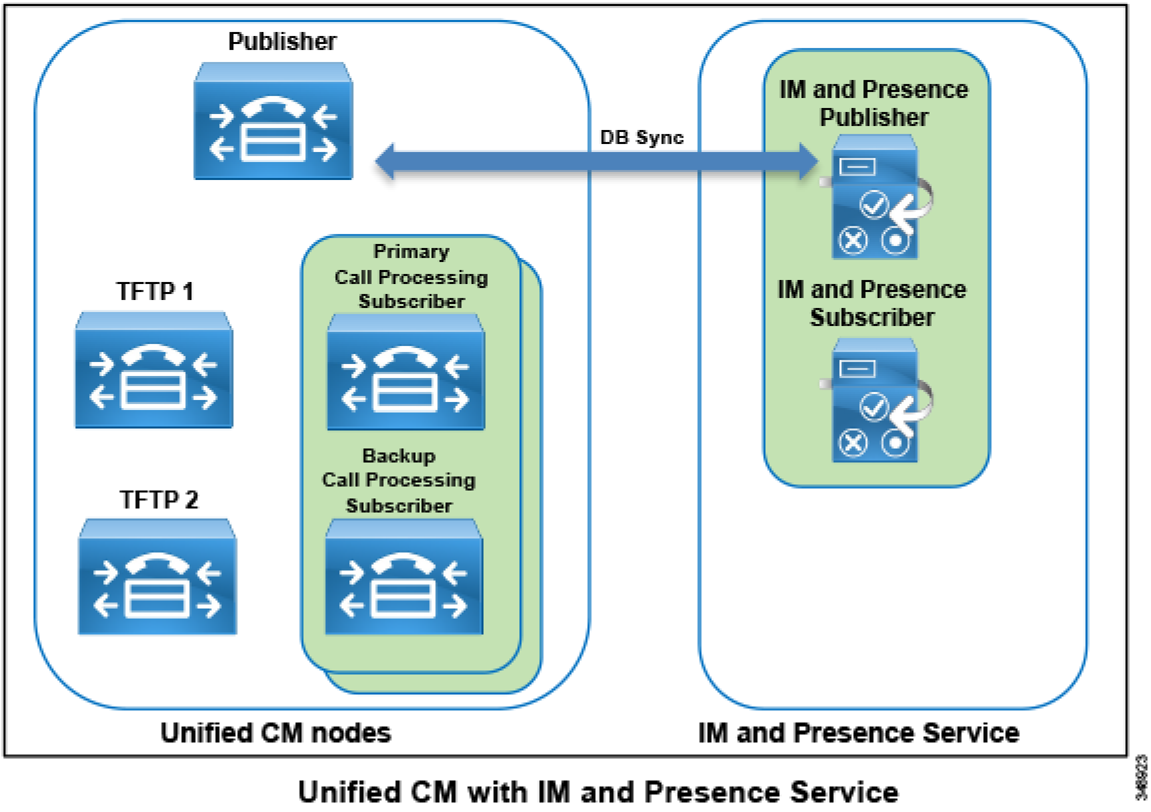
The Cisco Unified CM and IM and Presence Service cluster consists of the following nodes:

- 1x Cisco Unified CM publisher
- 2x (1 pair) Cisco Unified CM TFTP server subscribers
- 2x (1 pair) Cisco Unified CM call processing subscribers (Add additional pairs to scale.)
- 2x (1 pair) Cisco Unified IM and Presence nodes (Add additional pairs, or subclusters, to scale.)

The number of Unified CM call processing pairs and of IM and Presence pairs to add in order to scale is discussed in the chapter on Sizing.

Figure 2-4 shows an example of a Unified CM and IM and Presence Service cluster deployment with up to 10,000 devices and 10,000 users. For more sizing information, refer to the Sizing chapter.

Figure 2-4 Unified CM and IM and Presence Service Cluster Deployment


Publisher
DB Sync
IM and Presence Publisher
TFTP 1
Primary Call Processing Subscriber
IM and Presence Subscriber
TFTP 2
Backup Call Processing Subscriber
Unified CM nodes
IM and Presence Service
Unified CM with IM and Presence Service
348623

Endpoints

Jabber

Cisco Jabber clients provides core collaboration capabilities for voice, video, and instant messaging to users. Cisco Jabber is available on a wide variety of platforms including Windows, Mac, and mobile devices such as smartphones and tablets.

Cisco Jabber can be deployed in either of two modes:

- Full UC and Cisco Jabber for Everyone (IM only) Mode

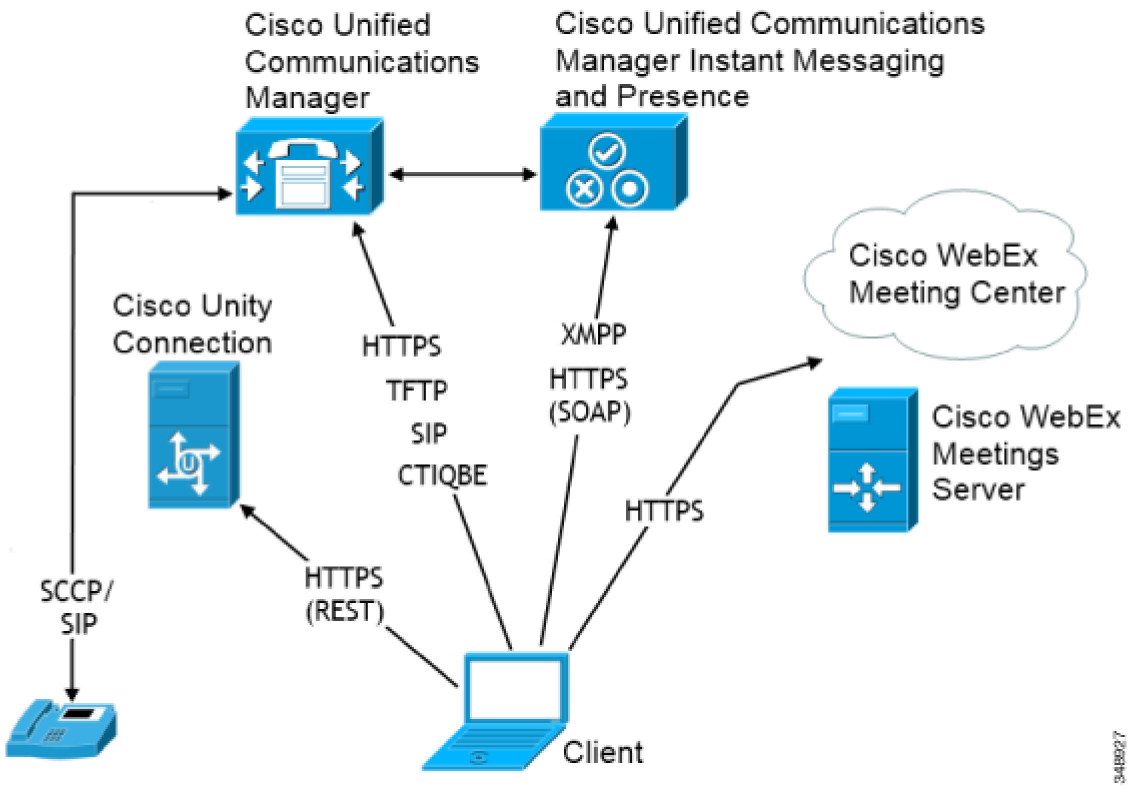
  This is the default mode. The user's primary authentication is to an IM and Presence server. This is the mode used in this Preferred Architecture design and cover in this document.

- Phone Mode

  In phone mode, the IM and Presence Service is not required.

Figure 2-5 illustrates the architecture of an on-premises deployment that includes Cisco Unified Communications Manager IM and Presence.

Figure 2-5 Cisco Unified Communications with IM and Presence Architecture


Cisco Unified Communications Manager
Cisco Unified Communications Manager Instant Messaging and Presence
Cisco WebEx Meeting Center
Cisco Unity Connection
HTTPS
TFTP
SIP
CTIQBE
XMPP
HTTPS (SOAP)
HTTPS
Cisco WebEx Meetings Server
SCCP/ SIP
HTTPS (REST)
Client
348927

To connect to services, Cisco Jabber requires the following information:

- Source of authentication that enables users to sign in to the client

  In full UC or IM-only modes, the source of authentication is the IM and Presence service. In phone-only mode, it is Unified CM.

- Location of services

  The services include IM and Presence, directory, CTI, voicemail, and conferencing.

To provide this information to the client, we recommend using the Service Discovery method over the Manual Connection method. With the Service Discovery method, the client automatically locates and connects to services.

In this design, the client automatically discovers services and configuration with the SRV record _cisco-uds that is retrieved when the user first enters his or her email address in the Jabber client.

The Jabber Contact Sources can be an LDAP contact Source with an Enhanced Directory Integration (EDI) for the Microsoft Windows desktops or a Basic Directory Integration (BDI) for other platforms such as OS X, iOS, or Android. Another source for the contacts can be the Unified CM User Data Service (UDS), but that will reduce the number of users supported on Unified CM.

Multi-Cluster Considerations

In a multi-cluster deployment, interconnect all the individual Unified CM clusters through SIP trunks. To avoid session traversal through individual clusters, deploy a full mesh of SIP trunks. With four or more clusters, deploy Cisco Unified CM Session Management Edition (SME) to centralize the dial plan and trunking and to avoid the complexity of a full-mesh SIP trunk topology. Cisco Unified CM SME is not covered in this document. For more information about SME, refer to the *Cisco Collaboration SRND*.

In multi-cluster deployments, use Global Dial Plan Replication (GDPR) to replicate dial plan information between clusters. GDPR can advertise a +E.164 number, one Enterprise Significant Number (ESN), and up to five alpha-numeric URIs per directory number. An ESN is the abbreviated inter-site dialing equivalent of a directory number. The information advertised and learned through GDPR enables deterministic intercluster routing for these dialing habits:

- +E.164 dialing based on the advertised +E.164 numbers
- Enterprise abbreviated inter-site dialing based on the advertised ESNs
- Alpha-numeric URI dialing based on the advertised URIs

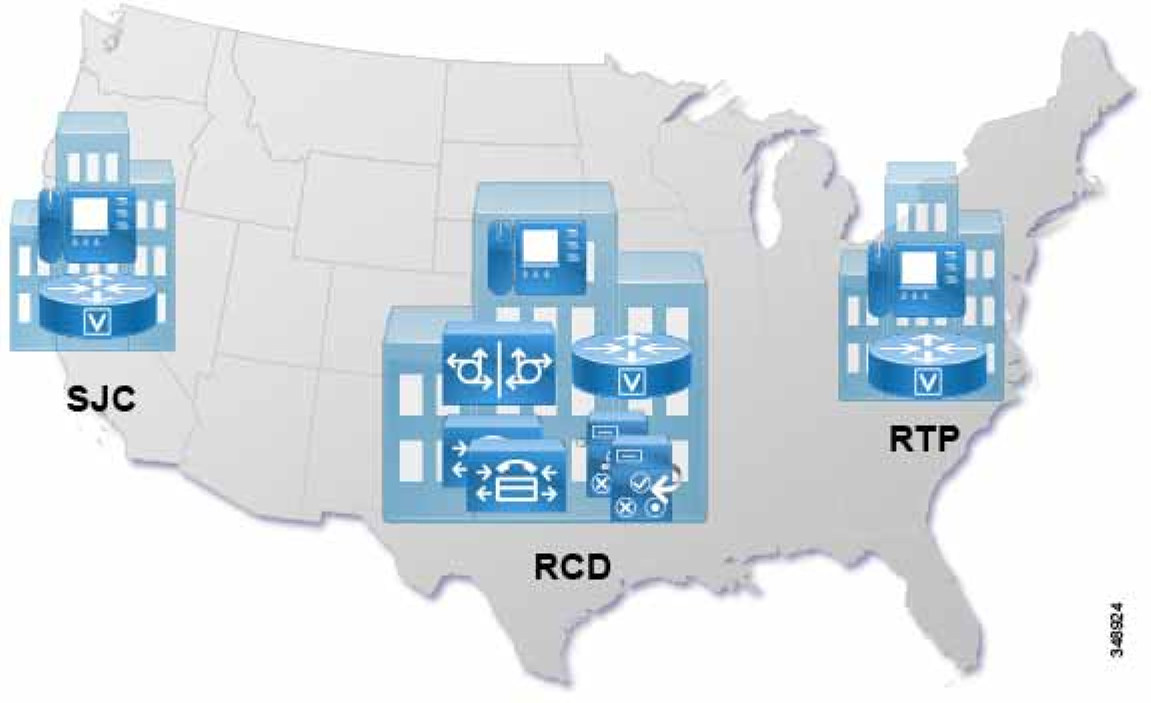
GDPR uses Intercluster Lookup Service (ILS) as the transport medium, therefore setting up ILS between all Unified CM clusters is required for multi-cluster deployments. In addition to GDPR, UDS-based service discovery used by Jabber also relies on the ILS exchange to detect the existence of UDS nodes on remote clusters to which /cucm-uds/homeCluster requests of non-local users can be forwarded to determine the home cluster of a user trying to log in to Jabber.

IM and Presence functionality is limited by having communications within a single cluster. To extend presence and instant messaging capability and functionality, these standalone clusters can be configured for peer relationships for communication between clusters within the same domain. This functionality provides the ability for users in one cluster to communicate and subscribe to the presence of users in a different cluster within the same domain. To create a fully meshed presence topology, each Cisco IM and Presence cluster requires a separate peer relationship for each of the other Cisco IM and Presence clusters within the same domain. The intercluster peer is configured as the IP address of the remote Unified CM cluster IM and Presence publisher node.

Topology Example

For the purpose of this document, we assume a centralized call processing deployment serving three sites in the US: SJC, RCD, and RTP. The Unified CM and IM and Presence Service servers are centrally located in RCD. Central PSTN access is in RCD as well. SJC and RTP are assumed to be small sites, with Survivable Remote Site Telephony (SRST) configured locally, with local PSTN access when the WAN connectivity to the RCD site is down. Figure 2-6 illustrates this topology example.

Figure 2-6 Example Topology



The topology example used in this document for multi-cluster considerations is a two-cluster deployment: the cluster in the United States as shown in Figure 2-6, and a second cluster to cover Europe, the Middle East, and Africa (EMEA).

Certificate Considerations

Whenever a certificate needs to be checked during session establishment, either between two servers or between a central service and a client application such as Cisco Jabber for Windows, the certificate must pass the following checks:

- Validity — The current time and date must be within the certificate's validity range.
- Trust — The certificate must be trusted. A certificate is considered trusted if the signing (issuing) party is trusted. Trust with signing parties typically is established by importing the certificate of the signing party into a store of trusted certificates.
- Identity — The subject or identity for which the certificate is issued must match the identity that the initiator of the session intended to reach.

By default all certificates used by Unified CM and IM and Presence are self-signed certificates. While the validity and identity aspect of the above checks does not create any special problems with the self-signed certificates, the trust aspect is an issue. To establish trust with a service based on a self-signed certificate, the self-signed certificate must be imported into the trusted certificates store of all entities requiring secure connections to the service. This can be handled if the set of communicating parties is small, but it becomes more difficult for large numbers of communication peers (for example, client applications such as Jabber).

If certificate validation fails on Jabber clients, then the user is prompted and can accept the certificate, which then is added to the trusted certificate store. This should be avoided because being prompted multiple times to accept a number of certificates during startup of the client is not the best user experience. Even more important is the fact that most users will not actually verify whether the presented certificate is correct by checking the certificate's fingerprint, and instead will just accept any certificate. This breaks the security concept of certificate-based authentication for secure session establishment.

For these reasons, the recommended deployment of Cisco Jabber clients requires that certificate validation during startup of the clients must not fail. This can be achieved in either of two ways:

- Use self-signed certificates and pre-distribute all required self-signed certificates to the devices' certificate stores.

  In Windows environments certificates can be added to devices' certificate stores via Microsoft group policies.

- Use certificates issued by a certificate authority (CA).

  In this case the self-signed certificates used by the infrastructure services are replaced by certificates issued and signed by a trusted CA. To establish trust with the CA, the CA's root certificate is added to the trusted certificates store of all clients. By default most client devices include all of the major public CA root certificates in their trusted certificate store.

The second option is the recommended approach because it allows issue of new service certificates without having to update all client trusted certificate stores as long as the signing CA's root certificate has already been added to the trusted certificates stores of all clients.

DNS Considerations

As explained in the previous section, the identity of server certificates presented during connection setup is validated. This implies that clients need to initiate connections based on fully qualified domain names (FQDNs) so that the subject in the presented certificate really can be checked against the identity to which the client intends to connect. The use of FQDNs for connection initiation implies that DNS is a fundamental requirement. The enterprise DNS needs to be set up so that name resolution is reliably available for all clients and servers in the network. In addition to providing reliable FQDN-to-IP-address (and reverse) resolution, DNS also is required for the automatic service discovery process used by Jabber clients.

During startup, Jabber clients locate the UDS service required for UDS-based service discovery by trying to resolve the _cisco-uds._tcp SRV using DNS. For best redundancy and load balancing, we recommend provisioning DNS SRV records with equal priority and weight for the Unified CM publisher and TFTP nodes.

Endpoint Addressing

All directory numbers on endpoints with DID address are provisioned as +E.164 numbers. The benefits of this approach include the following:

- +E.164 directory numbers are unique by definition.
- +E.164 directory numbers enable one dialing habit (+E.164) directly without requiring any further dial plan configuration.
- +E.164 directory numbers simplify the implementation of forced on-net routing.
- Configuration of Automated Alternate Routing (AAR) is greatly simplified. There is no need to provision multiple AAR groups and AAR PSTN prefixes because the target on-net destination can be used directly as an alternate PSTN address; it is a +E.164 number.
- Correct caller ID is automatically achieved for all call flows (direct, forwarded, on-net, and off-net).

Unique addresses are also required for endpoints without an associated DID (for example, lobby phones) and enterprise services (for example, call pickup, call park, and so forth). Since no +E.164 number exists for them, we recommend the use of an alternate enterprise specific numbering (ESN) schema to address them. The recommended format for the ESN schema is an access code chosen so that no overlap between ESN dialing and other dialing habits is created, followed by a site code and the intra-site extension. The length of the site code and extension is a trade-off between providing a large enough number space and keeping the ESN dialing as short as possible.

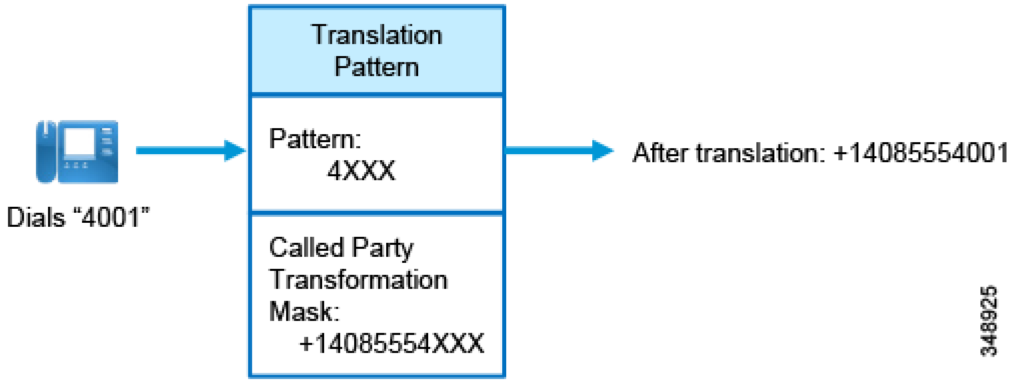
+E.164 Routing and Dialing Normalization

To achieve the intended forced on-net routing (calls to any on-net destination dialed using any of the supported numeric dialing habits has to be routed on-net), the recommended dial plan design uses a two-step routing approach. In the first step, the dialed digit string is normalized to +E.164, if possible (calls to non-DIDs obviously cannot be normalized to +E.164), and then in the second step the resulting +E.164 digit string is matched against a +E.164 numeric plan that includes directory numbers and route patterns.

The dialing normalization is achieved by provisioning translation patterns matching on the non+E.164 dial strings, and then the dialed string is transformed to +E.164 through the called party transformations on the translation patterns.

Figure 2-7 shows an example of a dialing normalization translation pattern that can be used to normalize abbreviated intra-site dialing in SJC to the full +E.164 number of the dialed destination. If a user in site SJC dials 4001, this dialed string is matched by a translation pattern 4XXX; and the called party transformation mask configured on the translation pattern, when applied to 4001, creates the resulting digit string +14085554001, which then can be routed in a +E.164 routing schema.

Figure 2-7 Example Dialing Normalization Translation Pattern


Translation Pattern
Dials "4001"
Pattern:
4XXX
Called Party Transformation Mask:
+14085554XXX
After translation: +14085554001
348925

After applying the called party transformations defined on a translation pattern, Unified CM then executes a secondary lookup of the resulting digit string using the calling search space (CSS) defined on the translation pattern. Unified CM enables definition of translation patterns that use the originator's CSS for this secondary lookup. This allows definition of dialing normalization translation patterns that can be reused in multiple context, because after applying the dialing normalization, the secondary lookup of the normalized digit string is executed, not based on a single fixed CSS, but based on the CSS in effect when the translation pattern was engaged.

Tip On dialing normalization translation patterns, set the option **Use Originator's Calling Search Space** so that the CSS used for the secondary lookup is identical to the CSS used for the primary lookup.

Tip On dialing normalization translation patterns that are fixed length (they do not end with a variable length wildcard), also set the option **Do Not Wait For Interdigit Timeout On Subsequent Hops** so that even if the secondary lookup matches on a variable length route pattern, the call is still routed without inter-digit timeout.

Classes of Service and Calling Search Spaces (CSSs)

Partitions and CSSs are the fundamental components in Unified CM used to build classes of service. Dialable patterns are grouped into equivalence classes by putting patterns belonging to the same class into the same partition. Each CSS then is a list of partitions that defines which partitions and, thus, which patterns a calling entity using the CSS can access. A CSS effectively enforces class of service by determining which destinations can be reached from a device using this CSS.

The number of classes of service defined is the major factor driving dial plan complexity, and thus the number of required classes of service should be as small as possible. In a well designed enterprise dial plan, re-use of patterns and partitions for multiple classes of service helps to simplify the dial plan deployment.

Outbound Gateway Selection Using Local Route Group

Following the maxim to avoid and eliminate redundancies in the dial plan as much as possible, the concept of Local Route Groups (LRGs) is used to define the egress gateway selection.

Route patterns using a local route group offer a unique characteristic: they allow for dynamic selection of the egress gateway based on the device originating the call. By contrast, calls routed by route patterns using static route groups will route the call to the same gateway, no matter which device originated the call. Route patterns configured to refer to a route list that makes use of LRGs will resolve to the actual route group configured as the LRG in the calling party's device pool.

This allows for re-use of route patterns that are not specific to each site, instead of requiring you to provision site-specific route patterns that are directly associated with the egress gateway of the respective site.

Outbound Calls: Called and Calling Number Localization

The dial plan design presented in this document uses local route groups for egress gateway selection based on the calling device. Hence, calling and called party transformations required to adapt to service provider requirements cannot be done on the route pattern or route list level. These transformations would be shared among all gateways. Instead, these service provider specific transformations to localize calling and called party information are configured either on the gateway using Cisco IOS voice translation rules or on Unified CM using calling and called party transformation patterns addressed by calling and called party transformation CSSs configured on the gateway or on the gateway's device pool.

Inbound Calls: Called and Calling Number Globalization

Because all call routing on Unified CM is based on +E.164 numbers for all incoming calls arriving at Unified CM, we need to make sure that called party information is globalized to +E.164 from the format received on the link from the provider. This is achieved through a combination of Cisco IOS translations on the SIP gateways (required to avoid loss of number type information received from the ISDN network when sending the request to Unified CM over SIP) and prefixes and possibly calling and called number transforms configured on Unified CM.

User Provisioning with LDAP Synchronization

Synchronization of Unified CM with a corporate LDAP directory allows the administrator to provision users easily by mapping Unified CM data fields to directory attributes. Critical user data maintained in the LDAP store is copied into the appropriate corresponding fields in the Unified CM database on a scheduled basis. The corporate LDAP directory retains its status as the central repository. Unified CM has an integrated database for storing user data and a web interface within Unified CM Administration for creating and managing user accounts and data. When LDAP synchronization is enabled, the local Unified CM database is still used, and additional local end-user accounts can be created. Management of end-user accounts is then accomplished through the interface of the LDAP directory and Unified CM Administration.

User Authentication with LDAP

The LDAP authentication feature enables Unified CM to authenticate LDAP synchronized users against the corporate LDAP directory. Locally configured users are always authenticated against the local database. Also, PINs of all end users are always checked against the local database only.

To enable authentication, a single authentication agreement is defined for the entire cluster.

The following statements describe Unified CM's behavior when authentication is enabled:

- End user passwords of users imported from LDAP are authenticated against the corporate directory by a simple bind operation.
- End user passwords for local users are authenticated against the Unified CM database.
- Application user passwords are authenticated against the Unified CM database.
- End user PINs are authenticated against the Unified CM database.

In environments that employ a distributed Active Directory topology with multiple domain controllers geographically distributed, authentication speed might be unacceptable. When the Domain Controller for the authentication agreement does not contain a user account, a search must occur for that user across other domain controllers. If this configuration applies to your deployment, and login speed is unacceptable, it is possible to set the authentication configuration to use a Global Catalog Server.

An important restriction exists, however. A Global Catalog does not carry the employeeNumber attribute by default. In that case either use Domain Controllers for authentication (beware of the limitations listed above) or update the Global Catalog to include the employeeNumber attribute. Refer to Microsoft Active Directory documentation for details.

To enable queries against the Global Catalog, 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.

Deployment Overview

Deployment begins with provisioning of the centralized Cisco Unified CM cluster followed by further configuration and provisioning tasks. The following sections describe how to set up and configure the call control according to the Preferred Architecture design in this document:

- DNS Requirements
- Provision the Cisco Unified CM and IM and Presence Service Cluster
- Cisco Unified CM and IM and Presence Certificate Management
- Initial Cisco Unified CM Configuration
- Other IM and Presence Settings
- Dial Plan Configuration
- LDAP System Configuration
- Cisco Unified CM Group Configuration
- Phone NTP References
- Date and Time Groups
- Media Resources
- Device Pools
- SIP Trunks
- Endpoint Provisioning
- ILS Configuration for Multi-Cluster Deployments
- GDPR Configuration (Multi-Cluster Only)
- Survivable Remote Site Telephony (SRST) Deployment
- Extension Mobility
- Busy Line Field (BLF) Presence
- Deploying Computer Telephony Integration (CTI)

DNS Requirements

Before deploying the solution, make sure DNS resolution is available for all servers to be deployed. Both forward (from DNS name to IP address) and reverse (from IP address to DNS name) lookups have to be configured in the enterprise DNS.

In addition to enabling UDS-based service discovery for Jabber clients, provision DNS SRV records for all Unified CM publisher and TFTP subscriber nodes, defining these as service locations for _cisco-uds. Example 2-1 shows an example of DNS SRV records defining a number of Unified CM nodes as _cisco-uds service locations.

Example 2-1 DNS SRV Record for UDS-Based Service Discovery

```
_cisco-uds._tcp.ent-pa.com     SRV service location:
          priority       = 10
          weight         = 10
          port           = 8443
          srv hostname   = us-cm-pub.ent-pa.com
_cisco-uds._tcp.ent-pa.com     SRV service location:
          priority       = 10
          weight         = 10
          port           = 8443
          srv hostname   = us-cm-tftp1.ent-pa.com
_cisco-uds._tcp.ent-pa.com     SRV service location:
          priority       = 10
          weight         = 10
          port           = 8443
          srv hostname   = us-cm-tftp2.ent-pa.com
```

In Example 2-1, three Unified CM nodes (publisher and two TFTP subscriber nodes) are defined as service locations for UDS service discovery to make sure that the load of the initial UDS requests from Jabber clients making use of UDS service discovery are evenly distributed among all active Unified CM nodes.

As part of the UDS service discovery process, after locating the home cluster using the /cucm uds/clusterUser resource, Jabber clients will use the /cucm-uds/servers resource to get a list of all UDS nodes in the user's home cluster, so that the actual UDS requests during the registration process are load balanced between all UDS nodes of the cluster even if the SRV records defined only the publishers as service locations.

Provision the Cisco Unified CM and IM and Presence Service Cluster

To deploy the Unified CM and IM and Presence Service cluster, perform the following tasks:

1. Determine the number of required call processing subscriber pairs based on the target number of users and devices.
2. Determine the number of required IM and Presence nodes based on the target number of users.
3. Determine the network parameters (DNS names, IP addresses, and so forth) for all required cluster members. Make sure to consider the TFTP servers also.
4. Deploy the required number of virtual machines on your compute infrastructure using the appropriate Cisco provided OVA template files. For information on how to obtain these OVA files, refer to the documentation at

   http://docwiki.cisco.com/wiki/Downloading_OVA_Templates_for_UC_Applications
5. In Cisco Prime Collaboration Deployment, define the Unified CM cluster with all its members, and map the nodes to the virtual machines created in task 4.
6. Deploy all nodes using Cisco Prime Collaboration Deployment.

For more information on how to provision a cluster using Cisco Prime Collaboration Deployment, refer to the *Cisco Prime Collaboration Deployment Administration Guide*, available at

http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-maintenance-guides-list.html

Cisco Unified CM and IM and Presence Certificate Management

As explained earlier, we recommend using certificates issued by a trusted CA to allow for easier client certificate validation of all certificates listed in Table 2-2.

Table 2-2 Certificates Validated by Cisco Jabber Clients

Service	Certificate	Description	Validated by
Cisco Unified CM	Tomcat	Unified CM web services certificate	Browser accessing GUI Jabber clients
Cisco Unified CM IM and Presence Service	Tomcat	Unified CM IM and Presence web services certificate	Browser accessing GUI Jabber clients
Cisco Unified CM IM and Presence Service	cup-xmpp	Unified CM IM and Presence XMPP service certificate	Jabber clients
Cisco Unity Connection	Tomcat	Unity Connection web services certificate	Browser accessing GUI Jabber clients

Steps required prior to replacing self-signed certificates with CA issued certificates:

1. Obtain the root certificate of the CA you plan to use to issue the certificates.
2. Navigate to the OS administration GUI of Unified CM.
3. Upload the CA root certificate as tomcat-trust.
4. Navigate to the OS administration GUI of Unified CM IM and Presence.
5. Upload the CA root certificate as xmpp-trust.
6. Navigate to the OS administration GUI of Cisco Unity Connection.
7. Upload the CA root certificate as tomcat-trust.

Steps to replace a self-signed certificate with a CA issued certificate:

1. Navigate to the OS administration GUI of the respective platform:
   - For Unified CM and Unified CM IM and Presence Tomcat certificate, use the Unified CM OS GUI.
   - For Unified CM IM and Presence cup-xmpp certificate, use the Unified CM IM and Presence OS GUI.
   - For Cisco Unity Connection Tomcat certificate, use Unity Connection OS GUI.
2. Generate a certificate signing request (CSR) for the desired certificate. Make sure to always set distribution to **Multi-Server (SAN)**.
3. Download the CSR.
4. Obtain a CA signed certificate from the trusted CA for the generated CSR.
5. On the OS administration GUI from step 1, upload the obtained CA issued certificate.

Tip A single multi-server certificate signing request should be generated for all certificates so that only a single certificate of a given type is required per cluster.

Tip Make sure that the X.509 key usage and X.509 extended key usage in the issued certificate match the request in the CSR (see Table 2-3).

As mentioned above, it is important to make sure that certificates issued by the CA have the required key usage and extended key usage. A typical problem is that the CA issuing the certificate based on the provided CSR does not simply issue a certificate with the key usage and extended key usage copied from the CSR, but instead sets the key usage and extended key usage of the issued certificate based on settings in a template selected for issuing the certificate. A certificate issued based on a typical Web Server template, for example, will not have the TLS Web Client Authentication extended key usage include. This creates problems with inter-server communications – for example, Intercluster Lookup Service (ILS) and User Data Store (UDS) – where the Tomcat certificate on the initiating side of the TLS connection is also used as a client certificate, and thus TLS connection setup fails due to the incorrect key usage (see the section Consider UDS Certificate Requirements).

Table 2-3 Key Usage Requirements for Tomcat and cup-xmpp Certificates

X.509 Key Usage	X.509 Extended Key Usage
Digital Signature	TLS Web Server Authentication
Key Encipherment	TLS Web Client Authentication
Data Encipherment	IPSec End System
Key Agreement	

Initial Cisco Unified CM Configuration

Immediately after installing the Unified CM cluster, perform the following basic configuration tasks:

- Node Name Configuration
- Enterprise Parameter Settings
- Service Activation
- Service Parameter Settings

Node Name Configuration

To allow for correct certificate validation and to ensure that references to Unified CM cluster members can always be resolved correctly, set the node names under System/Server in the Unified CM administration GUI to fully qualified domain names (FQDNs) for all cluster members. To achieve this, navigate to System/Server in the Cisco Unified CM administration GUI and verify that all servers show up in the first column as FQDNs. Change the entries of servers showing up as only a hostname without a DNS domain, to FQDNs.

Enterprise Parameter Settings

Check and update the Enterprise Parameters listed in Table 2-4.

Table 2-4 Enterprise Parameters

Enterprise Parameter	Description	Value
Cluster ID	Used to uniquely identify the Unified CM cluster in a number of intercluster features, including Intercluster Lookup Service (ILS) and intercluster call admission control	Example: USCluster
Auto Registration Phone Protocol	Signaling protocol provisioned for auto-registering phones	SIP
BLF For Call Lists	Specifies whether call lists in phones supporting this feature should show presence	Enabled
Advertise G.722 Codec	Allows G.722 between compatible devices	Enabled
URI Lookup Policy	According to RFC 3261, when determining SIP URI equivalence, the check on the left-hand side (user portion) of the URI has to be case-sensitive. The default behavior of Unified CM is to adhere to this standard, but to avoid potential issues with URIs using mixed capitalization, it is typically better to change the default.	Case Insensitive
Enable Dependency Records	Dependency records simplify the administration of Unified CM.	True
Auto select DN on any Partition	Simplifies administration. If enabled, the directory number configuration page automatically gets populated with the data of the first matching directory number.	True
CDR File Time Interval	Determines the time interval for call detail record (CDR) file updates	10
Enable Dependency Records	URLs used by endpoints for various purposes	True
Organization Top Level Domain		Example: ent-pa.com
Cluster Fully Qualified Domain Name	When routing numeric SIP URIs, Unified CM considers SIP URIs with the right-hand side (host portion) of the URI matching the configured Cluster Fully Qualified Domain Name (CFQDN) as destinations to be routed according to the configured local numeric dial plan. If no match is found for the numeric left-hand side of the URI in the configured numeric dial plan, then Unified CM rejects the call. For more details, see the section on *Routing of SIP Requests in Unified CM* in the *Dial Plan* chapter of the *Cisco Collaboration System 11.x SRND*.	Space-separated list of all Unified CM call processing nodes in the cluster. Example: us-cm-sub1.ent-pa.com us-cm-sub2.ent-pa.com

Service Activation

Table 2-5 summarizes the services to be activated on the Unified CM publisher node, the dedicated Unified CM TFTP server subscriber nodes, and the Unified CM call processing subscriber nodes.

Table 2-5 Unified CM Node Service Activation

Service	Publisher	Dedicated TFTP Subscriber	Call Processing Subscriber
CM Services			
Cisco CallManager			Yes
Cisco IP Voice Media Streaming App			Yes
Cisco CTIManager			Yes
Cisco Intercluster Lookup Service	Yes		
Cisco Location Bandwidth Manager			Yes
Cisco Dialed Number Analyzer Server	Yes		
Cisco Dialed Number Analyzer	Yes		
Cisco Tftp		Yes	
CTI Services			
Cisco WebDialer Web Service			Yes
Database and Admin Services			
Cisco Bulk Provisioning Service	Yes		
Cisco AXL Web Service	Yes		
Performance and Monitoring Services			
Cisco Serviceability Reporter	Yes		
Cisco CallManager SNMP Service	Yes	Yes	Yes
Security Services			
Cisco CTL Provider	Yes	Yes	Yes
Cisco Certificate Authority Proxy Function	Yes		
Directory Services			
Cisco DirSync	Yes		

Table 2-6 lists the services to be activated on Cisco Unified CM IM and Presence publisher and subscriber nodes.

Table 2-6 Unified CM IM and Presence Node Service Activation

Service	Publisher	Subscriber
Cisco AXL Web Service	Yes	Yes
Cisco Bulk Provisioning Service	Yes	
Cisco Serviceability Reporter	Yes	
Cisco SIP Proxy	Yes	Yes
Cisco Presence Engine	Yes	Yes
Cisco XCP Connection Manager	Yes	Yes
Cisco XCP Authentication Service	Yes	Yes

Service Parameter Settings

Some service parameters of the Cisco CallManager service are global in nature and need to be set only once in Unified CM Administration. lists the global service parameter settings for Cisco CallManager service are listed in Table 2-7.

Note Only non-default Service Parameter and other configuration field values are specified in this document. If a field configuration value is not mentioned, then the default value should be assumed.

Note Some of the service parameters listed are advanced service parameters.

Table 2-7 Global Service Parameters

Service Parameter	Value	Description
Apply Transformations On Remote Number	True	Makes sure that calling party transformations are also applied mid-call; for example, if a call is transferred from one party to another.
T302 Timer	5000	Whenever a destination is dialed digit-by-digit and based on the numeric dial plan provisioned in Unified CM, no immediate deterministic decision can be made about which provisioned pattern has to be considered for the dialed destination. Because a potential longer match (could be variable length) exists, the T302 inter-digit timeout has to expire before Unified CM selects the best route and routes the call. The default of 15,000 milliseconds (ms) typically is too long.
Automated Alternate Routing Enable	Enable AAR	This service parameter globally enables automated alternate routing (AAR).
Call Diagnostics Enabled	Enable Only When CDR Enabled Flag is True	This parameter determines whether call management records (CMR), also called diagnostic records, are generated.

Table 2-7 Global Service Parameters (continued)

Service Parameter	Value	Description
G.722 Codec Enabled	Enabled to All Devices Except Recording-Enabled Devices	G.722 disabled on recording-enabled devices to avoid problems with G.722 not being supported by the recorder.
Stop Routing on Q.931 Disconnect Cause Code	3 21 27 28 38 42 63	Allows Unified CM to stop hunting down the configured hunt list when receiving specific Q.850 cause codes.

Other service parameters of the Cisco CallManager service must be set explicitly as shown in Table 2-8 for each Unified CM call processing node.

Table 2-8 Per-Node Service Parameters

Service Parameter	Value	Description
CDR Enabled Flag	True	This parameter enables the generation of call detail records (CDR).
CDR Log Calls with Zero Duration Flag	True	This parameter enables or disables the logging of call detail records (CDRs) for calls that never connected or that lasted less than 1 second.
Digit Analysis Complexity	TranslationAndAlternatePatternAnalysis	This parameter specifies the amount of digit analysis information that CCM trace files will provide.

Other IM and Presence Settings

Previous sections discussed the IM and Presence service activation, certificates management, and the IM and Presence SIP trunk configuration. In addition to that, configure settings on IM and Presence servers:

- Configure a Unified CM domain in the **IM&P Cisco SIP Proxy** Service Parameter.
- In **Cisco Unified CM IM and Presence Administration** > **Presence** > **Settings** > **Standard Configuration**:
  - Configure a Cluster ID value.
  - Enable availability sharing. If not enabled, users can view only their own availability status.
  - Check **Enable ad-hoc presence subscriptions** to turn on ad-hoc presence subscriptions for Cisco Jabber users.
- In **Cisco Unified CM IM and Presence Administration** > **Presence** > **Routing** > **Settings**:
  - Configure **Proxy Server Settings**: **Enable Method/Event Routing Status**
- In **Cisco Unified CM IM and Presence Administration** > **Messaging** > **Settings**:
  - Enable instant messaging.

Also configure UC services for Jabber clients, as described in the section on Jabber Provisioning.

Dial Plan Configuration

A structured, well-designed dial plan is essential to successful deployment of any call control system. The design of an enterprise dial plan needs to cover these main areas:

- Endpoint addressing
- General numbering plan
- Dialing habits
- Routing
- Classes of service

The recommended dial plan design follows the design approach documented in the *Dial Plan* chapter of the *Cisco Collaboration System 11.x SRND*.

Example Topology

For the purpose of this document, we assume a centralized call processing deployment serving three sites in the US: SJC, RCD, and RTP. Table 2-9 provides the DID (direct inward dial) ranges for these sites.

Table 2-9 DID Ranges for Example Sites

Site	DID range
SJC	+1 408 555 4XXX
RCD	+1 972 555 5XXX
RTP	+1 919 555 1XXX

Endpoint Addressing

For endpoints with DID addresses, directory numbers are provisioned as full +E.164 numbers, where +E.164 represents a leading "+" followed by the full global E.164 phone number. To provision a +E.164 directory number in Unified CM, the leading "+" has to be escaped; for example, extension 4001 in SJC would have to be provisioned as \+14085554001.

Some endpoints will not have DIDs because not enough DIDs are available from the provider or because the associated devices do not need to be reachable from the PSTN (for example, lobby phones). For these endpoints no DIDs (E.164 numbers) exist, and thus an address format other than +E.164 is required for these endpoints. This address format is discussed in the section on General Numbering Plan.

Addressing Enterprise Services for External Access

Some services have assigned PSTN numbers. An example of this might be a voicemail pilot number that has to be reachable from the outside to enable users to call into voicemail from the PSTN. PSTN E.164 numbers for these services have to be reserved from the DID ranges assigned by the PSTN providers.

General Numbering Plan

In addition to endpoints with associated DIDs for which +E.164 addresses can be used, a number of additional destinations exist for which no DIDs exist:

- Lobby phones
- Regular endpoints for which no DIDs could be assigned by the provider
- Services (call pickup numbers, call park numbers, conferences, and so forth)

In this document we refer to these types of destinations as *non-DIDs*.

Addresses for these non-DIDs, similar to +E.164 addresses, must be unique system-wide to avoid site-specific partitions for non-DIDs. The recommended solution is to introduce an enterprise specific numbering (ESN) schema for all non-DIDs. This ESN schema follows the structure of typical abbreviated inter-site dialing:

- Access-code

  A single-digit access code for abbreviated inter-site dialing. In the design phase, choose the access code so that there is no overlap with any other enterprise dialing habit (see below).

- Site-code

  A digit sequence uniquely identifying a site in the network. In the design phase, choose the length of the site code so that it not only covers all existing sites, but also allows for growth.

- Extension

  A digit sequence uniquely identifying the respective entity within the site.

In this document we use 8 as the access-code for abbreviated inter-site dialing, and thus all ESNs start with 8 and use a three-digit site code and a four-digit extension. Table 2-10 indicates an ESN range for the DID and non-DID numbers for each site in our example.

Table 2-10 ESN Ranges for DIDs and Non-DIDs

Site	+E.164 Range	Site Code	ESN Range for DIDs	ESN Range for Non-DIDs
SJC	+1 408 555 4XXX	140	8-140-4XXX	8-140-5XXX
RCD	+1 972 555 5XXX	197	8-197-5XXX	8-197-6XXX
RTP	+1 919 555 1XXX	191	8-191-1XXX	8-191-2XXX

The plan is to use the same site code for DIDs and non-DIDs, but the first digit of the extension for non-DIDs is different from the first digit of the DID extensions. This also allows for abbreviated four-digit intra-site dialing to non-DIDs and DIDs.

While the ESN ranges in Table 2-10 leave room in the ESN plan for site-specific numbers, there is also a requirement to assign number ranges for non-site-specific services such as, for example, scheduled conferences. Table 2-11 shows an example of how this requirement can be addressed by reserving a dedicated site code (in this case 099).

Table 2-11 ESN Ranges for Conferences

ESN Range	Usage
8099[12]XXX	Scheduled conferences

Dialing Habits

Dialing habits describe what end users must dial to reach various types of destinations. Dialing habits can first be classified as numeric dialing (for example, 914085550123) or alphanumeric dialing (for example, bob@ent-pa.com).

In this design, in addition to alpha URI dialing, the numeric dialing habits shown in Table 2-12 are supported.

Table 2-12 Supported Numeric Dialing Habits

Dialed Pattern	Example (site SJC)	Type of Destination
XXXX	4001 (DID) 5001 (non-DID)	Abbreviated intra-site dialing to reach a destination at the same site. The called destination can be a DID, a non-DID, or a service number.
+E.164	+14085554001 (on-net, SJC) +19195551001 (on-net, RTP) +1212551001 (off-net)	Full +E.164 dialing for example from directories. The dialed destination can be on-net or off-net. The implemented dial plan makes sure that calls to on-net destinations dialed as +E.164 are routed on-net. Non-DIDs obviously cannot be called as +E.164.
Access code–site code–extension	8-140-4001 (DID, SJC) 8-140-5001 (non-DID, SJC) 8-191-1001 (DID, RTP) 8-191-2001 (non-DID, RTP)	Abbreviated inter-site dialing to reach a destination at the same site or a different site. The called destination can be a DID, a non-DID, or a service number. The access code (8 in the example) has to be selected so that it does not overlap with any other dialing habit; for example any abbreviated intra-site dialing: access code 8 for inter-site dialing prohibits four digit intra-site dialing starting with 8.
*E.164	*12125551567	Dialing of a video call through dedicated video ISDN gateways. The * is used to create a specific dialing habit with no overlap to any other numeric(!) dialing habit. To avoid the use of * also a number area starting with the abbreviated inter-site access code 8 can be used: for example 8000-<E.164>.
91-<10 digits>	914085554001 (on-net, SJC) 919195551001 (on-net, RTP) 912125551001 (off-net)	US specific habitual PSTN dialing of national destinations. The implemented dial plan ensures that if the dialed destination is on-net then the call is routed on-net. The leading 9 here is the PSTN access code typically used in the US.
9011-<E.164 number>	90114961007739764	US specific habitual PSTN dialing of international destinations. The implemented dial plan makes sure that if the dialed destination is on-net then the call is routed on-net.

In general, using fewer supported dialing habits simplifies the design. Starting the design process with an overview of all dialing habits makes sure that overlaps between any two dialing habits leading to inter-digit timeouts are detected and can be resolved before starting the dial plan deployment. Avoiding overlaps with any other (typically on-net) dialing habit is the key reason for using a PSTN access code (typically 9 in the US, as shown above).

Partitions

When defining the partitions and CSSs provisioned to build an enterprise dial plan, one goal is to avoid replication of duplicate configuration as much as possible. Following this maxim, Table 2-13 shows the global (that is, not site or country specific) partitions required.

Table 2-13 Global Partitions

Partition	Description
DN	Holds all +E.164 directory numbers and other local on-net +E.164 destinations (for example, pilot numbers reachable from the PSTN). All +E.164 patterns are provisioned as urgent patterns.
ESN	Holds all Enterprise Specific Numbers (ESNs). This includes ESN directory numbers (for example, for non-DID phones) as well as dialing normalization translation patterns transforming from abbreviated inter-site dialing of DIDs to +E.164.
PSTNInternational	Holds +E.164 route patterns required to provide PSTN access to international destinations.
URI	Holds manually provisioned URIs.
onNetRemote	Holds all patterns of remote on-net destinations. In environments with multiple Unified CM clusters, this includes all remote number ranges learned via Global Dial Plan Replication (GDPR).
B2B_URI	Holds SIP route patterns required for business-to-business (B2B) URI dialing through the Internet.
Directory URI	System Partition where all auto-generated URIs are put. This partition does not need to be created. It is listed here for reference to introduce the partition, which is used again later in this document.

All of the partitions Table 2-13 except the Directory URI partition must be created. In addition to the pattern classes represented by these global partitions, several site, country, or class-of-service specific pattern classes are required, as show in Table 2-14.

Table 2-14 Country or Site Specific Partitions

Partition	Description
USPSTNNational	Holds +E.164 route patterns required to provide PSTN access to national destinations in the US. To support other countries, and thus other country-specific dialing habits, a country appropriate xxPSTNNational partition (where xx represents the country; for example, DEPSTNNational, UKPSTNNational, ITPSTNNational) also needs to be provisioned, which then holds the +E.164 route patterns required to provide PSTN access to national destinations of that country. The reason we differentiate between international PSTN access (see Table 2-13) and national PSTN access is that we need to be able to build differentiated classes of service allowing calls to reach national only, or national and international destinations.
USToE164	Holds dialing normalization translation patterns to transform US specific habitual PSTN dialing (for example, 91-<*10 digits*>) to +E.164. To support other countries, and thus other country-specific dialing habits, a country appropriate xxToE164 partition (where xx represents the country; for example, DEToE164, UKToE164, ITToE164) also needs to be provisioned, which then holds the dialing normalization translation patterns required to transform the country specific habitual PSTN dialing to +E.164.
USEmergency	Holds route patterns required to provide access to emergency calls using the US specific emergency dialing habits.
USPhLocalize	Holds calling party transformation patterns to localize +E.164 calling party numbers for abbreviated display on phones in the US.
<site>Intra	Site-specific intra-site dialing. For example: SJCIntra. Holds dialing normalization patterns to transform site-specific abbreviated intra-site dialing to DIDs, or non-DIDs to +E164 or ESN, respectively.
<site>PhLocalize	Site-specific. For example: SJCPhLocalize. Holds calling party transformation patterns to localize +E.164 calling party numbers for abbreviated display on phones in a given site.

As emergency calls are placed using country specific dialing habits, partition USEmergency with the route patterns enabling the US dialing habit for emergency calls also is country specific. To also support other dialing domains (countries), the equivalent partitions for these other dialing domains (for example, DEEmergency, ITEmergency, DEPhLocalize, ITPHLocalize, for Germany and Italy respectively) would need to be created.

Dialing Normalization Translation Patterns

Table 2-15 summarizes which dialing normalization translation patterns need to be provisioned using the partitions from the previous section. All dialing normalization translation patterns are provisioned as urgent patterns and have **Use Originator's Calling Search Space** set as described in section on Partitions so that, after applying the called party transformation defined in the dialing normalization translation pattern, the original CSS is used to find the final match for the dialed destination.

Table 2-15 Summary of Dialing Normalization Translation Patterns

Partition	Pattern	Called Party Transformation Mask	Note
ESN	81404XXX	+14085554XXX	Abbreviated inter-site dialing to site SJC
ESN	81975XXX	+19725555XXX	Abbreviated inter-site dialing to site RCD
ESN	81911XXX	+19195551XXX	Abbreviated inter-site dialing to site RTP
SJCIntra	4XXX	+14085554XXX	Abbreviated intra-site dialing in site SJC to DID in SJC
SJCIntra	5XXX	81405XXX	Abbreviated intra-site dialing in site SJC to non-DID in SJC
RCDIntra	5XXX	+19725554XXX	Abbreviated intra-site dialing in site RCD to DID in RCD
RCDIntra	6XXX	81976XXX	Abbreviated intra-site dialing in site RCD to non-DID in RCD
RTPIntra	1XXX	+19195551XXX	Abbreviated intra-site dialing in site RTP to DID in RTP
RTPIntra	2XXX	81912XXX	Abbreviated intra-site dialing in site RTP to non-DID in RTP
UStoE164	9.1[2-9]XX[2-9]XXXXXX	No Mask, strip pre-dot, prefix +	US specific habitual PSTN dialing to national destinations in the US
UStoE164	9011.!#	No Mask, strip pre-dot, prefix +	US specific habitual PSTN dialing to national destinations in the US.
UStoE164	9011.!	No Mask, strip pre-dot, prefix +	US specific habitual PSTN dialing to national destinations in the US **Note** This is the only pattern for which **Do Not Wait For Interdigit Timeout On Subsequent Hops** is not set.

For dialing domains other than the US, other country specific dialing normalization translation patterns must be defined if the installation has to support those country specific dialing habits. Table 2-16 shows the required dialing normalization for Germany (DE) and Italy (IT) as examples.

Table 2-16 Dialing Normalization for Germany and Italy

Partition	Pattern	Called Party Transformation	Note
DEtoE164	000.!	strip pre-dot, prefix +	Germany: international call (000-E.164). **Note** **Do Not Wait For Interdigit Timeout On Subsequent Hops** is not set.
DEtoE164	000.!#	strip pre-dot trailing #, prefix +	Germany: international call (000-E.164).
DEtoE164	00.[^0]!	strip pre-dot, prefix +49	Germany: national call (00-national significant number). **Note** The numbering plan in Germany is variable length and this pattern needs to cover this. **Note** **Do Not Wait For Interdigit Timeout On Subsequent Hops** is not set.
DEtoE164	00.[^0]!#	strip pre-dot trailing #, prefix +49	Germany: national call (00-national significant number).
ITtoE164	000.!	strip pre-dot, prefix +	Italy: international call (000-E.164). **Note** **Do Not Wait For Interdigit Timeout On Subsequent Hops** is not set.
ITtoE164	000.!#	strip pre-dot trailing #, prefix +	Italy: international call (000-E.164)
ITtoE164	0.0[^0]!	strip pre-dot, prefix +39	Italy: national call (0-national significant number (NSN) where NSN starts with 0). **Note** The numbering plan in Italy is variable length and this pattern needs to cover this. **Note** **Do Not Wait For Interdigit Timeout On Subsequent Hops** is not set.
ITtoE164	0.0[^0]!#	strip pre-dot trailing #, prefix +39	Italy: national call (0-NSN where NSN starts with 0).
ITtoE164	0.[^0]!	strip pre-dot, prefix +39	Italy: national call (0-NSN where NSN does not start with 0). **Note** The numbering plan in Italy is variable length and this pattern needs to cover this. **Note** **Do Not Wait For Interdigit Timeout On Subsequent Hops** is not set.
ITtoE164	0.[^0]!#	strip pre-dot trailing #, prefix +39	Italy: national call (0-NSN where NSN does not start with 0).

The example in Table 2-16 shows that in Italy and Germany the ITU recommended 0 is used to access a trunk from inside the enterprise, and then 0 and 00 are used for national and international access. Since 1998, geographic numbers in Italy start with 0, and digits 1 to 9 as the first digit of a national significant number indicate different types of numbers. Hence, dial strings starting with exactly two 0s (00) need to be treated differently in Italy than in Germany. In Italy the second zero has to be considered part of the NSN and hence has to be kept in the resulting +E.164 digit string, while a second zero in Germany would need to be removed because geographic numbers in Germany do not start with a zero.

The example of the dialing normalization required for these two countries shows how country specific dialing habits can be modeled in the design approach presented.

For more information on international numbering plans, see the *International Numbering Resources* page of the ITU-T at http://www.itu.int/en/ITU-T/inr/Pages/default.aspx. There you can find links to various resources, including E.164 country codes and national numbering plans. An overview of dialing procedures used in various countries can be found in *Operational Bulletin No.994 (15.XII.2011) and Annexed List: Dialling procedures (international prefix, national (trunk) prefix and national (significant) number) (in accordance with ITU-T Recommendation E.164 (11/2010)) (Position on 15 December 2011)*, available at http://www.itu.int/pub/T-SP-OB.994-2011. The actual list of dialing procedures starts at page 25 of that document and is also available for download at http://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-E.164C-2011-PDF-E.pdf.

Classes of Service and Calling Search Spaces (CSSs)

As mentioned before, a CSS is a list of partitions that defines which partitions, and thus patterns, a calling entity using the CSS can access. In this document we use a dial plan approach that uses only the line CSS to define class of service.

Table 2-17 lists the classes of service considered in this design. The classes of service chosen for this design are only examples. If further classes of services are required, then these can be defined equivalently.

Tip The number of classes of service is one of the key parameters driving the complexity of enterprise dial plan designs. Therefore, it is good practice to define as few classes of service as possible for the dial plan.

The recommended design makes use of only the CSS provisioned on the line and does not use the device CSS to define class of service. The device CSS can be used to implement general dialing habits that need to be available for everyone. An example for this is emergency calling; see the section on Emergency Call Considerations in Multi-National Environments for more details on when to use the device CSS to implement emergency calls.

Table 2-17 Classes of Service

Class of Service	Access to
International	All on-net destinations National PSTN destinations International PSTN destinations Business-to-business URI dialing Emergency calls
National	All on-net destinations National PSTN destinations Emergency calls
Internal	All on-net destinations Emergency calls

Adding business-to-business URI dialing to only the International class of service is an example based on the assumption that business-to-business (B2B) calls consume limited edge resources. Also we are trying to avoid increasing the number of classes of service by a factor of two by introducing classes of service International, InternationalB2B, National, NationalB2B, Internal, and InternalB2B.

Because only the line CSS is used to define both class of service and the set of dialing habits available to a given caller, a CSS needs to be provisioned per site and class of service.

Table 2-18 shows how class of service International for a user in site SJC would be defined based on the partition set previously defined (see Table 2-13 and Table 2-14).

Table 2-18 Class of Service International for SJC User

CSS Name	Partitions
SJCInternational	DN Directory URI URI ESN onNetRemote SJCIntra UStoE164 USPSTNNational PSTNInternational B2B_URI USEmergency

As depicted in Table 2-19, the remaining classes of service are created equivalently by selectively removing access to B2B URI dialing, international, and national PSTN destinations.

Table 2-19 Classes of Service National and Internal for SJC User

CSS Name	Partitions	CSS Name	Partitions
SJCNational	DN Directory URI URI ESN onNetRemote SJCIntra UStoE164 USPSTNNational USEmergency	SJCInternal	DN Directory URI URI ESN onNetRemote SJCIntra UStoE164 USEmergency

CSSs for classes of services for users in other sites are created equivalent to the above CSSs, with the only difference being a different partition used with the site-specific dialing normalization patterns. Table 2-20 shows an example of the RTP site National and Internal classes of service.

Table 2-20 Classes of Service National and Internal for RTP User

CSS Name	Partitions
RTPNational	DN Directory URI URI ESN onNetRemote RTPIntra UStoE164 USPSTNNational USEmergency

CSS Name	Partitions
RTPInternal	DN Directory URI URI ESN onNetRemote RTPIntra UStoE164 USEmergency

These examples clearly show that the chosen partition scheme allows for optimal reuse of patterns and partitions when creating CSSs to implement classes of service for multiple sites.

For sites in other dialing domains (countries), the same CSS and partition schema as shown above can be used, with the only difference being that the dialing normalization partition for the specific dialing domain and the partition with the country specific route to national PSTN destinations would be used instead of the US partitions used above. For example, Table 2-21 shows the CSS for class of service International for a site FRA in Germany (DE).

Table 2-21 Class of Service International for Users in site FRA in Germany (DE)

CSS Name	Partitions
FRAInternational	DN Directory URI URI ESN onNetRemote FRAIntra DEtoE164 DEPSTNNational PSTNInternational B2B_URI DEEmergency

Special CSSs

In addition to classes of service for users, calling search spaces (CSSs) also are used to define classes of service for applications connected through trunks, such as Cisco Unity Connection, for example. Assuming that Unity Connection should have access only to on-net destinations and that, in addition to ESN and +E.164 dialing, also US dialing habits should be supported from Unity Connection, Table 2-22 shows the CSS to implement this class of service.

Table 2-22 Class of Service for Voicemail

CSS Name	Partitions
VoiceMail	DN ESN URI onNetRemote Directory URI UStoE164

In scenarios where Cisco Unity Connection needs to serve multiple countries, then implementing the country specific dialing normalization as defined in partition UStoE164 in the above example is not an option. The only dialing habits that can be supported in that case are the globally significant dialing habits ESN and +E.164.

To use Unified CM presence, a subscribe CSS has to be provisioned, among other things, to allow access to all presentities that a presence user subscribes to. To allow for a very simple provisioning of Unified CM presence without further differentiation of presence access, a single CSS needs to be provisioned that allows access to all possible on-net destinations. Table 2-23 shows the settings for this default subscribe CSS.

Table 2-23 Default Subscribe CSS

CSS Name	Partitions
DefaultSubscribe	DN ESN URI onNetRemote Directory URI

This subscribe CSS ensures access to all types of on-net destinations.

Table 2-24 shows the (trivial) CSS "DN" to be used as the incoming CSS on PSTN trunks. To avoid loops, a PSTN trunk can address only +E.164 directory numbers. A PSTN trunk would not need access to ESN patterns, dialing normalization patterns, or URIs because only a single number format is supported by the PSTN, and this is normalized to +E.164 on ingress.

Table 2-24 Inbound CSS for PSTN Gateways

CSS Name	Partitions
DN	DN

Cisco TelePresence Servers and the TelePresence Conductor require access to all on-net destinations, and at the same time need to be able to place calls to any PSTN destination. On the other hand, they do not require access to any dialing domain-specific or site-specific dialing normalization patterns. CSS TelePresenceConferencing shown in Table 2-25 implements this class of service.

Table 2-25 Inbound CSS for Trunk from TelePresence Conferencing

CSS Name	Partitions
TelePresenceConferencing	DN ESN URI onNetRemote Directory URI PSTNInternational

Table 2-26 shows the CSS ICTInbound to be used as an incoming CSS on trunks to other Unified CM clusters. To avoid loops, the incoming CSS on these intercluster trunks should not provide access to remote on-net destinations (partition onNetRemote), but the trunks (inbound CSS) need to support all valid on-net addressing modes (+E.164, ESN, and URIs). Dialing normalization is not part of this CSS because dialing habits other than +E.164 and ESN would already have been normalized to +E.164 or ESN on the remote Unified CM cluster prior to landing on the incoming intercluster trunk.

Table 2-26 Inbound CSS for Trunks to Other Unified CM Clusters

CSS Name	Partitions
ICTInbound	DN ESN URI Directory URI

Local Route Groups for Call Type Specific Outbound Gateway Selection

To allow for flexible egress gateway selection based on the calling device, we recommend using local route groups (LRGs). Using LRGs for egress gateway selection avoids the need for site-specific route patterns.

To allow for differentiated LRG selection for different call types, set up multiple LRG names as shown in Table 2-27.

Table 2-27 Local Route Group Names

Local Route Group Name	Description
LRG_PSTN_1	Local route group referring to primary PSTN resources to be used for PSTN calls
LRG_PSTN_2	Local route group referring to secondary PSTN resources to be used for PSTN calls
LRG_VIDEO_1	Local route group referring to primary PSTN resources to be used for video PSTN calls
LRG_VIDEO_2	Local route group referring to secondary PSTN resources to be used for video PSTN calls
LRG_Emergency_1	Local route group referring to primary PSTN resources to be used for emergency calls
LRG_Emergency_2	Local route group referring to secondary PSTN resources to be used for emergency calls

With these LRG definitions, dedicated route lists can be created for both "normal" PSTN calls and emergency calls so that different PSTN resources (gateways) are used for emergency calls than for normal PSTN calls. This makes sense in scenarios where centralized PSTN resources are provisioned for normal PSTN calls, but emergency calls should still use dedicated small gateways local to the site to allow for local emergency call routing to the correct Public Safety Answering Point (PSAP).

The video LRGs are provisioned for video-enabled ISDN gateways and treat them as separate resources.

Route Lists Using Local Route Groups

Using the LRGs as defined in the previous section, route lists should be created as depicted in Table 2-28.

Table 2-28 Route List Definitions

Route List	Members	Description
RL_PSTN	LRG_PSTN_1 LRG_PSTN_1 Standard Local Route Group	Normal PSTN calls should make use of the primary and secondary site-specific PSTN resources defined for normal PSTN calls. The last member, Standard Local Route Group, allows for fallback to PSTN resources not specific to a call type.
RL_Emergency	LRG_Emergency_1 LRG_Emergency_2 LRG_PSTN_1 LRG_PSTN_1 Standard Local Route Group	For emergency calls, the first call-specific resources for emergency calls should be used, then the second, then the PSTN resources defined for normal PSTN calls, and lastly the non-specific PSTN resources.
RL_VIDEO	LRG_VIDEO_1 LRG_VIDEO_2 LRG_PSTN_1 LRG_PSTN_2 Standard Local Route Group	For video calls, first the video-specific gateway resources are used, then the regular PSTN resources are considered as a fallback (audio only), and lastly the Standard Local Route Group is used if the others fail.

With the above LRG and route list definition on each device pool, up to seven route groups can be selected for the defined LRGs to allow for very specific outbound gateway selection. The actual PSTN resources to be used for certain call types are defined during device pool provisioning. If selecting different outbound PSTN resources based on call type is not required for a given set of devices, and only a single PSTN resource is needed for all call types, then it is sufficient to define only an actual route group for the Standard Local Route Group on the respective device pool and leave all other LRGs in that device pool set to **<None>**. Having **Standard Local Route Group** as the last entry in all route lists is a good way to achieve this.

Route Patterns for PSTN Access and Emergency Calls

PSTN access is achieved through PSTN route patterns. As described in the section about Classes of Service and Calling Search Spaces (CSSs), the route to international destinations needs to be provisioned in the PSTNInternational partition, while national PSTN routes are provisioned in the dialing domain specific partitions xxPSTNNational (where xx represents dialing domain USPSTNNational, for example). Table 2-29 shows the configured PSTN route patterns.

Table 2-29 PSTN Route Patterns

Pattern	Partition	Gateway or Route List	Description
\+!	PSTNInternational	RL_PSTN	Variable length to allow for dialing of arbitrary international destinations.
\+!#	PSTNInternational	RL_PSTN	Alternative pattern for international destinations to allow terminating variable length dialing with #. Discard Digits set to **Trailing-#**
\+1[2-9]XX[2-9]XXXXXX	USPSTNNational	RL_PSTN	Explicit pattern for national destinations in the US. **Urgent Priority** checked to avoid overlap with variable length PSTN route pattern \+! defined for international destinations.
911	USEmergency	RL_Emergency	US emergency calling **Urgent Priority** checked
9911	USEmergency	RL_Emergency	US emergency calling **Urgent Priority** checked

Apart from the route pattern settings explicitly shown in Table 2-29, all other settings are left with default values as shown in Table 2-30. This especially includes the calling, connected, and called party transformations, which are left empty (apart from stripping a trailing # as mentioned above) because the calling and called party transformations required to match the PSTN requirements are configured as explicit calling and called party transformations. This is described in the sections on Outbound Calls: Called and Calling Number Transformations on ISDN Gateways and Outbound Calls: Called and Calling Number Transformations on SIP Trunks.

Table 2-30 Route Pattern Default Settings

Setting	Value
Pattern Definition	
Numbering Plan	-- Not Selected --
Route Filter	<None>
MLPP Precedence	Default
Apply Call Blocking Percentage	Unchecked
Resource Priority Namespace Network Domain	<None>
Route Class	Default
Route Option	Route this pattern

Table 2-30 Route Pattern Default Settings (continued)

Setting	Value
Call Classification	OffNet
External Call Control Profile	<None>
Allow Device Override	Unchecked
Provide Outside Dial Tone	Checked
Allow Overlap Sending	Unchecked
Require Forced Authorization Code	Unchecked
Authorization Level	0
Require Client Matter Code	Unchecked
Calling Party Transformations	
Use Calling Party's External Phone Number Mask	Unchecked
Calling Party Transform Mask	Leave empty; do not enter any value
Prefix Digits (Outgoing Calls)	Leave empty; do not enter any value
Calling Line ID Presentation	Default
Calling Name Presentation	Default
Calling Party Number Type	Cisco CallManager
Calling Party Numbering Plan	Cisco CallManager
Connected Party Transformations	
Connected Line ID Presentation	Default
Connected Name Presentation	Default
Called Party Transformations	
Discard Digits	<None>
Called Party Transform Mask	Leave empty; do not enter any value
Prefix Digits (Outgoing Calls)	Leave empty; do not enter any value
Called Party Number Type	Cisco CallManager
Called Party Numbering Plan	Cisco CallManager
ISDN Network-Specific Facilities Information Element	
Network Service Protocol	-- Not Selected --
Carrier Identification Code	Leave empty; do not enter any value
Network Service	-- Not Selected --

While the international PSTN route patterns in partition PSTNInternational are not dialing domain (country) specific, the route patterns in partitions USPSTNNational and USEmergency are country specific. If the dial plan needs to support other countries, then the route patterns for these countries need to be created as shown in Table 2-31.

Table 2-31 Non-US Route Patterns for National Destinations

Pattern	Partition	Gateway or Route List	Description
\+49!	DEPSTNNational	RL_PSTN	Variable length because the German numbering plan with country code 49 is variable length.
\+49!#	DEPSTNNational	RL_PSTN	Alternative pattern for national destinations to allow terminating variable length dialing with #. Discard Digits set to **Trailing-#**
\+33XXXXXXXXX	FRPSTNNational	RL_PSTN	Explicit pattern for national destinations in France. **Urgent Priority** checked to avoid overlap with variable length PSTN route pattern \+! defined for international destinations.
112	DEEmergency	RL_Emergency	German emergency calling **Urgent Priority** checked
0112	DEEmergency	RL_Emergency	German emergency calling **Urgent Priority** checked
112	FREmergency	RL_Emergency	French emergency calling **Urgent Priority** checked
0112	FREmergency	RL_Emergency	French emergency calling **Urgent Priority** checked

Table 2-31 shows the difference between fixed and variable length numbering plans. The national numbering plan in Germany is variable length and thus the route pattern to match on national destinations in Germany has to match on variable length digit strings, and we also need to provision an alternate route pattern ending on # to enable users to explicitly terminate dial strings with # to avoid inter-digit timeouts when dialing national destinations. In contrast to this, the national numbering plan in France is fixed length (as in the US), and thus a single urgent fixed-length route pattern is enough to cover all national numbers in France.

Because Germany and France use the same emergency dialing habit, the emergency routing can be simplified by combining both emergency partitions DEEmergency and FREmergency into a single partition 112Emergency and by using that partition instead in the CSS definitions.

Emergency Call Considerations in Multi-National Environments

Independent from individual classes of service, access to emergency numbers is required from all endpoints at all times. As shown previously, this is easily achieved by adding the partition with the emergency calling route patterns to all CSSs. This approach is problematic if multiple countries need to be supported, those countries require different emergency dialing habits, and mobility features such as extension mobility and device mobility are used.

In this case, if a user roams between countries with different emergency dialing habits, then the device this user is using inherits the emergency dialing habits specific to the visiting user. For example, if a user from Germany logs into a phone in the US, then the line CSS as defined on the German user's extension mobility profile gets assigned to the visited phone in the US, so that on this phone emergency calls now need to be placed using the German emergency calling dialing 112, and the US emergency call dialing habit 911 is not supported any longer.

To make sure that phones in a given country always support the national emergency call dialing habit independent of whether a foreign user logged into the phone, a different approach for emergency calls can be implemented. Instead of adding the USEmergency to all CSSs, create a dedicated USEmergency CSS and assign that CSS as the device CSS on all devices in the US. Then if a foreign user logs into a phone in the US, the visiting user's "home" dialing habits as defined by the line CSS will be combined with the visited countries emergency dialing habit. In the above case of a German user logging into a US phone, that user's German PSTN dialing habits will be supported together with the US specific emergency dialing habit 911. Keep in mind that this combination of dialing habits between different countries might create overlaps between the visited sites' emergency dialing and the visiting user's regular dialing habits. For example, if a site in Germany has four-digit extensions starting with 9 (such as +E.164 range +49 6100 773 9XXX), then the abbreviated four-digit intra-site dialing defined for that site through a 9XXX dialing normalization translation pattern creates an overlap with the US emergency dialing 911 if a user from that German site logs into a phone in the US. As long as the emergency dialing habit is more specific, then creating the emergency calling route pattern as urgent pattern makes sure that no delay is experienced when placing an emergency call. On the other hand, the 911 US emergency pattern would "block" all four-digit dialing starting with 911, affecting four-digit intra-site dialing to directory numbers +49 6100 773 911X, for example.

Moving the emergency dialing from the line to the device CSS also avoids the problem that visiting users' emergency dialing habits (112 in case of a user from Germany) need to be transformed to the visited countries emergency dialing habit (911 in the US).

Route Patterns for Video PSTN (ISDN) Calls

Video ISDN gateways require special treatment from the dial plan perspective because it is unfeasible from the cost perspective to use ISDN video gateways for regular voice calls. In this design the selection of video ISDN gateways is explicitly tied to a special video PSTN dialing habit (see Table 2-12). Table 2-32 shows the required route patterns to enable this dialing habit.

Table 2-32 Route Patterns for Video PSTN (ISDN) Calls

Pattern	Partition	Gateway or Route List	Description
*!	PSTNInternational	RL_VIDEO	Variable length because we need to support E.164 behind the *
*!#	PSTNInternational	RL_VIDEO	Alternative pattern to allow termination of variable length dialing with #. Discard Digits set to **Trailing-#**
*1XXXXXXXXXX	PSTNInternational	RL_VIDEO	Supplementary route pattern to allow dialing to US destinations (fixed length) without inter-digit timeout. **Urgent Priority** checked.

Putting the video ISDN route patterns into partition PSTNInternational effectively adds video dialing capabilities to class of service International.

Outbound Calls: Called and Calling Number Transformations on ISDN Gateways

On ISDN trunks, calling and called party number information is sent and received in calling and called party information elements. These information elements are a triplet consisting of numbering plan, number type, and number. How these fields need to be set depends on the trunk service definition of the provider. As an example, for a call to E.164 destination 4961007739764 on a trunk in Germany in the same area code 6100, the called party number in the outgoing ISDN SETUP message could be sent as (plan/type/number) ISDN/national/61007739764, ISDN/subscriber/7739764, or unknown/unknown/061007739764.

If gateways terminating ISDN trunks are connected to Unified CM using SIP, then number types cannot be sent from Unified CM to the gateway because SIP does not know the concept of number types. Whether different ISDN number types need to be supported for different call types depends on the provider's SIP trunk service definition. On ISDN trunks, some providers always allow called and calling party numbers independent of called destination to be sent using the same ISDN plan and type indication.

Table 2-33 shows an example of alternate called party number formats that an ISDN provider in the US might accept.

Table 2-33 Alternate ISDN Number Format for Calls on US ISDN Trunk

Type of Call	Destination	Called Party Plan/Type/Number to Be Sent to PSTN	Digit String Sent to Gateway
National	+12125551234	unknown/unknown/12125551234	*12125551234
International	+4961007739764	unknown/unknown/0114961007739764	*0114961007739764

The digit string sent to the gateway is prefixed with a "*" to simplify the dial peer definition on the gateway. Prefixing called party numbers sent to the gateway with a "*" enables easy non-colliding destination-pattern based outbound dial-peer selection on the gateway for inbound and outbound calls because called party numbers received from the PSTN never start with a "*". The leading "*" prefixed by Unified CM needs to be removed on the gateway before sending the call to the PSTN. The leading "*" on all called party numbers sent from Unified CM to the gateway allows the use of "destination-pattern *" on the POTS dial peers on the gateway. The default digit stripping behavior of Cisco IOS will then automatically strip the leading "*".

The transformation from the called +E.164 destination to the digit string to be sent to the PSTN can be achieved on Unified CM, and on the gateway the ISDN plan and type can be enforced easily using Cisco IOS voice translation rules as shown in Example 2-2.

Example 2-2 Cisco IOS Voice Translations to Force Single ISDN Plan and Type

```
voice translation-rule 1
    rule 1 /^\*/ // type any unknown plan any unknown
    rule 2 // // type any unknown plan any unknown
voice translation-profile ISDNunknown
    translate called 1
    translate calling 1
dial-peer voice 1 pots
    translation-profile outgoing ISDNunknown
```

The Cisco IOS configuration piece shown in Example 2-2 demonstrates how to force a single ISDN plan and type for calling and called party information to be sent to the PSTN through a given POTS dial-peer. Rule 1 of voice-translation-rule 1 matches all numbers starting with "*" and simply removes this leading "*". Rule 2 of voice translation-rule 1 matches on all numbers with any plan and type, and it forces both plan and type to unknown while not changing the actual digit string of the number. With this Cisco IOS voice translation-rule applied to the POTS dial peer pointing to the ISDN, all called and calling party numbers sent from Unified CM to the gateway will be forwarded to the PSTN unchanged, with plan and type forced to unknown.

With this translation logic in place on the gateway, the piece that still needs to be provisioned on Unified CM is the transformation of the +E.164 called party information to the digit string to be sent to the PSTN according to Table 2-33. Table 24 depicts the required called party transformation patterns for localizing +E.164 for ISDN dialing.

Table 2-34 Called Party Transformation Patterns to Localize +E.164 for ISDN via SIP

Pattern	Partition	Transformation	Description
\+.1!	USGWLocalizeCd	Strip pre-dot, prefix *	+12125551234 –> *12125551234
\+.!	USGWLocalizeCd	Strip pre-dot, prefix *011	+4961007739764 –> *0114961007739764

To apply the called party transformations defined by the called party transformation patterns shown in Table 2-34 to a gateway, a CSS USGWLocalizeCd with only partition USGWLocalizeCd in it needs to be defined, and this CSS is then set as **Called Party Transformation CSS** in the **Device Mobility Related Information** section on the gateway's device pool. Configuring these transformations on the device pool enables sharing the same settings with multiple gateways in the same site sharing the same called party transformation requirements. To achieve this, the **Use Device Pool Called Party Transformation CSS** option needs to be checked in the **Outbound Calls** section on the gateway configuration page.

Also, we need to provision the transformation required to force the calling party number from +E.164 to whatever needs to be sent to the service provider. Here we need to consider how to treat calling party information for a call originating from a non-DID or a call originating from a DN that is not part of the DID range associated with the given gateway. The most common option is to set the caller ID to a site-specific main extension. This site specificity requires creation of site-specific calling party transformations as illustrated by Table 2-35.

Table 2-35 Calling Party Transformation Patterns to Localize +E.164 for ISDN via SIP

Pattern	Partition	Transformation	Description
\+.19195551XXX	RTPGWLocalizeCn	Strip pre-dot	+19195551001 –> 19195551001 Forward caller ID from the DID range associated with the gateway, but strip the leading plus (+), assuming that the calling party number can be sent to the provider as 1 plus 10 digits
\+!	RTPGWLocalizeCn	Mask 19195551888	Force everything to 19195551888
!	RTPGWLocalizeCn	Mask 19195551888	Force everything to 19195551888

The calling party transformation patterns in Table 2-35 perform the required transformations that make sure any calling party number, whether in the form of a +E.164 number or an enterprise specific number not matching the trunks DN range, is forced to a main number (19195551888).

To enable these transformations equivalent to the above method to apply outbound called party transformations, a CSS RTPGWLocalizeCn needs to be created using only partition RTPGWLocalizeCn, and this CSS needs to be applied as the calling party transformation CSS in the **Outbound Calls** section on the gateway configuration page or in the **Device Mobility Related Information** section on the gateway's device pool.

If a specific called or calling party transformation is needed per gateway, then using the device pool level settings for the called party transformations is overly complicated. In that case uncheck the **Use Device Pool Called/Calling Party Transformation CSS** options in the **Outbound Calls** section on the gateway configuration page, and set the called or calling party transformation CSS there.

Outbound Calls: Called and Calling Number Transformations on SIP Trunks

As mentioned earlier, SIP does not have the concept of "typed" numbers. Usually on SIP trunks all called and calling party numbers need to be sent in a single format independent of the type of called destination. The most common options are +E.164 or E.164. To enable easier dial-peer configuration with non-overlapping destination patterns for inbound and outbound calls, again we want to prefix all E.164 called party information with "*" when sent to the Cisco Unified Border Element terminating the SIP trunk.

If E.164 needs to be sent (without the +), then the above approach using called party transformation patterns can be reused. The single called party transformation shown in Table 2-36 is enough to make sure that the leading + of all +E.164 numbers gets stripped. Again we also need to create a CSS (for example, GWNoPlus) that addresses only partition GWNoPlus, and then apply this called party transformation pattern as **Called Party Transformation CSS** on either the gateway or the gateway's device pool.

*Table 2-36 Called Party Transformation Pattern to Localize +E.164 to *E.164 for SIP*

Pattern	Partition	Transformation	Description
\+.!	GWNoPlus	Strip pre-dot, prefix *	+4961007739764 –> *4961007739764 +12125551234 –> *12125551234

Even if no format transformation is required for calling party information sent on a SIP trunk, some filtering still needs to be applied to the calling party information to make sure that only valid numbers are sent to the provider. The same calling party transformations as described before in section on Outbound Calls: Called and Calling Number Transformations on ISDN Gateways and summarized in Table 2-35 can be used. Cisco IOS voice translations on Cisco Unified Border Element then make sure that the calling party information is sent to the provider according to the format requirements of the provider. Example 2-3 shows Cisco IOS voice translations to be applied on the VoIP dial-peer on the Cisco Unified Border Element (CUBE) pointing to the provider. These translations transform called party information from *E.164 to +E.164 and the calling party information from E.164 to +E.164.

Example 2-3 Cisco IOS Voice Translations to Force +E.164 Calling and Called Party Number on CUBE

```
voice translation-rule 2
   rule 1 /^\*/ /+/
   rule 2 // /+/
voice translation-profile SIPtoE164
   translate called 2
   translate calling 2
dial-peer voice 2 voip
   translation-profile outgoing SIPtoE164
```

Rule 1 in Example 2-3 replaces a leading "*" with a leading "+" while rule 2 just prefixes a "+" to all numbers.

Inbound Calls: Called and Calling Number Transformations on ISDN Gateways

Because all call routing on Unified CM is based on +E.164 for all incoming calls arriving at Unified CM, we need to make sure that called party information is transformed to +E.164 from the format received on the link from the provider. As mentioned earlier in the section on Outbound Calls: Called and Calling Number Transformations on ISDN Gateways, calling and called party information sent and received on ISDN trunks is a triplet consisting of numbering plan, number type, and number. Because SIP does not support number types, the semantics of the number type as received from the provider is lost if only the actual number is forwarded by the gateway over the SIP trunk to Unified CM. To avoid this, Cisco IOS voice translation needs to be deployed on the gateway to create a +E.164 digit string to be sent to Unified CM based on the received number plan, type, and number. Example 2-4 shows the Cisco IOS voice translation configuration to achieve this.

Example 2-4 Cisco IOS Voice Translations to Map from ISDN to +E.164

```
voice translation-rule 3
    rule 1 /^\(.+\)$/ /+1\1/ type national unknown plan any unknown
    rule 2 /^\(.+\)$/ /+\1/ type international unknown plan any unknown
voice translation-profile ISDNtoE164
    translate called 3
    translate calling 3
dial-peer voice 1 pots
    translation-profile incoming ISDNtoE164
```

The Cisco IOS translation shown in Example 2-4 assumes that we received called party information as type national and that the number in this case has only 10 digits. Rule 1 matches on any number (/^\(.+\)$/) with type international and simply prefixes +1 (/+\1/) while forcing plan and type to unknown because both are irrelevant when forwarded on the SIP trunk to Unified CM. The same translation rule is applied to both calling and called party information in translation profile ISDNtoE164, so that the calling party information as a 10-digit number with type national will be transformed correctly to +E.164 by rule 1. Rule 2 does not really apply to received called party information because the provider will typically send called party information using only a single format. Hence, rule 2 is relevant only for calls received from international destinations for which we expect to receive calling party information as type international with the number set to the full E.164 number of the calling party.

Different number formats might be used, depending on the provider, and this will require use of different transformations on the gateway or on Unified CM. For a detailed explanation of voice translation rules, see the document on *Number Translation using Voice Translation Profiles*, available at

> http://www.cisco.com/c/en/us/support/docs/voice/call-routing-dial-plans/64020-number-voice-translation-profiles.html

If for some reason the same rules cannot be used for calling and called party information transformation, then separate voice translation rules need to be provisioned for calling and called party information and associated with translation of calling and called party information in one translation profile.

Using inbound Cisco IOS voice translation rules is required only if different number types are sent by the provider. If the number type for calling or called party information is always unknown, for example, then the digit transformation to globalized +E.164 can happen on Unified CM either by using the inbound prefixes for calling and called party information or by using calling and called party transformation CSSs. Both prefixes and calling and called party transformations can be defined either on the trunk level or on the device pool level. Keep in mind that, because SIP does not support different number types, inbound calling and called prefixes or CSSs need to be set for number type **unknown** if set on the device pool level.

Inbound Calls: Called and Calling Number Transformations on SIP Trunks

Inbound call number information treatment on PSTN SIP trunks is generally simpler than the number handling in the ISDN case described before. The main reason is that number information on SIP trunks is not typed, and thus transformations are less complex and need to consider only the received digit string. Typically both calling and called party information on a SIP trunk is already in +E.164 format, and thus no transformations are needed.

If calling and called parties are received in E.164 format, then the easiest way to transform to +E.164 is to simply configure a prefix "+" on the SIP trunk in Unified CM or on the trunk's device pool. This prefix can be configured in the Incoming Calling Party Settings or Incoming Called Party Settings on the trunk or the trunk's device pool. Remember that for SIP trunks the setting for number type **Unknown Number** is relevant on the device pool level.

Calling Party Information Display on Phones

Because all directory numbers are provisioned as +E.164 numbers for calls originating from these +E.164 directory numbers, calling party information is in +E.164 format automatically. To simplify and provide consistent calling party presentation for all possible call flows, all calling party information received from outside networks such as the PSTN is normalized to +E.164 as discussed earlier. When a call is presented to a phone or to an outside network, the calling party information presented for that call sometimes needs to be transformed to the format expected by the network in case of the call being sent to a gateway or the format expected by the user in case of the call being sent to a phone.

Of special consideration are calls originating from phones with non-DIDs. In this case the only available calling party information is identical to the provisioned non-DID in the format of an enterprise specific number (ESN). Table 2-10 summarizes the ESN ranges used in the example topology.

On phones, sometimes +E.164 is not the preferred calling party display information, although keeping this information as +E.164 simplifies the deployment and is preferred. In that case the desired format typically depends on both the calling and called entities. Table 2-37 shows an example of the expected calling party display on a phone in site SJC for calls from various sources.

Table 2-37 Expected Calling Party Display on SJC Phone

Calling Entity "Native" Calling Party Information	Expected Display	Comment
+12125551234	912125551234	Call from US; display follows PSTN dialing habit.
+14085554001	4001	Call from +E.164 DN in the SJC DID range; display follows abbreviated intra-site dialing habit.

Table 2-37 Expected Calling Party Display on SJC Phone (continued)

Calling Entity "Native" Calling Party Information	Expected Display	Comment
81405001	5001	Call from non-DID in the SJC ESN range (see Table 2-10); display follows abbreviated four-digit intra-site dialing to non-DIDs in site SJC.
+4961007739764	90114961007739764	Call from international PSTN destination; display follows US PSTN dialing habit for international destinations.

To achieve the display format depicted in Table 2-37, calling party transformation patterns need to be provisioned in adequate partitions, and calling party transformation CSSs based on these partitions have to be configured on the phones, to enable the transformations.

Table 28 summarizes all calling party transformation patterns that must be provisioned to achieve the abbreviated calling party number display shown in Table 2-37 for all US sites based on the number ranges shown in Table 2-10.

Table 2-38 Phone Localization Calling Party Transformation Patterns

Pattern	Partition	Transformation	Description
\+.1!	USPhLocalize	Strip pre-dot, prefix 9	Any US destination: +12125551234 –> 912125551234
\+.!	USPhLocalize	Strip pre-dot, prefix 9011	Any international destination: +4961007739764 –> 90114961007739764
\+14085554XXX	SJCPhLocalize	Mask 4XXX	Call from local DN range: +14085554001 –> 4001
81405XXX	SJCPhLocalize	Mask 5XXX	Call from local non-DID range: 81405001 –> 5001
\+19725555XXX	RCDPhLocalize	Mask 5XXX	Call from local DN range: +19725555001 –> 5001
81976XXX	RCDPhLocalize	Mask 6XXX	Call from local non-DID range: 81976001 –> 6001
\+19195551XXX	RTPPhLocalize	Mask 1XXX	Call from local DN range: +19195551001 –> 1001
81912XXX	RTPPhLocalize	Mask 2XXX	Call from local non-DID range: 81912001 –> 2001

Table 2-39 shows the calling party transformation CSSs to enable calling party localization for phones in all US sites. The schema allows the reuse of dialing domain (country) specific calling party localization transformation patterns for all sites in that dialing domain (country). The country specific calling party localization patterns basically map national and international numbers to the country specific national and international dialing habit.

Table 2-39 Phone Localization Calling Party Transformation CSSs for US Sites

CSS	Partitions
SJCPhLocalize	SJCPhLocalize USPhLocalize
RCDPhLocalize	RCDPhLocalize USPhLocalize
RTPPhLocalize	RTPPhLocalize USPhLocalize

Table 2-40 shows an example of the country specific phone localization calling party transformation patterns that would need to be provisioned for Italy and Germany.

Table 2-40 Phone Localization Calling Party Transformation Patterns for Italy and Germany

Pattern	Partition	Transformation	Description
\+49.!	DEPhLocalize	Strip pre-dot, prefix 00	Any German destination: +4941001234 –> 0041001234
\+.!	DEPhLocalize	Strip pre-dot, prefix 000	Any international destination: +14085551234 –> 00014085551234
\+39.!	ITPhLocalize	Strip pre-dot, prefix 0	Any Italian destination: +390730123456 –> 00730123456 +393012345678 –> 03012345678
\+.!	ITPhLocalize	Strip pre-dot, prefix 000	Any international destination: +14085551234 –> 00014085551234

Automated Alternate Routing

Automated alternate routing (AAR) is a mechanism that reroutes calls to registered endpoints via an alternate route through the PSTN in case sufficient bandwidth is not available (call admission control does not allow the call) between the originating endpoint, gateway, or trunk and the called endpoint. AAR applies only to calls to endpoints. Insufficient bandwidth for calls to other destinations such as gateways and trunks does not trigger AAR. For those cases, the alternate routing mechanism is based on route lists and route groups. The following steps are required to activate AAR:

- Set the Automated Alternate Routing Enable service parameter (see the section on Service Parameter Settings).
- Configure a single AAR group **Default** without any Dial Prefix (default).
- Define a CSS PSTNReroute with access only to +E.164 PSTN route patterns. Based on the examples in this design, this CSS would need to include only partition PSTNInternational.
- On all endpoints, trunks, and other devices initiating calls that potentially might be subject to AAR:
  - Set the AAR Calling Search Space to PSTNReroute.
  - Set AAR Group to **Default**.
- On all device pools, set the AAR Calling Search Space to PSTNReroute.
- On all device pools, set AAR Group to **Default**

- On +E.164 directory numbers, configure the AAR masks so that the resulting number is the +E.164 number of the directory number. In a country with a fixed length numbering plan, the mask can be set to some identical value on all directory numbers (such as +1XXXXXXXXXX in the US). If variable length directory numbers need to be covered, more specific masks covering a single site or, as a worst case scenario, a fully qualified +E.164 AAR mask identical to the respective directory number have to be provisioned. For non-DIDs the AAR mask is left empty. This effectively disables AAR if a non-DID is called. This makes sense because a non-DID does not have an equivalent E.164 address and thus cannot be reached via the PSTN.

The above list shows one of the advantages of a dial plan using +E.164 directory numbers. In this case the called +E.164 address can be reused directly for alternate dialing over the PSTN without applying any other modifications.

Alternate Routing for Unregistered Endpoints

In case of a WAN failure in a multi-site deployment with centralized call processing, endpoints in the affected lose connectivity with the centralized Unified CM and register with a local SRST gateway instead (see the section on Survivable Remote Site Telephony (SRST) Deployment). This allows the affected phones to still place and received calls to and from phones in the same site and the PSTN. Calls from phones registered with the central Unified CM will fail, though, because from the central Unified CM’s perspective the called device is unregistered and thus unreachable. To enable automatic rerouting of calls to unregistered endpoints over the PSTN, perform the following tasks on each directory number that requires automatic rerouting:

- Set the Forward Unregistered Internal and Forward Unregistered External destination to the same value as the +E.164 directory number.
- Set the Forward Unregistered Internal and Forward Unregistered External CSS to PSTNReroute. This is the same CSS as defined in the section on Automated Alternate Routing, and it allows access to PSTN route patterns.

Alternate routing over the PSTN for unregistered endpoints makes sense only for endpoints with +E.164 directory numbers. For endpoints without a DID (endpoints with an ESN as directory number), the only meaningful rerouting for unregistered endpoints is to forward incoming calls to voicemail. To forward calls to unregistered endpoints to voicemail, perform these tasks:

- Select the Voicemail options for Forward Unregistered Internal and Forward Unregistered External.
- Set the Forward Unregistered Internal and Forward Unregistered External CSS to a CSS implementing class of service Internal (for example, SJCInternal). Effectively this CSS has to provide access to only the voicemail pilot number.

LDAP System Configuration

Before defining the actual synchronization agreements, the LDAP system has to be enabled. In the LDAP System Configuration menu, do the following:

- Select (check) the **Enable Synchronizing from LDAP Server** option
- Select the correct LDAP Server Type for your deployment.
- Select the correct LDAP Attribute for User ID for your deployment.

In an environment where users are synchronized from Microsoft Active Directory, use the settings shown in Table 2-41.

Table 2-41 LDAP System Settings for Microsoft Active Directory

Setting	Value
LDAP Server Type	Microsoft Active Directory
LDAP Attribute for User ID	sAMAccountName

LDAP Custom Filter

If a Unified CM based directory search is used on phones, then it does make sense to synchronized the full corporate LDAP directory to Unified CM. In that case we need to be able to differentiate between users who actually use UC services of the local cluster and users who are synchronized only to reflect the complete corporate LDAP directory on Unified CM.

To achieve this goal, custom LDAP filters can be used to define two groups of users: local and remote. Remote here means that these users do not use any UC services on the local Unified CM cluster. Table 2-42 shows two custom LDAP filters, assuming that our deployment has users in the US and Europe and that only the US users are considered as local users.

Table 2-42 Custom LDAP Filter Settings

LDAP Filter Name	Filter

Local	`(&`
	`    (objectclass=user)`
	`    (!(objectclass=Computer))`
	`    (!(UserAccountControl:1.2.840.113556.1.4.803:=2))`
	`    (telephoneNumber=+1*)`
	`)`

Remote	`(&`
	`    (objectclass=user)`
	`    (!(objectclass=Computer))`
	`    (!(UserAccountControl:1.2.840.113556.1.4.803:=2))`
	`    (|`
	`        (telephoneNumber=+3*)`
	`        (telephoneNumber=+4*)`
	`    )`
	`)`

For better readability, the LDAP filter strings in Table 2-42 are separated into multiple lines, with the indentation levels reflecting the structure of the LDAP filter strings. To provision these LDAP filters in Unified CM, you have to concatenate all lines of a given filter into a single line.

Both LDAP filters are extensions of the default LDAP filter for Microsoft Active Directory. Default LDAP filters for other directory types can be found in the chapter on *Directory Integration and Identity Management* in the *Cisco Collaboration System 11.x SRND* and in the Unified CM online help for the LDAP directory settings.

The LDAP filters in Table 2-42 use the beginning of the phone numbers as criteria to determine whether the individual user is a local or a remote user.

When using multiple LDAP synchronization agreements, you have to make sure that the LDAP filters used by these synchronization agreements are disjunct so that no single user is matched by both filters.

Feature Group Templates

Capabilities of users synchronized from LDAP are defined in a feature group template (FGT). Table 2-43 summarizes the settings for the FGT defining the capabilities of users with active devices on the Unified CM cluster.

Table 2-43 Feature Group Template for Local Users

Setting	Value	Comment
Name	FGTlocal	Name should indicate that this is an FGT used for local users.
Description	FGT for local users	
Home Cluster	Checked	Make sure that UDS-based service discovery for this user resolves to the local Unified CM cluster.
Enable User for Unified CM IM and Presence	Checked	Enable the user for IM and Presence
BLF Presence Group	Standard Presence Group	Single BLF presence group for all users, to simplify the deployment.
SUBSCRIBE Calling Search	DefaultSubscribe	Use the default subscribe CSS described in the section on Special CSSs.

All other settings can be left as default values.

Because remote users are also synchronized from LDAP (see the section on LDAP Custom Filter), an FGT for remote users must also be provisioned. The key difference is that in that FGT the **Home Cluster** and **Enable User for Unified CM IM and Presence** options are not checked. Table 2-44 summarizes these settings.

Table 2-44 Feature Group Template for Remote Users

Setting	Value	Comment
Name	FGTremote	Name should indicate that this is an FGT used for remote users.
Description	FGT for remote users	

Table 2-44 Feature Group Template for Remote Users (continued)

Setting	Value	Comment
Home Cluster	Not checked	Make sure that UDS-based service discovery for this user does not resolve to the local Unified CM cluster.
Enable User for Unified CM IM and Presence	Not checked	Do not enable the user for IM and Presence.

All other settings can be left as default values.

LDAP Synchronization Agreements

To synchronize all local users to Unified CM, an LDAP synchronization agreement needs to be configured. Table 2-45 shows the required settings to be configured under **System/LDAP/LDAP Directory**.

Table 2-45 LDAP Synchronization Agreement for Local Users

Setting	Value	Comment
LDAP Configuration Name	Local	Indicates that this LDAP synchronization agreement synchronizes local users.
LDAP Manager Distinguished Name	Name of admin users	Can be in the form of ldapaccess@ent-pa.com or cn=ldapaccess,cn=users,dc=ent-pa,dc=com
LDAP Password	Password of the LDAP admin	
LDAP User Search Base	LDAP Search base	Example: dc=ent-pa,dc=com
LDAP Custom Filter	Local	Refers to the custom LDAP filter described in the section on LDAP Custom Filter.
Perform Sync Just Once	Unchecked	LDAP synchronization is executed periodically.
Perform a Re-sync Every	Reasonable interval	Make sure to set the interval small enough to pick up corporate directory changes in a reasonable time, but keep in mind that executing the LDAP synchronization creates significant load on the Unified CM publisher. Synchronizing once every 24 hours probably is a good default.
Directory URI	mail	Typically directory URIs of users are identical to their email addresses.
Access Control Groups	Standard CCM End Users Standard CTI Enabled	Add or remove other access control groups as needed, but keep in mind that without Standard CCM End Users, the users will not be able to log into the self-service portal.
Feature Group Template	Local	Refers to the FGT described in the section on Feature Group Templates.
LDAP Server Information	References to corporate LDAP servers to be uses as source	Make sure to provision redundant servers, if possible.

The LDAP synchronization agreement in Table 2-45 ties together the FGT and custom LDAP filter defined before. This makes sure that, for all users in the corporate directory matching the custom LDAP filter, a user on Unified CM is created with the capabilities defined in the FGT.

A dedicated LDAP synchronization agreement is also required to synchronize the remote users who do not use UC services on the local Unified CM cluster. Table 2-46 summarizes the settings for this LDAP synchronization agreement.

Table 2-46 LDAP Sync Agreement for Remote Users

Setting	Value	Comment
LDAP Configuration Name	Remote	Indicates that this LDAP synchronization agreement synchronizes remote users.
LDAP Manager Distinguished Name	Name of admin users	Can be in the form of ldapaccess@ent-pa.com or cn=ldapaccess,cn=users,dc=ent-pa,dc=com
LDAP Password	Password of the LDAP admin	
LDAP User Search Base	LDAP Search base	Example: dc=ent-pa,dc=com
LDAP Custom Filter	Remote	Refers to the custom LDAP filter described in the section on LDAP Custom Filter.
Perform Sync Just Once	Unchecked	LDAP synchronization is executed periodically.
Perform a Re-sync Every	Reasonable interval	Make sure to set the interval small enough to pick up corporate directory changes in a reasonable time, but keep in mind that executing the LDAP synchronization creates significant load on the Unified CM publisher. Synchronizing once every 24 hours probably is a good default.
Directory URI	mail	Typically directory URIs of users are identical to their email addresses.
Access Control Groups	No access control groups selected	Remote users are not members of any access control group.
Feature Group Template	Remote	Refers to the FGT described in the section on Feature Group Templates.
LDAP Server Information	References to corporate LDAP servers to be uses as source	Make sure to provision redundant servers, if possible.

Using the above LDAP synchronization agreements, all users can be identified from the corporate directory, and the FGTs associated with the LDAP synchronization agreements make sure that capabilities are configured correctly for all users.

User Authentication with LDAP

Table 2-47 shows an example of LDAP authentication settings.

Table 2-47 LDAP Authentication Settings

Setting	Example	Comment
LDAP Authentication for End Users		
Use LDAP Authentication for End Users	Checked	Enables LDAP authentication for the Unified CM cluster.
LDAP Manager Distinguished Name	cn=ldapmanager,dc=ent pa,dc=com	Distinguished name of an AD account with read access rights to all user objects in the desired user search base.
LDAP Password	Some password	
Confirm Password	Same as above	
LDAP User Search Base	ou=enterprise,dc=ent-pa,dc=com	
LDAP Server Information		
Host Name or IP Address for Server	ent-dc1.ent-pa.com	Server with global catalog role
LDAP Port	3268	Port to access global catalog (recommended)

Cisco Unified CM Group Configuration

Cisco Unified CM groups allow you to define groups of Unified CM instances in the cluster that determine which Unified CM instances should be used by devices to register to the Unified CM cluster. If only a single Unified CM call processing pair is deployed (see the section on Provision the Cisco Unified CM and IM and Presence Service Cluster for more information), then a single Unified CM group named Default also needs to be deployed, and both Unified CM instances running on the single pair of Unified CM call processing subscribers in the cluster have to be members of this single Unified CM group.

If more than one pair of Unified CM call processing subscribers exists, then additional Unified CM groups need to be provisioned (one for each pair of Unified CM call processing subscribers), and in each Unified CM group the two Unified CM instances running on that specific pair are added to the group.

For a Unified CM cluster with two pairs of Unified CM call processing subscribers named ucm1a.ent-pa.com and ucm1b.ent-pa.com in the first pair and ucm2a.ent-pa.com and ucm2b.ent-pa.com in the second pair, with ucm1a and ucm2a being the primary Unified CM call processing subscribers in each pair, Table 2-48 lists the Unified CM groups to be provisioned.

Table 2-48 Example Unified CM Group Definition

Unified CM Group	Unified CM Group Members
CM_1	CM_ucm1a.ent-pa.com CM_ucm1b.ent-pa.com
CM_2	CM_ucm2a.ent-pa.com CM_ucm2b.ent-pa.com

All registrations have to be equally balanced between Unified CM groups. This is achieved by assigning devices to Unified CM groups via device pool configuration as discussed in the section on Device Pools.

Phone NTP References

If you want to do so, you can configure phone Network Time Protocol (NTP) references in Cisco Unified Communications Manager Administration to ensure that a phone running SIP gets its date and time from the NTP server. If all NTP servers do not respond, the phone that is running SIP uses the date header in the 200 OK response to the REGISTER message for the date and time.

After you add the phone NTP reference to Cisco Unified CM Administration, you must add it to a date/time group.

To define phone NTP references, get the IP addresses of the NTP servers you plan to use, and configure the settings according to Table 2-49.

Table 2-49 Phone NTP Reference Settings

Setting	Example	Comment
IP Address	66.228.35.252	IP address of NTP server to be used
Description	0.pool.ntp.org	Should refer to the hostname of the IP address being entered
Mode	Unicast	Unicast limits devices to using only NTP response from listed servers

Make sure to provision multiple phone NTP references for redundancy.

Date and Time Groups

Date and time groups allow you to define the time zone and the date and time format to be used for sets of devices registering with Unified CM. The date/time group configuration is specified in the device pool, and the device pool is specified on the phone page. For more information on device pools, see the section on Device Pools.

If you want SIP phones to get their date and time from NTP servers, then in the date/time group you prioritize the phone NTP references, starting with the first server that you want the phone to contact.

Create one named Date/Time Group for each of the time zones in which you will deploy endpoints, as illustrated in Table 2-50.

Table 2-50 Example Date/Time Group Definitions

Date and Time Group	Time Zone
RCD_Time	America/North_Dakota/New_Salem
RTP_Time	America/New_York
SJC_Time	America/Los_Angeles

Media Resources

A media resource is a software-based or hardware-based entity that performs media processing functions on the data streams to which it is connected. Media processing functions include mixing multiple streams to create one output stream (conferencing), passing the stream from one connection to another (media termination point), converting the data stream from one compression type to another (transcoding), streaming music to callers on hold (music on hold), echo cancellation, signaling, voice termination from a TDM circuit (coding/decoding), packetization of a stream, streaming audio (annunciation), and so forth. The software-based resources are provided by the Cisco Unified CM IP Voice Media Streaming Application.

Media Resource Manager

The Media Resource Manager (MRM), a software component in the Unified CM, determines whether a media resource needs to be allocated and inserted in the media path. When the MRM decides and identifies the type of the media resource, it searches through the available resources according to the configuration settings of the media resource group list (MRGL) and media resource groups (MRGs) associated with the devices in question. MRGLs and MRGs are constructs that hold related groups of media resources together for allocation purposes

Media Resource Selection and Avoiding the Default MRG

Media resource groups (MRGs) and media resource group lists (MRGLs) provide a method to control how resources are allocated, which could include rights to resources, location of resources, or resource type for specific applications. MRGs are used to group together media resources of similar characteristics, and MRGLs define a set of MRGs to be considered when selecting a required media resource for a session. If the Media Resource Manager does not find a required resource by searching through a configured MRGL, considering all media resources being members of MRGs of that list, then the Media Resource Manager checks a default media resource group for media resources. All media resources by default are members of this default MRG unless they are explicitly configured to be members of any specific MRG.

In this design we will not use the default MRG because it makes troubleshooting of media resource selection more complicated. To make sure that the default MRG is empty, you have to assign all media resources to at least one MRG.

Cisco IP Voice Media Streaming Application

The Cisco IP Voice Media Streaming Application provides the following software-based media resources:

- Conference bridge
- Music on Hold (MoH)
- Annunciator
- Media termination point (MTP)

When the IP Voice Media Streaming Application is activated on a node in the Unified CM cluster, one of each of the above resources is automatically configured. For service activation recommendations, see Table 2-5.

In this design only unicast MoH is used, with media being streamed from the Cisco IP Voice Media Streaming Application running on the Unified CM cluster subscriber nodes.

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.

All MOH and annunciator media resources created by the Cisco IP Voice Media Streaming Application running on Unified CM are combined in a single MRG by performing the following tasks:

- Create an MRG named Software.
- Assign all annunciator resources created by the Cisco IP Voice Media Streaming Application to MRG Software.
- Assign all MoH resources created by the Cisco IP Voice Media Streaming Application to MRG Software.

The software-based conferencing and media termination points created by the Cisco IP Voice Media Streaming Application are not used in this design. To disable them, perform the following tasks:

- Create an MRG named Unused.
- Assign all software-based conference bridges created by the Cisco IP Voice Media Streaming Application to MRG Unused.
- Assign all software-based media termination points created by the Cisco IP Voice Media Streaming Application to MRG Unused.

This makes sure that these resources are not part of the default MRG any longer and are never considered in the Media Resource Manager media resource selection process.

MRG and MRGL Definitions

Its good practice to keep the number of provisioned MRGLs to a minimum. Factors contributing to the number of required MRGLs include:

- Site specificity

  If site-specific media resources exist, then site-specific MRGs for those resources need to be configured, and typically also site-specific MRGLs are required to allow for site-specific selection of (typically local) media resources.

- Different types of media resources of the same class

  Unified CM does not differentiate between audio-only and audio/video conferencing resources. If both audio and audio/video conferencing media resources are provisioned, then an MRG (and MRGL) is required per type of media resource to allow configuration of differential access policies to these resources. See the Conferencing chapter for more details on conferencing resources.

If no site-specific media resources and no differentiation of media resource types is required, then at least a single MRGL named Standard needs to be configured.

For each required MRGL based on site specificity and media resource type provision, create an MRGL by performing the following tasks:

- Set the MRGL name so that it reflects the site specificity and media resource type of the MRGL.
- Select the desired MRGs for the MRGL. Make sure to always include the Software MRG so that access to MoH and Annunciator is ensured.

Table 2-51 shows example MRGL definitions that provide differentiated treatment of audio and video conferencing. MRGL Audio would need to be assigned to devices requiring access to audio conferencing media resources only, while MRGL Video would allow access to video conferencing resources.

Table 2-51 Example MRGL Definition with Audio and Video Conferencing

MRGL Name	MRGs	Comment
Audio	Audio Software	MRGL with access to audio conferencing media resources in MRG Audio. MRG Software added to provide access to MoH and annunciator.
Video	Video Software	MRGL with access to video conferencing media resources in MRG Video. MRG Software added to provide access to MoH and annunciator.

Device Pools

Device pools define sets of common characteristics for devices. Characteristics defined on the device pool include the settings shown in Table 2-52.

Table 2-52 Device Pool Settings

Setting	Description
Cisco Unified Communications Manager Group	Unified CM groups are needed to distribute registrations equally among Unified CM call processing subscriber pairs (see the section on Cisco Unified CM Group Configuration). The Unified CM Group provisioned on the device pool determines the Unified CM call processing subscribers to which devices associated with the given device pool will try to register.
Local Route Groups	As described in the section on Local Route Groups for Call Type Specific Outbound Gateway Selection, multiple LRGs are defined to allow for call type specific egress gateway selection based on LRGs. For each defined LRG name, the route group selected for that LRG name defines which devices will be considered for a call of the selected type (defined by the route pattern matching on the called number and pointing to a route list referring to specific LRGs). It is important to set route groups for all defined LRG names to avoid call failures due to route lists not containing any valid PSTN resources.
Roaming Sensitive Settings	
Date/Time Group	Defines date and time format and phone NTP references. See the section on Phone NTP References.
Media Resource Group List	MRGL defining the media resources available for a group of devices. See the section on MRG and MRGL Definitions.
Device Mobility Related Information	
AAR Calling Search Space	The CSS used to route calls to an alternate PSTN destination. The dial plan design in this document allows use of the same AAR CSS (PSTNReroute) in all cases (see the section on Automated Alternate Routing).
AAR Group	To enable AAR, an AAR group has to be defined. Using +E.164 directory numbers allows you to deploy AAR using a single AAR group, Default (see the section on Automated Alternate Routing).
Calling Party Transformation CSS	This CSS defines the calling party transformations applied to calling party information sent in the direction of the affected device. For gateways this CSS is tied to the calling party transformation CSS defined in the Outbound Calls section on the gateway configuration page. For phones this CSS is tied to the calling party transformation CSS defined in the Remote Number section on the phone configuration page.
Called Party Transformation CSS	This CSS defines the called party transformations applied to called party information sent in the direction of the affected device. For gateways this CSS is tied to the called party transformation CSS defined in the Outbound Calls section on the gateway configuration page. For phones this CSS has no equivalent on the phone configuration page and does not have any effect when configured on a device pool used for phones.
Call Routing Information	This setting allows you to define incoming calling and called party transformations per numbering type to be applied to incoming calls on gateways. The same settings also can be configured in the gateway configuration page if individual gateway-specific settings are required.

All other device pool level settings are not used in this design.

Whenever the same settings for the configuration options listed in Table 2-52 need to be applied to a group of devices, we recommend creating a device pool with these settings and then assigning all devices to this device pool. If one of the settings needs to be changed for all of the devices, the device pool level configuration allows you to change the setting for all devices at one point.

To minimize the number of device pools, create a device pool only if multiple devices share the same characteristics. An example of this is phones in the same site. Table 2-53 shows an example of device pool settings for phones with video conferencing capabilities in site RTP.

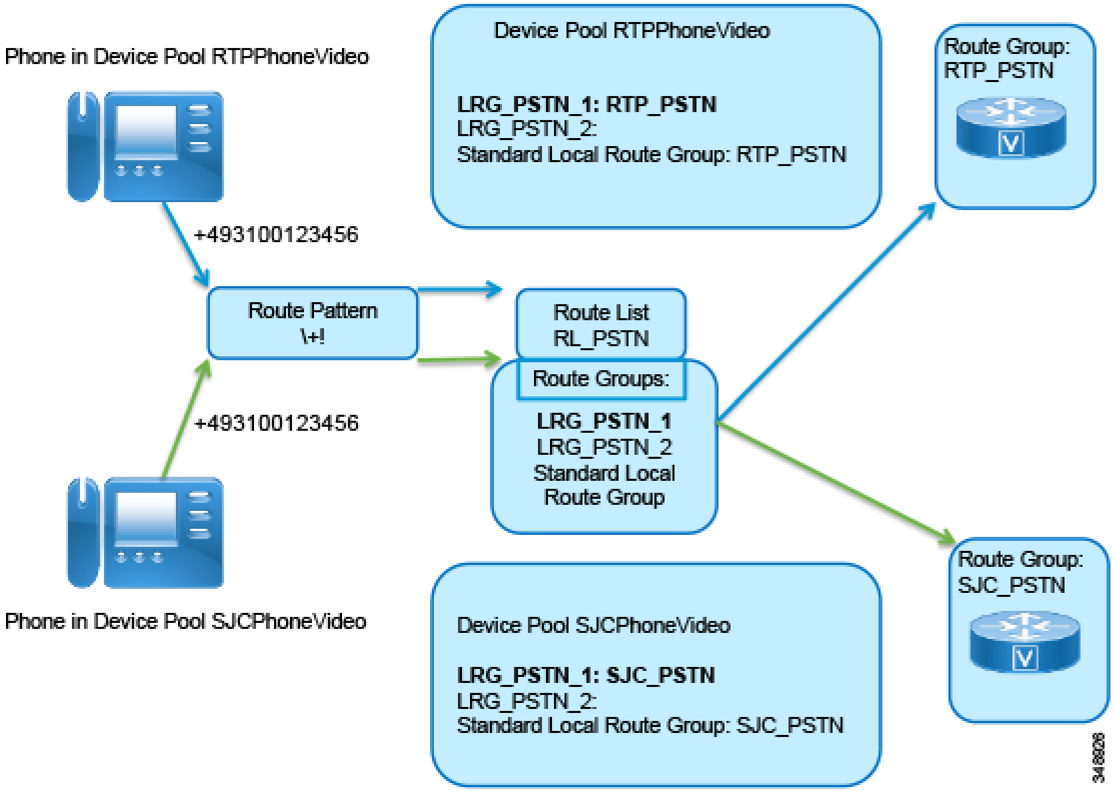
Table 2-53 Device Pool Settings for Phones with Video Conferencing Capabilities in Site RTP

Setting	Value	Comment
Device Pool Name	RTPPhoneVideo	Name should uniquely identify the devices (type and further classification) this device pool is used for. In this case we use this device pool for phones in site RTP with video conferencing capabilities
Cisco Unified Communications Manager Group	CM_1	
Local Route Group Settings		
Standard Local Route Group	RTP_PSTN	All route lists use Standard Local Route Group as last option. Always set Standard Local Route Group to the local PSTN gateways' route group.
LRG_PSTN_1	RTP_PSTN	First option for PSTN calls is to use local RTP gateways.
LRG_PSTN_2	SJC_PSTN	Use HQ gateways as fallback.
LRG_VIDEO_1	SJC_VIDEO	No site-specific video gateways exist. We use the video gateway in site SJC.
LRG_VIDEO_2	<None>	
LRG_EMERGENCY_1	<None>	No setting; fallback to Standard Local Route Group.
LRG_EMERGENCY_2	<None>	No setting; fallback to Standard Local Route Group.
Roaming Sensitive Settings		
Date/Time Group	RTP_Time	See the section on Date and Time Groups.
Media Resource Group List	Video	Provide access to video conferencing media resources (see Table 2-51).
Device Mobility Related Information		
AAR Calling Search Space	PSTNReroute	Same for all devices and device pools.
AAR Group	Default	Same for all devices and device pools.
Calling Party Transformation CSS	RTPPhLocalize	Site-specific calling party transformations (see Table 2-38 and Table 2-39).
Called Party Transformation CSS	<None>	Does not apply to phones.

Table 2-53 shows how the actual site-specific PSTN gateways are assigned to the LRG names to achieve the site-specific egress gateway selection for phones in different sites.

Figure 2-8 shows how different LRG selections for the same LRG name LRG_PSTN_1 on the device pools for phones in site RTP and SJC make sure that PSTN calls from phones in site RTP and SJC egress to the PSTN through different gateways although the same route pattern and route list are used for calls from both sites.

Figure 2-8 Site-Specific Egress Gateway Selection



From the example in Table 2-53 we can see that, following the same schema, we would need to provision two device pools per site to be able to differentiate between devices with and without video conferencing capabilities. If video conferencing capabilities were the exception, we could decide to use only one device pool per site with MRGL set to Audio and then on the few video-enabled devices set the MRGL to Video in the device configuration.

Table 2-54 summarizes the device pool settings of the device pool used for gateways in a specific site. Site RTP is used as an example here.

Table 2-54 Device Pool Settings for PSTN Gateways in Site RTP

Setting	Value	Comment
Device Pool Name	RTP_PSTN	Name should uniquely identify the devices (type and further classification) this device pool is used for. In this case we use this device pool for PSTN gateways in site RTP.
Cisco Unified Communications Manager Group	CM_1	
Local Route Group Settings		
Standard Local Route Group	RTP_PSTN	There actually is no call flow for which a PSTN trunk would need a PSTN resource. Also see the note on configuration order in the section on Route Groups. When you create the device pool, the required route group does not exist yet. Hence, initially you need to configure the device pool and leave the LRG mapping set to <None>. After configuring the SIP trunks and route groups, you can come back and set the LRG mapping.
LRG_PSTN_1	<None>	
LRG_PSTN_2	<None>	
LRG_VIDEO_1	<None>	
LRG_VIDEO_2	<None>	
LRG_EMERGENCY_1	<None>	
LRG_EMERGENCY_2	<None>	
Roaming Sensitive Settings		
Date/Time Group	RTP_Time	See the section on Date and Time Groups.
Media Resource Group List	Audio	Calls coming in from the PSTN would not require access to video conferencing resources.
Device Mobility Related Information		
AAR Calling Search Space	PSTNReroute	Same for all devices and device pools, although not really required for a PSTN trunk.
AAR Group	Default	Same for all devices and device pools, although not really required for a PSTN trunk.
Calling Party Transformation CSS	RTPGWLocalizeCn	Site-specific calling party transformations to make sure that only valid calling party information is sent (all numbers not belonging to the RTP DID range are masked). Also, the digit string is set to a format suitable for the ISDN gateway (see Table 2-35).
Called Party Transformation CSS	USGWLocalizeCd	See Table 2-34. This transformation makes sure that called party numbers are transformed from +E.164 to the format that can be sent as plan **unknown** and type **unknown**.
Call Routing Information		
Incoming Calling Party Settings	Nothing is configured here. We assume that the transformation from ISDN number format to +E.164 is achieved using Cisco IOS voice translation rules on the gateway (see the section on Inbound Calls: Called and Calling Number Transformations on ISDN Gateways).	
Incoming Called Party Settings		

Table 2-55 summarizes the device pool settings for a SIP trunks to other Unified CM clusters and application servers. SIP trunks to other Unified CM clusters do not require any transformations on calling and called part information because the called party number already is globalized to +E.164 by the dialing normalization translation patterns provisioned in the dial plan, and calling party information internal to Unified CM based on the provisioned dial plan is either +E.164 or an ESN and both formats make sense in the context of on-net intercluster calls.

Table 2-55 Device Pool Settings for Central Trunks and Applications

Setting	Value	Comment
Device Pool Name	Trunks_and_Apps	Name should uniquely identify the devices (type and further classification) this device pool is used for.
Cisco Unified Communications Manager Group	CM_1	
Local Route Group Settings		
Standard Local Route Group	RTP_PSTN	Trunks actually do not need PSTN access, but applications might require PSTN access. So PSTN resources of one site are selected via the Standard Local Route Group configuration. Other site's PSTN resources can be used as failover.
LRG_PSTN_1	RTP_PSTN	
LRG_PSTN_2	SJC_PSTN	
LRG_VIDEO_1	<None>	
LRG_VIDEO_2	<None>	
LRG_EMERGENCY_1	<None>	
LRG_EMERGENCY_2	<None>	
Roaming Sensitive Settings		
Date/Time Group	RTP_Time	See the section on Date and Time Groups.
Media Resource Group List	Video	Intercluster calls could potentially require video media resources.
Device Mobility Related Information		
AAR Calling Search Space	PSTNReroute	Same for all devices and device pools.
AAR Group	Default	Same for all devices and device pools.
Calling Party Transformation CSS	<None>	No transformations on intercluster trunks and trunks to application servers.
Called Party Transformation CSS	USGWLocalizeCd	No transformations on intercluster trunks and trunks to application servers.
Call Routing Information		
Incoming Calling Party Settings	Nothing configured. We assume that inbound calling and called party numbers already are normalized.	
Incoming Called Party Settings		

SIP Trunks

All connections to other entities, including call controls, applications, and conferencing resources, use SIP trunks.

SIP Profiles

A SIP profile comprises the set of SIP attributes that are associated with SIP trunks and SIP endpoints. To keep the number of SIP profiles to a minimum, follow these rules:

- Consider the default profiles first.
- Then consider already defined non-default profiles.
- Create a new SIP profile only if none of the default profiles match.
- Avoid defining profiles per trunk.

Table 2-56 shows the settings for a SIP profile to be used for all SIP IP phones and SIP trunks to other Unified CM clusters or SIP gateways.

Table 2-56 SIP Profile for SIP Phones and Standard Trunks

Setting	Value	Comment
Copy of Standard SIP Profile		
Name	FQDN	
Use Fully Qualified Domain Name in SIP Requests	Checked	Prevents IP address of Unified CM server from showing up in SIP calling party information sent by Unified CM.
Early Offer support for voice and video calls	Best Effort (no MTP inserted)	This is the recommended configuration for all Unified CM trunks. Best Effort Early Offer trunks never use MTPs to create an Early Offer and, depending on the calling device, can initiate an outbound SIP trunk call using either Early Offer or Delayed Offer. In the context of this design, outbound calls always use Early Offer.
Enable OPTIONS Ping to monitor destination status for Trunks with Service Type "None (Default)"	Checked	Allows monitoring of the reachability of SIP trunk peers; applies to SIP trunks only.
Ping Interval for In-service and Partially In-service Trunks (seconds)	10	One ping every 10 seconds, combined with a retry count of 6, makes sure that SIP trunk unavailability is detected within a minute.
Ping Interval for Out-of-service Trunks (seconds)	60	If a trunk is out of service, then we do not have to try to reach the peer as often.
Ping Retry Timer (milliseconds)	500	
Ping Retry Count	6	

SIP Trunk Security Profiles

Cisco CallManager Administration groups SIP trunk security-related settings – for example, device security mode, digest authentication, and incoming/outgoing transport type settings – so you can apply all configured settings to a SIP trunk when you choose the profile in the SIP Trunk Configuration window.

Table 2-57 shows the default settings on the system generated SIP trunk security profile Non Secure SIP Trunk Profile. This SIP trunk security profile is used for the SIP trunks to ISDN PSTN gateways, for example.

Table 2-57 Non Secure SIP Trunk Profile SIP Trunk Security Profile Settings

Setting	Value
Name	Non Secure SIP Trunk Profile
Device Security Mode	Non Secure
Incoming Transport Type	TCP+UDP
Outgoing Transport Type	TCP
Enable Digest Authentication	Not Checked
Incoming Port	5060
Enable Application level authorization	Not Checked
Accept presence subscription	Not Checked
Accept out-of-dialog refer	Not Checked
Accept unsolicited notification	Not Checked
Accept replaces header	Not Checked
Transmit security status	Not Checked
Allow charging header	Not Checked
SIP V.150 Outbound SDP Offer Filtering	Use Default Filter

Table 2-58 shows the settings for a SIP Trunk Security Profile used for a SIP trunk to the IM and Presence nodes, differing from the default settings in Table 2-57.

Table 2-58 SIP Trunk Security Profile for IM and Presence Trunk

Setting	Value	Comment
Name	IM and Presence	Meaningful name describing the use of the SIP Trunk Security Profile.
Accept Presence Subscription	Checked	
Accept Out-of-Dialog REFER	Checked	
Accept Unsolicited Notification	Checked	
Accept Replaces Header	Checked	

Table 2-59 shows the settings on the SIP trunk security profile to be used for intercluster trunks to other Unified CM clusters. On these trunks we want to accept presence subscriptions to enable intercluster Busy Lamp Field (BLF) presence.

Table 2-59 SIP Trunk Security Profile for Intercluster Trunks

Setting	Value	Comment
Name	ICT	Name describing the use of the SIP Trunk Security Profile.
Accept Presence Subscription	Checked	
Transmit Security Status	Checked	

SIP Trunk Connections

SIP trunks are the preferred way to set up connectivity between Unified CM clusters and between Unified CM and other systems such as gateways, applications, and media resources. Depending on the type of connected system, the parameters configured on each SIP trunk differ slightly. Table 2-60 summarizes the settings for a SIP trunk to a PSTN gateway in site RTP.

Table 2-60 SIP Trunk Settings for Trunk to ISDN Gateway in Site RTP

Setting	Value	Comment
Name	ST_RTP_PSTN_1	Prefix ST_ to avoid name collisions with other devices stored in the same table internally. The remainder of the name identifies the location of the gateway and allows numbers for multiple gateways.
Description		Some meaningful description
Device Pool	RTP_PSTN	Common device pool for all RTP PSTN gateways. Allows sharing of site-specific settings between all RTP gateways.
Media Resource Group List	<None>	Use the MRGL defined on the device pool.
AAR Group	Default	Same everywhere
PSTN Access	Checked	
Run On All Active Unified CM Nodes	Checked	This settings is recommended on all SIP trunks. This makes sure that outbound calls to SIP do not require intra-cluster control signaling between Unified CM call processing subscribers.
Inbound Calls		
Calling Search Space	DN	Inbound calls have +E.164 called party numbers, and only local destinations can be called from the PSTN. Hence, no access is required to ESN numbers and intercluster destinations.
AAR Calling Search Space	PSTNReroute	
Outbound Calls		
Use Device Pool Called Party Transformation CSS	Checked	

Table 2-60 SIP Trunk Settings for Trunk to ISDN Gateway in Site RTP (continued)

Setting	Value	Comment
Use Device Pool Calling Party Transformation CSS	Checked	
SIP Information		
Destination	X.X.X.X	IP address of ISDN gateway
SIP Trunk Security Profile	Non Secure SIP Trunk Profile	Default SIP trunk security profile
SIP Profile	FQDN	

The key here is that the inbound CSS provides access to local +E.164 destinations only. These include voicemail pilots or other services that need to be reachable from the PSTN, but no access is required to PSTN route patterns, dialing normalization translation patterns, ESNs, URIs, and intercluster destinations.

Settings for SIP trunks to other Unified CM clusters differ somewhat from the settings on SIP trunks to ISDN gateways. Table 2-61 summarizes these settings.

Table 2-61 SIP Trunk Settings for Intercluster Trunk to Other Unified CM Cluster

Setting	Value	Comment
Name	ST_UCM_EMEA	Prefix ST_ to avoid name collisions with other devices stored in the same table internally. The remainder of the name identifies the trunk's purpose.
Description		Some meaningful description
Device Pool	Trunks_and_Apps	Common device pool for central trunks (see Table 2-55)
Media Resource Group List	<None>	Use the MRGL defined on the device pool.
AAR Group	Default	Same everywhere
PSTN Access	Not checked	
Run On All Active Unified CM Nodes	Checked	This settings is recommended on all SIP trunks. This makes sure that outbound calls to SIP do not require intra-cluster control signaling between Unified CM call processing subscribers.
Inbound Calls		
Calling Search Space	ICTInbound	Incoming calls on trunks need to support +E.164, ESN, and URI dialing. This special CSS supports all three dialing habits but does not provide access to PSTN or remote on-net destinations (see Table 2-26 in the section on Special CSSs). For applications requiring PSTN access, another special class of service (CSS) is required to also provide access to the partitions with PSTN access route patterns (see Table 2-29 in the section on Route Patterns for PSTN Access and Emergency Calls).
AAR Calling Search Space	PSTNReroute	Same CSS everywhere

Table 2-61 SIP Trunk Settings for Intercluster Trunk to Other Unified CM Cluster (continued)

Setting	Value	Comment
Outbound Calls		
Use Device Pool Called Party Transformation CSS	Checked	
Use Device Pool Calling Party Transformation CSS	Checked	
Calling and Connected Party Info Format	Deliver URI and DN in connected party, if available	On intercluster trunks to other Unified CM clusters, blended identity with both numeric and URI information should be delivered to the remote cluster. If both types of identity exist, then based on the capabilities of the called endpoint, the cluster terminating the call can decide which piece of the identity information can be displayed on the final called party.
SIP Information		
Destination	X.X.X.X	List IP addresses of all Unified CM call processing subscribers of remote Unified CM cluster. The order of the IP addresses is not relevant because outbound calls are randomly distributed among the defined destinations.
SIP Trunk Security Profile	ICT	See Table 2-59
SUBSCRIBE Calling Search Space	ICTInbound	Subscriptions on +E.164, ESN, and URIs should be accepted. For the definition of the CSS, see the section on Special CSSs.
SIP Profile	FQDN	See Table 2-56

In contrast to the SIP trunk to a PSTN ISDN gateway, inbound calls from other Unified CM clusters in addition to +E.164 numbers also need access to ESNs and URIs. However, to avoid routing loops and transit-routing, intercluster trunks do not have access to intercluster destinations (partition onNetRemote, see Table 2-13).

For the SIP trunk to the IM and Presence nodes, configure a SIP trunk between Unified CM and IM and Presence. For this SIP trunk, configure the destination IP addresses of all IM and Presence nodes. Select the SIP Trunk Security Profile that you just created for the IM and Presence Service. Also select the Standard SIP Profile.

Route Groups

All SIP trunks are assigned to route groups. Route groups combine trunks with common characteristics. Table 2-62 shows the route group definition for the PSTN gateways in site RTP.

Table 2-62 Route Group for RTP PSTN Gateways

Setting	Value	Comment
Route Group Name	RTP_PSTN	Meaningful name
Distribution Algorithm	Circular	We want to make sure to balance the load over all gateways.
Route Group Members	ST_RTP_PSTN_1 ST_RTP_PSTN_2 ST_RTP_PSTN_3	Add all SIP trunks to all SIP gateways in site RRP.

Note Route groups can be configured only after the SIP trunks have been created, and these can be added only after the respective device pool have been configured. This means that at the time of creating the device pool for PSTN gateways, route groups do not yet exist. Thus the configuration order is:

1. Configure the device pool for the PSTN gateway without defining the LRG mapping in the device pool.
2. Configure SIP trunks.
3. Create the route group.
4. Go back to the device pool and add LRG mapping (if required).

For intercluster trunks to other Unified CM clusters, a route group per trunks also needs to be defined. Table 2-63 shows an example of a route group for an intercluster trunks to a remote Unified CM cluster.

Table 2-63 Route Group for Intercluster Trunk to Other Unified CM Cluster

Setting	Value	Comment
Route Group Name	UCM_EMEA	Meaningful name; in this case, for the route group holding only the intercluster trunk to the EMEA Unified CM cluster.
Distribution Algorithm	Circular	Irrelevant as long as only one route group member exists.
Route Group Members	ST_UCM_EMEA	SIP trunk to remote Unified CM cluster.

Similar trivial route groups must be created for each non-PSTN SIP trunk provisioned on Unified CM.

Specific Non-LRG Route Lists

The section on Route Lists Using Local Route Groups introduces route lists for PSTN access using local route groups only. For non-PSTN trunks, specific route lists need to be created using the route groups referring to these non-PSTN trunks. The reason for defining trivial route groups with only a single member and trivial route lists with only a single non-LRG route group as member, is that route patterns in Unified CM should never point to a trunk directly, because whenever a route pattern is changed in Unified CM, then the device the route pattern is pointing to is reset, and pointing route patterns to a route list instead of trunks makes sure that editing the route patterns will not reset the trunk itself but rather the route list. Examples for such trunks include trunks to other Unified CM clusters and applications.

Table 2-64 shows the trivial route list for an intercluster trunk to another Unified CM cluster.

Table 2-64 Route List for Intercluster Trunk to Another Unified CM Cluster

Route List	Members	Description
RL_UCM_EMEA	UCM_EMEA	Only a single member: the actual trunk to the remote Unified CM cluster. The leading RL makes sure to avoid naming collisions with trunks. Internally route lists are treated as devices, and the names of route lists cannot be identical to names of SIP trunks, for example.

Similar trivial route lists have to be created for each non-PSTN SIP trunk provisioned on Unified CM.

Endpoint Provisioning

When provisioning a new endpoint, these minimum tasks are required:

- Configure the Device
- Configure the Line
- Add the Device to Devices Controlled by the User
- Configure the Line Association for Presence

Configure the Device

When adding a new endpoint to Unified CM, the design described in this document requires the settings summarized in Table 2-65. Settings not mentioned here are either left as default or have to be configured according to device-specific requirements:

Table 2-65 Endpoint Device Settings

Setting	Value	Description
Device Information		
Device Pool	RTPPhoneVideo	Site-specific device pool for endpoints (see Table 2-53). In this case this is the device pool for endpoints in site RTP with access to video conferencing media resources.
Calling Search Space	USEmergency	Access to emergency routing in multi-national environments is implemented on the device level (see the section on Emergency Call Considerations in Multi-National Environments). If only one country (dialing domain) such as the US needs to be supported, then this CSS can be left as <None>.
AAR Calling Search Space	PSTNReroute	Same everywhere (see the section on Automated Alternate Routing).
Media Resource Group List	<None>	Use device pool level settings.
AAR_Group	Default	Same everywhere (see the section on Automated Alternate Routing).
Owner	Select "User"	If the device is a phones without user association (for example a lobby phone), then select "Anonymous (Public/Shared Space)" and do not set the "Owner User ID"
Owner User ID	Select the user ID of the owner of this phone.	
Allow Control of Device from CTI	Checked	
Number Presentation Transformation		
Caller ID For Calls From This Phone	"Use Device Pool Calling Party Transformation CSS (Caller ID For Calls From This Phone)" checked	
Remote Number	"Use Device Pool Calling Party Transformation CSS (Device Mobility Related Information)" checked	
Protocol Specific Information		
SIP Profile	FQDN	See Table 2-56

Configure the Line

On each endpoint, at least the first line needs to be provisioned. Table 2-66 summarizes the line settings specific to the design described in this document.

Table 2-66 Line Settings

Setting	Value	Description
Directory Number Information		
Directory Number	\+14085554146	Full +E.164 directory number matching the phone number of the user this DN is provisioned for. The leading + has to be escaped with \. If a non-DID is provisioned, then the directory number is set to the ESN (for example, 81405001).
Route Partition	DN	If a non-DID is provisioned, then the partition is ESN.
Alerting Name	Aristotle Boyle	Full name of the user associated with the number. If the number is not associated with a user, then provision a meaningful name (for example, Bldg. 31 Lobby).
Allow Control of Device from CTI	Checked	
Directory Number Settings		
Calling Search Space	SJCInternational	CSS defining the effective class of service for calls from this line. The CSS is specific to site and class of service (see the section on Classes of Service and Calling Search Spaces (CSSs) for other CSSs).
BLF Presence Group	Standard Presence Group	Same for all lines
+E.164 Alternate Number		
Number Mask	Leave mask empty	An empty mask creates the +E.164 alternate number identical to the directory number configured above. If a non-DID is provisioned, no +E.164 alternate number is added because no PSTN address exists for non-DIDs, by definition.
Add to Local Route Partition	Not checked	The +E.164 alternate number is not added to a local route partition because the directory number itself already is a +E.164 number
Advertise Globally via ILS	Not checked	The +E.164 alternate number is not advertised via ILS. Instead, summary routes for each DID range are advertised (see Table 2-70). The only reason to create the +E.164 alternate number is to be able to advertise this +E.164 alternate number as the GDPR PSTN failover number for URIs associated with this directory number.
PSTN Failover for Enterprise Alternate Number, +E.164 Alternate Number, and URI Dialing		
Advertised Failover Number	+E.164 Number	The +E.164 number is advertised as GDPR PSTN. If a non-DID is provisioned, then set to <None>.

Table 2-66 Line Settings (continued)

Setting	Value	Description
AAR Settings		
Voice Mail	Not checked	If a non-DID is provisioned, then check this option.
AAR Destination Mask	+14085554XXX	This DID range mask makes sure that the alternate PSTN destination for AAR is equal to the directory number. If a non-DID is provisioned, then leave this mask empty.
AAR Group	Default	Same everywhere
Call Forward and Call Pickup Settings		
Calling Search Space Activation Policy	Use System Default	
Forward All	"Voicemail" not checked Calling Search Space: SJCInternational	CSS might be set to a more restricted CSS
All other Forward settings other than "Forward Unregistered Internal" and "Forward Unregistered External"	"Voicemail" checked Calling Search Space: SJCInternational	CSS might be set to a more restricted CSS
"Forward Unregistered Internal" and "Forward Unregistered External"	Destination: +14085554146 Calling Search Space: PSTNReroute	Forward Unregistered implements an alternate route through the PSTN in case the endpoint is unregistered. This makes sense only for endpoints in remote sites with local PSTN access for which an alternate route through the PSTN can be established. If a non-DID is provisioned or a DN for which PSTN reroute does not make sense, then check "Voicemail" and set the CSS to SJCInternational or some other CSS that can reach the voicemail pilot.
Line 1 on Device		
Display (Caller ID)	Aristotle Boyle	Full name of the user associated with the number. If the number is not associated with a user then, provision a meaningful name (for example, Bldg. 31 Lobby).
Line Text Label	4146	Makes sure that the last four digits of the directory number are displayed next to the line button on the phone. This setting exists only for lines on devices supporting line text labels.
External Phone Number Mask	+14085554XXX	The external phone number mask is not referenced anywhere in the provisioned dial plan and can be set to anything. For phones on which the external phone number mask determines the text in the first line on the phone display, the mask can be set to something that creates a meaningful label.

Add the Device to Devices Controlled by the User

For devices associated with users, after provisioning the device in the End User Configuration of the respective user in the Device Information section in Unified CM Administration, make sure that the device is associated with the user. The recommended way to achieve this is to select **Device Association** and search for devices where the directory number matches the phone number of the user.

Configure the Line Association for Presence

To determine the presence state of a user, only the line appearances (per DN and device) explicitly associated for presence are considered. To make sure that all line appearances of a user's directory numbers are considered for presence, in the End User Configuration of the respective user in the section on Device Information in Unified CM Administration, select **Line Appearance Association for Presence** and associate all line appearances.

Verify the User's Primary Extension

To make sure that the user's directory URI synchronized from LDAP propagates to the directory number, select the Primary Extension in the Directory Number Associations section in the End User Configuration of the respective user in Unified CM Administration.

Jabber Provisioning

Service Discovery enables Jabber to establish configuration automatically. The Jabber client gets its configuration through Unified CM User Discovery Service (UDS). It is the recommended configuration and is preferred over the older manual configuration.

The services are configured through UC services. A Service profile specifies which UC services to use. Each user is associated with a service profile.

Table 2-67 shows the UC services that can be made available to Jabber clients. Those services are configured in **User Management** > **User Settings** > **UC service**.

Table 2-67 UC Services

UC Service Type	Comment
IM and Presence	Create an IM and Presence service for each IM and Presence node.
Directory	Create a Directory service for each active directory server. Do not select "Use UDS for Contact Resolution" when integrating with LDAP directly. When using UDS for Contact Resolution, the Unified CM user scalability decreases.
CTI	Create a CTI service for each Unified CM running the CTI Manager service. This is used for desk phone control mode. Load balance the CTI load across all Unified CM call processing nodes.
Voicemail	Create a Voicemail service for each Unity Connection node.
Conferencing	Jabber can be integrated with Cisco WebEx Meetings Server or Cisco WebEx Meeting Center. In this design, we cover the integration with Cisco WebEx Meetings Server.

Associate the UC services to a Service Profile. A Service Profile is then associated to each user. For deployments with more than two Unified CM call processing subscribers, spread the CTI load equally across all Unified CM call processing subscribers and ensure that the CTI scalability limit is not exceeded on any single Unified CM call processing subscriber running the CTI Manager service. To associate Jabber clients with another Unified CM call processing subscriber running the CTI Manager service, configure another Service Profile with the relevant CTI UC service settings.

For users connected to the internal enterprise network (not using Cisco Collaboration Edge), directory search Contact Sources can be provided through UDS or through LDAP. With LDAP, Enhanced Directory Integration (EDI) for Windows desktops and Basic Directory Integration (BDI) for Mac, iOS, and Android are available. BDI and EDI can co-exist. The Contact Source or directory can be configured through the jabber-config.xml file or through the directory UC service which takes precedence. The recommendation is to configure a jabber-config.xml file that is uploaded onto the Unified CM TFTP server. The jabber-config.xml file is also used to enable URI dialing for Jabber clients. Example 2-5 shows a jabber-config.xml file to enable URI dialing for Jabber clients. This is the recommended minimum. Additional configuration options can be added.

Example 2-5 jabber-config.xml File to Enable URI Dialing

```
<?xml version="1.0" encoding="utf-8"?>
<config version="1.0">
    <Policies>
        <EnableSIPURIDialling>true</EnableSIPURIDialling>
    </Policies>
</config>
```

For more details, refer to the following documentation:

- *Configuration and Administration of IM and Presence on Cisco Unified Communications Manager*

  http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-installation-and-configuration-guides-list.html

- *Deployment and Installation Guide for Cisco Jabber*

  http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/jabber/10_5/CJAB_BK_D6497E98_00_deployment-installation-guide-ciscojabber.html

ILS Configuration for Multi-Cluster Deployments

When the Intercluster Lookup Service (ILS) is configured on multiple clusters, ILS updates Unified CM with the current status of remote clusters in the ILS network.

The ILS cluster discovery service allows Unified CM to learn about remote clusters without the need for an administrator to manually configure connections between each cluster.

The ILS cluster discovery service enables UDS-based service discovery for Jabber clients in multi-cluster environments. In addition, ILS is the foundation for global dial plan replication (GDPR), which allows the exchange of reachability information for both alphanumeric URIs and numeric destinations between Unified CM clusters to enable deterministic intercluster routing for those destinations.

To create an ILS network of multiple Unified CM clusters, perform the following tasks:

- Assign Unique Cluster IDs for Each Unified CM Cluster in the Network
- Activate ILS on the First ILS Hub Cluster in the Network
- Activate ILS on the Remaining ILS Clusters in the Network
- Consider UDS Certificate Requirements

Assign Unique Cluster IDs for Each Unified CM Cluster in the Network

The cluster IDs defined in the Unified CM cluster enterprise parameters have to be unique. See Table 2-4 for details.

Activate ILS on the First ILS Hub Cluster in the Network

Forming an ILS network starts with activating ILS on the first Unified CM cluster. This done by changing the role from Standalone Cluster to Hub Cluster in the ILS Configuration menu in Unified CM Administration.

Table 2-68 shows the settings to be applied when activating ILS on the first Unified CM cluster.

Table 2-68 ILS Activation on First Unified CM Cluster

Setting	Value	Comment
Role	Hub Cluster	ILS is activated by changing the role from Standalone Cluster to Hub Cluster.
Exchange Global Dial Plan Replication Data with Remote Clusters	Checked	Makes sure that URI and numeric reachability information is exchanged with remote clusters.
Advertised Route String	us.route	The advertised route string is the location attribute tied to all URI and numeric reachability information advertised by this Unified CM cluster. Remote clusters trying to reach any of the destinations advertised by this cluster will establish the route to this destination by matching the learned SIP route string against SIP route patterns provisioned on the remote cluster.
Synchronize Clusters Every	2	Setting the synchronization interval to a reasonably small interval makes sure that changes are picked up by remote clusters after a short period of time. The overhead of a short synchronization interval is limited because GDPR uses an incremental update algorithm that exchanges only delta information if any changes occurred since the last update.
ILS Authentication		
Use Password	Checked	Password based authentication is selected.
Password	<some password>	Choose a secure password. This password is shared among all Unified CM clusters participating in the ILS network.

When activating ILS by changing the role from Standalone Cluster to Hub Cluster in Unified CM Administration, an ILS Cluster Registration pop-up appears and asks you to input a Registration Server. When activating ILS on the first Unified CM cluster, no registration server information is available, so the input in that pop-up should be left empty.

Activate ILS on the Remaining ILS Clusters in the Network

Adding more Unified CM clusters to the ILS network requires the same process as activating ILS on the first Unified CM cluster: changing the role from Standalone Cluster to Hub Cluster in the ILS Configuration menu in Unified CM Administration.

Table 2-69 shows the settings to apply when activating ILS on the remaining Unified CM clusters.

Table 2-69 ILS Activation on Additional Unified CM Clusters

Setting	Value	Comment
Role	Hub Cluster	ILS is activated by changing the role from Standalone Cluster to Hub Cluster.
Exchange Global Dial Plan Replication Data with Remote Clusters	Checked	Makes sure that URI and numeric reachability information is exchanged with remote clusters.
Advertised Route String	emea.route	Make sure that the SIP route string for each cluster is unique to allow for deterministic routing based on these route strings. The example here indicated that this is a Unified CM cluster serving EMEA destinations.
Synchronize Clusters Every	2	Make sure to use the same synchronization interval on all clusters for consistency.
ILS Authentication		
Use Password	Checked	Password based authentication is selected.
Password	<some password>	Choose a secure password. This password is shared among all Unified CM clusters participating in the ILS network.

Consider UDS Certificate Requirements

To enable UDS-based service discovery, the UDS process on each Unified CM cluster tries to establish connectivity with the UDS processes running on remote Unified CM clusters to learn about the remote clusters' UDS nodes. For this server-to-server communication, TLS connections between the Unified CM clusters' nodes are established and the remote peers' certificates are validated during TLS connection setup. To prevent this validation from failing, the Tomcat certificates of the Unified CM publisher and call processing subscriber nodes of all Unified CM clusters must be exchanged.

Also, this server-to-server communication is one of the reasons why **TLS Web Client Authentication** has to be in the X.509 extended key usage when issuing Tomcat certificates on an external CA (see Table 2-3).

To exchange the Unified CM cluster publisher certificates, perform the following tasks:

- On each Unified CM cluster in the Cisco Unified Operating System Administration, use Bulk Certificate Management to export the cluster's Tomcat certificate to a central SFTP server.
- After Tomcat certificates from all clusters have been exported, use the Consolidate option in Cisco Unified Operating System Administration on one of the clusters to consolidate all Tomcat certificates into one file.
- On each Unified CM cluster in the Cisco Unified Operating System Administration, use Bulk Certificate Management to import the consolidated file of all Tomcat certificates.

This process makes sure that all Tomcat certificates of all Unified CM cluster nodes are imported in the local certificate stores of all Unified CM clusters as Tomcat trust, so that the certificate validation happening as part of TLS connection setup as part of the UDS communication between clusters does not fail. The use of multi-server certificates greatly reduces the number of individual Tomcat certificates that need to be exchanged.

GDPR Configuration (Multi-Cluster Only)

When Global Dial Plan Replication (GDPR) is enabled across an ILS network, remote clusters in an ILS network share global dial plan data, including the following:

- Directory URIs
- +E.164 and ESN patterns
- PSTN failover numbers

GDPR allows you to create a global dial plan, including intercluster dialing of directory URIs and alternate numbers, that spans across an ILS network. GDPR allows you to quickly configure the global dial plan across the ILS network without the need to configure each dial plan component on each cluster separately.

Configuring GDPR requires the following steps in addition to activating ILS as described in the previous section:

- Advertise URIs
- Configure Advertised Patterns
- Configure Partitions for Learned Numbers and Patterns
- Configure Intercluster Trunks
- Configure SIP Route Patterns

Advertise URIs

In this document we assume that URIs for users are automatically provisioned based on the directory URI synchronized for each user from the email attribute of the corporate directory (see Table 2-45) and the primary extension configure for the user. By default the **Advertise Globally via ILS** option is set for these URIs automatically created in partition Directory URI. Also make sure to set the **Advertise Globally via ILS** option on all URIs you have provisioned in addition to the ones created automatically.

Configure Advertised Patterns

To keep the route plan small on remote clusters, in this design only summary patterns are advertised for each +E.164 and ESN range hosted on each cluster. For the example cluster hosting the sites RTP, RCD, and SJC, the patterns shown in Table 2-70 need to be configured as GDPR advertised patterns. For information on the DID ranges and ESN ranges used in the example, refer to Table 2-10 and Table 2-11.

Table 2-70 Patterns Advertised via GDPR

Pattern	Pattern Type	PSTN Failover Setting	Comment
+14085554XXX	+E.164 Number	Use Pattern as PSTN Failover Number	Site SJC DID range
81404XXX	Enterprise Number	Apply Strip Digits and Prepend Digits to Pattern and Use for PSTN Failover PSTN Failover Strip Digits: 4 PSTN Failover Prepend Digits: +1408555	ESN range of SJC DIDs. Strip digits and prefix to transform from ESN to PSTN failover number.
81405XXX	Enterprise Number	Don't use PSTN Failover	ESN range of SJC non-DIDs. No PSTN failover possible.
+19195551XXX	+E.164 Number	Use Pattern as PSTN Failover Number	Site RTP DID range
81911XXX	Enterprise Number	Apply Strip Digits and Prepend Digits to Pattern and Use for PSTN Failover PSTN Failover Strip Digits: 4 PSTN Failover Prepend Digits: +1919555	ESN range of RTP DIDs. Strip digits and prefix to transform from ESN to PSTN failover number.
81912XXX	Enterprise Number	Don't use PSTN Failover	ESN range of SJC non-DIDs. No PSTN failover possible.
+19725555XXX	+E.164 Number	Use Pattern as PSTN Failover Number	Site RCD DID range
81975XXX	Enterprise Number	Apply Strip Digits and Prepend Digits to Pattern and Use for PSTN Failover PSTN Failover Strip Digits: 4 PSTN Failover Prepend Digits: +1972555	ESN range of RCD DIDs. Strip digits and prefix to transform from ESN to PSTN failover number.
81976XXX	Enterprise Number	Don't use PSTN Failover	ESN range of RCD non-DIDs. No PSTN failover possible.
8099XXXX	Enterprise Number	Don't use PSTN Failover	ESN ranges for conferences on this cluster (see Table 2-11).

Advertising both the +E.164 range and the ESN range for each site makes sure that both formats can be used as the intercluster dialing habit on the remote clusters learning this information.

Configure Partitions for Learned Numbers and Patterns

Numeric patterns (+E.164 and ESN) learned from remote clusters are added to the local route plan into predefined partitions. The **Partitions for Learned Numbers and Patterns** menu in Unified CM Administration allows you to define differentiated partitions for each type of learned information. In this design we do not need this differentiation and simply configure GDPR to learn all remote numeric patterns in a single partition, onNetRemote (see Table 2-13).

Table 2-71 summarizes the settings for the GDPR partitions.

Table 2-71 GDPR Partition Settings

Setting	Value	Comment
Partition for Enterprise Alternate Numbers	onNetRemote "Mark Learned Numbers as Urgent" checked	
Partition for +E.164 Alternate Numbers	onNetRemote "Mark Learned Numbers as Urgent" checked	Marked urgent to avoid inter-digit timeout on +E.164 on-net intercluster calls.
Partition for Enterprise Patterns	onNetRemote "Mark Fixed Length Patterns as Urgent" checked "Mark Variable Length Patterns as Urgent" unchecked	
Partition for +E.164 Patterns	onNetRemote "Mark Fixed Length Patterns as Urgent" checked "Mark Variable Length Patterns as Urgent" unchecked	Marked urgent to avoid inter-digit timeout on +E.164 on-net intercluster calls.

Configure Intercluster Trunks

The GDPR exchange only makes sure that all URI and numeric reachability information is exchanged between Unified CM clusters and associated with a SIP route string as the location attribute. Sessions between clusters need SIP trunks to be established. In this design we assume full-mesh SIP trunks between all Unified CM clusters, with a maximum of three Unified CM clusters. The maximum of three Unified CM clusters makes sure that the topology of the full mesh of SIP trunks is manageable. If more than three Unified CM clusters are required, then adding Unified CM Session Management Edition (SME) is recommended to simplify the topology to a hub-and-spoke topology with SME as the hub and all other Unified CM clusters as spokes or leaf clusters.

Regular SIP intercluster trunks are used for GDPR routing. SIP trunk ST_UCM_EMEA, as with the settings shown in Table 2-61, is an example of an intercluster trunk provisioned for GDPR routing.

Configure SIP Route Patterns

SIP route patterns tie together the SIP route strings learned via GDPR and the SIP trunk topology. Think of it as if a GDPR route strings tells us "where" a learned URI or numeric pattern is located, and we need route patterns matching on these route strings to tell how to get to this destination.

To achieve full GDPR reachability, we need to make sure that each SIP route string advertised via GDPR can be routed according to the provisioned SIP route patterns. Table 2-72 summarizes the trunks, route groups, route lists, and SIP route patterns that need to be provisioned to enable full intercluster GDPR routing between two Unified CM clusters.

Table 2-72 GDPR Routing with Two Unified CM Clusters

Component	US Cluster	EMEA Cluster	Comment
SIP Trunk	ST_UCM_EMEA	ST_UCM_US	SIP trunk on each cluster to the other Unified CM cluster (see Table 2-61)
Route Group with above SIP trunk as member	UCM_EMEA	UCM_US	Dedicated route group for the intercluster trunk (see Table 2-63)
Route List with above route group as member	RL_UCM_EMEA	RL_UCM_US	Dedicated non-LRG route list for the intercluster trunk (see Table 2-64)
SIP Route String	us.route	emea.route	SIP route string advertised by the Unified CM cluster
SIP Route Pattern pointing to above route list	emea.route in partition onNetRemote	us.route in partition onNetRemote	Provisioned SIP route pattern matches on the SIP route string advertised by the other Unified CM cluster

Example GDPR Call Flow

With the above configuration, this section describes how a call would be routed if +14085554001 is dialed on an endpoint with class of service "international" registered to the EMEA cluster in the above example.

1. The dialed digits (+14085554001) are matched against the dial plan on the EMEA cluster, using the calling device’s CSS XXXInternational, where XXX represents a site code of a site provisioned on the EMEA cluster. The actual site-specific dialing normalization is irrelevant here.

   The important point is that CSS XXXInternational contains at least the following partitions (see Table 2-18; again XXX represents a site code while XX represents some dialing domain identifier):

   - DN
   - Directory URI
   - URI
   - ESN
   - onNetRemote
   - XXXIntra
   - XXtoE164
   - XXPSTNNational
   - PSTNInternational
   - B2B_URI
   - USEmergency

   The dialed digits (+14085554001) in these partitions have three matches:

   - +14085554XXX in partition onNetRemote learned from the US cluster with SIP route string us.route (see Table 2-70)
   - \+! in partition PSTNInternational (see Table 2-29)
   - \+!# in partition PSTNInternational (see Table 2-29)

2. Because +14085554XXX in partition onNetRemote is inserted into the route plan as urgent pattern (see Table 2-71) and this pattern at this point is the best match, digit collection is stopped immediately and the call is routed based on this best match.
3. +14085554XXX in partition onNetRemote is a GDPR learned pattern and is associated with SIP route string us.route. Hence, us.route is matched against the configured SIP route patterns on the EMEA cluster, again using the calling device's CSS XXXInternational.

   The only match is SIP route pattern us.route in partition onNetRemote.
4. The call on the EMEA cluster is extended to SIP trunk ST_UCM_EMEA, dereferencing the route list RL_UCM_EMEA the matched SIP route pattern us.route points to and route group RG_UCM_EMEA (see Table 2-72)
5. On the US cluster, the inbound CSS ICTInbound of SIP trunk ST_UCM_EMEA (see Table 2-61) is used to route the inbound call to destination +14085554001.
6. CSS ICTInbound has these partitions:
   - DN
   - ESN
   - URI
   - Directory URI

   In these partitions the only (potential) match is on a +E.164 directory number \+14085554001 (marked urgent) in partition DN. If this directory number exists, then the call is extended to all associated devices.

Routing of remotely dialed ESN destinations follows the exact same flow, with the only exception being that the final lookup on the US cluster using CSS ICTInbound in that case would find a match on an ESN in partition ESN.

IM and Presence Intercluster

To create a fully meshed presence topology, each Cisco IM and Presence cluster requires a separate peer relationship for each of the other Cisco IM and Presence clusters within the same domain. The address configured in this intercluster peer is the IP address of the remote Unified CM cluster IM and Presence publisher node.

The interface between each Cisco IM and Presence cluster is two-fold: an AXL/SOAP interface and a signaling protocol interface (SIP or XMPP). The AXL/SOAP interface, between publisher-only servers of an IM and Presence cluster, handles the synchronization of user information for home cluster association, but it is not a full user synchronization. The signaling protocol interface (SIP or XMPP) is a full mesh encompassing all servers within the deployment. It handles the subscription and notification traffic, and it rewrites the host portion of the URI before forwarding if the user is detected to be on a remote Cisco IM and Presence cluster within the same domain.

When Cisco IM and Presence is deployed in an intercluster environment, a presence user profile should be determined. The presence user profile helps determine the scale and performance of an intercluster presence deployment and the number of users that can be supported. The presence user profile helps establish the number of contacts (or buddies) a typical user has, as well as whether those contacts are mostly local cluster users or users of remote clusters.

Survivable Remote Site Telephony (SRST) Deployment

Configure SRST at each remote site in order to provide call processing survivability in case the WAN to the remote site fails. With SRST, if the WAN fails, phone calls can still be made within the remote site or out to the PSTN.

Deployment

Deploy one Cisco Integrated Services Router (ISR) for each remote sites that you want to enable for SRST.

Provisioning

To configure SRST, you must perform the configuration on both Unified CM and the SRST router.

On Unified CM:

- Configure an SRST Reference for each remote site, and associate this SRST Reference in the device pool of the remote phones.
- Configure Call Forwarding Unregistered (CFUR) on the DNs of the remote phones to use the +E.164 number and the AAR CSS. In case the WAN fails, the call will use this information to get routed via the PSTN.

On the SRST router:

- Configure SRST on each remote branch router. Since our recommendation is to use SIP phones, use the **voice register global** and **voice register pool** commands. Use the **voice service voip/sip** command to bind the IP addresses of the source interface and enable the registrar capability. Configure DHCP for the phones in the remote branch. The DHCP server may be configured on the SRST router or on other network service resources.
- If the WAN fails, the SIP phones will register with their +E.164 extensions. In order to allow users to call other local users by their four-digit extensions, configure a voice translation profile that is referenced as an incoming profile in the voice register pool configuration. This voice translation profile transforms the called number from four digits to the complete +E.164 number.
- Configure POTS dial-peers to allow local access to the PSTN in case the WAN is down. Configure translation voice profiles in order to comply with the service provider's PSTN dialing requirements. For more details on dial-peer configuration, refer to the section that describes how to Deploy Cisco Unified Border Element.

The SRST configuration in Example 2-6 is just a partial configuration to illustrate some of the concepts discussed in the previous paragraphs. It does not cover the full SRST configuration. For instance, configuration to reach the Cisco Unity Connection server in the main site is covered in the chapter on Core Applications.

Example 2-6 SRST Partial Configuration

```
voice service voip
 allow-connections sip to sip
sip
  bind control source-interface GigabitEthernet0/0.241
  bind media source-interface GigabitEthernet0/0.241
  registrar server
!
voice register global
 mode srst
 max-dn 100
 max-pool 100
!
voice register pool  1
 translation-profile incoming 4-digit-rtp
 id network 10.0.94.0 mask 255.255.255.0
!
voice translation-rule 1
 rule 1 /\(^1...\)$/ /+1919555\1/
!
voice translation-profile 4-digit-rtp
 translate called 1
!
```

For more details on configuring SRST, refer to the *Cisco Unified SCCP and SIP SRST System Administrator Guide*, available at

> http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cusrst/admin/sccp_sip_srst/configuration/guide/SCCP_and_SIP_SRST_Admin_Guide.html

Extension Mobility

Cisco Extension Mobility allows users to temporarily access their Cisco Unified IP Phone configuration – such as line appearances, services, and speed dials – from other Cisco Unified IP Phones.

One or two Unified CM call processing nodes can actively handle Extension Mobility requests. The benefits of adding a second Unified CM call processing node for Extension Mobility are resiliency and increased capacity. In this scenario, a load balancer is required to send the requests to both Unified CM nodes. Cisco IOS Server Load Balancing can be used, for example.

Extension Mobility Cross Cluster (EMCC) provides the ability to perform Extension Mobility logins between clusters within an enterprise. This feature is not covered in this guide. For more details on EMCC, refer to the *Cisco Collaboration System 11.x SRND* and the EMCC product documentation.

Deploying Extension Mobility

To deploy Extension Mobility, perform the following tasks:

- Ensure that the Cisco Extension Mobility service is activated on one or two Unified CM call processing servers.
- Add an IP Phone Service for Extension Mobility. A secure IP Phone Services URL using HTTPS in addition to a non-secure URL can be configured. The non-secure URL is

  > http://<*IPAddress*>:8080/emapp/ EMAppServlet?device=#DEVICENAME#

  You can either make this service available to all phones in the cluster by selecting Enterprise Subscription or make it available to selected phones by subscribing those phones to this service.

- For each user that will use Extension Mobility, create at least one Device Profile. Since a Device Profile is tied to a specific user, the Device Profile is usually referred to as a User Device Profile. If a Device Profile is not created for a user, that user will not be able to log in with extension mobility.
- Associate the device profile to a user for extension mobility. If CTI is needed, also associate the profile to be a CTI controlled device profile.
- For each phone that can be used for users to log in, enable Extension Mobility. For Cisco DX Series endpoints, also enable Multi-User (the phone will reset). For Cisco TelePresence endpoints using the TC software (for instance, Cisco TelePresence EX and SX Series endpoints), ensure that the TelePresence endpoints are not provisioned in Cisco TelePresence Management Suite (TMS), otherwise the sign-in button will not be available on the endpoint.
- On the DN configuration, configure the association of the appropriate user to the line. This allows the DN to send presence information for that user if the line of that phone is in use. For example:

  User B is using Jabber and is monitoring user A. User A logs into a phone with Extension Mobility and has a User Device Profile with the DN associated to himself/herself. When user A goes off-hook, this presence information will be reported on the Jabber client of user B.

For more details on Extension Mobility, refer to the *Features and Services Guide for Cisco Unified Communications Manager*, available at

http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/admin/10_0_1/ccmfeat/CUCM_BK_F3AC1C0F_00_cucm-features-services-guide-100.html

Busy Line Field (BLF) Presence

The BLF Presence feature allows a user (watcher) to monitor the real-time status of another user at a directory number or SIP URI from the device of the watcher. A watcher can monitor the status of the user by using the following options:

- BLF/SpeedDial buttons
- Missed call, placed call, or received call lists in the directories window
- Shared directories, such as the corporate directory

BLF Presence is not based in Cisco Unified IM and Presence.

Deploying BLF Presence

- Enable the **BLF for Call List** enterprise parameter (see Table 2-4).
- Configure the cluster-wide service parameters for BLF presence.
- To use BLF presence group authorization, configure BLF presence groups and permissions.
- Apply a BLF presence group to the directory number, SIP trunk, phone that is running SIP, phone that is running SCCP, end user, and application user (for application users that are sending BLF presence requests over the SIP trunk) in Cisco Unified Communications Manager Administration.
- To allow BLF presence requests from a SIP trunk, select the **Accept Presence Subscription** option in the SIP Trunk Security Profile Configuration window (see Table 2-59).
- Configure the SUBSCRIBE calling search space and apply the calling search space to the phone, trunk, or end user, if required.
- For BLF/SpeedDial buttons on the phones, customize phone button templates for the BLF/SpeedDial buttons or add them directly to the phones.

Deploying Computer Telephony Integration (CTI)

- Activate the CTI Manager service on the Unified CM call processing nodes that need the CTI Manager service.
- For redundancy, through the CTI application administration, select a primary and backup Unified CM node running the CTI Manager service,
- Download the TAPI client software for applications using TAPI.
- If possible, for a given CTI-enabled endpoint, configure the same Unified CM call processing node for CCM registration and for CTI Manager monitoring and control.
- Ensure the CTI load is spread across all Unified CM nodes running the CTI Manager and that the CTI capacity limits are not exceeded. For example, with Jabber clients, if two Unified CM call processing pairs are required, spread the registration across the two pairs; also, if the Jabber clients are configured with the ability to be in deskphone mode, spread the CTI Manager connectivity across the two pairs. This can be achieved with multiple Service Profiles with different CTI profiles associated. Ensure the number of Jabber clients in deskphone mode monitored and controlled by each Unified CM running the CTI Manager service does not exceed the CTI capacity limit.

Conferencing

Revised: November 20, 2015

This chapter describes the components and deployment of video and audio conferencing in an enterprise deployment. The chapter describes the Architecture for conferencing and then outlines the major tasks involved in the Conferencing Deployment Process.

Each major task of the Conferencing Deployment Process starts with an *Overview* section listing the steps required for that task, followed by a section on the important *Deployment Considerations* for that task, and then a section (or sections) detailing the deployment tasks listed in the *Overview* section.

What's New in This Chapter

Table 3-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 3-1 New or Changed Information Since the Previous Release of This Document

New or Revised Topic	Described in:	Revision Date
Cisco TelePresence Management Suite (TMS) information has been moved to this chapter (previously covered in the Conferencing chapter)	Role of Cisco TMS, page 3-6 5. Deploy TelePresence Management Suite, page 3-38	November 20, 2015
Multiparty Licensing in Cisco TelePresence Conductor for TelePresence Server license management	Licensing, page 3-16	November 20, 2015

Core Components

The core architecture contains these key conferencing elements:

- Cisco TelePresence Server for audio and video conference resources
- Cisco TelePresence Conductor for management of audio and video resources as well as TelePresence Server resource licenses
- Cisco TelePresence Management Suite (TMS) for conference provisioning, monitoring, and scheduling
- Cisco TelePresence Management Suite Extension for Microsoft Exchange (TMSXE) for interfacing with Microsoft Exchange room and resource calendars
- Cisco TelePresence Management Suite Provisioning Extension (TMSPE) for configuring Cisco Collaboration Meeting Rooms (CMRs) as well as provisioning multiparty licenses for users
- Cisco WebEx Software as a Service (SaaS) for scheduled audio and web conferences and hybrid video and web conferences

In addition, Cisco TMS architecture includes these non-Cisco components:

- Microsoft SQL database
- Microsoft Active Directory
- Microsoft Exchange or Microsoft Office 365
- Network Load Balancer

Key Benefits

- Simplified, optimal user experience for conference participants
- Flexible, extendable architecture that supports deployment of one or more permanent, scheduled, and/or instant conference resources
- Dynamic optimization of conference resources on the TelePresence Server for inbound calls
- High availability of conference resources and processes
- Resilience in the video network
- A single tool for hosts to schedule participants and conference rooms for a meeting
- Ability to integrate the WebEx and video participants into a single meeting without additional end-user intervention
- Multiparty licensing that enables full access to all conference resources and simplifies the deployment process of the TelePresence Server

Conference Types

The conferencing solution supports the conference types and conferencing features listed in Table 3-2 and Table 3-3.

Table 3-2 Types of Conferences

Conference Type	Description
Instant conferences	Manually escalated from a point-to-point call hosted on Unified CM, to a three-party call hosted on a conference bridge. (Also referred to as ad hoc conference.) Instant conferences are not scheduled or arranged prior to the conference.
Permanent conferences	Predefined addresses that allow conferencing without previous scheduling. The conference host shares the address with other users, who can call in to that address at any time. (Also referred to as meet-me, static, or rendezvous conferences.) Permanent conferences covered in this chapter use Cisco Personal Collaboration Meeting Rooms.
Scheduled conferences	Conferences booked via Cisco TMS and/or integration using Cisco TMS with a start and end time, optionally with a predefined set of participants.
Cisco Personal Collaboration Meeting Rooms (CMR)	Permanent conferences that are provisioned from Cisco TMS with a portal to allow users to manage items such as their conference name, layouts, and PIN.
Cisco CMR Hybrid	Similar to scheduled conferences, but with a link to Cisco WebEx Meeting Center to allow TelePresence and WebEx participants to join the same meeting and share voice, video, and content.

Table 3-3 Alternative Solutions

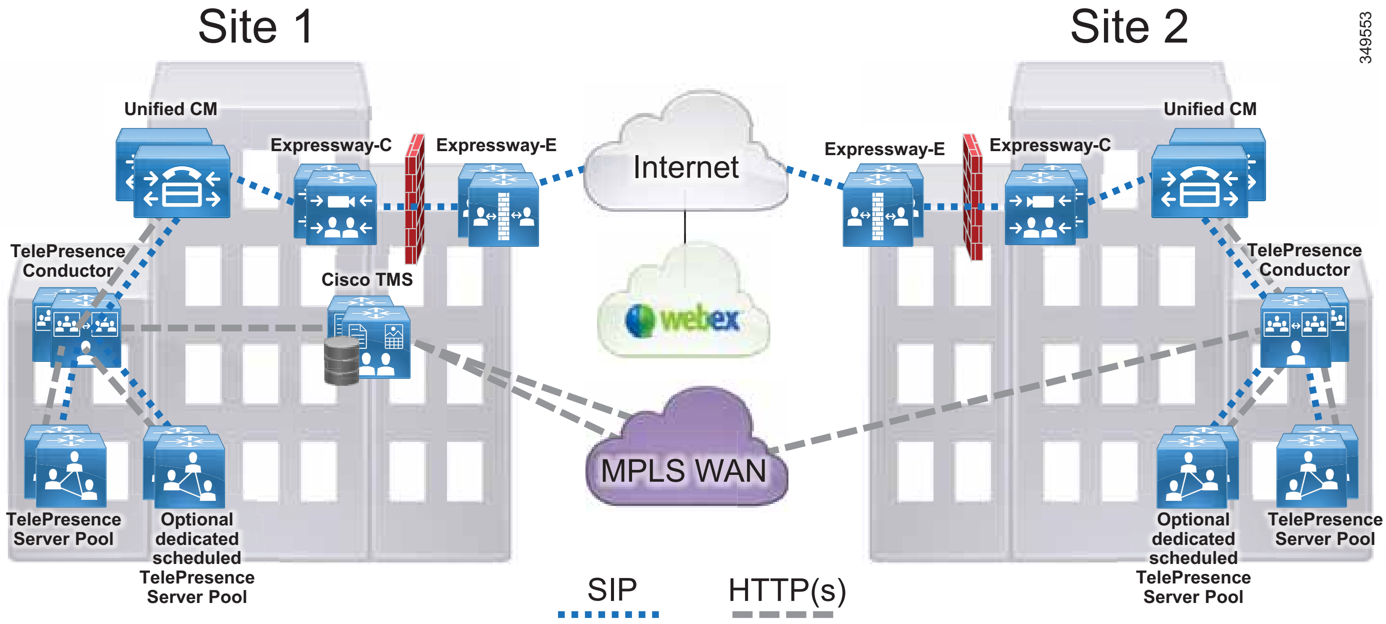
Solution Type	Description
Cisco WebEx Meetings Server	If cloud-based web and audio conferencing is not suitable, it is possible to use the on-premises WebEx Meetings Server solution. This product does not offer hybrid conferencing with TelePresence, but it offers a standalone audio, video, and collaboration web conferencing platform.
Cisco CMR Cloud	Cisco CMR Cloud is an alternate conferencing deployment model that negates the need for any on-premises conferencing resources or management infrastructure. It supports standards-based video including TelePresence, audio, and WebEx participants in a single call, all hosted in the cloud.

Architecture

The conferencing architecture consists of TelePresence Conductor for TelePresence Server resource management; TelePresence Management Suite (TMS) for resource provisioning, scheduling and monitoring; and Unified CM for call processing. SIP call control is used exclusively in this architecture. TelePresence Conductor manages the conference bridges for all types of conferences. Use SIP trunks to connect the bridges to the TelePresence Conductor and to connect the TelePresence Conductor to Unified CM. (Figure 3-1)

Unified CM communicates with TelePresence Conductor using XML Remote Procedure Call (RPC), and it uses Conductor's application programming interface (API) over HTTP to control the conference bridges. Cisco TMS also uses XML-RPC connections to link to the TelePresence Conductor for provisioning and scheduled conference management. (Figure 3-1)

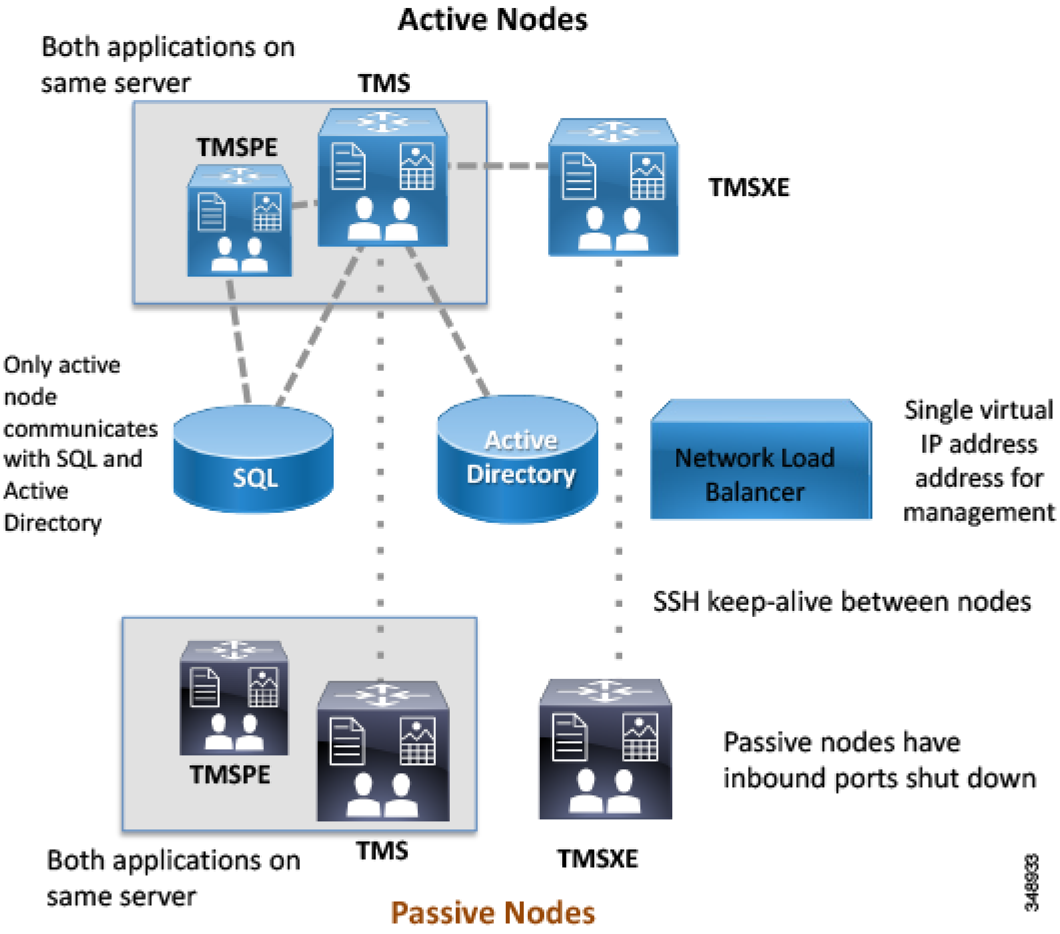
Figure 3-1 Architecture Overview



In addition, TelePresence Conductor can centrally manage the licenses for all TelePresence Servers under its control when Multiparty Licensing mode is enabled. In Multiparty Licensing mode, licenses are not required in the TelePresence Servers, and TelePresence Conductor handles the license checking at conference start.

The scheduling architecture consists of an active and a passive node for both Cisco TMS and TMSXE, which are deployed behind a network load balancer. Cisco TMS and TMSPE are installed on the same virtual machine while TMSXE must be installed on a separate virtual machine, as indicated in Figure 3-2. The TMS servers are installed in the customer data center that also hosts the organization's SQL deployment. All the server nodes function from an external Microsoft SQL database. Additionally, endpoints, TelePresence Conductor, TelePresence Servers, and Unified CM are involved in a successful scheduled conference. (Figure 3-2)

Figure 3-2 High-Level View of the Scheduling Architecture


Active Nodes
Both applications on same server
TMS
TMSPE
TMSXE
Only active node communicates with SQL and Active Directory
SQL
Active Directory
Network Load Balancer
Single virtual IP address address for management
SSH keep-alive between nodes
TMSPE
Passive nodes have inbound ports shut down
Both applications on same server
TMS
TMSXE
Passive Nodes
348933

Role of the TelePresence Conductor

TelePresence Conductor manages the TelePresence Server resources for all types of conferences. It selects which TelePresence Server to host a specific conference, and it balances the conference load across the TelePresence Servers in the defined pools. Unified CM is unaware of the individual TelePresence Servers in the network and communicates only with the TelePresence Conductor. When Multiparty Licensing mode is enabled in TelePresence Conductor, it can centrally manage licenses for all TelePresence Servers.

Using TelePresence Conductor for conferences and license management has several benefits, including:

- Increased efficiency by sharing resources from TelePresence Servers for conferences
- Better user experience with the advanced TelePresence Server features such as ActiveControl and dynamic optimization of resources
- Simpler deployment options through provisioned CMRs
- Single deployment model for all types of conferences
- Simplify TelePresence Server deployment process

TelePresence Conductor optimizes TelePresence Server resources dynamically when the **Optimize resources** setting is enabled in the TelePresence Conductor conference template. This enforces maximum resource usage of a participant based on the maximum receive bandwidth advertised by them at conference join. This can reduce the amount of resources conference calls use and allow more concurrent connections to take place. For more information, see the TelePresence Server release notes.

Role of TelePresence Servers

TelePresence Servers are conference bridges that operate in remotely managed mode and are grouped into pools in TelePresence Conductor. TelePresence Conductor applies service preferences to prioritize the pools for conferencing. The bridge pool can be shared between scheduled and non-scheduled conferences or dedicated to either type of conference.

Role of Cisco TMS

Cisco TMS integrates the conference room endpoints, TelePresence Conductor, and connections to the WebEx cloud in a manner that provides an end user with a unified experience for scheduled conferencing. Unified CM maintains the configuration control for endpoints, and TMS is then able to push the calendar to those endpoints. Administrators are able to set the parameters for the default conference for their organization, and then individual conferences will be created according to this template.

Some of the TMS features are not used in the Preferred Architecture – for example, phone books, software management, and reporting functions.

Cisco TelePresence Management Suite Provisioning Extension (TMSPE) supports automated bulk provisioning by administrators of personal Collaboration Meeting Rooms (CMRs) and a user portal for individuals to define and manage their own personal CMRs. CMRs are used for non-scheduled conferencing where specific endpoints are not identified, and the user simply dials in to the CMR number. In addition, TMSPE provides an interface for Multiparty License provisioning.

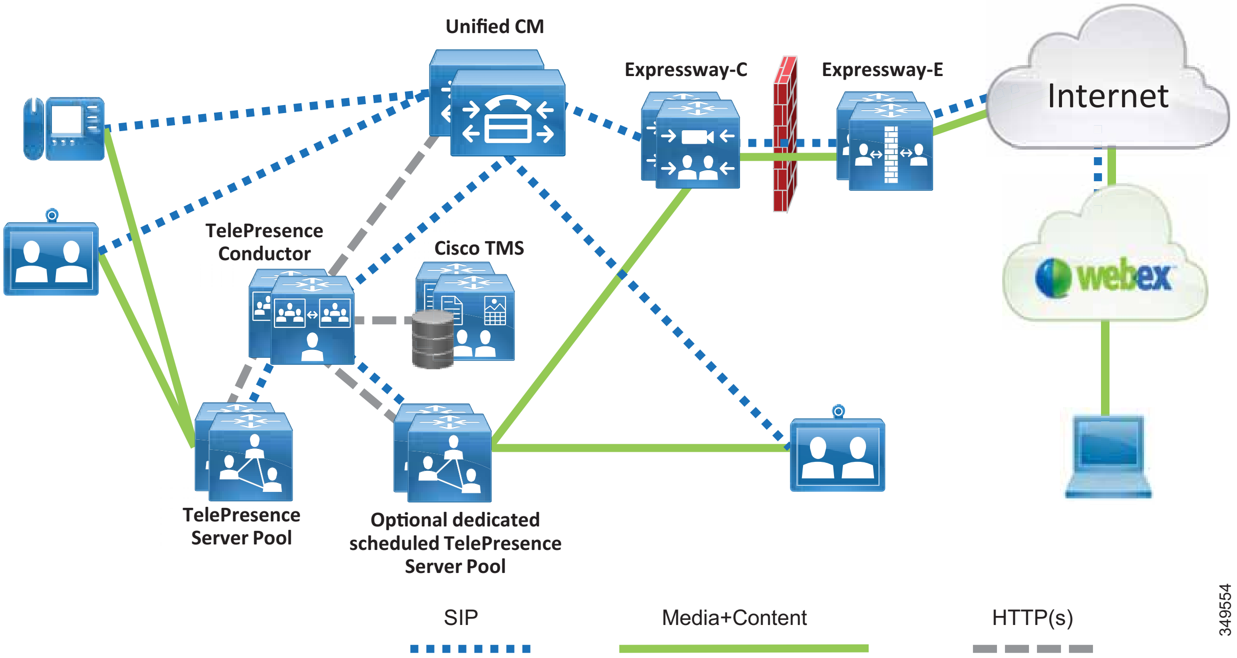
Role of Cisco TMS Extensions for Microsoft Exchange

When end users schedule a meeting in Microsoft Outlook with multiple conference room resources, the Exchange Web Services (EWS) feature of Exchange synchronizes that event into TMS as a scheduled conference. This synchronization is bidirectional, allowing an administrator or support staff to update meetings as well without the need to access the meeting organizer's Outlook event. All endpoint resources within the organization that are intended to be in the conference must be listed on a single Exchange meeting request.

Deployment Overview

The standard deployment uses multiple Unified CM nodes for call control. The TelePresence Conductor is connected to Unified CM with SIP trunks that manage TelePresence Servers to provide conference resources. (Figure 3-3) TelePresence Servers are used to bridge calls, and Cisco TMS provides conference management facilities and scheduling.

Figure 3-3 Standard Deployment



The standard deployment has several TelePresence Servers behind TelePresence Conductor, combined with Unified CM for call control and endpoint management, and Cisco TMS for conference management. The same conferencing infrastructure can be used for both non-scheduled and scheduled conferencing as well as CMR Hybrid services. WebEx provides scheduled audio and video web conferencing. These elements together provide voice and video conferencing for the local enterprise.

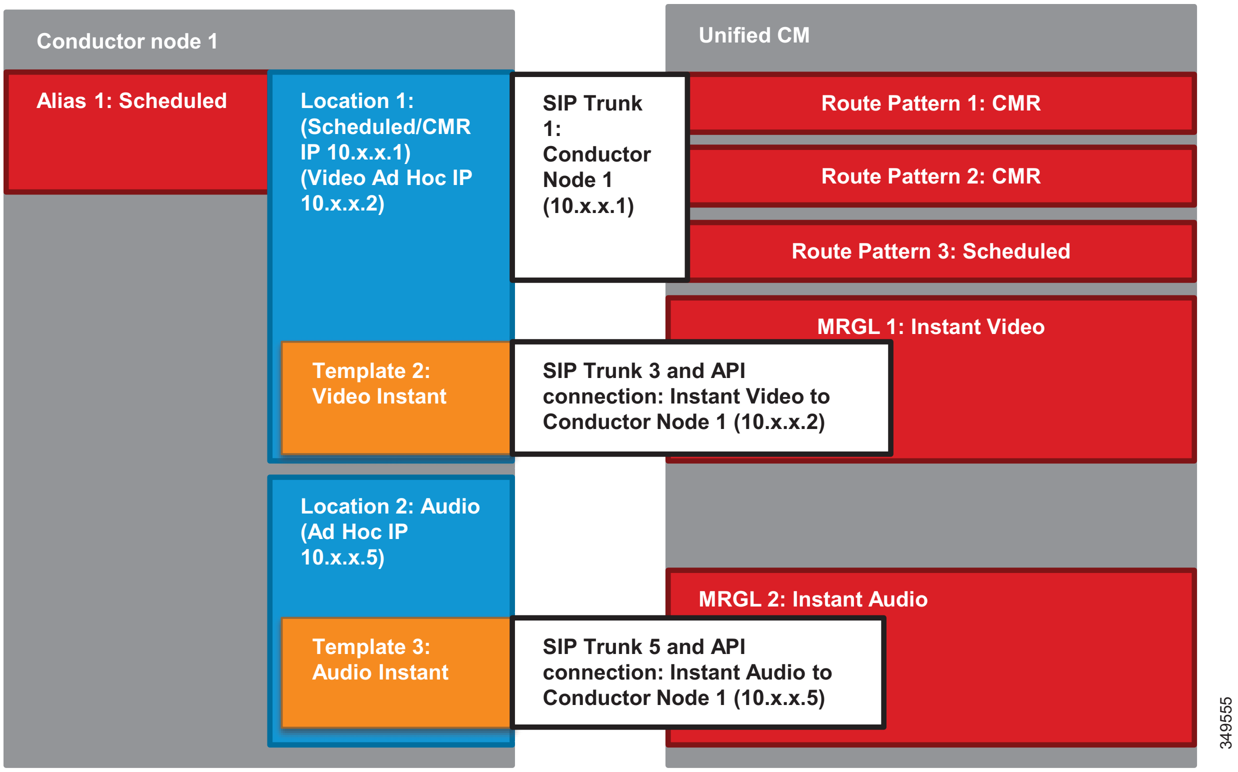
Requirements and Recommendations

- Early Offer is required for CMR Hybrid calls and for some third-party services such as Microsoft Lync.
- Early Offer messaging is recommended for all SIP trunks connected to Unified CM that carry TelePresence calls.
- All TelePresence Servers should be configured in remotely managed mode and placed behind TelePresence Conductor as conferencing resources for all conference types.
- Enable Multiparty Licensing in TelePresence Conductor for TelePresence Server resource license management.
- The Multiway^{TM} method of escalated conferencing is not recommended.
- Audio conferencing on the TelePresence Server does not support join or leave tones.

Conference Call Flows

Unified CM provides device registration and routing of voice and video calls between the connected endpoints. Permanent, instant, and scheduled conference calls are carried over SIP trunks. XML-RPC connections are configured on the Unified CM publisher and connect over HTTP between Unified CM call processing subscriber nodes and TelePresence Conductor nodes.

Figure 3-4 Unified CM, and TelePresence Conductor SIP Trunks


Conductor node 1
Alias 1: Scheduled
Location 1: (Scheduled/CMR IP 10.x.x.1) (Video Ad Hoc IP 10.x.x.2)
Template 2: Video Instant
Location 2: Audio (Ad Hoc IP 10.x.x.5)
Template 3: Audio Instant
SIP Trunk 1: Conductor Node 1 (10.x.x.1)
SIP Trunk 3 and API connection: Instant Video to Conductor Node 1 (10.x.x.2)
SIP Trunk 5 and API connection: Instant Audio to Conductor Node 1 (10.x.x.5)
Unified CM
Route Pattern 1: CMR
Route Pattern 2: CMR
Route Pattern 3: Scheduled
MRGL 1: Instant Video
MRGL 2: Instant Audio
349555

CMR and scheduled conference calls are routed over the same trunk from Unified CM. In contrast, instant conference calls route directly to a TelePresence Conductor template from the Unified CM that created the conference, so multiple trunks may exist for instant conferences, one for each type. Each trunk has an associated XML-RPC connection. Their originating Unified CM controls instant conferences, so a SIP API trunk pair is required per instant conference type from each Unified CM cluster that supports instant conferencing to the local TelePresence Conductor.

Instant call flows that are managed by Unified CM cannot be used to add participants to conferences created by any other method, such as scheduled conferences. Other call flows cannot be used to add participants to instant conferences. The instant call escalation method is supported only in an instant conference that was created by it, and conferences generated by other methods cannot be extended by the instant mechanism. This avoids any potential for chained conferences.

Note Unified CM delivers instant conferences to different IP addresses than CMR and scheduled conferences on TelePresence Conductor. Multiple Unified CM clusters can access the same IP address on TelePresence Conductor. However, this document assumes that a TelePresence Conductor cluster is dedicated to each Unified CM cluster.

Instant Conferences

Instant conferences are configured on the TelePresence Conductor, so the conference is never statically defined on a single TelePresence Server. TelePresence Conductor load-balances the conferences across the available TelePresence Servers in a pool, increasing conference resilience. Instant conferences require a SIP trunk with an associated XML-RPC connection between Unified CM and each TelePresence Conductor node. Unified CM routes the instant conference participants to the IP addresses of these SIP trunks.

Permanent Conferences with Collaboration Meeting Rooms

Permanent conferences are deployed using Personal Collaboration Meeting Rooms (CMRs). Personal CMRs provide a permanent-type conference that is created with Cisco TMSPE in conjunction with the Conductor Provisioning API. Users can dial the conference alias at any time to start a meeting.

Individual end-users create their own CMRs through the Cisco TMSPE user portal, based on group-level templates provisioned by their administrator. Each CMR created through the user portal has a corresponding conferencing bundle entity (ConfBundle) on TelePresence Conductor, which is created and managed through the Conductor Provisioning API and contains the data required to create a conference for one or more users, including conference template information, a set of conference aliases, a set of auto-dialed participants, and a conference name. ConfBundles are reported in the web UI as "Collaboration meeting rooms." CMRs created using Cisco TMSPE cannot be modified through the TelePresence Conductor web UI. Conversely, conference templates and aliases created using TelePresence Conductor cannot be modified through Cisco TMSPE. CMR conferences require a SIP trunk between Unified CM and each TelePresence Conductor node. Unified CM routes the CMR conference participants to the IP addresses of these SIP trunks.

Configuration Information

- For details about Cisco TMSPE settings, see the *Cisco TelePresence Management Suite Provisioning Extension with Cisco Unified CM Deployment Guide*, available at

  http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-provisioning-extension/model.html

- For details about the Conductor Provisioning API, see *Cisco TelePresence Conductor API Guide*, available at

  http://www.cisco.com/c/en/us/support/conferencing/telepresence-conductor/products-programming-reference-guides-list.html

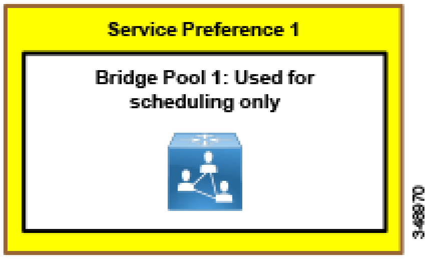
Scheduled Conferences

This solution supports scheduling of conferences through TelePresence Conductor. Scheduling is performed with Cisco TMS. Scheduled conferences require a SIP trunk between Unified CM and each TelePresence Conductor node. Unified CM routes the scheduled conference participants to the IP addresses of these SIP trunks. These trunks are the same trunks used for CMR conferences.

Two deployments for scheduling TelePresence Server resources are possible when using TelePresence Conductor: Conferencing resources can be dedicated for scheduled conferencing or they can be shared resources for both scheduled and non-scheduled conferencing. Some advantages and disadvantages are covered in Table 3-4.

- Dedicated TelePresence Servers — Deploy one or more TelePresence Servers that are dedicated just for scheduled conferences, with each TelePresence Server in a separate bridge pool and service preference of its own (Figure 3-5). Optionally, a second bridge and pool combination can be used as a backup.

Figure 3-5 Dedicated Scheduling TelePresence Server



- Shared TelePresence Servers — Allow TelePresence Servers to be used for non-scheduled as well as scheduled conferences (Figure 3-6). In this case, resource availability for scheduled conferences cannot be guaranteed because the necessary resources might already be in use by non-scheduled conferences. All TelePresence Servers can be configured into a single bridge pool if they are of similar size.

Figure 3-6 Shared Scheduling TelePresence Server

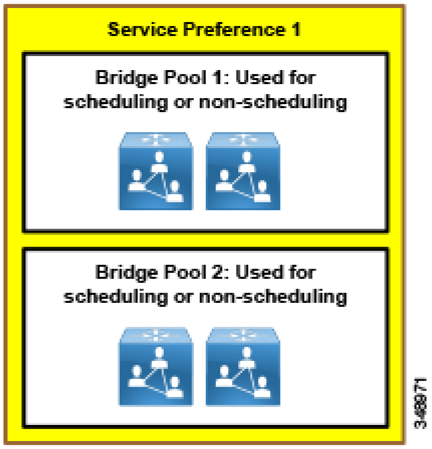

Service Preference 1
Bridge Pool 1: Used for scheduling or non-scheduling
Bridge Pool 2: Used for scheduling or non-scheduling
348971

Table 3-4 Dedicated versus Shared TelePresence Server Resources

Type of Resource	Deployment Details	Advantages	Disadvantages
Dedicated	Service preference: Dedicated TelePresence Server pool for scheduled conferences. Single pool, with a single TelePresence Server in the pool. Pool marked to be used for scheduling in the TelePresence Conductor Service Preference. Pool is reported to TMS in capacity information requests. Service Preferences can optionally have a non-scheduled bridge pool for high availability, which contains additional dedicated TelePresence Server resources. (For more information see High Availability for Conferencing, page 3-13.)	By dedicating conferencing resources to scheduled conferences, there is no risk of non-scheduled conferences using resources and, provided that enough resources are provisioned, causing scheduled conferences to fail due to lack of resource availability.	Takes up a TelePresence Server exclusively for scheduling. Non-scheduled conferences would need to use a different TelePresence Server. Inefficient use of conferencing resources when users use more non-scheduled conferencing over a given period. More resources are necessary to handle the fluctuation in usage patterns. Benefits of optimized resources are negated because no non-scheduled conferences can use the available resources. Cascaded conferencing does not occur. And to avoid wasting resources, cascading should be disabled.
Shared	Service preference: Shared-use TelePresence Servers for scheduled and non-scheduled conferences. One or more pools shared for scheduled and non-scheduled conferences. One or more Service Preferences. Each Service Preference has all bridge pools enabled for scheduling within TelePresence Conductor and can contain multiple TelePresence Server conference bridges. Service Preferences can optionally have a non-scheduled bridge pool for high availability that contains additional dedicated or shared TelePresence Server resources. (For more information see High Availability for Conferencing, page 3-13.)	More efficient utilization of resources. When users use more or fewer scheduled resources, there is no impact on the number of resources left idle, as can happen with dedicated resources that are not used. Cascaded conferencing is available (if enabled). Targeted management of TelePresence Server resources is possible. Non-scheduled conferences are able to use resources available due to resource optimization of scheduled conferences where this occurs. Over time, monitoring of use patterns can identify the most appropriate pool configuration.	Conference resources are not guaranteed to be available for scheduled conferences because non-scheduled conferences could use up all of the resources. Ensuring that enough resources are provided to service high utilization can reduce this risk.

Tip TelePresence Servers can be set up in the TelePresence Conductor to host instant conferences only, CMR conferences only, scheduled conferences only, or all three. Using a single TelePresence Server pool can minimize the number of TelePresence Servers needed because you can provision for the overall maximum number of conference participants.

Multiway

Multiway is a method for instant conferencing, but is not used in this Enterprise Collaboration Preferred Architecture. In Unified CM deployments as described in this document, instant conferences should be initiated using the conference button on the device instead. Using both types of instant conferencing simultaneously is not supported.

Third-Party Endpoints

Endpoints from other equipment providers can participate in instant, scheduled, and CMR conferences using standard SIP. Only endpoints registered to Unified CM that support the conference button can initiate an instant conference. Cisco Expressway or Cisco VCS can be used to interwork H.323 calls to SIP, allowing H.323 endpoints to join conferences.

Cisco WebEx Software as a Service (SaaS)

Cisco WebEx Software as a Service (SaaS) is a cloud hosted solution that provides audio and video web conferencing with rich content collaboration. We recommend using this service when hosting scheduled audio conferences. It is also used to create CMR Hybrid conferences, as described in the next section.

Cisco CMR Hybrid

Cisco WebEx and TelePresence users can participate jointly in scheduled meetings or the users' personal Collaboration Meeting Rooms (CMRs). Both SIP and PSTN-based audio are supported for the audio portion of the call between Cisco WebEx and the TelePresence Servers. (The audio connections between WebEx participants and the WebEx conference can be PSTN audio, SIP audio, or computer telephony.)

Cisco WebEx Meetings Server

For some deployments, a cloud-based collaboration service might not be suitable. For these customers we recommend using Cisco WebEx Meetings Server instead as an on-premises solution. This server does not integrate with TelePresence the way the Software as a Service version of WebEx does, but instead acts as a standalone audio, video, and collaboration web conferencing service. For installation instructions, refer to the WebEx Meetings Server product documentation at

> http://www.cisco.com/c/en/us/support/conferencing/webex-meetings-server/products-installation-guides-list.html

Collaboration Meeting Rooms (CMR) Cloud

Cisco CMR Cloud is an alternate conferencing deployment model that negates the need for any on-premises conferencing resource or management infrastructure. CMR Cloud is a simple-to-use cloud hosted meeting room solution that is offered as an add-on option to a Cisco WebEx Meeting Center subscription, and it is delivered through the Cisco WebEx Cloud. The solution enables meetings in the cloud that can scale to support up to 25 standards-based video endpoints and up to 500 video-enabled and 500 audio-only WebEx Meeting Center users, with a total of up to 1,025 participants in a single meeting. This guide does not cover deployment of CMR Cloud. For CMR Cloud deployment details, refer to Cisco Collaboration Meeting Rooms (CMR) Cloud Enterprise Deployment Guide, available at

http://www.cisco.com/c/en/us/support/conferencing/webex-meeting-center/products-installation-and-configuration-guides-list.html

High Availability for Conferencing

High availability must be considered at several levels with the conferencing solution and is achieved in different ways depending on the service being considered.

For both scheduled and non-scheduled conferences, high availability involves Unified CM, the TelePresence Server, and TelePresence Conductor.

TelePresence Server High Availability

In the event of a TelePresence Server failure, the conference can take place on another TelePresence Server as long as there is an available resource within the Service Preference in TelePresence Conductor. The conference number remains identical, and therefore this is seamless as far as the end user is concerned. However, if the TelePresence Server fails while hosting a live conference, users need to redial the same number to join the conference on the new TelePresence Server.

TelePresence Conductor High Availability

A full-capacity standard TelePresence Conductor can be part of a cluster of up to three Conductor nodes. Each node must be within 30 ms round trip delay to every other node. Calls can flow through any nodes in the cluster. If a node becomes unavailable, the remaining TelePresence Conductor nodes can be used to join conferences.

TelePresence Conductor clusters are used for redundancy; when the primary conductor fails, a secondary or tertiary conductor is available to manage the current and future conferences. This redundancy works seamlessly for instant and permanent conferencing. For scheduled conferencing, however, this redundancy is not automatic but requires manual intervention to fail-over to a secondary or tertiary TelePresence Conductor. The reason for this is that Cisco TMS recognizes only one TelePresence Conductor node. If that cluster node should fail, Cisco TMS scheduling will be unavailable until that node is brought back up or Cisco TMS is updated to communicate with a different TelePresence Conductor node in the cluster.

Cisco TMS considers each alias configured on a TelePresence Conductor as a separate set of TelePresence Server resources. If an alias has no available resources to schedule a call, then the next aliases are considered in priority order. Where multiple TelePresence Conductors are configured (one per TelePresence Conductor cluster), then these are also considered by Cisco TMS for scheduled conferences. Cisco TMS uses the IP Zone configuration of the TelePresence Conductor and the endpoints scheduled at the time of the booking to decide which TelePresence Conductor should be preferred in the booking.

TMS High Availability

High availability deployment of Cisco TMS includes: Two TMS front-end servers that also host the TMS Provisioning Extension (TMSPE) application, two servers running TMSXE, a network load balancer, and an external Microsoft SQL database (see Figure 3-2). TMS resiliency supports only two servers – one active node and one passive node – and this model does not increase or decrease the capacity of the TMS deployment. The network load balancer (NLB) is deployed in front of the TMS servers. Inbound traffic to TMS goes through the NLB, which forwards it to the active node. Outbound traffic from TMS is sent directly to the destination without going through the NLB. If the NLB detects a failure on the existing active node, it automatically switches to the new active node without any user intervention.

This document assumes that the WebEx Cloud is used for scheduled audio conferences. WebEx Cloud has built-in high availability and therefore has no additional resiliency considerations beyond normal telephony and web connectivity. If Cisco WebEx Meetings Server is used instead, then high availability must be considered for the servers, but this approach is not discussed in this document.

Security for Conferencing

The Preferred Architecture fully supports media and signaling encryption; but for simplicity, the solution presented in this document implements non-secure SIP trunks between Unified CM and TelePresence Conductor for all conferences. An exception to this is the solution requirement that API communications between TelePresence Conductor and the TelePresence Server must be encrypted, and therefore HTTPS must be used in this case. It is also a requirement for communications between Cisco Expressway and the WebEx Cloud to use valid Certificate Authority signed certificates to enable encrypted media and signaling.

Another level of security can be added to restrict access to the conferences themselves with PINs or passwords. Any scheduled conference or CMR conference can have a PIN set so that all participants are challenged to enter the PIN before being allowed to connect. With CMR Hybrid, a PIN can be used to protect the TelePresence portion of the conference, and a password can be added or auto-generated for the WebEx portion of the conference.

Scaling the Conferencing Solution

You can scale the conferencing solution primarily by increasing the number of available TelePresence Servers.

The total number of TelePresence Servers is limited by the capacity of TelePresence Conductor. Each full-capacity TelePresence Conductor (or each cluster) can manage up to 30 TelePresence Servers or 2,400 concurrent conference calls. Clustering does not increase the maximum number of TelePresence Servers or concurrent calls that can be supported.

If a deployment grows beyond the capacity of a single TelePresence Conductor, it is possible to create a new independent cluster and continue to add TelePresence Servers there.

Use an independent TelePresence Conductor cluster for each regional Unified CM cluster. For example, if your deployment has three Unified CMs clusters, then you should also deploy three TelePresence Conductor clusters, one in each region.

Conferencing Deployment Process

To deploy the conferencing solution, perform the following major tasks in the order listed here:

1. Plan the Conferencing Deployment
2. Deploy TelePresence Servers
3. Deploy TelePresence Conductor
4. Enable Unified CM for Scheduled and Non-Scheduled Conferences
5. Deploy TelePresence Management Suite
6. Deploy Cisco Collaboration Meeting Rooms
7. Deploy Collaboration Meeting Rooms (CMR) Hybrid

1. Plan the Conferencing Deployment

Before deploying the conferencing solution, plan for the following aspects:

Requirements

- Provision IP addresses; each TelePresence Conductor node requires an IP address for itself and up to 64 additional IP addresses, depending on the conference services deployed.
- Order and install encryption keys on all TelePresence Servers.
- CMR Hybrid has a number of requirements that should be understood in advance to ensure everything that is required is in place. For example, an option key must be ordered for Cisco TMS, and Cisco Expressway-E must have a publicly signed certificate from an appropriate Certificate Authority and must trust the WebEx root certificate in order to allow CMR Hybrid to function.

Product Models

Cisco makes several different models of the TelePresence Conductor and TelePresence Server. It is important to understand the differences between these models so you can choose the best option for your deployment.

There are three models of TelePresence Conductor. Use the full-capacity version for enterprise deployments because it supports up to 2,400 concurrent calls and up to three Conductor nodes in a cluster.

Any TelePresence Server may be used for scheduled or non-scheduled conferences behind TelePresence Conductor. TelePresence Servers on virtual machines are available in several capacities. If larger capacities are required, the MSE 8000 chassis-based Multiparty Media 820 (MM820) blade can be used. The MM820 supports two blades per cluster, which doubles the capacity of the MM820. A cluster behaves as if it is a single device.

Any of these devices can be configured to run in remotely managed mode and therefore work with TelePresence Conductor.

Licensing

Licenses must be installed on various products:

- Cisco TMS must have enough device licenses installed for the deployment.
- A full version of TelePresence Conductor comes with the required licenses installed.
- TelePresence Conductor manages TelePresence Servers using Multiparty licensing.

Multiparty is a user-based licensing model that is centrally managed by TelePresence Conductor. It comes with two variations: Personal and Shared. Personal Multiparty (PMP) is for specific named hosts while Shared Multiparty (SMP) is for conference room systems or for sharing between users. Each license entitles a user to host a conference with unlimited participants and up to 1080p video resolution. Table 3-5 summarizes the features included in the Personal and Shared Multiparty licenses.

Table 3-5 Cisco Personal and Shared Multiparty License Features

Feature	Personal Multiparty	Shared Multiparty
Tied to a named host	Yes	No
Availability	Included in Cisco UWL Professional	A la carte or discounted with room system
Minimum order	25	1
Maximum conference size	Unrestricted, within the limit of available hardware capacity	
Maximum resolution	1080p30 (full HD) video and content Single or multiple screen endpoints	
Rich media sessions for business-to-business or business-to-customer	Included	Included
Cisco TMS, TMSXE, and Skype for Business and Lync Interoperability Licenses	Included	New customers buy with Starter Pack
Support for instant, personal CMR, and scheduled conferences	Yes	Yes

Multiparty licensing is the license model used in the Preferred Architecture. For more information on Multiparty licenses, refer to *Cisco Multiparty Licensing At-a-Glance*, available at

http://www.cisco.com/c/dam/en/us/solutions/collateral/collaboration/pervasive-conferencing/at-a-glance-c45-729835.pdf

Cisco TelePresence Management Suite

Before beginning the installation and configuration process, you must decide on several items to align with the specific structure and preferences of your organization. Additionally, some specific settings must be used during the configuration process and should be gathered prior to beginning the install process.

Microsoft SQL

Cisco TMS utilizes an external Microsoft SQL database to store all data regarding meetings, users, and systems. During the installation process, TMS and associated software extensions create a number of specific databases. The TMS application does not allow users to log into the web page if communication is not currently active with the tmsng database. This dependency on constant communication with the SQL database requires the SQL database to utilize Microsoft's methods for making the database resilient as well. The databases will vary in size depending upon the deployment size and number of scheduling events; but as a general guideline, 1 GB of initial storage will suffice for most organizations.

Table 3-6 lists the Microsoft SQL 2012 specifics required to support Cisco TMS, TMSXE, and TMSPE.

Table 3-6 Microsoft SQL 2012 Specifics Required to Support Cisco TMS, TMSXE, and TMSPE

Requirement	Parameter
SQL user account permissions for account used by TMS	**dbcreator** and **security admin** roles
Authentication	SQL Server and Windows authentication (mixed mode)
Default language	English
Time zone	Must match the time zone on TMS server
Databases created	tmsng (CiscoTMS) tmspe (CiscoTMSPEmain) mspe_vmr (Cisco TMSPE Collaboration Meeting Rooms) tmspe_userportal (Cisco TMSPE self-service portal)
Resiliency model	**AlwaysOn Failover Cluster** instances through Windows Server Failover Clusters (WSFC)

Note While other modes of SQL resiliency are supported by TMS, any method besides **AlwaysOn Failover Cluster** requires manual adjustments by the TMS administrator during an SQL outage situation.

Active Directory

Cisco TMS functions using many aspects of Microsoft Active Directory, and the server must be added to the organization's domain,. All TMS users must be imported from and authenticated with Active Directory.

During the configuration process, you must enter an **AD Service account username and password** for TMS to import users. This is a read-only account, and TMS does not modify any information in Active Directory. This account should have access to the highest level of the AD structure that enables all subsequent end users to access its functionality. In organizations with multiple domains, the TMS user account must be associated with the top level domain. An additional service account is required for the TMSXE application for end-user booking of Exchange resources. This should also be a read-only service account, and end user credentials are used for the actual event booking. TMSXE user account permits only the TMSXE application to authenticate and communicate with the Exchange Servers through Exchange Web Services.

Additionally, identify existing, or create new, Groups with AD that will serve to synchronize TMS administrators and end users with scheduling access to TMS.

Note Local machine accounts on the TMS server should not be used because they are not duplicated between front-end servers, and the user credentials would not be available if the other node became active.

Email Integration

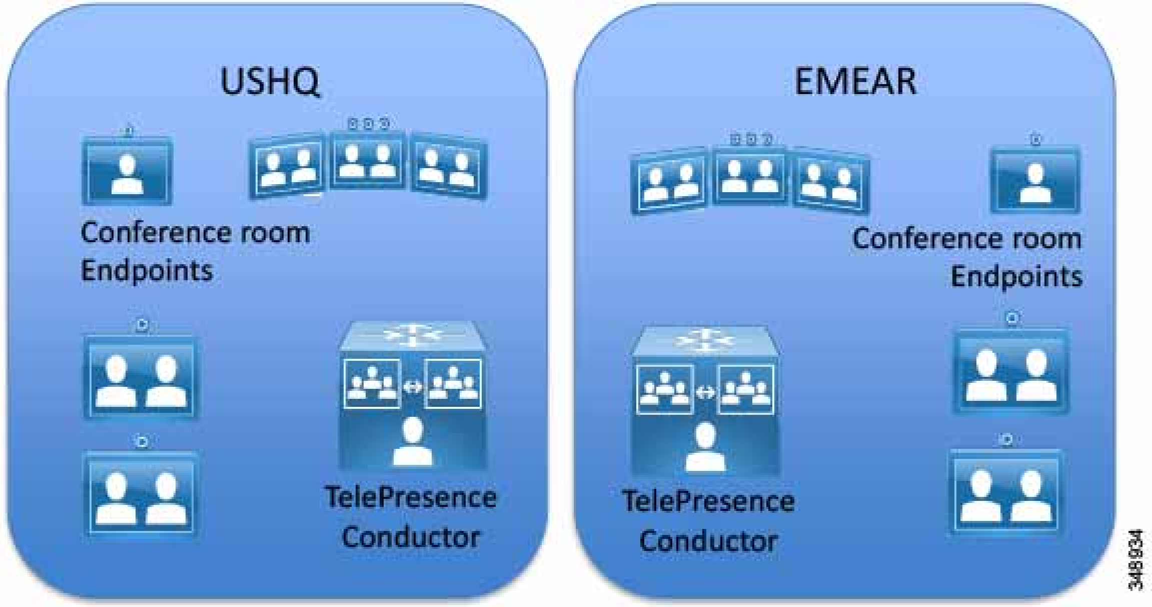
TMS sends automated emails to users when they schedule meetings, with all connection information included for the participants. During the installation process, you must enter the "from" address that end users will see as the originating for these emails, so select an address such as collabconferencing@ent-pa.com or a similar address not currently used in your organization.

You will also need to enter the SMTP address of the outgoing mail server.

Zones

Cisco TMS uses a concept called *zones* to provide guidance to the scheduling engine on how to build the calls and keep the traffic localized as much as possible. Endpoints, conferencing resources, and ISDN gateways are all assigned to these zones. The zones define where to use which kind of network connections. The Preferred Architecture is based upon all endpoints being able to use a single IP network for connections, and ISDN would be used only for connecting outside of the organization. (See Figure 3-7.)

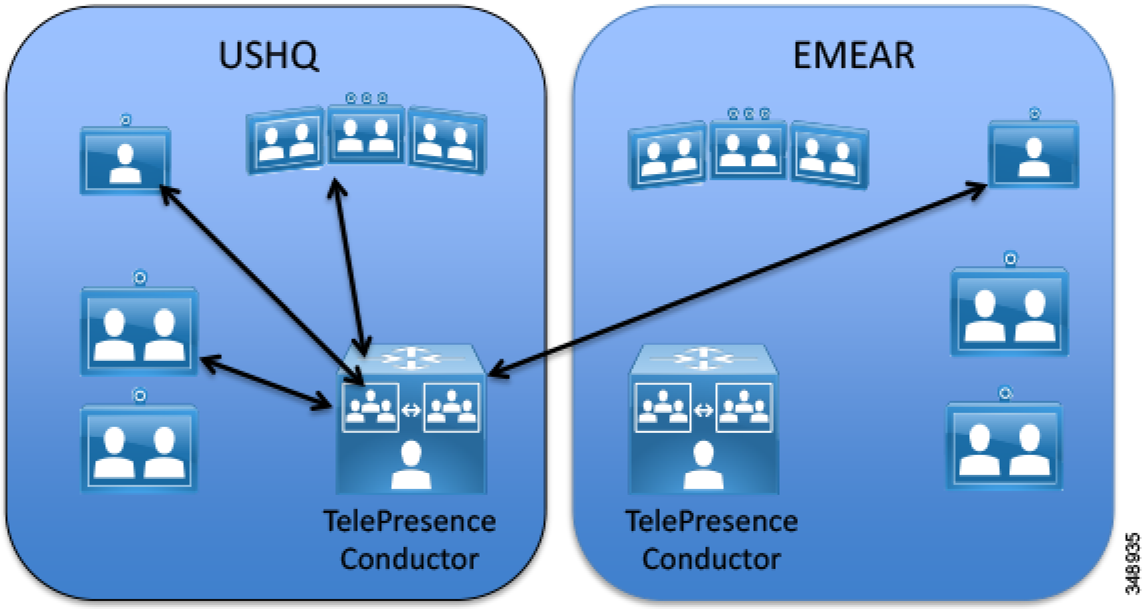
Figure 3-7 Cisco TMS IP Zones



IP Zones

IP zones refer to areas of the network that share a common primary data center, and they are used primarily to identify "local" conferencing resources. During the configuration process, add IP zones for each location where conferencing resources will be placed. As shown in Figure 3-8, the USHQ TelePresence Conductor is selected because most participants are connected from that zone.

Figure 3-8 Cisco TMS Uses IP Zones to Select Best TelePresence Conductor for a Conference



ISDN Zones

ISDN zones are similar to IP zones, but specific to any deployment of ISDN gateways in the organization. If no ISDN functionality is needed, you still have to configure a single ISDN zone for the entire enterprise during the installation.

Endpoint Naming Conventions

Endpoints are added to Cisco TMS for two reasons:

- Correlation with Exchange resources for conference resource allocation
- Enabling TMS to provide One Button to Push connection information on the endpoint user interface

As endpoints are added to TMS, use the same character string as the room or resource name in Exchange. This provides uniformity and consistency to end users when system names appear in the call history and fill the text of on-screen labels from conferencing resources.

An organized plan for how to use the folder structure of TMS Systems Navigator will also assist the administrator in having a simplified interface.

Default Conference Parameters for Your Organization

These settings are customizable for each organization and should be used in accordance with your own network considerations, meeting flows, and corporate culture. The default conference settings are used for all meetings scheduled by end users through Outlook. For all possible settings of the default conference, refer to the *Cisco TelePresence Management Suite Administrator Guide*, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-tms/products-maintenance-guides-list.html

WebEx Site for CMR Hybrid

Be sure that you have the hostname and site name for your WebEx site. Then configure this WebEx site for integration between the on-premises TelePresence conferencing infrastructure and the WebEx cloud. For details, refer to *Cisco Collaboration Meeting Rooms (CMR) Hybrid Configuration Guide*, available at

http://www.cisco.com/c/en/us/support/conferencing/collaboration-meeting-rooms-cmr/products-installation-and-configuration-guides-list.html

CMR Provisioning

An understanding of how the organization plans to utilize Collaboration Meeting Rooms leads to an understanding of the workflow that end users expect for meetings. Some organizations may choose to leverage CMRs instead of scheduled resources for certain meeting types, especially in cases where workers are in separate locations and not able to gather in a common conference room.

Location of Servers

Both the active and passive nodes for a redundant TMS deployment must be configured with the same time zone within the server operating system. In addition, this must be the same time zone as the SQL server. Support of TMS redundancy is limited to the same local network for both the active and passive nodes, along with the SQL server.

2. Deploy TelePresence Servers

This section describes the major tasks required to deploy TelePresence Servers and prepare them for use with TelePresence Conductor. These TelePresence Servers can be used for scheduled and non-scheduled conferences.

Overview

Deployment Tasks for TelePresence Servers:

1. Install the media encryption feature key.
2. Create a new API access user for TelePresence Conductor to use.
3. Enable the HTTPS and Encrypted SIP (TLS) services, and disable H.323 services.
4. Set SIP settings to Call Direct and use TLS, and disable H.323 gatekeeper registration.
5. Set the server to remotely managed mode, if it has that option, and reboot.

Deployment Considerations

The physical location of a TelePresence Server is important to consider because media traffic flows between it and each participant in the conference. To provide the best experience for participants, centralize the location of the TelePresence Servers in each region where they will be deployed.

Set the TelePresence Servers to remotely managed mode, which is a system-wide setting that enables a more advanced API and requires that API to be used for all operations. Remotely managed mode is the only mode available on the Cisco TelePresence Server on Virtual Machine, while other variants of the TelePresence Server can use either remotely managed mode or locally managed mode (which cannot be used with TelePresence Conductor).

⚠ **Caution** In remotely managed mode certain features are not available from the TelePresence Server interface, and management of conferences and pre-configuration of endpoints are done at the TelePresence conductor level instead. Changing the operating mode requires the TelePresence Server to be rebooted, and any conferences configured on the TelePresence Server in locally managed mode are lost when the unit reboots.

TelePresence Conductor manages the TelePresence Servers via an XML-RPC API. TelePresence Conductor also routes SIP signaling to the TelePresence Servers via its Back-to-Back User Agent (B2BUA).

The TelePresence Server has the ability to use a secure connection for communications. These security features are enabled with the activation key. Encrypted communication between TelePresence Server and TelePresence Conductor is mandatory. The media encryption key allows TelePresence Server to use encrypted media via SRTP.

Instant, Permanent, and Scheduled Conferences

Figure 3-9 Instant Conference Call Flow

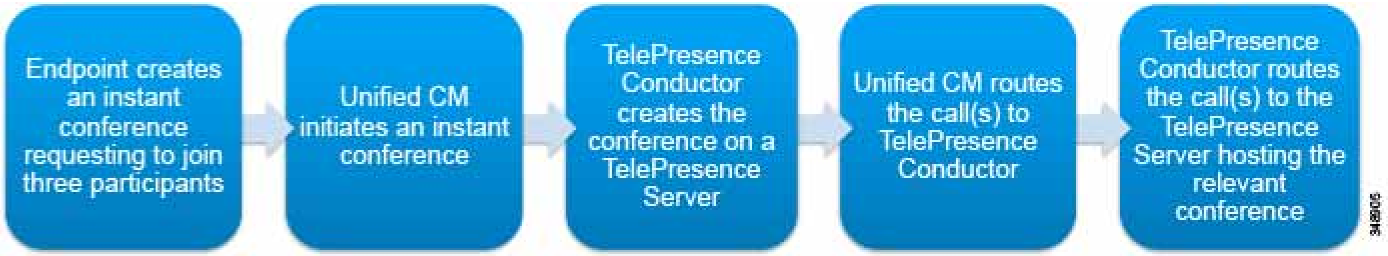
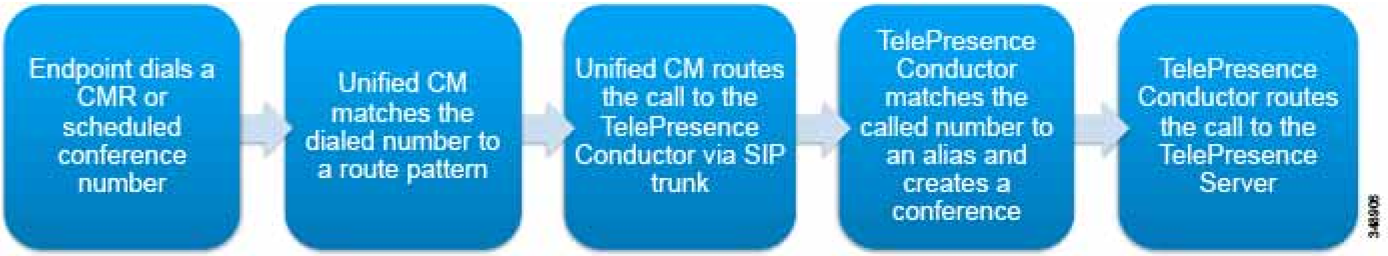

Endpoint creates an instant conference requesting to join three participants
Unified CM initiates an instant conference
TelePresence Conductor creates the conference on a TelePresence Server
Unified CM routes the call(s) to TelePresence Conductor
TelePresence Conductor routes the call(s) to the TelePresence Server hosting the relevant conference
348905

Figure 3-10 Permanent or Scheduled Conference Call Flow


Endpoint dials a CMR or scheduled conference number
Unified CM matches the dialed number to a route pattern
Unified CM routes the call to the TelePresence Conductor via SIP trunk
TelePresence Conductor matches the called number to an alias and creates a conference
TelePresence Conductor routes the call to the TelePresence Server
348906

Deployment Tasks for TelePresence Servers

First apply an activation key to all TelePresence Servers. This allows the use of encrypted signaling with HTTPS and TLS. Then enable and configure the HTTPS and encrypted SIP (TLS) services in the TelePresence Servers. If the deployment requires encrypted media, apply the media encryption feature key.

To enable TelePresence Conductor to communicate with the TelePresence Server, create an API access user account on the TelePresence Server that TelePresence Conductor uses for API authentication.

Disable all H.323 settings on the TelePresence Server, and enable call direct mode in SIP settings using TLS for outbound transport. This is required because TelePresence Conductor uses SIP only.

Then set the TelePresence Server to remotely managed mode. This enables TelePresence Conductor to communicate with the server. This task is not relevant for Cisco TelePresence Server on Virtual Machine that always runs in remotely managed mode.

Summary

After completing the above tasks, TelePresence Servers will be ready to add to TelePresence Conductor.

3. Deploy TelePresence Conductor

This section describes the major tasks required to deploy TelePresence Conductor for scheduled and non-scheduled conferences.

Overview

Deployment Tasks Common to Instant and Scheduled Conferences on TelePresence Conductor:

1. Enable Multiparty Licensing for TelePresence Servers in TelePresence Conductor.
2. Create an API user in TelePresence Conductor.
3. Create a shared bridge pool and add the TelePresence Servers to the bridge pool. In the case of a dedicated model, create several bridge pools and add one TelePresence Server to each bridge pool.
4. Create a service preference and add the bridge pool(s) to the service preference.
5. Assign one IP address to TelePresence Conductor for each conference type:
   - Instant audio conferences
   - Instant video conferences
   - CMR and scheduled conferences

   Repeat this process on each TelePresence Conductor node in the cluster. Each node requires a unique set of IP addresses.

Deployment Tasks for Instant Conferences on TelePresence Conductor:

6. To configure TelePresence Conductor for instant conferences, create a template for each conference type:
   - Instant audio template
   - Instant video template

   In video template, set Optimize resources to **yes**.
7. Create a unique location for each instant conference template. Only one template can be assigned per location:
   - Instant audio location
   - Instant video, CMR, and scheduled location
8. Assign the relevant instant template to the relevant location, and assign an Ad hoc IP address to each. Each Ad hoc IP address is used by Unified CM to communicate with TelePresence Conductor via SIP and the API.

Deployment Tasks for Scheduled Conferences on TelePresence Conductor:

9. Edit the previously created instant video, CMR, and scheduled location. Change its type to **Both** and assign it a Rendezvous IP address. The Rendezvous IP address is used by Unified CM to communicate with TelePresence Conductor via SIP.

   Using Unified CM call processing subscriber node IP addresses, add up to three trunk IP addresses to the instant video, CMR, and scheduled location.

10. Edit the previously created bridge pool(s), and assign the location to the shared bridge pool. If there is more than one pool, add the same location to every pool.

11. To configure TelePresence Conductor for scheduled conferences, create template:

   – Scheduled 720p HD video template

   Assign the service preference to the template. In the template, set Optimize resources to **yes** and set Scheduled conference to **yes**.

12. The unique incoming alias for the scheduling template is created during Cisco TMS configuration. Edit the service preference and ensure that the bridge pool has the **Pools to use for scheduling** radio button selected. This should be set only on bridge pools that should report their capacity to TelePresence Conductor.

Deployment Considerations

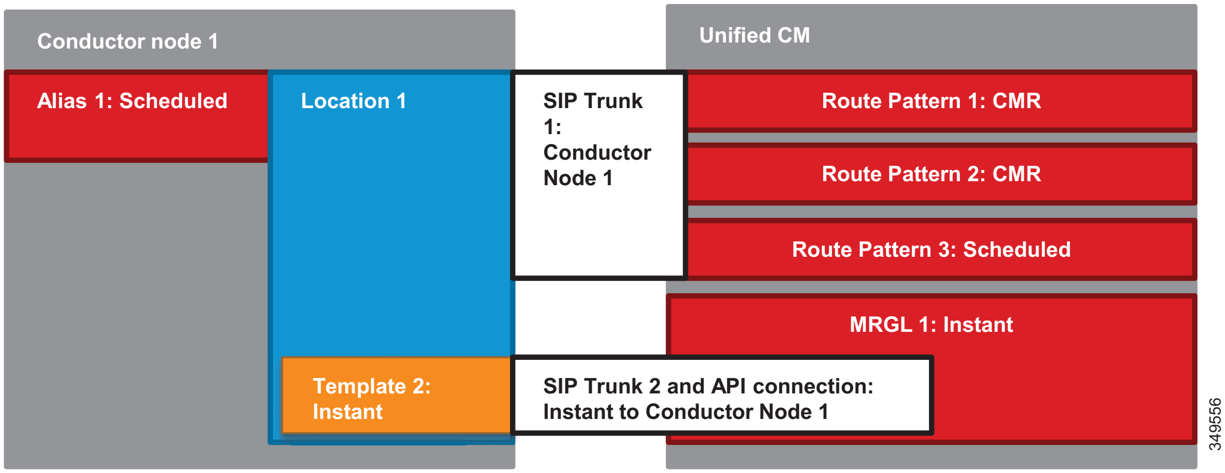
Note Before configuring TelePresence Conductor, clear all alarms to allow the product to function.

To enable instant, CMR, and scheduled conferencing, the system administrator needs to complete the configuration on TelePresence Conductor and Unified CM. These two products share the responsibility for conference creation logic. While TelePresence Conductor maintains exclusive connectivity to the TelePresence Servers, it acts similarly to an MCU as far as Unified CM is concerned, and it makes decisions on conference placement based on requests from Unified CM.

TelePresence Conductor should be deployed regionally in a centralized manner so that each regional Unified CM cluster has a TelePresence Conductor cluster associated with it that manages all the TelePresence Servers in the region, as illustrated in Figure 3-1.

TelePresence Conductor should be deployed using its Back-to-Back User Agent (B2BUA). Unified CM communicates with TelePresence Conductor using an XML-RPC API and SIP trunks, similar to the way TelePresence Conductor communicates with the TelePresence Server. (Figure 3-11)

Figure 3-11 API and SIP Trunk Connections Between Unified CM and TelePresence Conductor



Instant and Scheduled Conferences

Figure 3-12 TelePresence Conductor Internal Configuration Flow for Instant Conferences

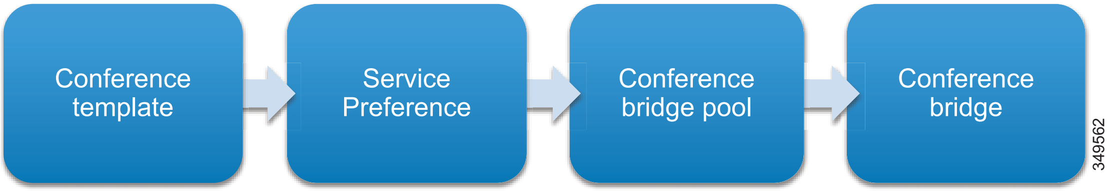
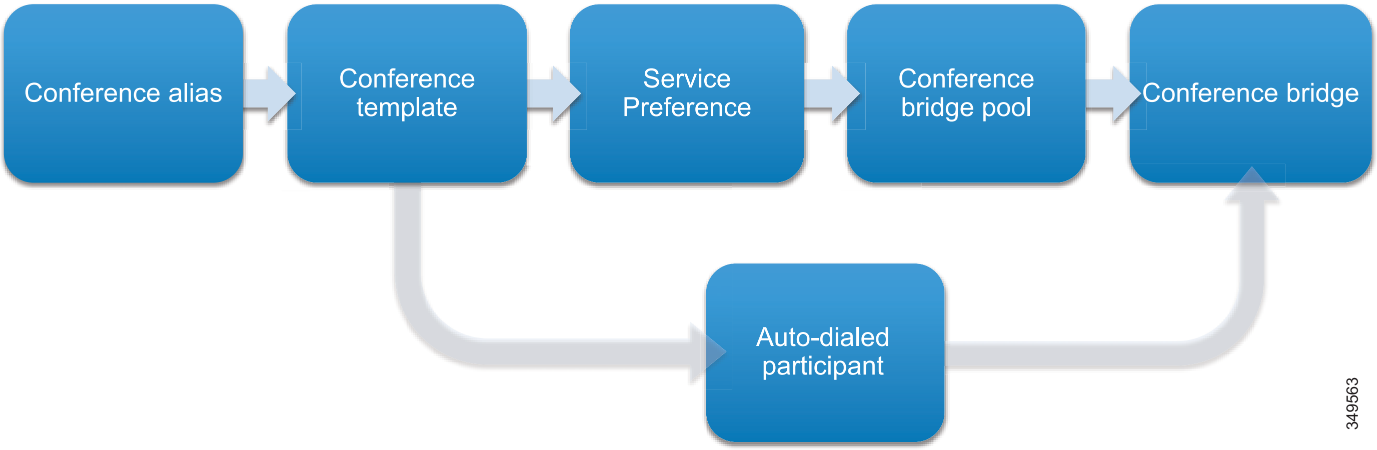


Figure 3-13 TelePresence Conductor Internal Configuration Flow for Scheduled Conferences

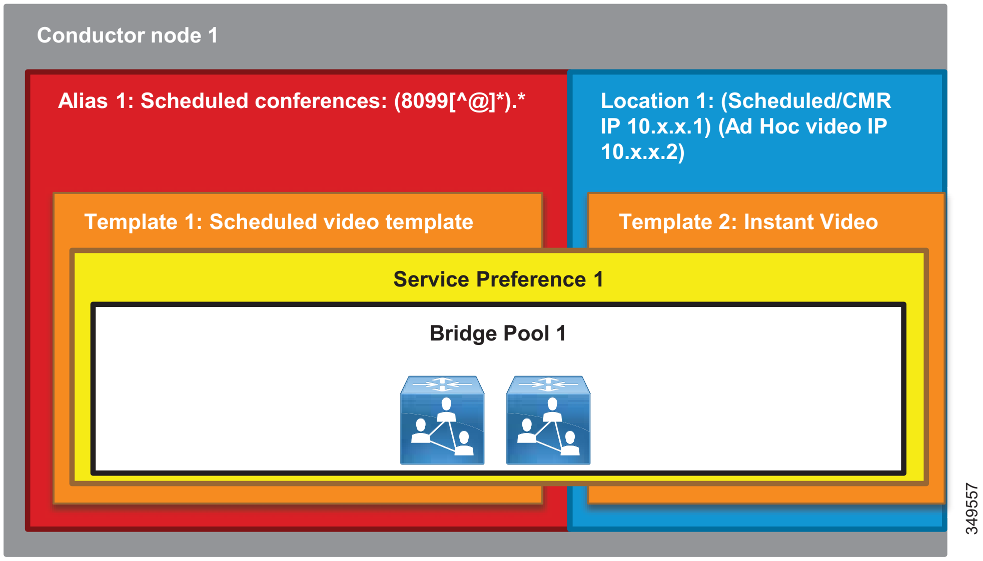


Deployment Tasks Common to Instant and Scheduled Conferences on TelePresence Conductor

Resolve all alarms within TelePresence Conductor before starting configuration. Unresolved alarms prevent TelePresence Conductor from functioning. Also change the root and administrator passwords, and create an API user that Unified CM will use to authenticate in order to access the TelePresence Conductor API. Enable Multiparty Licensing for TelePresence Servers in TelePresence Conductor.

Figure 3-14 shows in detail what elements must be configured within TelePresence Conductor for scheduled and instant conferences that use the TelePresence Server pool.

Figure 3-14 TelePresence Conductor Configuration


Conductor node 1
Alias 1: Scheduled conferences: (8099[^@]*).*
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
Template 1: Scheduled video template
Template 2: Instant Video
Service Preference 1
Bridge Pool 1
349557

Each TelePresence Server in the deployment must be assigned to a bridge pool in TelePresence Conductor. A TelePresence Server can belong to only one bridge pool. Bridge pools must reflect the physical location of the TelePresence Server, and if a bridge pool contains multiple TelePresence Servers, they should all be of a similar size. Each TelePresence Conductor is dedicated to a region; therefore, in this document we assume that all TelePresence Servers are in the same physical location. If there are multiple physical locations, you must configure a location and pool for each one in order to maintain call admission control by Unified CM. As discussed previously, if you are using a dedicated scheduling model, only a single TelePresence Server dedicated for scheduling should appear in a bridge pool, whereas a shared model can have multiple TelePresence Servers in the bridge pool.

A service preference is a prioritized list of bridge pools set up through TelePresence Conductor, and it defines the order for using pools if conference resources are limited. For any particular conference, the administrator can determine the order of preference for the pools that TelePresence Conductor will attempt to use to host that conference. If no TelePresence Servers in the first bridge pool can be used to host a conference (for example, insufficient resources are available to meet the conference requirements), TelePresence Conductor will check the second bridge pool in the list. The scheduled conferences and instant conferences shown in Figure 3-14 all use the same service preference and therefore use the same bridge pool. It is possible that each could use a different service preference and therefore a different bridge pool, or that multiple bridge pools could be defined in a service preference.

A service preference can contain 1 to 30 conference bridge pools, and a single conference bridge pool can be used in any number of service preferences.

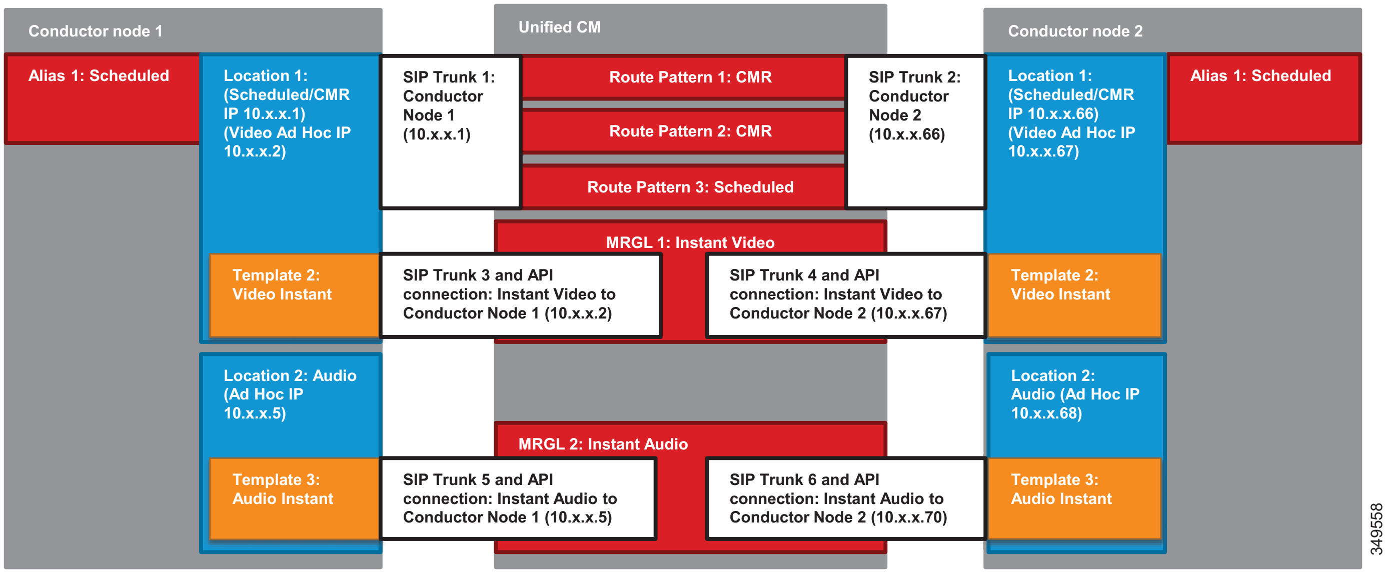
As with pools, all TelePresence Servers configured in a TelePresence Conductor service preference must be in the same physical location to ensure that media is always sent to the same physical location and therefore bandwidth usage is understood.

Each SIP trunk between Unified CM and TelePresence Conductor must use a unique IP address. The IP addresses must correspond to the IP addresses configured in the locations within TelePresence Conductor. A location may be dedicated to instant, CMR/scheduled, or both conference types. If both conference types are enabled for a location, that location requires two IP addresses. (Figure 3-15)

Up to 64 IP addresses can be configured on TelePresence Conductor to accommodate many different conference types and locations.

Note Each TelePresence Conductor node must have a unique set of IP addresses.

Figure 3-15 TelePresence Conductor Location and Unified CM IP Addresses

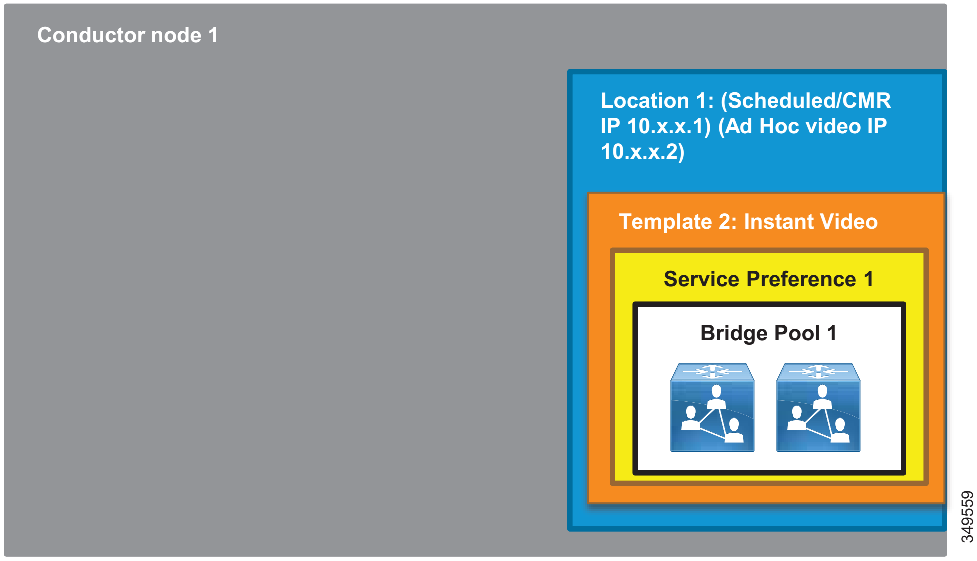

Conductor node 1
Alias 1: Scheduled
Location 1: (Scheduled/CMR IP 10.x.x.1) (Video Ad Hoc IP 10.x.x.2)
Template 2: Video Instant
Location 2: Audio (Ad Hoc IP 10.x.x.5)
Template 3: Audio Instant
Unified CM
SIP Trunk 1: Conductor Node 1 (10.x.x.1)
Route Pattern 1: CMR
Route Pattern 2: CMR
Route Pattern 3: Scheduled
SIP Trunk 2: Conductor Node 2 (10.x.x.66)
MRGL 1: Instant Video
SIP Trunk 3 and API connection: Instant Video to Conductor Node 1 (10.x.x.2)
SIP Trunk 4 and API connection: Instant Video to Conductor Node 2 (10.x.x.67)
MRGL 2: Instant Audio
SIP Trunk 5 and API connection: Instant Audio to Conductor Node 1 (10.x.x.5)
SIP Trunk 6 and API connection: Instant Audio to Conductor Node 2 (10.x.x.70)
Conductor node 2
Location 1: (Scheduled/CMR IP 10.x.x.66) (Video Ad Hoc IP 10.x.x.67)
Alias 1: Scheduled
Template 2: Video Instant
Location 2: Audio (Ad Hoc IP 10.x.x.68)
Template 3: Audio Instant
349558

Deployment Tasks for Instant Conferences on TelePresence Conductor

Once IP addresses have been added to TelePresence Conductor, TelePresence Servers have been assigned to bridge pools, and bridge pools have been added to service preferences, you can create the first template for instant conferences. Each template should have the relevant quality settings to provide correct resource usage by participants. For example, the audio template should be set to mono audio only, with no video or content.

Figure 3-16 shows the elements that must be configured to enable instant video conferencing. As illustrated in Figure 3-15, each type of instant service (video and audio) must have a unique location and template.

Figure 3-16 TelePresence Conductor Instant Video Conference Location and Template


Conductor node 1
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
Template 2: Instant Video
Service Preference 1
Bridge Pool 1
349559

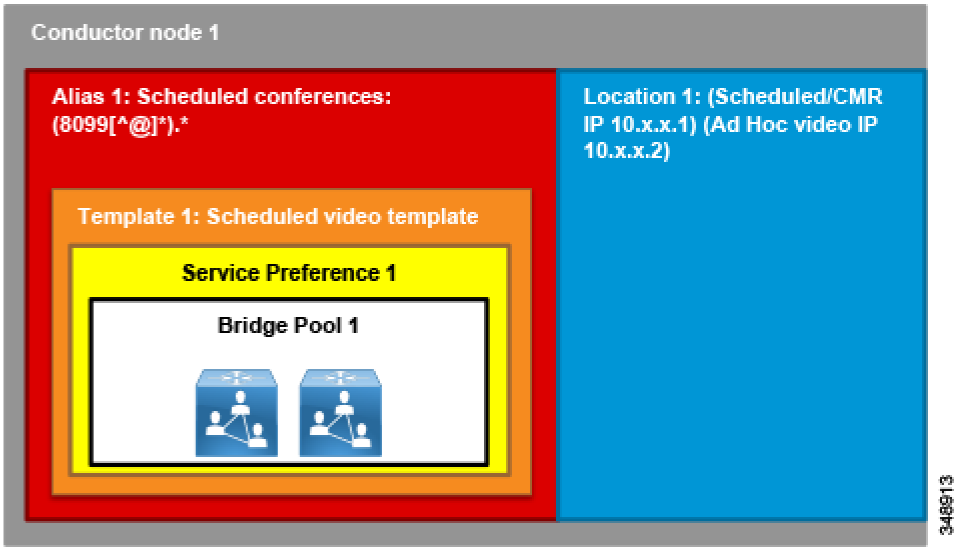
Create a unique location for each instant conference template. A single location may be used for both an instant template and scheduled/CMR conferences. For instant conferences, the two most important location configuration parameters are the instant conference template and the IP address dedicated to instant conferences. Unified CM uses this IP address to send requests to TelePresence Conductor to create an instant conference, and TelePresence Conductor uses the template to determine the maximum quality for the created conference.

Deployment Tasks for Scheduled Conferences on TelePresence Conductor

Scheduled conferences are enabled through a template that is configured to allow scheduling. Multiple templates may be configured; for instance, it might be necessary to create a second template configured for multi-screen systems. Each of these templates should have the relevant quality settings to provide correct resource usage by participants.

Figure 3-17 shows the elements that must be configured to enable scheduled conferencing.

Figure 3-17 TelePresence Conductor Scheduled Template


Conductor node 1
Alias 1: Scheduled conferences: (8099[^@]*).*
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
Template 1: Scheduled video template
Service Preference 1
Bridge Pool 1
348913

TelePresence Conductor uses conference aliases to choose the relevant conference template for an incoming call to a location. Aliases use Regex and can match based on the dialed number. The incoming number range is limited by the Unified CM route pattern assigned to the SIP trunk using the TelePresence Conductor Rendezvous IP address (configured later). Although Unified CM can point a large range of numbers to the SIP trunk, TelePresence Conductor aliases may be used to divide this range into smaller segments, each of which may use a different template. This is done by using Regex to match only a portion of the number range for each alias.

Create a location within TelePresence Conductor for conferences. A single configured location may be used for both an instant template and scheduled conferences. For scheduled conferences, the most important location configuration parameter is the **Rendezvous IP address**. Unified CM uses this IP address to send call signaling to TelePresence Conductor for CMR and scheduled calls. To facilitate outbound dialing from TelePresence Conductor, add up to three trunk IP addresses to the location, using Unified CM call processing subscriber node IP addresses that should receive these calls.

To route outbound calls to the correct SIP trunk, assign the relevant location to the bridge pool used for CMR and scheduled conferences.

Summary

After you complete the deployment tasks outlined above, TelePresence Conductor will be ready to add to Unified CM.

4. Enable Unified CM for Scheduled and Non-Scheduled Conferences

This section describes the major tasks required to enable Unified CM for scheduled and non-scheduled conferences.

Overview

Deployment Tasks to Enable Unified CM for Instant Conferences:

1. Create a new SIP profile named **Standard SIP Profile for TelePresence Conductor**.
2. Create three SIP trunks and point each SIP trunk to the relevant IP address configured in the TelePresence Conductor location for each type of instant conference:
   - Instant audio SIP trunk
   - Instant video SIP trunk

   This step must be repeated for each TelePresence Conductor node. For example, if there are three TelePresence Conductor nodes, there should be three sets of two SIP trunks configured, one set of two for each TelePresence Conductor node.
3. Create two conference bridges and add a SIP trunk (configured in task 2) to each. Each conference bridge should contain the relevant trunk:
   - Instant audio conference bridge
   - Instant video conference bridge

   Configure each conference bridge with the username and password created on TelePresence Conductor for API usage.

   This step must be repeated for each TelePresence Conductor node. For example, if there are three TelePresence Conductor nodes, there should be three sets of two conference bridges configured, one set of two for each TelePresence Conductor node.
4. Create two media resource groups (MRGs):
   - Instant audio MRG
   - Instant video MRG

   Add all conference bridges of the matching type (one per TelePresence Conductor node) to the relevant MRG. If you are using three TelePresence Conductor nodes, then each MRG should have three conference bridges in it, each pointing to the same instant template at the relevant IP address on each node.
5. Create two media resource group lists (MRGLs) and add one MRG to each:
   - Instant audio MRGL
   - Instant video MRGL

   To allow an endpoint to use instant conferencing, assign the appropriate MRGL to the device pool or the device itself.

Deployment Tasks to Enable Unified CM for CMR and Scheduled Conferences:

6. Create a SIP trunk and point it to the relevant IP address configured in the Rendezvous IP field in the TelePresence Conductor location:

   CMR and Scheduled conferences (ST_CONDUCTOR_CMR_SCHED)

   This step must be repeated for each TelePresence Conductor node. For example, if there are three TelePresence Conductor nodes, there should be three SIP trunks configured, one for each TelePresence Conductor.

7. Create a route group:

   CMR and Scheduled route group (RG_CMR_SCHED)

   Add all SIP trunks to the route group. If you are using three TelePresence Conductor nodes, then the route group should have three SIP trunks in it, each pointing to the relevant IP address of each TelePresence Conductor node.

8. Create a route list and add the route group to it:

   CMR and Scheduled route list (RL_CMR_SCHED)

9. Create a route pattern that matches the scheduled alias configured in TelePresence Conductor created earlier (8099[12]XXX). Further route patterns are required to configure CMR, and they are discussed in section 6. Deploy Cisco Collaboration Meeting Rooms, page 3-54.

Deployment Considerations

Unified CM is the first point of logic that decides how to route a call and that chooses the TelePresence Conductor template used for a conference based on its configuration. Unified CM has different configuration procedures for instant and CMR or scheduled conferences because the mechanism for joining each type of conference is different.

Note The endpoint used to initiate an instant conference must have a conference button. Endpoints that do not have a conference button can still be participants in an instant conference, but they must be added to the conference by an endpoint that has a conference button.

Instant and Permanent Conferences

Figure 3-18 Unified CM Internal Configuration Flow for Instant Conferences

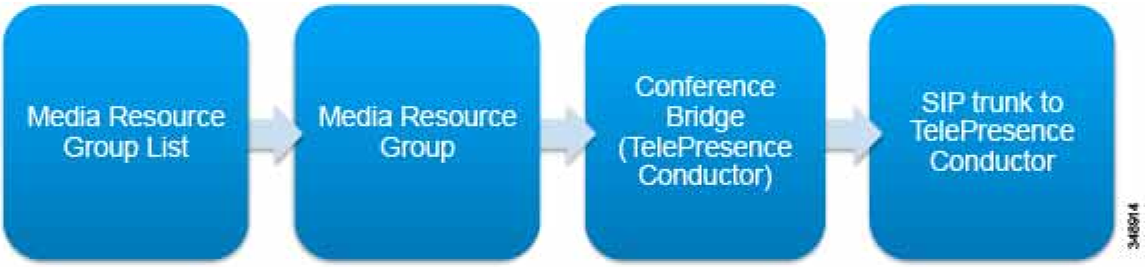
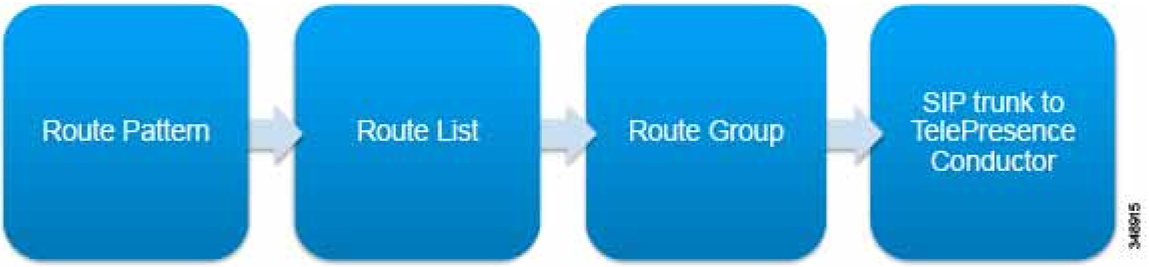

Media Resource Group List
Media Resource Group
Conference Bridge (TelePresence Conductor)
SIP trunk to TelePresence Conductor
348914

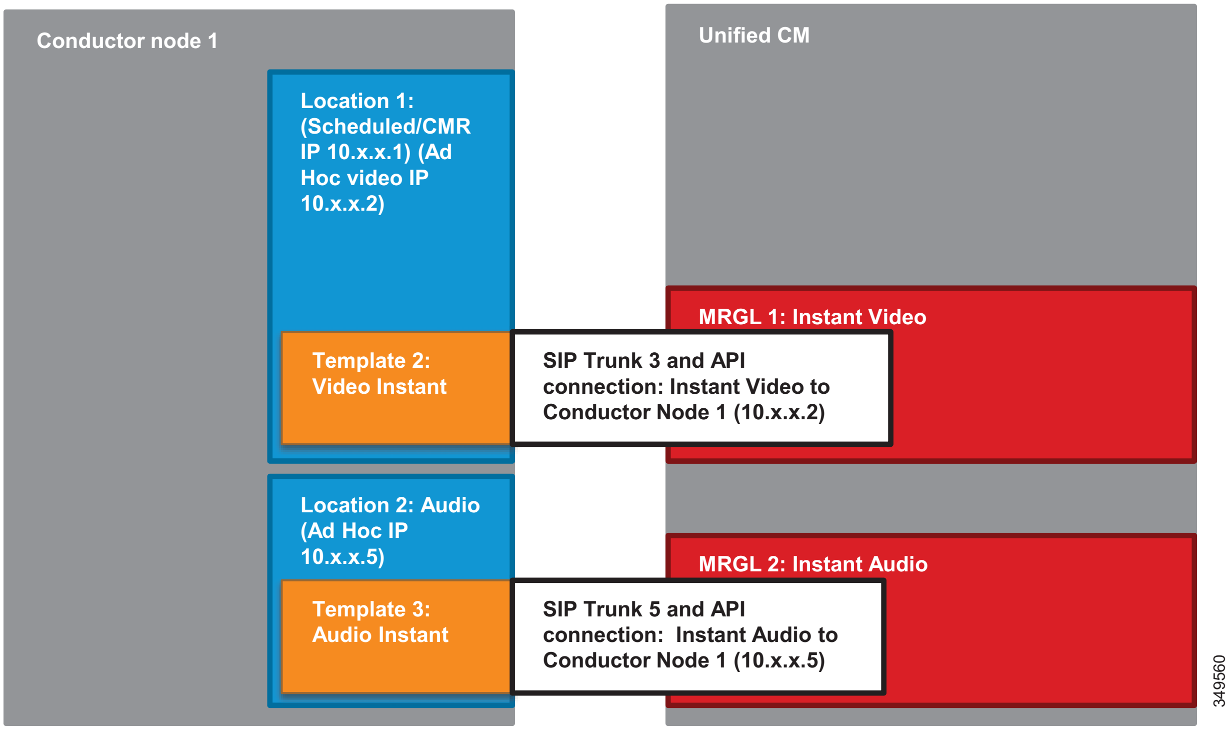
Figure 3-19 Unified CM Internal Configuration Flow for CMR and Scheduled Conferences



Deployment Tasks to Enable Unified CM for Instant Conferences

It is important to understand the relationship between the Unified CM configuration and the TelePresence Conductor configuration. For this deployment two locations were configured in TelePresence Conductor, and in each an Ad hoc IP address was added for instant conferences. Any calls sent to one of these IP addresses uses the conference template configured with it in the location. To route calls correctly, it is important to understand which SIP trunks in Unified CM should point to which IP addresses configured within TelePresence Conductor.

Figure 3-20 Unified CM and TelePresence Conductor Instant Relationship


Conductor node 1
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
Template 2: Video Instant
Location 2: Audio (Ad Hoc IP 10.x.x.5)
Template 3: Audio Instant
Unified CM
MRGL 1: Instant Video
SIP Trunk 3 and API connection: Instant Video to Conductor Node 1 (10.x.x.2)
MRGL 2: Instant Audio
SIP Trunk 5 and API connection: Instant Audio to Conductor Node 1 (10.x.x.5)
349560

Each type of instant conference requires a unique SIP trunk and conference bridge configured in Unified CM to communicate with TelePresence Conductor. Each SIP trunk can be used for only one instant conference, and CMR or scheduled conferences require their own SIP trunks. Each TelePresence Conductor node also requires a unique set of SIP trunks and conference bridges.

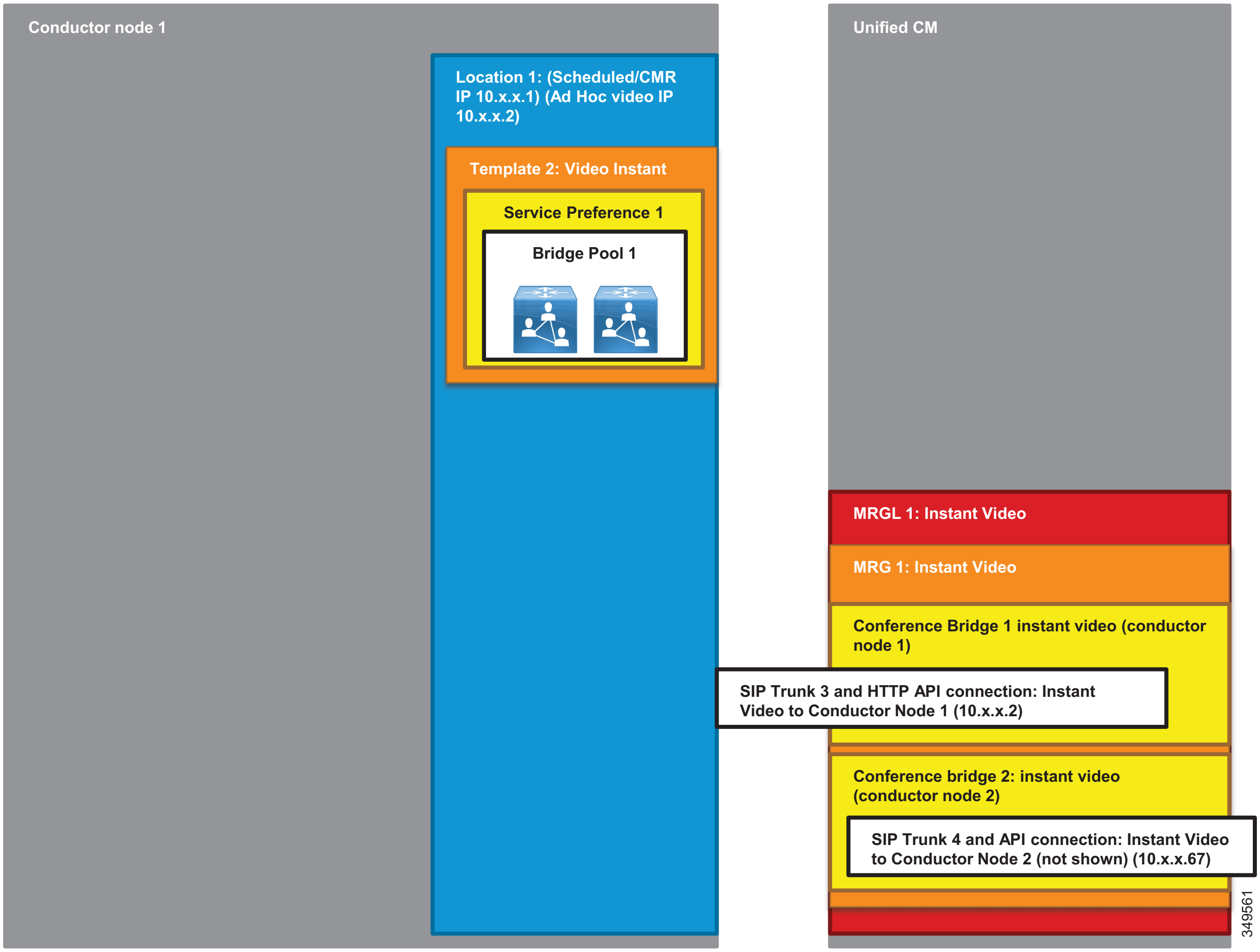
SIP trunks to TelePresence Conductor require a customized SIP profile in order to support calls in all scenarios. To create the SIP profile, copy the **Standard SIP Profile for TelePresence Conferencing** and name the copy **Standard SIP Profile for TelePresence Conductor**, then change the settings as indicated in Table 3-7.

Table 3-7 Settings for SIP Profile

Setting	Value	Comment
Timer Invite Expires (seconds)	30	Causes the trunk timer to expire at the same time as TelePresence Conductor's timer
Early Offer support for voice and video calls	Best Effort (no MTP inserted)	This is the recommended configuration for all Unified CM trunks. Best Effort Early Offer trunks never use MTPs to create an Early Offer and, depending on the calling device, may initiate an outbound SIP trunk call using either Early Offer or Delayed Offer. In the context of this design, outbound calls always use Early Offer.

SIP trunks inform Unified CM where to route SIP traffic. In the case of instant conferences, the SIP trunks also inform Unified CM where to direct API requests, and they are used in the conference bridge configuration. SIP trunks connected to TelePresence Conductor can be configured to be secure; but for the purpose of this guide, they are assumed to be configured as non-secure.

Figure 3-21 Unified CM Instant Configuration


Conductor node 1
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
Template 2: Video Instant
Service Preference 1
Bridge Pool 1
Unified CM
MRGL 1: Instant Video
MRG 1: Instant Video
Conference Bridge 1 instant video (conductor node 1)
SIP Trunk 3 and HTTP API connection: Instant Video to Conductor Node 1 (10.x.x.2)
Conference bridge 2: instant video (conductor node 2)
SIP Trunk 4 and API connection: Instant Video to Conductor Node 2 (not shown) (10.x.x.67)
349561

Conference bridge configuration provides two key pieces of information to Unified CM: the API credentials to communicate with TelePresence Conductor and the destination address for that communication. The username and password can be the same for each conference bridge that points to TelePresence Conductor. These credentials should match those configured in the TelePresence Conductor configuration. The SIP trunk configured in the conference bridge indicates to Unified CM where to send HTTP API traffic. Configure each SIP trunk with the settings indicated in Table 3-8.

Table 3-8 SIP Trunk Settings for Instant Conferences

Setting	Value	Comment
Name	ST_CONDUCTOR_ADHOC_AUDIO1-1	Prefix "ST_" to avoid name collisions with other devices stored in the same table internally. The remainder of the name identifies the conference type and TelePresence Conductor node that the trunk points to.
Description		Some meaningful description
Device Pool	Trunks_and_Apps	Common device pool for central trunks
Media Resource Group List	<None>	Use the MRGL defined on the device pool
AAR Group	Default	Same everywhere
Transmit UTF-8 for Calling Party Name	Checked	This will allow the ASCII Alerting Name to be transmitted to devices that support UTF-8 characters
PSTN Access	Not checked	
Run On All Active Unified CM Nodes	Checked	This settings is recommended on all SIP trunks. It makes sure that outbound calls to SIP do not require intra-cluster control signaling between Unified CM call processing subscribers.
Inbound Calls		
Calling Search Space	TelePresenceConferencing	As defined in the Call Control chapter
AAR Calling Search Space	PSTNReroute	
Outbound Calls		
Use Device Pool Called Party Transformation CSS	Checked	
Use Device Pool Calling Party Transformation CSS	Checked	
SIP Information		
Destination	10.X.X.2	TelePresence Conductor audio Ad hoc IP address
SIP Trunk Security Profile	Non Secure SIP Trunk Profile	Default SIP trunk security profile
SIP Profile	Standard SIP Profile for TelePresence Conductor	Use the SIP profile created above

Once all conference bridges are configured, they can be added to media resource groups (MRG). Each media resource group should represent one type of instant conference and should contain one conference bridge from each TelePresence Conductor node, so that if communication with one TelePresence node is not possible, then calls can be routed to another node.

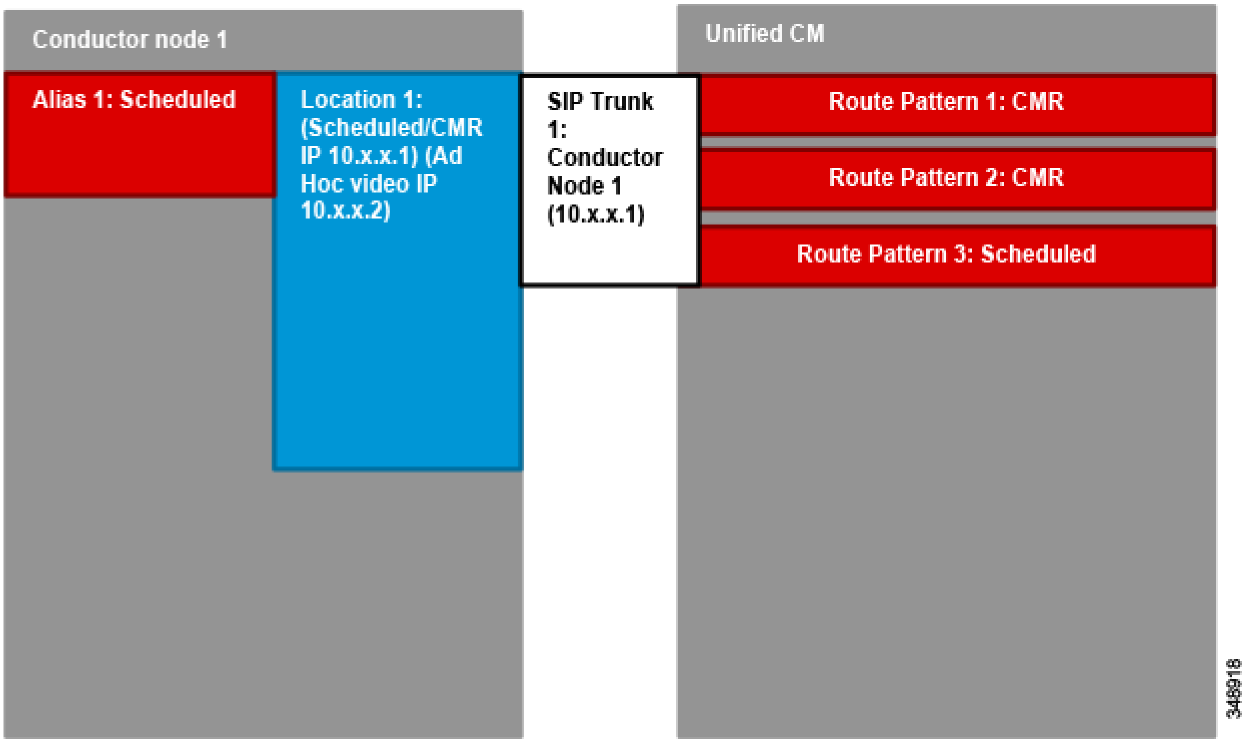
Each media resource group can then be added to its own media resource group list (MRGL). The media resource group list can be assigned to devices or the device pool in Unified CM and used when those devices escalate a point-to-point call to a conference call using the conference button.

In total, two MRGLs should be configured: one for instant audio conferences and one for instant video conferences. These two MRGLs should each point to a different location configured within TelePresence Conductor and reference the instant template configured within the location.

Deployment Tasks to Enable Unified CM for CMR and Scheduled Conferences

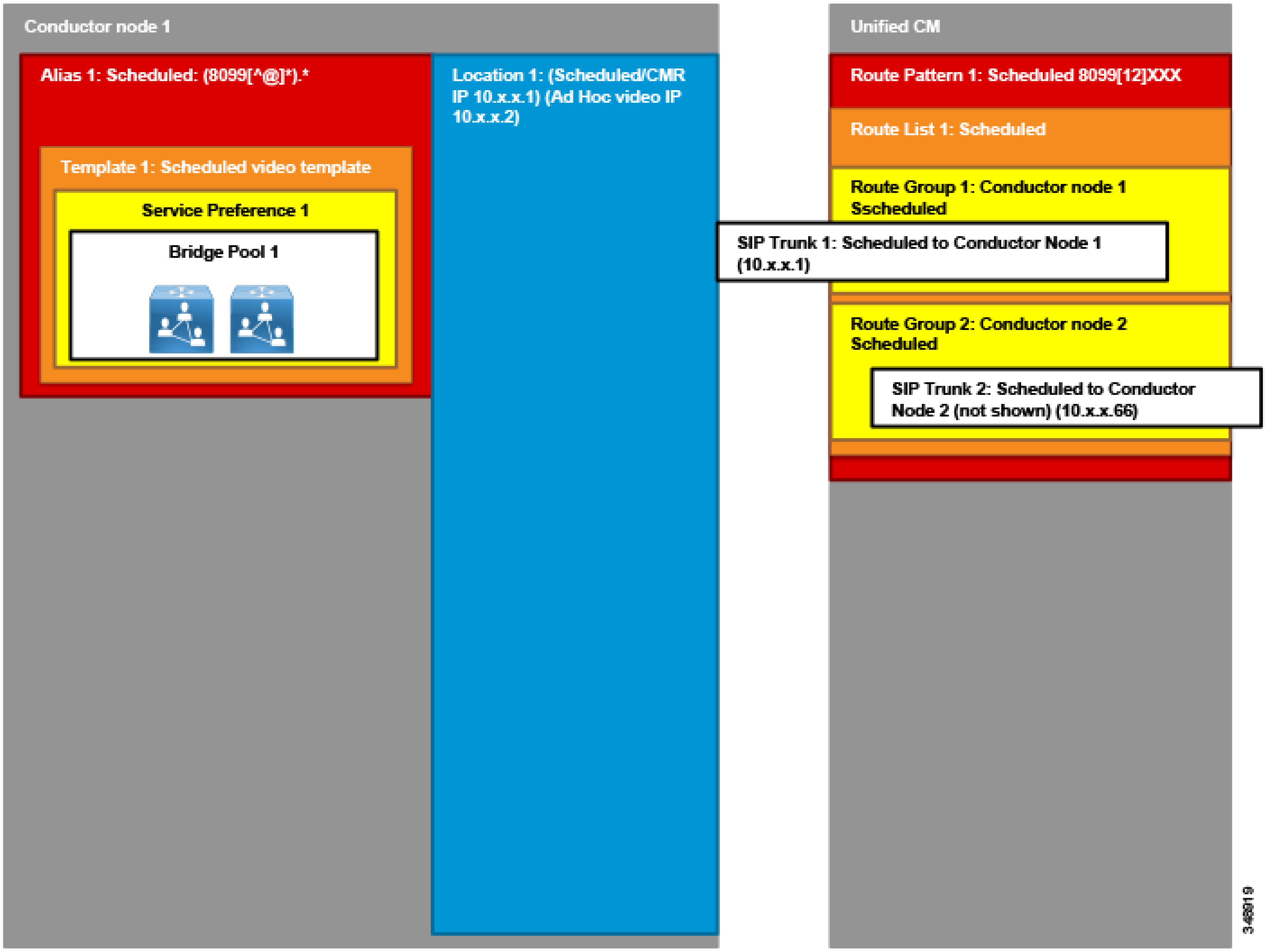
CMR and scheduled conferences are configured on Unified CM in a similar way to instant conferences, but they require a dial plan to be configured rather than media resources.

Figure 3-22 Unified CM and TelePresence Conductor CMR and Scheduled Relationship


Conductor node 1
Alias 1: Scheduled
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
SIP Trunk 1: Conductor Node 1 (10.x.x.1)
Unified CM
Route Pattern 1: CMR
Route Pattern 2: CMR
Route Pattern 3: Scheduled
348918

First configure a SIP trunk to connect to TelePresence Conductor, using the relevant IP address configured in the TelePresence Conductor locations created for your deployment. Each TelePresence Conductor node requires a unique SIP trunk. Use the same SIP profile for CMR and scheduled conferences that you created for instant conferences, with the settings indicated in Table 3-8.

Figure 3-23 Unified CM CMR and Scheduled Configuration


Conductor node 1
Alias 1: Scheduled: (8099[^@]*).*
Template 1: Scheduled video template
Service Preference 1
Bridge Pool 1
Location 1: (Scheduled/CMR IP 10.x.x.1) (Ad Hoc video IP 10.x.x.2)
Unified CM
Route Pattern 1: Scheduled 8099[12]XXX
Route List 1: Scheduled
Route Group 1: Conductor node 1 Sscheduled
SIP Trunk 1: Scheduled to Conductor Node 1 (10.x.x.1)
Route Group 2: Conductor node 2 Scheduled
SIP Trunk 2: Scheduled to Conductor Node 2 (not shown) (10.x.x.66)
348019

After configuring all the SIP trunks, create a route group for each of them.

Add each route group for a particular conference type into a unique route list. There should be one route list per TelePresence Conductor location. The route list is chosen when a call matches a route pattern that points to it.

To route calls through the relevant SIP trunk to the relevant location within TelePresence Conductor, configure a route pattern for the route list. The route pattern should match with the alias range configured for the required template(s), as indicated in Table 3-9.

Table 3-9 Route Pattern for Scheduled Conference Route List

Pattern	Partition	Gateway or Route List	Description
8099[12]XXX	ESN	RL_CMR_SCHED	Pattern to match scheduled alias range

Summary

After you complete the deployment tasks outlined above, Unified CM should be able to communicate with TelePresence Conductor.

5. Deploy TelePresence Management Suite

This section describes the deployment tasks for Cisco TMS for scheduled conferences using TelePresence Conductor.

Overview

Cisco TMS can provide the following services within a conferencing deployment:

- Performs scheduling and conference control functions for scheduled conferences
- Enables the provisioning and monitoring of personal collaboration meeting rooms (CMRs) in conjunction with TelePresence Conductor (See the section 6. Deploy Cisco Collaboration Meeting Rooms.)
- Manages and allocates resources for CMR Hybrid (See section 6. Deploy Cisco Collaboration Meeting Rooms.)

Deployment Tasks for Cisco TMS High Availability:

1. Install and configure Cisco TMS on active and passive nodes.
2. Install and configure the network load balancer (NLB).
3. Configure file sharing between active and passive node servers.

Deployment Tasks for Cisco TMS Basic Configuration:

4. Configure ISDN and IP zones.
5. Configure Active Directory integration, group structure, and users.
6. Create the TMS System Navigator folder structure.
7. Configure default conference setting.

Deployment Tasks for Cisco TMS for Scheduled Conferences:

8. Integrate TelePresence Conductor with TMS.
9. Integrate Unified CM with TMS.
10. Add conference room endpoints to TMS.
11. Install and configure TMS Extensions for Microsoft Exchange (TMSXE).

Deployment Tasks for Cisco TMS High Availability

This section describes the tasks required to deploy Cisco TMS with high availability.

Install and Configure Cisco TMS on Active and Passive Nodes

Cisco TelePresence Management Suite (TMS) should be installed for redundant deployments according to the guidelines in the *Cisco TelePresence Management Suite Installation and Upgrade Guide*, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-tms/products-installation-guides-list.html

- Install the application on the primary server.
- Point to the external SQL resource configured in the planning stage.
- Make note of the encryption key.
- Verify basic operation by logging into the web portal and enabling TMS redundancy.
- Install the application on the second server using the encryption key from the first server, and using the same SQL credentials as the first server.

Both servers will access the single SQL database that holds all conferencing and configuration data. In the active and passive node configuration, a single encryption key and certificate are used for both servers. Having this encryption key and certificate on each server allows for all communications from end users to TMS, and from TMS to managed devices, to be done using secure protocols.

Install and Configure Network Load Balancer (NLB)

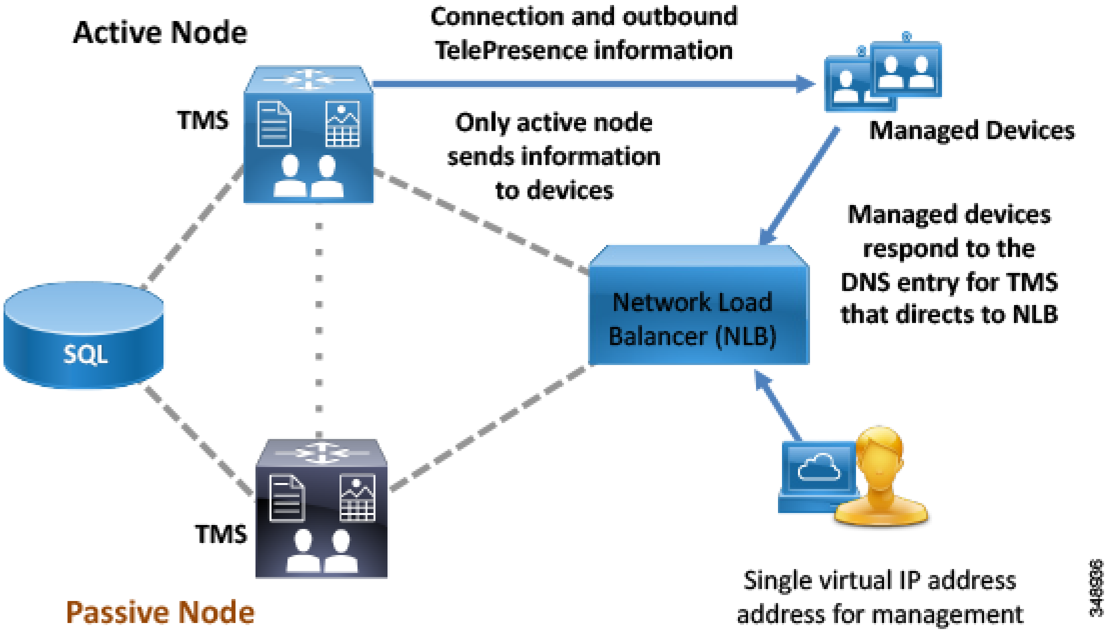
The specifics of the network load balancing configuration are left to the instructions of the load balancer chosen by the customer. The following are functional requirements that must be configured:

- Forward HTTP, HTTPS, and SNMP traffic to the active node.
- Configure the network load balancer probe to the Probe URL within Cisco TMS.
- Push all traffic to the active node.

The Cisco TMS server sends outbound communications directly to managed devices without routing that traffic through the NLB. However, all return communications from managed devices and all web portal requests must be routed through the NLB. The communication path permits end users and endpoints to use a single address, regardless of which TMS server node is in active mode.

Configure TMS Network Settings to the FQDN of the TMS address configured on the network load balancer. This setting within TMS will populate the address that the managed devices use to initiate communications to TMS. By using a FQDN of tms.company.com that resolves to the load balancer, all inbound traffic from endpoints or end user web clients will be directed through the NLB and resolve to the active node. (See Figure 3-24.)

Figure 3-24 NLB Directs Communications from Managed Devices to the Active TMS Node


Active Node
Connection and outbound TelePresence information
TMS
Only active node sends information to devices
Managed Devices
Managed devices respond to the DNS entry for TMS that directs to NLB
SQL
Network Load Balancer (NLB)
TMS
Passive Node
Single virtual IP address address for management
348936

Configure File Sharing Between Active and Passive Node Servers

While the SQL database is used for all operational data, some application specific files are stored within the file structure of the host server. These customizable files are added by the TMS application and must be synchronized between the two servers when using a redundant deployment. The files include software and images that can be uploaded to Cisco TMS, and images created by Cisco TMS.

In a default installation the files are located at:

C:\Program Files\TANDBERG\TMS\Config\System\

C:\Program Files\TANDBERG\TMS\Data\GenericEndpoint\

C:\Program Files\TANDBERG\TMS\Data\SystemTemplate\

C:\Program Files\TANDBERG\TMS\wwwTMS\Data\CompanyLogo\

C:\Program Files\TANDBERG\TMS\wwwTMS\Data\ExternalSourceFiles\

C:\Program Files\TANDBERG\TMS\wwwTMS\Public\Data\SystemSoftware\

Use the Distributed File System (DFS) function within the Windows Server operating system to complete this replication process between the two servers. DFS will keep these folds in sync between the two servers when the "Full mesh" configuration is used.

Deployment Tasks for Cisco TMS Basic Configuration

Perform the following additional configuration tasks during the installation of Cisco TMS to make the deployment function as intended in the Preferred Architecture:

- ISDN and IP Zones
- Active Directory Integration, Group Structure, and Users
- System Navigator Folder Structure
- Default Conference Settings
- Default Conference Settings
- Modify Email Templates within TMS

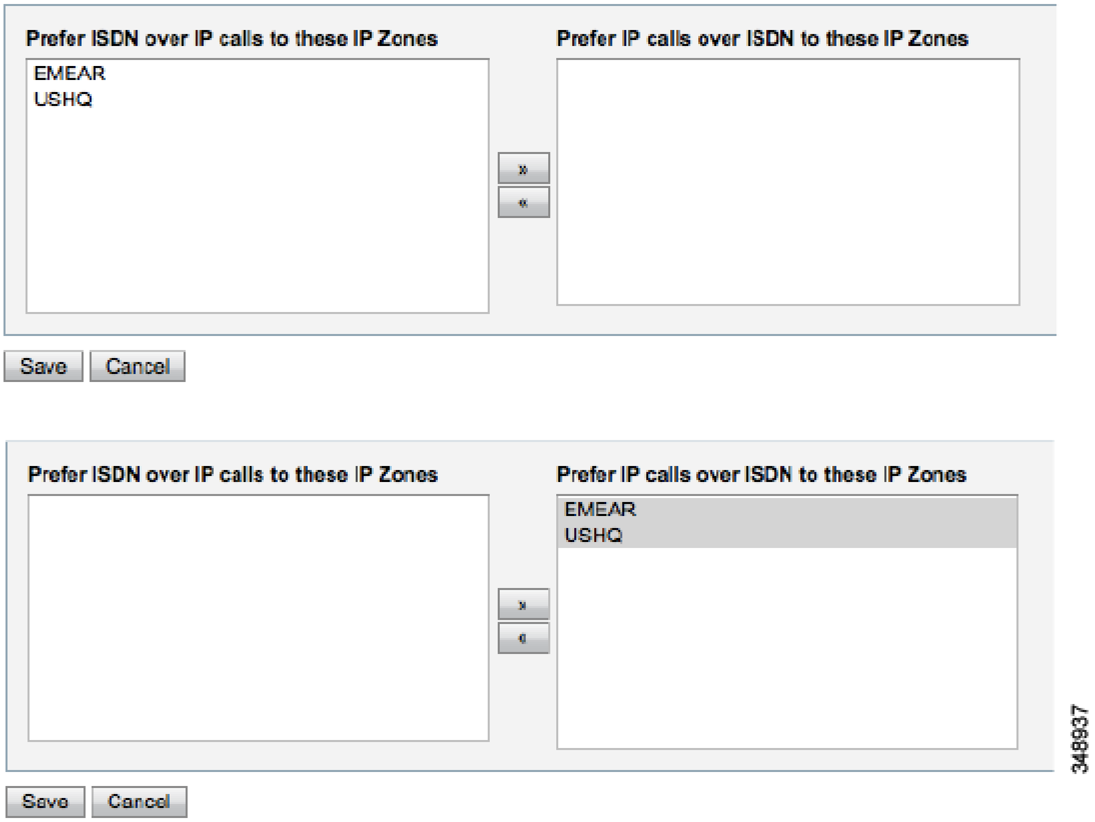
ISDN and IP Zones

Configure additional IP zones for each location that will have conferencing resources.

Each location that has conferencing resources should be identified with a unique IP Zone.

Note Be sure to populate the **Prefer IP calls over ISDN to these IP Zones** according to your network configuration (see Figure 3-25). By default, all new IP Zones will "prefer" ISDN over IP to all other existing IP Zones. Failure to make this selection could cause a failure in conference configuration by TMS because TMS will not see any way to route calls between zones.

Figure 3-25 Configure IP Zones to Prefer IP Calls over ISDN


Prefer ISDN over IP calls to these IP Zones
EMEAR
USHQ
Prefer IP calls over ISDN to these IP Zones
»
«
Save
Cancel
Prefer ISDN over IP calls to these IP Zones
Prefer IP calls over ISDN to these IP Zones
EMEAR
USHQ
»
«
Save
Cancel
348937

Configure any additional ISDN zones.

For each intended ISDN gateway, create an additional ISDN Zone in TMS. This will make endpoints use your intended ISDN gateway for all scheduled calls, as defined in the Collaboration Edge chapter. Each of your ISDN gateways will have a prefix for dialing externally, and that prefix must be configured into TMS.

Active Directory Integration, Group Structure, and Users

Verify that all of the information is correctly entered for your Active Directory service account.

Note Make sure all of your settings for AD connectivity are correct, and test the connection. Other AD interfacing commands within TMS might not display errors, even if AD synchronization is not functioning.

Build a group structure to match your organizational needs using Active Directory Groups.

Three different groups are created by default during the TMS installation:

- Users
- Video Unit Administrator
- Site Administrator

These groups may be modified to meet customer needs, but they cannot be removed. By default, all groups have the same access permissions as Site Administrator.

These default groups are limited to manual entry of users; therefore groups should be imported from Active Directory, and existing Active Directory Groups should be used to manage end user access to TMS functions. Be sure to consider groups for support desk personnel and technical administrators as well as end users who schedule conferences.

For additional information about groups, see the *Cisco TelePresence Management Suite Administrator Guide*, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-tms/products-maintenance-guides-list.html

Using the **Import from AD** feature allows for a single point of end user job function management. When employees are added or removed, or job functions change and organizational Active Directory groups are modified, TMS permissions are automatically updated.

Once you have imported groups from Active Directory, assign appropriate permissions to each group. On the screen that appears, simply uncheck any permissions that you do not want that group to have. Failure to restrict these permissions can result in unintended configuration changes.

Also, be sure to select the appropriate default group for all users.

Note Anyone accessing Cisco TMS will be added automatically to the Users group, and this cannot be unselected. De-select any permissions that the administrator does not want everyone within the organization to have.

Import Users

Once permissions are set for groups, import users using the **Synchronize All Users with AD** function. Depending upon organization size and number of groups involved, the synchronization can take many minutes to complete.

Note Users will not appear in the list of users until they log into TMS for the first time.

System Navigator Folder Structure

The TMS System Navigator utilizes a folder structure to group devices logically for the administrator. Build a folder structure to match your organization’s physical deployment. These folders are visible only to the administrators, not to end users. Arrange the folders according to the logical flow for your organization. For example, create a folder for each geography, and then create a sub-folder for the infrastructure and another folder for conference room endpoints. Folders within the System Navigator may contain endpoints and/or infrastructure devices that receive connection instructions from TMS.

Default Conference Settings

Before scheduling conferences, the administrator should understand the end user community usage model as well as any endpoint limitations. Important Cisco TMS settings to consider include:

- One Button to Push
- Bandwidth
- Add WebEx to all Conferences
- Allow Participants to Join 5 minutes Early

One Button to Push

One Button to Push enables end users to see a calendar of the day's meetings for a particular room and to launch the connection to the conference. Cisco TMS gives users 72 hours worth of calendar information per request.

Bandwidth

This setting is per endpoint. Adjust the bandwidth to the desired setting for your network. To allow for HD main channel and maximum resolution of content, the default bandwidth for non-immersive systems should be set to 2048 kbps. Any endpoint that has a lower setting for maximum bandwidth will join at its maximum bandwidth.

Add WebEx to all Conferences

This setting should be selected to provide a full collaboration experience for each meeting. Select the option to set the Method of User Access to **WebEx (Username)**. This setting will cause the username of the person scheduling the meeting to appear in the WebEx portion of the invitation, and it allows the meeting to populate that end user's Jabber or WebEx mobile client agenda.

Allow Participants to Join 5 minutes Early

This setting should be selected to allow for slight variations of end-user time interfaces. Allowing users to join prior to the exact time of the TMS server provides a more consistent end-user experience and prevents end users from receiving an "unable to connect" message if they attempt to connect to a meeting a few minutes before the meeting start time.

Modify Email Templates within TMS

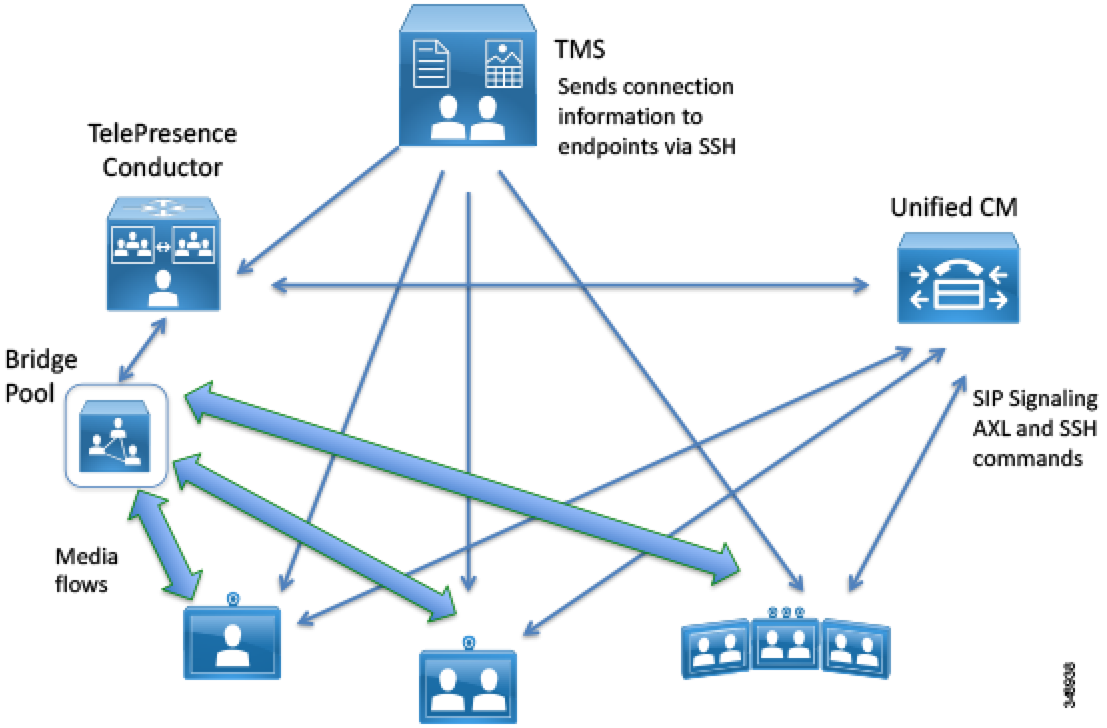
Cisco TMS contains the templates used to notify conference organizers. However, Cisco TMSXE can inject errors, warnings, and informational text into email messages sent by Cisco TMS. These messages can be modified by the administrator. Avoid removing or changing text in curly brackets – for example, {MEETING_TITLE}, {CONTACT_HOST}, and so forth – because these are variables that embed other specific content from the scheduled event.

Look at all email templates to ensure that communications automatically generated by TMS align with your intended procedures. Many of these templates might be rather simplistic and are intended to be enhanced by individual organizations. The templates may be modified using any standard HTML editor.

Deployment Tasks for Cisco TMS for Scheduled Conferences

For Cisco TMS to build scheduled conferences, you must add the needed components into TMS as systems. Unified CM is added to TMS to allow the TMS scheduling mechanisms be aware of the call control entity for all devices. TMS does not control any settings on Unified CM, but it does communicate directly to conference room endpoints managed by Unified CM. (See Figure 3-26.)

Figure 3-26 Cisco TMS Communicates Directly with Unified CM Managed Endpoints


TMS
Sends connection information to endpoints via SSH
TelePresence Conductor
Unified CM
Bridge Pool
SIP Signaling AXL and SSH commands
Media flows
348938

Integrate TelePresence Conductor with TMS

To allow Cisco TMS to perform scheduling and conference control for scheduled conferences, add one TelePresence Conductor node from each TelePresence Conductor cluster. The TelePresence Servers should then be added to Cisco TMS also.

Cisco TMS and TelePresence Conductor must be configured with similar aliases, and these are used by Cisco TMS to determine where a scheduled call is placed. Conferences cannot exceed the size of any single TelePresence Server if cascading is not enabled, which it would not be if you are using a dedicated TelePresence Server deployment.

Add one TelePresence Conductor from each TelePresence Conductor cluster to Cisco TMS. Add them to the appropriate folder and assign them the correct IP zone, using an administrator account configured on the TelePresence Conductor. For each TelePresence Conductor configured in Cisco TMS, set the parameters as listed in Table 3-10.

Table 3-10 Cisco TMS Parameter Settings for TelePresence Conductor

Setting	Value	Comment
IP Zone	Region IP zone	This setting tells Cisco TMS the physical location of this device
Username	TMSadmin	This setting should match the username configured on the TelePresence Conductor
Password	<password>	
Usage Type	Other	

After adding TelePresence Conductor, add each TelePresence Server to Cisco TMS in a way similar to the above method for TelePresence Conductor. Once all TelePresence Servers are added, perform a forced refresh on TelePresence Conductor and then on all the added TelePresence Servers.

As part of preparing to install the TelePresence Conductor for scheduled calls, you created conference alias and identified a numeric range for the TelePresence Conductor to use as part of the dial plan and designated in the SIP trunks. Table 3-11 lists the TelePresence Conductor dial plan parameter settings.

Table 3-11 TelePresence Conductor Dial Plan Settings

Parameter	Value
Numeric ID Base	This is the first number in the scheduled conferencing range of the dial plan.
Numeric ID Step	Keep the default value of 1 to utilize every number in the range.
Numeric ID Quantity	Specify the number of digits allowed in the dial plan for this TelePresence Conductor.
Register with Gatekeeper	This should be set to **off** because the Preferred Architecture uses SIP connections, not H.323.

Leave all other settings as defaults, and save the configuration to add the TelePresence Conductor to TMS.

Cisco TMS will populate the dial plan numbers provided in the previous steps into both E.164 aliases and SIP URIs. However, the implementation of E.164 logic within TMS differs from its use elsewhere in the Preferred Architecture. TMS associates an E.164 alias with H.323 communication only. It is therefore necessary to adjust the integrated ticket system of TMS to ignore certain warnings for the TelePresence Conductor.

Once the TelePresence Conductor has been added to TMS, adjust the Ticket Filters for this entry by adding the filter for **Gatekeeper Mode Off**.

Create an alias in Cisco TMS under the TelePresence Conductor tab to use for scheduled calls. As shown in Table 3-12, the lower settings are visible only after clicking **Save** because they are read-only.

Table 3-12 Cisco TMS Alias Settings for TelePresence Conductor

Setting	Value	Comment
Name	Scheduled meeting	Give the alias a name; for example, Scheduled meeting.
Alias Pattern	8099%	The pattern can be fixed or can contain a variable, which is denoted by %. The alias pattern must match the route pattern on Unified CM. Note that the variable part of the alias will be generated by Cisco TMS from the Numeric ID Base configured for the TelePresence Conductor in **Systems** > **Navigator** > *select the TelePresence Conductor* > **Extended Settings**.
Priority	1	Give the alias a priority. The alias with the lowest number has the highest priority, and it will be used first when Cisco TMS creates a conference. If that alias has no available resources, the alias with the next highest number will be used, and so on. The priority can be any number between 0 and 65535.
Description	Default scheduled conference	Enter a description of this alias.
Prefer for Multiscreen	No	Cisco TMS uses aliases with this field checked when selecting aliases for conferences including immersive TelePresence systems. The alias with the highest priority will be chosen first. If all the immersive aliases are in use, a non-immersive alias will be used for the conference If checked, ensure that the TelePresence Conductor conference template that this alias is configured to use has **Allow multiscreen** set to **Yes**.
Allow Booking	Yes	If **No** is selected, this alias will not be used by Cisco TMS in any bookings. This setting could be used if you want to stop using a particular alias that has a number of future bookings and therefore cannot be deleted. Disabling booking by using this setting will enable the alias to be deleted once the final booking has taken place.
Allow Booking with WebEx	Yes	Set to **Yes** to allow this alias to be used in bookings including Cisco Collaboration Meeting Rooms (CMR) Hybrid.
Max Participants per Conference	N/A	When booking a conference with this alias, if more than this number of participants is selected, it will not be possible to save the conference. The number is a theoretical maximum; the actual number of possible participants in a conference could be much lower, depending on how the associated TelePresence Conductor conference templates are set up. This field is populated only if there is a corresponding alias on the TelePresence Conductor.
Max Screens per Participant	N/A	The maximum number of screens per participant for this alias. This field is populated only if there is a corresponding alias on the TelePresence Conductor.
Regular Expression	N/A	The regular expression of the alias that must be configured on TelePresence Conductor.
Service Preference	N/A	The Service Preference that this alias is linked to. This field will display **None** if there is no corresponding alias on the TelePresence Conductor.

Create the corresponding alias in TelePresence Conductor. First copy the regular expression generated in Cisco TMS in the alias **Regular Expression** field under the TelePresence Conductor tab. In the case shown in Table 3-12, the generated alias would be **(8099[^@]*).***.

In TelePresence Conductor, create a new conference alias and configure it as shown in Table 3-13.

Table 3-13 TelePresence Conductor Alias

Setting	Value	Comment
Name	Default scheduled meeting	Give the alias a name; for example, Scheduled meeting.
Incoming alias	(8099[^@]*).*	Paste the regular expression you copied from Cisco TMS.
Conference name	\1	Enter a regular expression (regex) replacement string that defines how the incoming alias will be modified to result in the conference name; for example, \1. Entering a static value here will not allow concurrent use of the same alias. Make sure that if your alias pattern has a dynamic part, the regex string you enter here reflects that. Note that the TelePresence Conductor Conference name as defined by this field is unrelated to the Conference Title in Cisco TMS.
Priority	1	Enter the priority for this conference alias. The priority is used when the alias that has been dialed matches more than one conference alias. In such cases, the conference alias with the highest priority (closest to 0) will be used.
Conference template	Scheduled Personal Multiparty 720p HD video template	Select one of the conference templates that you assigned for scheduling in the TelePresence Conductor configuration section.
Role type	Participant	This setting determines the privileges that will be assigned to a caller dialing in to the conference using this conference alias. The options that are available are determined by the template that has been selected.
Allow conference to be created	No	This setting determines whether participants dialing this conference alias can create the conference. Scheduled conferences are always created by Cisco TMS and therefore this setting should not be enabled.

Select the TelePresence Conductor within Cisco TMS and **Force refresh** of the device. Additional information is then listed under Service Preferences within the TelePresence Conductor tab.

TelePresence Conductor reports the total capacity of a service preference to Cisco TMS. The Capacity Adjustment setting allows you to specify what percentage of the total capacity will be available for scheduling conferences with this Service Preference.

If you are using bridge pools that are reserved only for scheduling, do not change this setting. If, however, instant and CMR conferences share the bridge pools being used for scheduling, setting the percentage to less than 100% reserves capacity for non-scheduled conferences.

It is also possible to set the percentage higher than 100%. If users regularly book more capacity than they use (for example, 10 dial-ins for a conference where only 5 are ever used), you should set the Capacity Adjustment to 120% or higher.

To use the TelePresence Conductor for scheduled calls, you must edit TelePresence Conductor settings within Cisco TMS. H.323 dialing should be disabled in both directions, Allow Booking should be enabled, and SIP dialing should be enabled in both directions.

Previously within the alias in Cisco TMS, a wild card was configured, and the number range used in place of this wild card must be configured so that the scheduled conference number range matches that configured on Unified CM. Edit the Extended Settings of the TelePresence Conductor in Cisco TMS as listed in Table 3-14. The numeric ID should match the route pattern configured for the trunk to TelePresence Conductor from Unified CM.

Table 3-14 Extended Settings for TelePresence Conductor

Setting	Value	Comment
Numeric ID Base	1000	The first number Cisco TMS uses when creating the variable part of the alias. The combination of the non-variable part of the alias and this number will give the dial string used by participants dialing into the scheduled conference; for example, 80991000.
Numeric ID Step	1	Cisco TMS will add this number to the Numeric ID Base to avoid duplicated aliases. As conferences finish, the alias will be made available to new conferences.
Numeric ID Quantity	1999	The number of times Cisco TMS will increase the number from the Numeric ID Base, using the Numeric ID Step increment. This number should be set so that the highest number does not exceed the allocated range for scheduling: 80991000 to 80992999.
Conference Layout	Default View Family	Set the default layout for all conferences.
Limit Ports to Number of Scheduled Participants	On	Limit ports to the number of scheduled audio and video participants for all conferences to the number scheduled at time of conference booking. No additional participants will be able to join conferences.

It is important to configure Cisco TMS to use TelePresence Conductor for scheduling, otherwise scheduling will fail. In **Administrative Tools** > **Configuration** > **Conference Settings**, edit the settings as shown in Table 3-15.

Table 3-15 Cisco TMS Conference Settings

Setting	Value	Comment
Preferred MCU Type in Routing	Cisco TelePresence Conductor	Prefers TelePresence Servers for scheduling over other devices

Integrate Unified CM with TMS

While Unified CM administers the conference room endpoints for all other aspects of configuration and management, the Unified CM cluster must be added into TMS to allow for booking and connection initiation. To add Unified CM to TMS, perform the following tasks:

- Create an Application User for Cisco TMS within Unified CM
- Add the Publisher for each Unified CM Cluster in Your Environment

Adding multiple Unified CM clusters requires adherence to the dial plan configuration outlined in the Call Control chapter.

Create an Application User for Cisco TMS within Unified CM

This application user allows TMS to communicate with endpoints controlled by Unified CM. This user must be assigned all of the conference room devices within Unified CM that will be scheduled. This user must also be added to a user group just for Cisco TMS, with the following roles:

- Standard AXL API Access
- Standard CTI Enabled
- Standard SERVICEABILITY
- Standard CCM Admin Users
- Standard RealtimeAndTraceCollection

For more information, refer to the *Cisco Unified Communication Manager Configuration Guide for the Cisco TelePresence System*.

Add the Publisher for each Unified CM Cluster in Your Environment

Adding the Unified CM publisher to TMS makes TMS aware of the call control authority for its endpoints. Without knowledge of Unified CM, the TMS scheduling engine cannot properly utilize the full functionality of your deployment, and connection failures could occur.

Add the publisher by the same method used for other devices, by using the application user you created in the above step for the user name and password when prompted by TMS.

Add Conference Room Endpoints to TMS

Rather than adding devices by IP address or DNS name, use the **From List** tab and then select Unified CM. Select all the conference room TelePresence devices that you wish to have available through the scheduling interfaces of TMS. Be sure to select the appropriate IP Zone for each endpoint. This IP Zone will be used to select the best TelePresence Server available for the endpoints in any conference. Make sure the DN for each endpoint in Unified CM complies with the E.164 guidelines listed in the Call Control chapter.

Do not add personal TelePresence devices (for example, Cisco TelePresence DX Series endpoints) that will not be scheduled through Exchange as resources to TMS.

Install and Configure TMS Extensions for Microsoft Exchange

Cisco TelePresence Management Suite Extension for Microsoft Exchange (Cisco TMSXE) is an extension for Cisco TelePresence Management Suite that enables videoconference scheduling via Microsoft Outlook, and it replicates Cisco TMS conferences to Outlook room calendars.

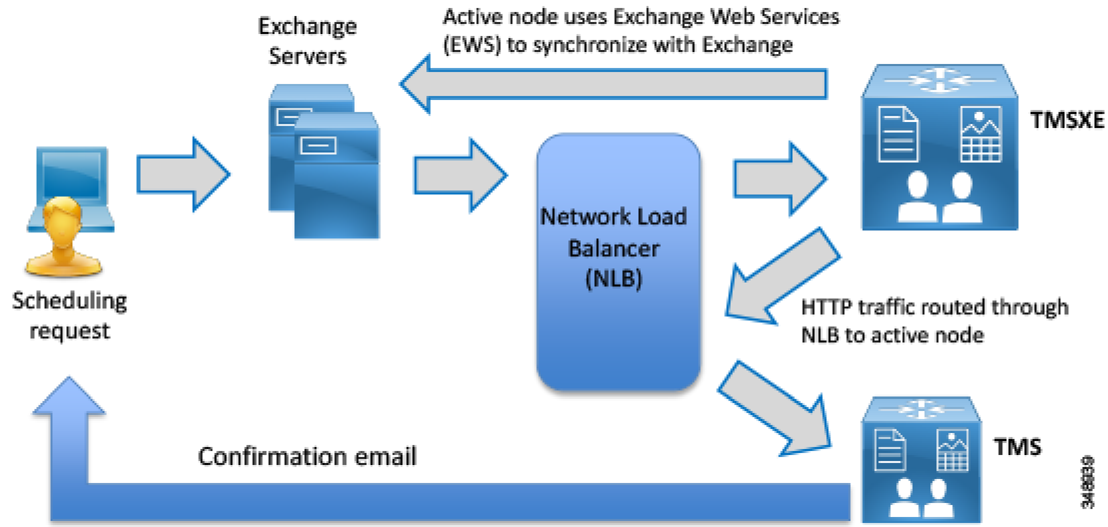
This software extension to TMS requires a license key to activate the functionality within TMS. This key must be installed in TMS before installing the TMSXE software. For deployments with more than 50 scheduled endpoints, TMSXE must be installed on its own server or virtual machine instance.

Prerequisites

Before installing Cisco TMSXE, make sure both Outlook and Exchange are already set up so that users are able to book meetings that include room mailboxes (see Figure 3-27). This integration is licensed either by groups of endpoints or as an Application Integration license key. The correct key must be procured and entered into TMS before proceeding with the installation. If both option keys are added, only the Application Integration Package option will be used by Cisco TMS

Cisco TMSXE may use Microsoft Exchange Resources that are either on-premises, Office 365 hosted deployments, or hybrid customer deployments. Consult the Microsoft Exchange administration and deployment guides for any guidelines or recommendations that might apply to specific customer environments.

Figure 3-27 Sample Flow for Scheduling a Conference by an End User


Exchange Servers
Active node uses Exchange Web Services (EWS) to synchronize with Exchange
TMSXE
Network Load Balancer (NLB)
Scheduling request
HTTP traffic routed through NLB to active node
TMS
Confirmation email
348039

Once the per-system option key has been activated in Cisco TMS, the **Allow Remote Bookings** setting determines whether each system is using a license. This setting allows the administrator to select which endpoints are able to be booked by end users and consume one of the individual endpoint licenses. This setting is void and hidden if the Application Integration Package option is used.

Before endpoints can be added to Cisco TMSXE, they must be represented by a room mailbox in Exchange. To simplify TMSXE setup, we recommend using the endpoint's Cisco TMS display name as the mailbox name (with any spaces removed). This provides commonality across all methods by which end users would see the system name appear.

Special Notes About Privacy Features of Exchange:

All room mailboxes added to Cisco TMSXE must be configured to handle booking subjects and privacy settings in the same way. This means that the following settings must be applied to either all or none of the mailboxes:

- Delete the subject

  We recommend not using this feature so that support staff is able to identify a particular meeting in the Conference Control Center. Also, this will allow the meeting title to appear on the One Button to Push interface of capable endpoints.

- Add the organizer's name to the subject

  Use of this setting should be considered very carefully, and will depend upon organizational culture and practices. Keep in mind that if one person schedules meetings for multiple groups, those meetings will be listed by that scheduler's user name and not by the meeting subject, which might be more beneficial. On the other hand, if meetings are scheduled by their respective hosts, then it would be easy to identify "Bob's meeting" instead of remembering the specific meeting title. For most organizations, we recommend not using this setting.

- Remove the private flag on an accepted meeting

  While the "private" flag is respected within the Outlook client, it is not supported by Cisco TMS, and meeting subjects will be freely viewable:

  - In Cisco TMS
  - On endpoints that support the Meetings calendar, if other individuals also have use of a room used for a meeting where the subject title should not be public within the organization. (For example, if a "Merger meeting" for the chief executive is scheduled in a room also used by lower-level employees who would not need to have knowledge of a pending merger, those lower-level employees would be able to see the meeting on a room system calendar.)
  - If a booking that has a "private" flag in Exchange has its participants or recurrence pattern modified in Cisco TMS, the "private" flag will be removed when these changes are replicated to Exchange.

Create TMSXE User

- Create a TMSXE user in Active Directory and import that user into TMS.
- In TMS, the user needs to be in a new or existing group with the following permissions enabled under Booking:
  - Read
  - Update
  - Book on Behalf of
  - Approve Meeting

Install Certificates

Cisco TMSXE and TMS communicate using HTTPS. The certificate also allows for secure communications between the TMSXE server and the Exchange environment. As with the TMS application server, the same certificate is loaded on both the active and passive nodes of TMSXE, and the certificate DNS entry points to the entry of the Network Load Balance address used for TMSXE.

Run Software Installer

- Be sure to select TMS Booking Service to allow use of WebEx Productivity Tools for scheduling.
- Select the appropriate redundancy option for active or passive nodes.
- Complete the software installation on both active and passive nodes.

Once both the active and passive nodes have been installed, configure the Network Load Balancer with the probe URL for each node.

Configure Cisco TMSXE

- Cisco TMS Connection Information

  Configure TMS connection information using the TMSXE account created in Active Directory to allow the TMSXE application to communicate with the TMS application.
- Configure Exchange Web Services

  Configure Exchange Web Services (EWS) to allow TMSXE to communicate with the Exchange servers for user and resource mailboxes. The credentials used for this connection are also the same TMSXE credentials used elsewhere.
- Align Exchange and TMS Resources

  Align Exchange resources to TMS System IDs. This may be done individually or by using a .csv file as outlined in the *Cisco TelePresence Management Suite Extension for Microsoft Exchanged Deployment Guide*.

Use of WebEx Productivity Tools for End Users

To provide the best experience for end-user collaboration tool usage, deploy WebEx Productivity Tools for all users. WebEx Productivity Tools for CMR Hybrid allow users to synchronously book and configure:

- CMR Hybrid meetings that include both WebEx and TelePresence
- WebEx-only meetings
- TelePresence-only meetings

The panel provides access to simple and advanced settings for both WebEx and TelePresence, including the option to add call-in and call-out TelePresence participants and to allow WebEx participants to join the meeting ahead of start time.

Note that all organizers must be set up with a WebEx user for Productivity Tools to work, even when booking TelePresence-only meetings.

Detailed instructions on configuration and deployment of WebEx Productivity Tools with TelePresence can be found in *Cisco WebEx Site Administration User's Guide*, which is available as web help and a PDF file from your WebEx site.

Summary

After you complete the deployment tasks outlined above, Cisco TMS will be configured to communicate with TelePresence Conductor for scheduled conferencing.

6. Deploy Cisco Collaboration Meeting Rooms

This section describes the major tasks required to deploy Cisco Collaboration Meeting Rooms (CMR).

Overview

Deployment Tasks for Unified CM for CMR:

1. Configure an Early Offer SIP trunk between Unified CM and TelePresence Conductor. The SIP trunk ST_CONDUCTOR_CMR_SCHED configured previously can be used here.
2. Set up new route pattern(s) for CMR numeric alias that point to the route list containing the relevant trunks. The route list RL_CMR_SCHED configured previously can be used here.
3. Create an Imported Global Dial Plan Catalog with the Route String cmr.route, for example, and use the Bulk Administration Tool (BAT) to import CMR SIP URIs into the global dial plan catalog.
4. Create a SIP route pattern with the global catalog's route string for CMR URI that points to the route list (RL_CMR_SCHED) used in task 2.

Deployment Tasks for TelePresence Conductor for CMR:

5. Configure one or more bridge pools and service preferences. Use either the same bridge pools used for scheduling if using a shared scheduling model, or create new bridge pools dedicated to non-scheduled conferences if using a dedicated model.
6. Create an administrator account within TelePresence Conductor with read/write access level and API access enabled.

Deployment Tasks for Cisco TMSPE for CMR:

7. Install and activate Cisco TMSPE in Cisco TMS.
8. Create the user groups **Executive** and **Marketing**, and import users from Active Directory who belong to the corresponding AD groups.
9. Add only a single TelePresence Conductor node from each cluster to TMSPE, using the **TelePresence Conductor setting** option.
10. Create CMR templates that define the conference settings and conference alias, and change the default settings as required. Also, assign multiparty licenses (PMP or SMP) to users within the group in the CMR template.
11. Assign a CMR template to each user group that entitles the users to have CMRs.

Deployment Considerations

Cisco Collaboration Meeting Rooms (CMRs) are similar to permanent conferences created in the TelePresence infrastructure that resides in the enterprise's data center. Each CMR has a unique set of video addresses that a user can call into to start a meeting at any time, and the video addresses can be in the format of numeric aliases or SIP URIs. Each CMR can be associated with an individual user and can be created through the user's Cisco TelePresence Management Suite Provisioning Extension (TMSPE) portal.

Cisco CMR provides an easy way for participants to join a conference using TelePresence, regardless of where those participants are located. Everyone dials into the same virtual meeting room from their laptop, telepresence room, desktop endpoint, or mobile device.

Deploying CMR involves the deployment of Unified CM, TelePresence Conductor, and Cisco TMSPE. The following sections describe the high-level process for deploying each component for CMR.

Tip Before starting CMR, decide on the format of the conference aliases (numeric or SIP URI).

Deployment Tasks for Unified CM for CMR

The main function of Unified CM is to handle call routing to and from TelePresence Conductor. Connect Unified CM to TelePresence Conductor with a SIP trunk enabled for Early Offer. (Use the same trunk as previously configured for scheduled conferences: ST_CONDUCTOR_CMR_SCHED.) When a user dials the CMR conference alias, the call is sent to TelePresence Conductor via the SIP trunk. Similarly, TelePresence Conductor can send calls to Unified CM through the SIP trunk for auto-dial participants. The conference alias has two formats: SIP URI or numeric. The dial plan design should include the call routing for both the numeric alias and SIP URI for CMR. For dial plan design details, refer to the Call Control chapter.

Cisco Collaboration Meeting Rooms (CMR) can be created for each individual user, and the CMR numeric alias can be based upon the user's DID number. Table 3-16 shows the CMR numeric alias ranges for a deployment using the dial plan example from Call Control chapter.

Table 3-16 CMR Numeric Alias Ranges

Site	+E.164 DID Range	CMR Numeric Alias Range
SJC	+1 408 555 4XXX	8-004-4XXX
RTP	+1 919-555 1XXX	8-005-1XXX
RCD	+1 972 555 5XXX	8-006-5XXX

For numeric aliases, configure a route pattern for each site that routes to the TelePresence Conductor route list for permanent conferences, as shown in Table 3-17.

Table 3-17 Route Patterns Configuration for CMR Numeric Alias

Pattern	Partition	Gateway or Route List	Description
80044XXX	ESN	RL_CMR_SCHED	Pattern to match SJC DID range
80051XXX	ESN	RL_CMR_SCHED	Pattern to match RTP DID range
80065XXX	ESN	RL_CMR_SCHED	Pattern to match RCD DID range

For SIP URIs for CMR, keep the domain part the same as the end user directory URI, and manipulate the user part of the URI; for example, *{username}*.cmr@ent-pa.com. Using a single domain for both directory and CMR URI dialing simplifies the dial plan design if CMRs are hosted in more than one Conductor system or cluster. In addition, users need to enter only the user part of the URI to enter the CMR, and Unified CM automatically appends the domain part to complete the dialing. This greatly enhances the user experience.

To set up the call routing for the URIs, create an Imported Global Dial Plan Catalog with the Route String cmr.route, for example. Then create a list of CMR SIP URIs for all users in the system (and whenever new users are added to the system), and use the Bulk Administration Tool (BAT) to import the URIs into the global dial plan catalog. After that, configure a Domain Routing SIP Route Pattern with the global catalog's route string that routes to the TelePresence Conductor route list for permanent conferences, as shown in Table 3-18.

Table 3-18 SIP Route Pattern Configuration for CMR URI

Pattern	Partition	Gateway or Route List
cmr.route	URI	RL_CMR_SCHED

Deployment Tasks for TelePresence Conductor for CMR

TelePresence Conductor should be configured with one or more bridge pools and service preferences. All CMRs created through the Cisco TMSPE portal will be hosted in this TelePresence Conductor. TelePresence Conductor and Unified CM are connected through an Early Offer SIP trunk. The bridge pools and service preferences configured previously in TelePresence Conductor for scheduled conferences can be reused for CMR if you are using a shared resource model. Otherwise, use the bridge pools and service preferences configured previously in TelePresence Conductor for instant conferences for CMR. In addition, create an administrator account with read/write access level and API access enabled inside the TelePresence Conductor. Cisco TMSPE uses this administrator account to access TelePresence Conductor for creating CMRs.

Tip The Aliases and templates configured in TelePresence Conductor are not used by TMSPE; instead, the aliases and templates are configured within TMSPE.

Deployment Tasks for Cisco TMSPE for CMR

Cisco TMSPE, together with Cisco TMS, provides a portal for users to create CMRs, and this requires Cisco TMSPE to be installed and activated in Cisco TMS. Cisco TMSPE enables the administrator to create or import users into one or more user groups that entitle the users to have CMRs. For example, the user groups can match with the groups in Active Directory where the users are members, or the user groups can match with an organization unit (OU) of Active Directory where the users reside. In Cisco TMSPE, create the user groups **Executive** and **Marketing**, and import users from Active Directory that belong to the corresponding AD groups. Table 3-19 lists the Active Directory search filters for importing users into user groups. The TelePresence Conductor settings option within TMSPE allows the administrator to specify which TelePresence Conductor will be used for the CMR deployment. For each TelePresence Conductor cluster, the administrator must create only one record pointing to one of the Conductor nodes. The administrator account created in TelePresence Conductor is used here to configure the TelePresence Conductor settings.

Table 3-19 Search Filter for Importing Users into User Groups

User Group	Search Filter
Executive	(&(objectClass=user)(memberof=cn=executive, ou=enterprise, dc=ent-pa, dc=com))
Marketing	(&(objectClass=user)(memberof=cn=marketing, ou=enterprise, dc=ent-pa, dc=com))

A CMR template corresponds to a conference template and a conference alias on TelePresence Conductor. The CMR template allows the administrator to specify the conference attributes (for example, conference quality, content quality, conference PIN, and so forth), conference alias (SIP URI and/or numeric alias), and TelePresence Conductor for CMR creation. Also, the CMR template allows the administrator to assign multiparty licenses (PMP or SMP) to users within the group. Assign one CMR template to each user group to enable users in that group to set up their own CMR through the Cisco TMSPE portal.

When a user creates a CMR, Cisco TMSPE applies the settings that have been defined in the CMR template associated with that user's group, and it makes a provisioning API call to create the conference on TelePresence Conductor. No further interaction is needed from the administrator.

Note CMRs created by using Cisco TMSPE cannot be modified through the TelePresence Conductor web interface. Conference templates and aliases created by using TelePresence Conductor cannot be modified through Cisco TMSPE.

Summary

After you complete the deployment tasks outlined above, users can log in to their Cisco TMSPE portal to create a CMR and generate the corresponding SIP URI and numeric alias. Users can then dial the SIP URI or numeric alias to start the meeting.

7. Deploy Collaboration Meeting Rooms (CMR) Hybrid

This section describes the major tasks required to enable Cisco CMR Hybrid.

Overview

Deployment Tasks for TelePresence Conductor for CMR Hybrid:

1. The previously configured TelePresence Conductor and alias for scheduled conferences can be used for CMR Hybrid conferences.

Deployment Tasks for Unified CM for CMR Hybrid:

2. Configure a SIP route pattern that matches the WebEx site and points to the Expressway-C route list.

Deployment Tasks for Expressway for CMR Hybrid:

3. Create a new DNS zone that uses TLS verify, forces encryption, and uses the TLS verify name **sip.webex.com**.
4. Create a search rule that matches any URI that contains the WebEx domain and removes any appended characters.

Deployment Tasks for Cisco TMS for CMR Hybrid:

5. Add the WebEx site to Cisco TMS.
6. Configure the following attributes in the Cisco TMS user profiles of the users who are enabled to use CMR Hybrid:
   - WebEx username
   - WebEx password (unless single sign-on is enabled)
   - The WebEx site on which they have an account

Deployment Tasks for Cisco TMSXE for CMR Hybrid:

7. Download and install the WebEx and TelePresence Integration to Outlook plug-in for relevant users.
8. Configure a WebEx scheduling Mailbox, such as webex@company.com, with the following settings:
   - Turn off the Calendar Attendant.
   - Set AddNewRequestsTentatively to **Disabled**.
9. Configure the WebEx scheduling mailbox inside TMSXE so that TMSXE monitors this mailbox for CMR Hybrid bookings.

Deployment Tasks for Audio for CMR Hybrid:

10. Chose the best audio connection type for the deployment requirements, and enable the relevant settings.

Deployment Tasks for WebEx Site Administration for CMR Hybrid:

11. Enable Cisco TelePresence Integration (Meeting Center only).
12. Configure the following additional settings:
    - Cisco TMS booking service URL
    - List Cisco TelePresence meetings on calendar
    - Send invitation email to meeting host
    - Display toll-free number to attendees
    - Set the VoIP and video connection to **Automatically encrypted UDP/TCP SSL**.
    - Set relevant audio settings based on the desired audio connectivity.
13. Assign the Meeting Center TelePresence session type to all users who use CMR Hybrid.

Deployment Considerations

Cisco CMR Hybrid provides the following key features:

- Two-way video with up to 720p screen resolution between the WebEx application and telepresence devices
- Integrated audio and presentation sharing, including application and desktop content sharing capability for all users in a meeting
- Network-based recording of meetings, including content sharing, chat, and polling
- Integrated conference scheduling using Cisco TelePresence Management Suite (Cisco TMS), which enables users to easily schedule Cisco CMR Hybrid meetings
- Secure call control and connectivity enabled by media encryption provided by Cisco Expressway-E
- Management and conference resource allocation of TelePresence Servers provided by Cisco TelePresence Conductor
- Interoperability with third-party telepresence devices

Each user who will schedule Cisco CMR Hybrid meetings in Cisco TMS, must have a host account on the WebEx site.

Tip We recommend using WBS29 or later version of Cisco WebEx Meeting Center for CMR Hybrid deployments.

CMR Hybrid can be deployed using either SIP audio or PSTN audio. It can be scheduled either by using integration to Microsoft Outlook, using the Cisco Smart Scheduler, or using the Cisco WebEx Scheduling Mailbox.

Deployment Tasks for TelePresence Conductor for CMR Hybrid

The previously configured TelePresence Conductor and alias for scheduled conferences can be used for CMR Hybrid conferences.

Deployment Tasks for Unified CM for CMR Hybrid

The previous configuration to enable Unified CM to communicate with Cisco Expressway and TelePresence Conductor for scheduled conferences, is also valid to enable communication with CMR Hybrid. One additional configuration is required to ensure calls made from TelePresence Conductor to the WebEx site are routed correctly: configure a SIP route pattern that matches the WebEx site (for example, yoursite.example.com) to route to the Expressway-C route list.

Deployment Tasks for Expressway for CMR Hybrid

Cisco Expressway-E must have a server certificate that is signed by specific Root Certificate Authorities and that trusts both the DST Root CA certificate for the WebEx cloud and the CA certificate of the CA that issued the server certificate. For a list of supported Root CAs, refer to the *Cisco Collaboration Meeting Rooms (CMR) Hybrid Configuration Guide*, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-tms/products-installation-and-configuration-guides-list.html

Configure a new DNS zone on the Expressway-E that has **TLS verify** enabled and that uses **sip.webex.com** as the TLS verify name. Configure a search rule to point to this DNS zone that matches the WebEx domain.

⚠ **Caution** Do not use static NAT with Expressway-E, due to a defect that will cause the media part of a call to fail. If static NAT is required, refer to the recommended workarounds in the *Cisco Collaboration Meeting Rooms (CMR) Hybrid Configuration Guide*.

Deployment Tasks for Cisco TMS for CMR Hybrid

To benefit from Cisco CMR Hybrid, connections must be made to the organization's WebEx site. Verify that the settings in Table 3-20 have been made within TMS.

Table 3-20 TMS Settings for WebEx

Parameter	Value
Enable WebEx	Yes
Add WebEx to All Conferences	Yes
Get WebEx Username from Active Directory	Username (samAccountName)
WebEx Site	Configured per customer account
Default site for new users	Yes
Enable SSO	Yes

The settings in Table 3-20 allow TMS to communicate with the organization's WebEx site on behalf of the end user during scheduling. By automatically adding WebEx to every conference, even if the administrator needs to generate a meeting through methods other than WebEx Productivity Tools, a WebEx link will be made available to end users. The SSO settings allow end users to use a single set of credentials to access all collaboration services, because WebEx services are able to leverage AD credentials for end users to access WebEx tools.

Add the WebEx site to Cisco TMS to initiate the connection to the site.

To schedule meetings using Cisco TMS, users must have a username and password that the server is configured to trust. Users from Active Directory are trusted but must have the following information stored in their Cisco TMS user profile:

- WebEx username
- WebEx password (unless single sign-on is enabled)
- The WebEx site on which they have an account.

Deployment Tasks for Cisco TMSXE for CMR Hybrid

To allow all client types to use Cisco TMSXE to book meetings, enable the following two options for scheduling CMR Hybrid conferences using Cisco TMSXE:

- Using the WebEx Productivity Tools Plug-In for Microsoft Outlook

  Users add WebEx to their meeting by using the WebEx Meeting Options panel in Microsoft Outlook.

- Using WebEx Scheduling Mailbox

  Users add WebEx to their meeting invitation directly from their email client by including a special invitee: the WebEx mailbox.

The TMSXE booking service should be configured as normal.

Meeting organizers who want to schedule meetings using the WebEx and TelePresence Integration to Outlook plug-in, must download and install the WebEx Productivity Tools with TelePresence from the WebEx site.

To enable a WebEx scheduling Mailbox, create a new user mailbox in Microsoft Exchange that is used as a scheduling participant whenever a meeting is enabled for CMR Hybrid. Turn off the Calendar Attendant for this mailbox, and disable the AddNewRequestsTentatively setting to prevent new requests from being marked as tentative. Also disable the AD User associated with the account.

Configure the new mailbox in TMSXE as the WebEx Scheduling Mailbox so that TMSXE knows which account to monitor for CMR Hybrid bookings.

Deployment Tasks for Audio for CMR Hybrid

CMR Hybrid supports the following audio connection options, and the choice of method depends on customer preference:

- SIP audio:
  - Configure the WebEx site in Cisco TMS to use SIP audio.
  - Enable hybrid mode on the WebEx Site.
- PSTN audio:
  - Configure the WebEx Site in Cisco TMS to use PSTN audio.
  - Enable hybrid mode on the WebEx site (optional).
  - Configure PSTN calls to pass through a PSTN gateway to WebEx.
- TSP audio:
  - Configure the MACC Domain Index and Open TSP Meeting Room WebEx settings.
  - Configure the TSP dial string.
  - Configure how the conference is opened.
  - Configure TSP audio for the meeting organizer.

Deployment Tasks for WebEx Site Administration for CMR Hybrid

Configure the WebEx Site to enable CMR Hybrid to function. The key settings are:

- Allow Cisco WebEx OneTouch meetings (Meeting Center only)
- Cisco TMS booking service URL
- List Cisco TelePresence meetings on calendar
- Send invitation email to meeting host
- Display toll-free number to attendees
- Set the VoIP and video connection to **Automatically encrypted UDP/TCP SSL**.
- The relevant audio settings based on the desired audio connectivity

The Meeting Center TelePresence session type must be assigned the host accounts in the WebEx Site to complete the setup. This can be done either by opening the Edit User screens for an individual user or by selecting the appropriate session type for each user from the Edit User List screen.

Summary

After you complete the deployment tasks outlined above, CMR Hybrid should be ready for users to create TelePresence and CMR Hybrid meetings.

Related Documentation

- Cisco Personal Multiparty At-A-Glance

  http://www.cisco.com/c/dam/en/us/solutions/collateral/collaboration/pervasive-conferencing/at-a-glance-c45-729835.pdf
- Cisco TelePresence Management Suite (TMS) and CMR Hybrid deployment guides

  http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-tms/products-installation-and-configuration-guides-list.html
- Cisco TelePresence Conductor and Collaboration Meeting Rooms (CMR) Premises (optimized conferencing) deployment guides

  http://www.cisco.com/c/en/us/support/conferencing/telepresence-conductor/products-installation-and-configuration-guides-list.html
- Cisco TelePresence Conductor API guides

  http://www.cisco.com/c/en/us/support/conferencing/telepresence-conductor/products-programming-reference-guides-list.html
- Cisco TelePresence Server release notes

  http://www.cisco.com/c/en/us/support/conferencing/telepresence-server/products-release-notes-list.html

CHAPTER 4

Collaboration Edge

Revised: November 20, 2015

This chapter describes the Collaboration Edge preferred architecture, which includes a series of servers and gateways defining access to services at the perimeter of a collaboration network. The Collaboration Edge preferred architecture provides access to public networks, including the Internet and PSTN.

The chapter presents a detailed Architecture description of Collaboration Edge, followed by a Deployment Overview section that describes how to deploy Cisco Expressway for Internet access and Cisco Unified Border Element for PSTN access. The chapter also covers High Availability for Collaboration Edge, Security for Collaboration Edge, and Scaling the Collaboration Edge Solution. Then the section on the Collaboration Edge Deployment Process presents more detailed information on deploying Cisco Expressway, Cisco Unified Border Element, Cisco voice gateways, and Cisco ISDN video gateways.

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:	Revision Date
A number of minor changes to clarify some of the content	Various sections of this chapter	November 20, 2015

Core Components

The core components of the Collaboration Edge architecture are:

- Cisco Expressway-C and Expressway-E, for Internet connectivity and firewall traversal for voice and video
- Cisco Unified Border Element, for audio PSTN connectivity via IP trunks
- PSTN voice gateway, for direct audio PSTN connectivity
- ISDN video gateway, for direct video ISDN connectivity

Key Benefits

- Connect to customers and partners, independent of the technology they are implementing and the public network they are using.
- Provide for a resilient, flexible and extendable architecture.
- Provide any hardware and software client with the ability to access any public network (Internet and PSTN).
- Provide secure VPN-less access to collaboration services for Cisco mobile and remote clients and endpoints.

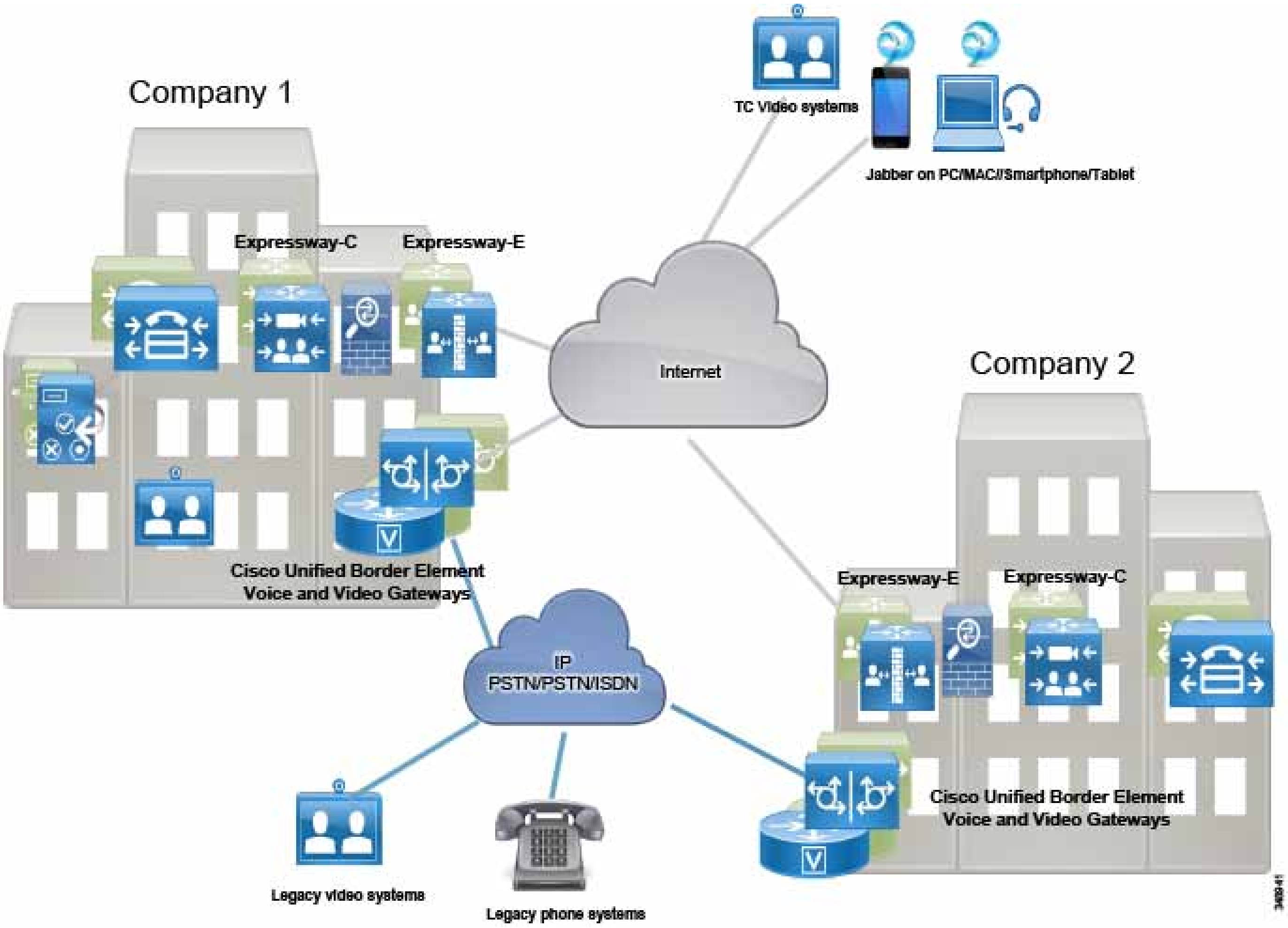
Architecture

The architecture for Collaboration Edge interfaces with two major networks: Internet and PSTN.

Internet connectivity enables VPN-less Mobile and Remote Access (MRA) and business-to-business communications. These services allow Jabber users and hardware video endpoints to access corporate collaboration services securely outside the organization’s network boundaries, and they provide business-to-business audio and video communications with external organizations.

Cisco Expressway-C and Expressway-E should be deployed as a pair and in almost all cases where a firewall boundary needs to be traversed. Expressway-C sits in the internal network and Expressway-E in the demilitarized zone (DMZ), one for each side of the firewall, in order to enable firewall traversal capabilities. In addition, each Expressway-C and Expressway-E can be clustered. (See Figure 4-1.) In most cases the firewall boundary crossed is an Internet connection, but it could also be a separate corporate WiFi network for Bring Your Own Device (BYOD) connections.

Figure 4-1 High-Level View of the Architecture



PSTN connectivity enables audio and video communications to Telecom carrier networks. The PSTN connection can be achieved in multiple ways:

- Through an IP trunk to a Telecom carrier, usually for voice-only services. This connectivity is provided by the Cisco Unified Border Element (CUBE) on an Integrated Service Router (ISR) G2/G3 or Aggregation Services Routers (ASR). Cisco Unified Border Element should be deployed in a central site where the Telecom carrier's network communicates with the enterprise network.
- Through voice gateways. Gateways include analog and ISDN interfaces on a variety of router platform, such as Cisco Integrated Service Routers (ISR) G2/G3. In this document only ISDN voice interfaces are considered. Voice gateways should be deployed locally in the sites where a PSTN connection is required
- Through Cisco TelePresence ISDN Gateways 3241 or MSE 8321, which enable legacy H.320 video access to PSTN. TelePresence ISDN gateways should be centralized anywhere an ISDN video connection is required. Due to the nature and cost of TelePresence ISDN gateways, they can be shared through multiple locations.

There are cost savings associated with deploying Internet communications for video calls (Expressway) and IP PSTN connections for audio-only calls (CUBE). However, it is worth noting that:

- Not all companies have enabled Internet communications for video systems (business-to-business). If some of the partners and customers are using ISDN only for video communications, video gateways are still recommended.
- Although IP network reliability is increasing over time, network connectivity problems might prevent remote sites from accessing centralized IP PSTN services. If such sites are heavily relying on PSTN connectivity to run daily business, a local PSTN connection used as backup for the centralized access is recommended.

The recommendations for PSTN are:

- Centralize PSTN, which will help reducing operational costs and expenses.
- Local PSTN connections maintained only for those sites highly relying on PSTN to run daily business. In these cases, the number of ISDN channels should be reduced because they will be used only in those situations where central PSTN access is not available. This would help save money by reducing hardware costs and simplifying the management.

Based on the above considerations, IP trunk connections to the PSTN for voice, with local PSTN breakout used as backup and Internet for video, satisfy the vast majority of connectivity requirements. However, to provide fully connectivity, ISDN video gateways are also recommended to reach partners and customers that are still not reachable on the Internet.

Cisco Collaboration Edge includes scenarios where users have access to the following options:

- Mobile and Remote Access (MRA) for teleworkers and mobile connectivity
- Business-to-business video communications between organizations
- PSTN for cellphones and access to landlines
- ISDN video access for communications to existing H.320 video systems

Under these scenarios any corporate user inside the company or on the Internet has access to PSTN voice calls, ISDN video calls, and business-to-business communications as if they were inside the enterprise. Services such as hold, transfer, and conference are also available in most cases. Independently from who is calling whom, the Collaboration Edge solution enables interconnectivity between mobile and remote access, business-to-business, PSTN voice, and video services.

Role of Expressway-C and Expressway-E for Internet Access

Use of the Internet for collaboration services continues to increase in popularity and is quickly replacing existing legacy ISDN video systems. The two primary protocols leveraged for Internet based collaboration services are SIP and H.323.

Moreover, the Internet is also used to connect remote and mobile users to voice, video, IM and presence, and content sharing services without the use of a virtual private network (VPN).

Mobile and remote access, as well as business-to-business services, can be enabled as part of the same Expressway-C and Expressway-E solution pair. Expressway-C is deployed inside the corporate network, while Expressway-E is deployed in the DMZ.

The Expressway-C and Expressway-E pair performs the following functions:

- Interworking — The capability to interconnect H.323-to-SIP calls for voice, video, and content sharing.
- Boundary communications services — While Expressway-C sits in the corporate network, Expressway-E is in the enterprise DMZ and provides a distinct connection point for communication services between the enterprise network and the Internet.
- Security — The capability to provide authentication and encryption for both mobile and remote access and business-to-business communications.

Mobile and remote access and business-to-business calls flow through Expressway-E and Expressway-C, which handle both call signaling and media as well as other collaboration data flows, including XMPP and HTTP.

Mobile and Remote Access

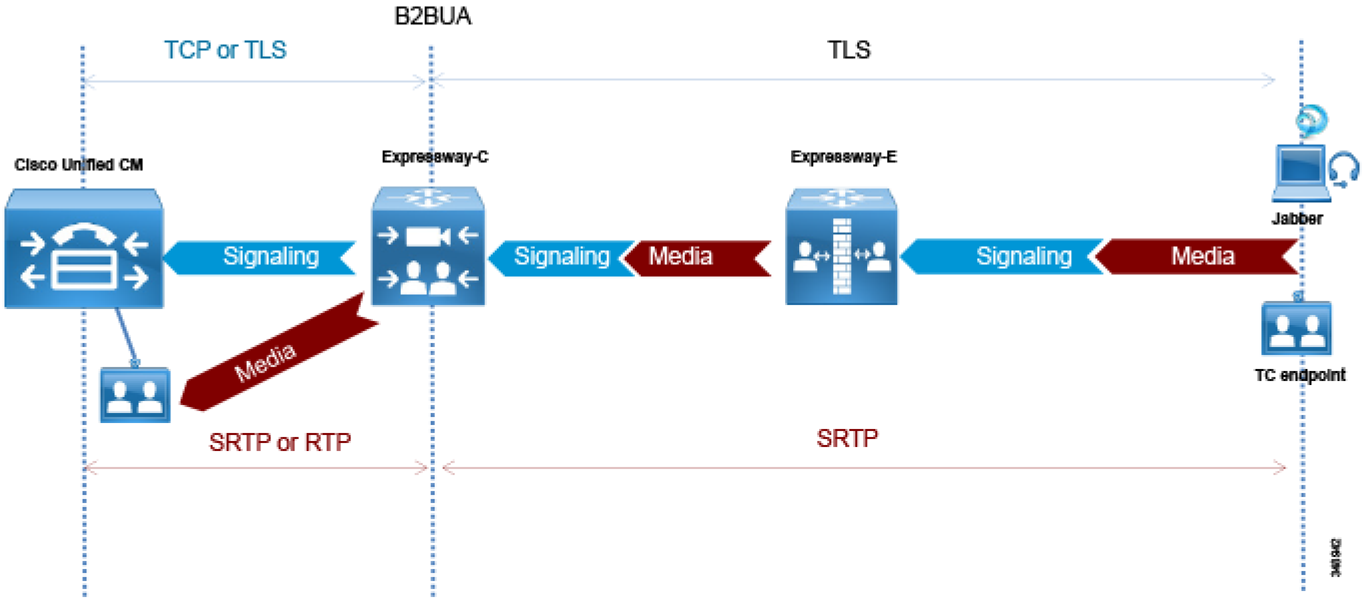
The mobile and remote access feature of the Cisco Expressway solution provides secure reverse proxy firewall traversal connectivity, which enables remote users and their devices to access and consume enterprise collaboration applications and services.

As shown in Figure 4-2, the Cisco Expressway solution encompasses two main components: the Expressway-E node and the Expressway-C node. These two components work in combination with Cisco Unified Communications Manager (Unified CM) to enable secure mobile and remote access. The Expressway-E node provides the secure edge interface to mobile and remote devices.

Expressway-C creates a secure TLS connection with the Expressway-E node. The Expressway-C node provides proxy registration to Unified CM for remote secure endpoint registration. The Expressway-C node includes a back-to-back user agent (B2BUA), which provides media termination capabilities.

Figure 4-2 shows that both signaling and media traverse Expressway-C and Expressway-E for all mobile and remote access calls.

Figure 4-2 B2BUA and Call Legs on Expressway

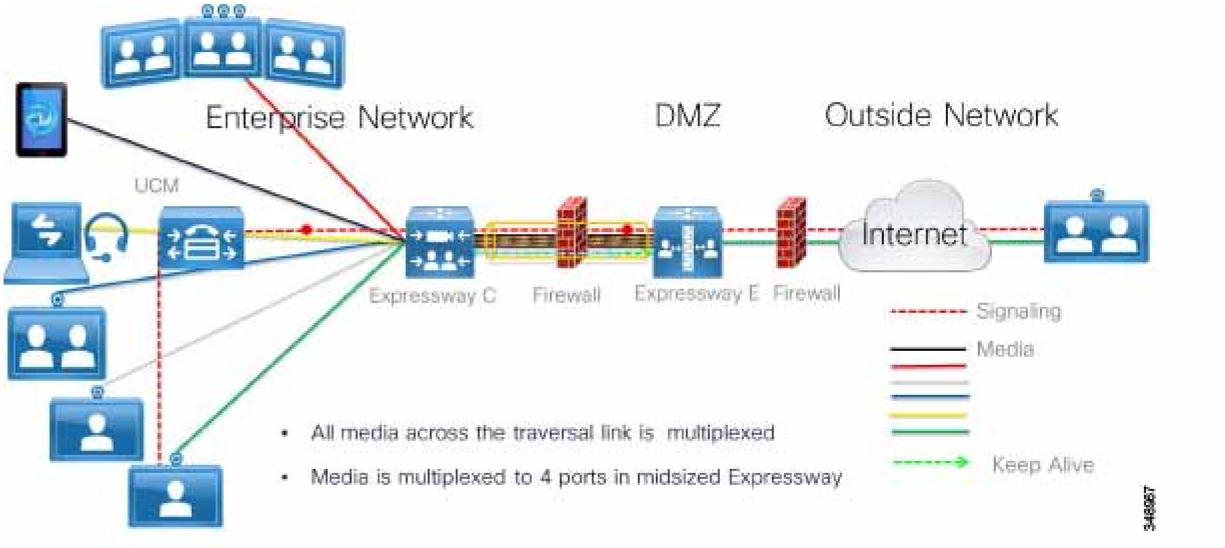


Business-to-Business Communications

Expressway-C and Expressway-E are designed to work together to form a firewall traversal solution that is the core component for business-to-business communications over the Internet.

Expressway-C sits on the inside (trusted side) of the enterprise network and serves the role of providing a secure, trusted, and standards-based way of connecting to Expressway-E. It acts as a traversal client to all devices behind it. This solves the problem for devices using a large number of media ports by multiplexing all of the media to a very small number of ports opened for outbound communications. It provides an authenticated and trusted connection from inside the enterprise to outside by sending a keep-alive for the traversal zone from Expressway-C to Expressway-E. Additionally, it provides a single point of contact for all Internet communications, thus minimizing the security risk. (See Figure 4-3.)

Figure 4-3 Expressway-C Multiplexing and Keep-Alive



Real-time and near real-time communication protocols such as SIP, H.323, and XMPP do not address the need to communicate with devices that might be behind a firewall. Typical communications using these protocols include the device IP address in the signaling and media, which becomes the payload of the TCP and UDP packets, respectively. When these devices are on the same internally routable network, they can successfully communicate directly with each other. The signaling IP address carried in the payload of the TCP packet is routable back to the initiating device, and vice versa. However, when the initiating device is on a different network behind a public or network edge firewall, two problems are encountered. The first problem is that the receiving device, after decoding the packet, will respond to the internal IP address carried in the payload. This IP address is typically a non-routable RFC 1918 address and will never reach the return destination. The second problem encountered is that, even if the return IP address is routable, the media (which is RTP/UDP) is blocked by the external firewall. This applies to both business-to-business and mobile and remote access communications.

Expressway-E sits at the network edge in the DMZ. It serves the role of solving both the signaling and media routing problems for SIP, H323, and XMPP, while maintaining standards interoperability. It changes the appropriate headers and IP addresses to process the media and signaling on behalf of the endpoints, devices, and application servers that are inside the network.

Instant Messaging and Presence Federation

Instant messaging and presence federation involves allowing users to send XMPP traffic through an organization's external firewall for chat and presence status information to and from users in another organization.

Prior Cisco architectures involved using the Cisco Adaptive Security Appliance (ASA) firewall as a TLS proxy and allowing inbound ports to be opened through the external firewall to directly access the internal IM and Presence servers. This is still the recommended solution for SIP federation.

XMPP federation uses the same Expressway-C and Expressway-E paired architecture for a trusted secure firewall traversal solution for XMPP traffic to and from external destinations. Expressway-E provides a secure DMZ-based termination point for XMPP to the Internet. Expressway-C provides a TLS-based authenticated secure connection to Expressway-E for firewall traversal, and as such does not require any port to be opened on the firewall.

Expressway-C also provides an AXL API connection to the IM and Presence server. The AXL API sends XMPP server-to-server information collected from Expressway-E to the IM and Presence database. This provides the IM and Presence server with the necessary connection information to initiate a federated connection to the other organization through Expressway-E without opening any other ports on the firewall. XMPP federation allows voice and video escalation. The same organization might implement both XMPP and SIP federation at the same time.

PSTN Access

This section describes the architecture for PSTN access using Cisco Unified Border Element as the session border controller (SBC).

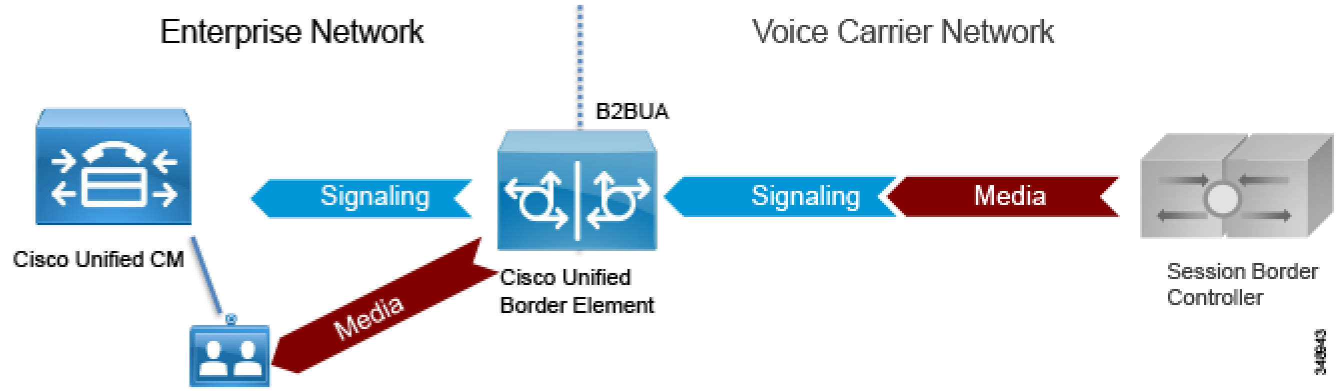
Role of the Cisco Unified Border Element

Voice connectivity using IP trunks to Telecom carriers, instead of traditional PSTN connections, is increasing in popularity and gradually replacing existing TDM-based PSTN access. SIP is commonly used as the access protocol for connectivity into provider networks, and today many Telecom carriers offer a voice-only service to the PSTN through a session border controller such as Cisco Unified Border Element. Session border controllers are SIP back-to-back user agents (B2BUAs) and are typically used in flow-through mode, where both the voice media and SIP signaling for each call flow through the Cisco Unified Border Element. (See Figure 4-4.)

Cisco Unified Border Element is a licensed Cisco IOS application available on a wide range of Cisco router and gateway platforms, and it is the recommended platform to connect to the PSTN through a SIP trunk to the Telecom carrier's border element.

Cisco Unified Border Element enables enterprise voice networks based on Cisco Unified Communications Manager (Unified CM) to connect to and interoperate with Telecom carriers through SIP trunk services. Cisco Unified Border Element terminates and re-originates both signaling and media streams to provide secure border interconnection services between IP networks. Using Cisco Unified Border Element, customers can save on their current network services, simplify their network architectures, and position their networks for ongoing enhancements in collaboration services.

Figure 4-4 Cisco Unified Border Element as B2BUA



Cisco Unified Border Element performs the following functions between the enterprise and Telecom carrier networks:

- Session control — The capability to offer flexible trunk routing, call admission control, resiliency, and call accounting for the SIP sessions.
- Interworking — The capability to offer media transcoding services for voice, and interoperability between SIP Delayed Offer and Early Offer.
- Demarcation — The capability to act as a distinct demarcation point between two networks for address and port translation and to facilitate troubleshooting.
- Security — The capability to intelligently allow or disallow real-time traffic between networks, and to encrypt the real-time traffic as appropriate for the application.

Role of Voice Gateways

We recommend using TDM gateways to connect to the PSTN if centralized PSTN access is not available. Cisco offers a full range of TDM gateways for analog and digital connections to the PSTN on ISR G2/G3 routers with appropriate interface cards enabling: low-density digital (BRI), high-density digital (T1, E1, and T3), and analog (FXS, FXO, and E&M) interfaces.

For more information on voice gateways, refer to the *Cisco 3900 Series, 2900 Series, and 1900 Series Software Configuration Guide*.

Role of Video ISDN Gateways

Video communications has a long history with ISDN. Videoconferencing first became commercially viable with ISDN as the communications protocol. Because of this, there is still the need to communicate inbound and outbound via ISDN with legacy video systems. The Cisco TelePresence ISDN Gateway performs the role of converting ISDN to SIP, and vice versa, for videoconferencing calls. The Cisco Preferred Architecture for Enterprise Collaboration includes ISDN gateway access trunked to Unified CM for the purpose of communicating with legacy videoconferencing systems.

Although Cisco ISDN video gateways manage both H.323 and SIP, only SIP is considered in the Preferred Architecture. While H.323 can still be used on networks based on the Cisco TelePresence Video Communication Server (VCS), networks based on Cisco Unified Communications Manager require SIP.

More information on the Cisco TelePresence ISDN Gateway, refer to the documentation available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-isdn-gateway/tsd-products-support-series-home.html

Deployment Overview

This section presents a general description of how to deploy Cisco Expressway for Internet connectivity and Cisco Unified Border Element for PSTN access.

Deployment of Expressway-C and Expressway-E for Internet Connectivity

The standard deployment of the Cisco Collaboration Edge architecture involves deploying at least one Expressway-C and Expressway-E pair for secure mobile device and remote VPN-less access back to enterprise collaboration services.

Both Expressway-C and Expressway-E should be deployed in a cluster to provide better resiliency. The number of servers for each cluster depends on the number of the concurrent proxied registrations to Unified CM and the number of concurrent calls. While the first takes into consideration mobile and remote users who register through Expressway to Unified CM, the second accounts for concurrent calls for business-to-business and for mobile and remote access (MRA). (See the Sizing chapter for details.)

This service is provided to Jabber clients and Cisco TelePresence System endpoints (C, EX, MX, DX, and SX Series models). Frequently, multiple pairs of Expressway-C and Expressway-E are deployed for geographic coverage and scale, providing access to multiple instances of collaboration services. GeoDNS should be used to balance remote client and endpoint access based on a variety of metrics from the Internet service provider.

This same Expressway can be leveraged for business-to-business communications as well. When the volume of calls exceeds the capacity of the Expressway cluster (600 concurrent calls for the Medium OVA template or 2,000 for the Large OVA template), business-to-business and MRA services have to be split on different boxes. See the Sizing chapter for further details.

When Expressway is used for both services, Unified CM is connected to Expressway-C through a SIP trunk for unified business communications access over the Internet. Expressway-C sits on the trusted side of the network, providing secure firewall traversal services to Expressway-E.

Based on the enterprise security policy, a number of different deployment models can be implemented. In this document we focus on a DMZ deployment with a dual interface because it is the most common and secure deployment model. For additional deployment models, refer to the *Cisco Expressway Basic Configuration Deployment Guide*.

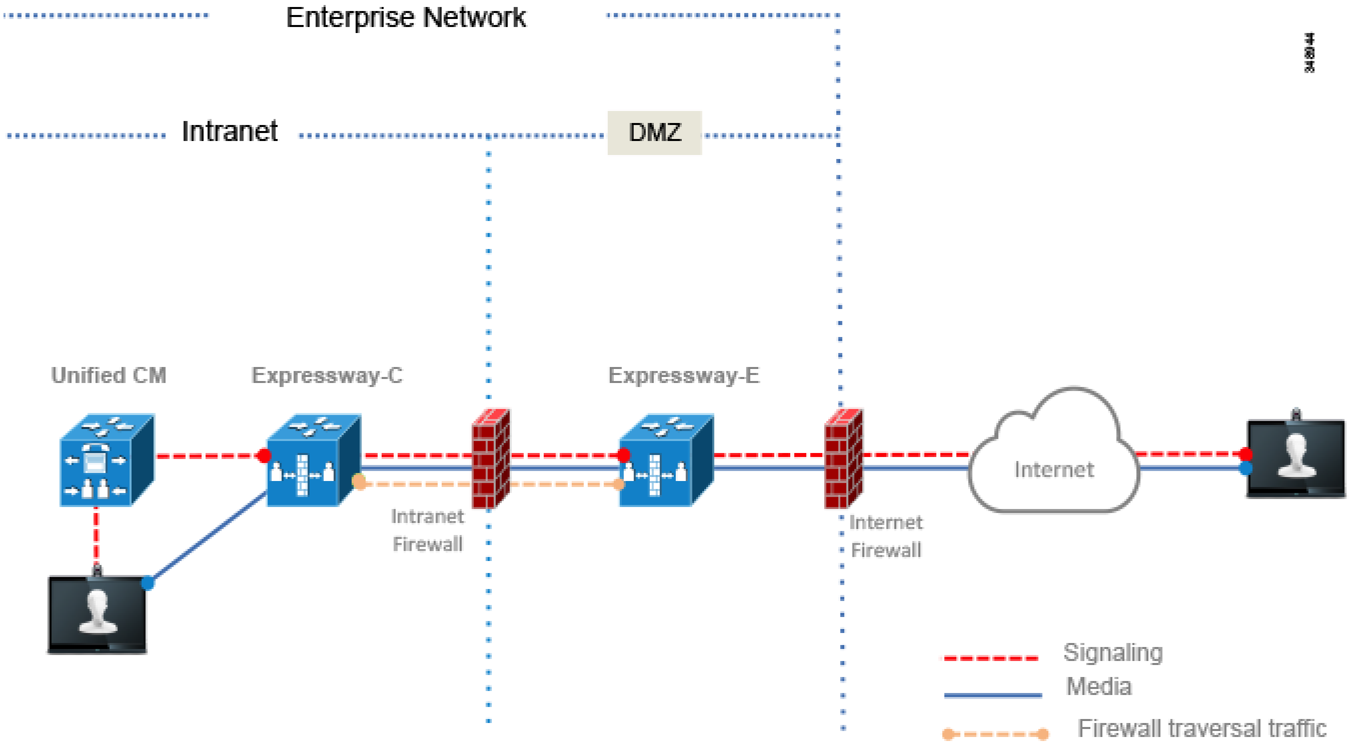
Expressway-C and Expressway-E provide firewall traversal capabilities. Firewall traversal works as follows:

1. Expressway-E is the traversal server installed within the enterprise DMZ. Expressway-C is the traversal client installed inside the enterprise network.
2. Expressway-C initiates traversal connections outbound through the firewall to specific ports on Expressway-E, with secure login credentials. If the firewall allows outbound connections, as it does in the vast majority of cases, no additional ports are required to be opened in the enterprise firewall. For ports details, refer to the *Unified Communications Mobile and Remote Access via Cisco Expressway Deployment Guide*.

   Mobile and remote access requires a separate traversal zones, called the Unified Communications traversal zone. The Unified Communications traversal zone works with SIP and requires TLS and media encryption, while the business-to-business traversal zone allows SIP and H.323 as voice and video signaling protocols. The Unified Communications traversal zone also allows XMPP and HTTPs, which are used to connect to IM and Presence servers and for provisioning purposes.
3. Once the connection has been established, Expressway-C sends periodic keep-alive packets to Expressway-E to maintain the connection.
4. When Expressway-E receives an incoming call or other collaboration service request, it issues an incoming request to Expressway-C.
5. Expressway-C then routes the request to Unified CM or other collaboration service applications.
6. The connection is established, and application traffic (including voice and video media) traverses the firewall securely over an existing traversal connection.

In order for firewall traversal to work, a traversal client zone has to be configured on Expressway-C and a traversal server zone has to be configured on Expressway-E. Figure 4-5 summarizes the firewall traversal process.

Figure 4-5 Expressway-C and Expressway-E Firewall Traversal Process



In the dual-interface deployment scenario, Expressway-E sits in the DMZ between two firewalls: the Internet firewall provides for NAT services toward the Internet, and the intranet firewall provides access to the corporate trusted network.

Expressway-E has two LAN interfaces: one toward the Internet firewall (also called the *external interface*), and the other toward the intranet firewall (also called the *internal interface*).

There is no need for the external interface to be assigned a public IP address because the address can be translated statically by NAT. In this case, the public IP address has to be configured on Expressway-E itself.

Expressway-C has an embedded B2BUA to terminate mobile and remote access as well as business-to-business calls. Expressway-E has an embedded B2BUA, used to terminate business-to-business calls. Expressway-C and Expressway-E have other B2BUAs dedicated to different services, such as Microsoft and H.323-to-SIP protocol interworking; however, the B2BUA term used in this chapter identifies the B2BUA used only in mobile and remote access and business-to-business call scenarios.

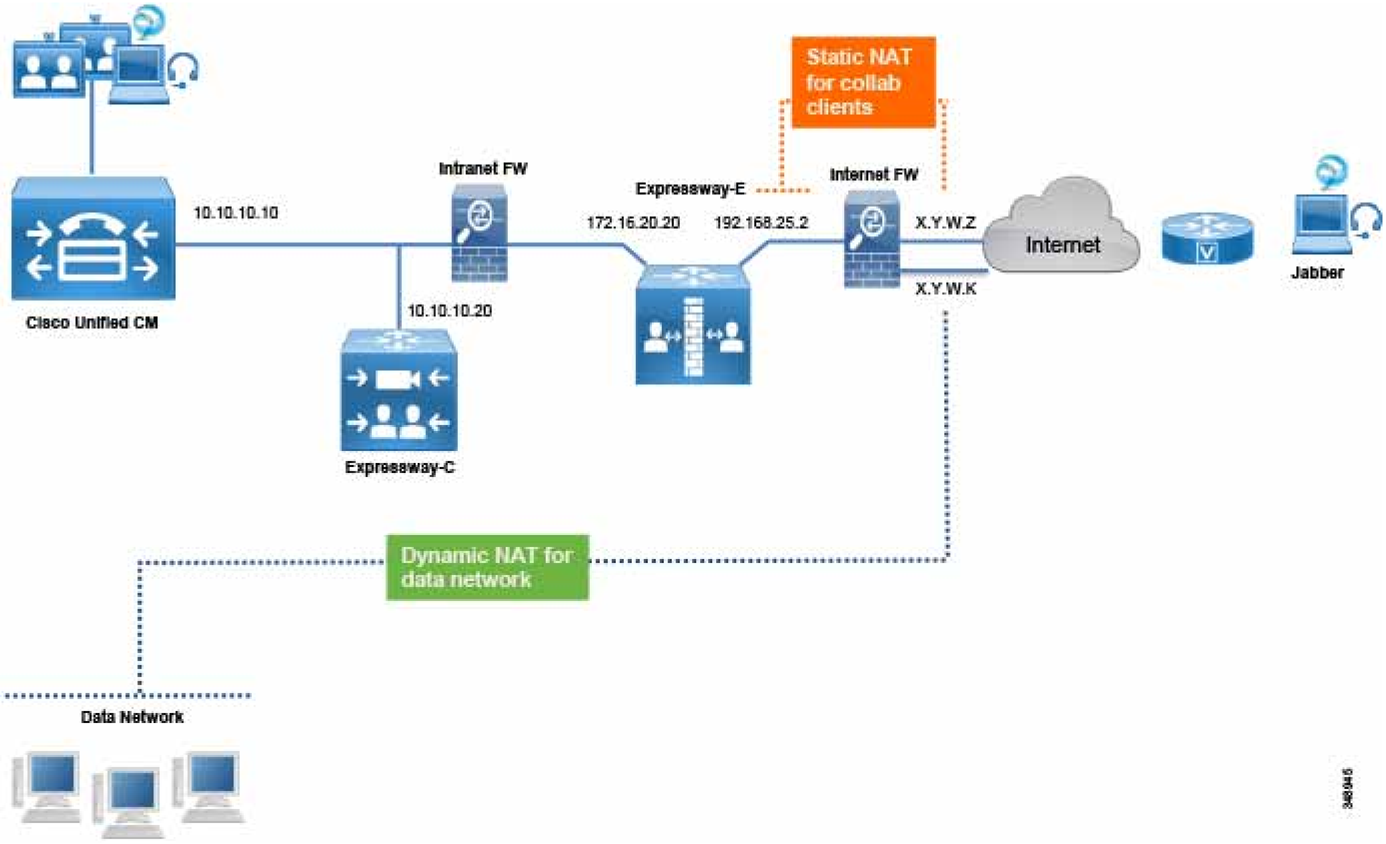
The B2BUA terminates collaboration application traffic. A connection from the Internet to Expressway-C via Expressway-E is always encrypted for mobile and remote access, while the connection between Expressway-C and the Unified Communications Manager endpoint can be encrypted or not based on the configuration. A connection from the Internet for business-to-business communications may or may not be encrypted, based on the configuration and dictated by the corporate policies. Note that in this case the communication will be encrypted on the Internet only if the remote business-to-business party supports encryption with public certificates. In all other cases, the video call will be sent unencrypted. This document focuses on encryption between the Internet and Expressway-C for mobile and remote access services only, while communications between Expressway-C and the internal back-end servers and clients are sent unencrypted. Business-to-business encryption capabilities are discussed further in the section on Security for Collaboration Edge.

Expressway-C proxies the registration of mobile and remote access Jabber clients or Cisco TelePresence System endpoints to Unified CM, which lists them as registered devices with the IP address of Expressway-C.

Figure 4-6 shows the deployment described above. The relevant IP addresses are shown in the figure. Public IP addresses, which will vary based on location and Internet service provider, are shown with letters instead of digits.

Expressway-E has two interfaces; the internal interface has IP address 172.16.20.20, while the external interface has IP address 192.168.25.2. The external interface IP address is statically translated to X.Y.W.Z. This address is also configured on Expressway-E. When Expressway-E sends an INVITE, it creates the Session Description Protocol (SDP) message with the IP address set to the translated interface address instead of using its own address, so that the called party can use the public routable address instead of the private one.

Figure 4-6 NAT Interfaces on the Internet Firewall



When an endpoint on the Internet connects to Unified CM or other collaboration application through Expressway, its IP address is first translated to a public IP address. On Expressway-E, the source IP address is replaced by the address of the internal IP LAN interface of Expressway-E. When the packet enters Expressway-C, Expressway-C replaces the source IP address of the packet with its own IP address before forwarding the packet to the collaboration service applications.

In the other direction, when traffic from internal endpoints traverses the Expressway toward the Internet, their source IP addresses are replaced by the Expressway-E external LAN interface address, which is later statically translated by NAT on the Internet firewall. Source IP addresses of data devices are dynamically translated to X.Y.W.K by using another interface of the Internet firewall.

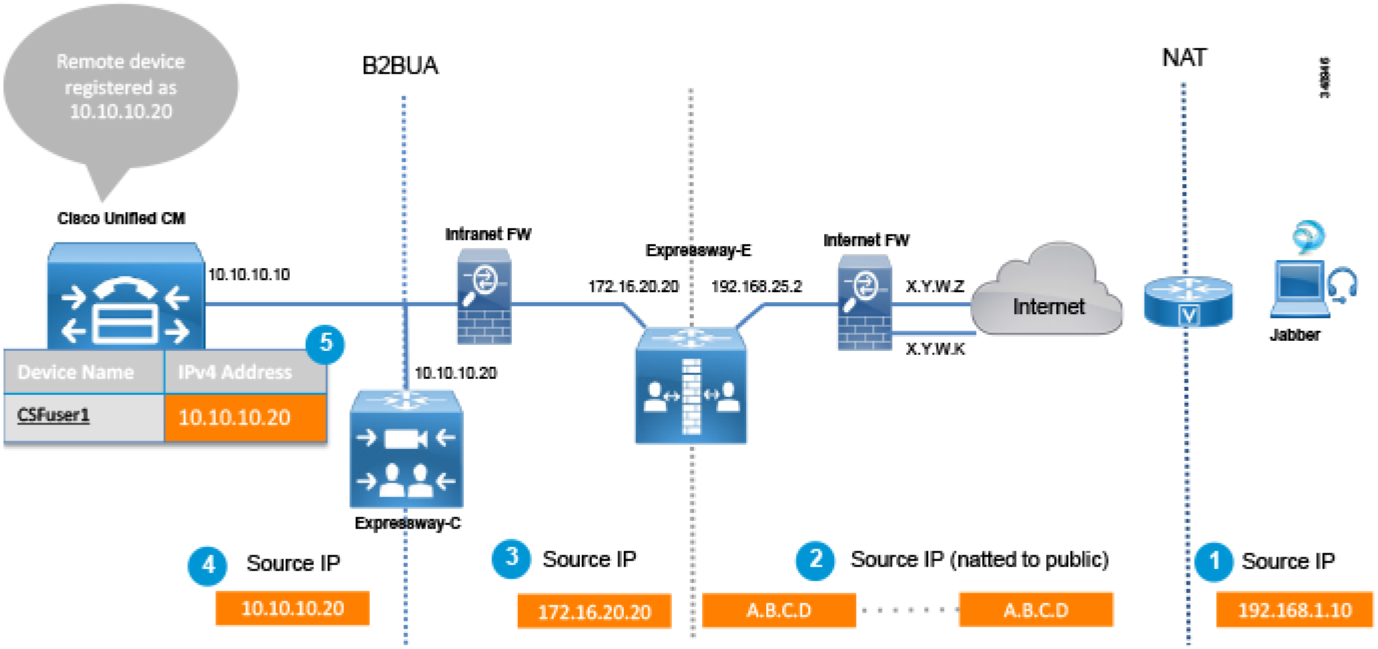
For a PC with data and a communication application, such as Jabber and a browser, the Jabber application address would be statically translated by NAT and the browser application address would be dynamically translated by NAT.

Even if the static NAT translation occurs in the firewall, the packet source IP address is transformed during its travel: it is translated to the IP address of Expressway-C when the packet goes from Expressway-C to Expressway-E, and it is translated to the IP address of Expressway-E when the packet goes from Expressway-E to the firewall. In the firewall, the packet is statically translated by NAT and sent to the Internet.

Mobile and Remote Access

In the case of call control services, Expressway-C proxy registers the endpoint to Unified CM using its own IP address, as shown in Figure 4-7.

Figure 4-7 NAT on the Life of a Packet



The address translation process shown in Figure 4-7 involves the following steps:

1. The source IP address of the endpoint is translated by NAT at the router that gives access to the Internet (192.168.1.10 to A.B.C.D.) if the endpoint does not have a public IP address.
2. The packet arrives at Expressway-E.
3. Expressway-E sends the packet to Expressway-C by using its own internal LAN interface address (A.B.C.D to 172.16.20.20).

4. Expressway-C receives the packet and terminates the connection. It re-originates another connection toward Unified CM by using its own IP address (172.16.20.20 to 10.10.10.20).
5. The endpoint is registered on Unified CM with the IP address of Expressway-C (10.10.10.20).

Registering the device to Unified CM with the IP address of the Expressway-C has some inherent benefits. For example, it is possible to limit the video bandwidth of remote devices when they are not connected directly to the corporate network, and assign them a different value when they are on-premises. Although we do not discuss it here, this can easily be achieved through the use of mobility features on Unified CM, which allow for definition of specific policies based on the IP address range.

When an endpoint is registered through the Internet, it cannot be managed remotely by the Cisco Collaboration architecture. This is because the endpoint IP address is dynamically translated and is behind a firewall. If remote management is required, deploy the endpoint through a VPN.

VPN technologies are not part of this architecture, but can be added as required.

Mobile and remote access has to be enabled on Expressway-E and Expressway-C. Expressway-C can then be configured to discover Unified CM and IM and Presence clusters by specifying the DNS name of the Unified CM and IM and Presence publisher nodes.

Expressway-E, deployed in the DMZ, provides a trusted point of entry for Jabber clients and TelePresence endpoints that use the mobile and remote access service. It also provides authentication, provisioning, registration, calling services, IM and presence, voice messaging, and directory services for remote Jabber clients and TelePresence endpoints, as well as business-to-business connectivity over the Internet.

Expressway-C connects to Unified CM and IM and Presence clusters and Cisco Unity Connection using HTTPs, SIP, and XMPP. (See Figure 4-8.)

Moreover, there are a number of cases where Jabber has to connect via HTTP to a specific server; for example, for Visual Voicemail, Jabber Update Server, custom HTML tabs and icons, and directory photo host. In these cases Jabber would connect directly to these servers without passing through Unified CM, and Expressway-C would need an HTTP allow list that specifies which servers the Jabber client is allowed to connect to.

Figure 4-8 Expressway Connection to Unified CM, IM and Presence Service, and Unity Connection

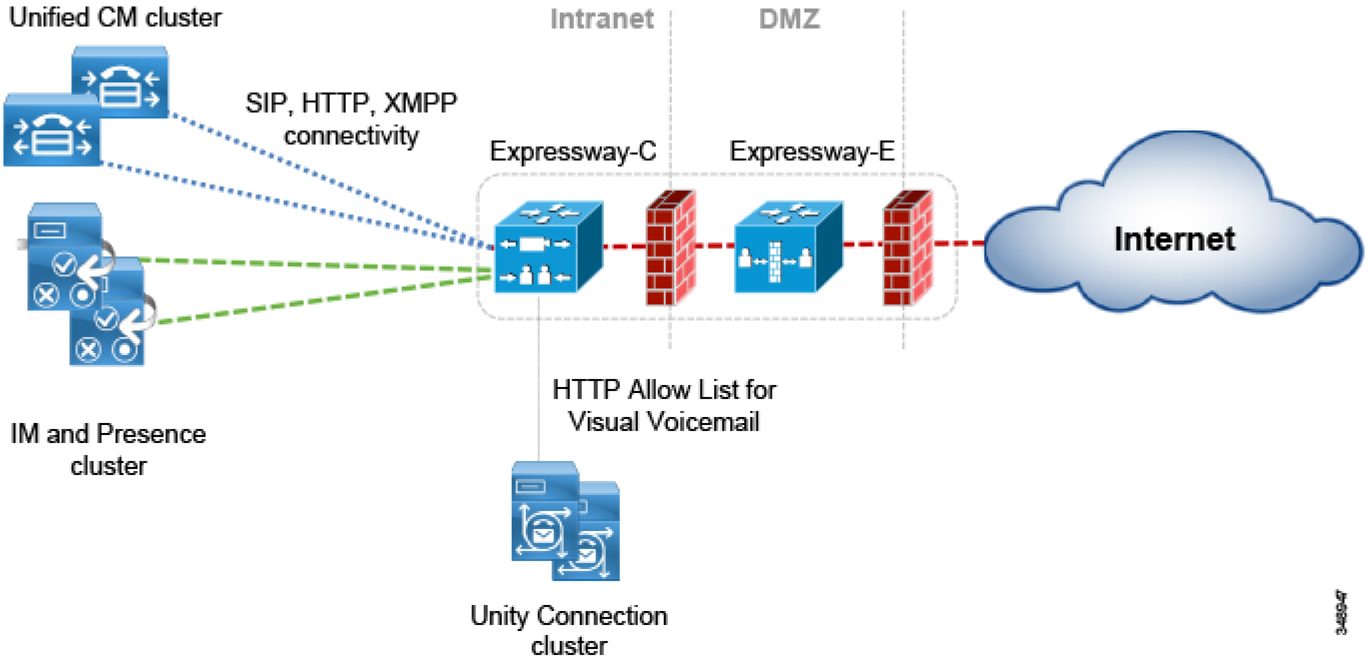


Table 4-2 summarizes the protocols used by Expressway for mobile and remote access.

Table 4-2 Expressway Protocols for Mobile and Remote Access

Protocol	Security	Service
SIP	TLS	Session establishment – register, invite, and so forth
HTTPS	TLS	Login, provisioning, configuration, contact search, visual voicemail
XMPP	TLS	Instant messaging, presence
RTP	SRTP	Audio, video, content sharing, advanced control

When a Jabber or TelePresence endpoint user logs in, they specify their fully qualified name (for example, user1@ent-pa.com). The client queries the public DNS server for specific SRV records:

- _cisco-uds._tcp.ent-pa.com, which is configured only on the corporate DNS server.
- _collab-edge._tls.ent-pa.com, which is configured only on the public DNS server and resolves to the public interfaces of the Expressway-E cluster. Note that this record always specifies TLS.

If the client is connected over the Internet, no answer will be provided by the public DNS server for _cisco-uds, but the client will receive an answer for the _collab-edge SRV record.

The DNS server will then send the A-record for Expressway-E (or multiple records if Expressway-E is clustered) to the client. Once the client knows the DNS name of Expressway-E, it can start the provisioning and registration procedure.

While provisioning takes place by using HTTPs, registration uses SIP and XMPP.

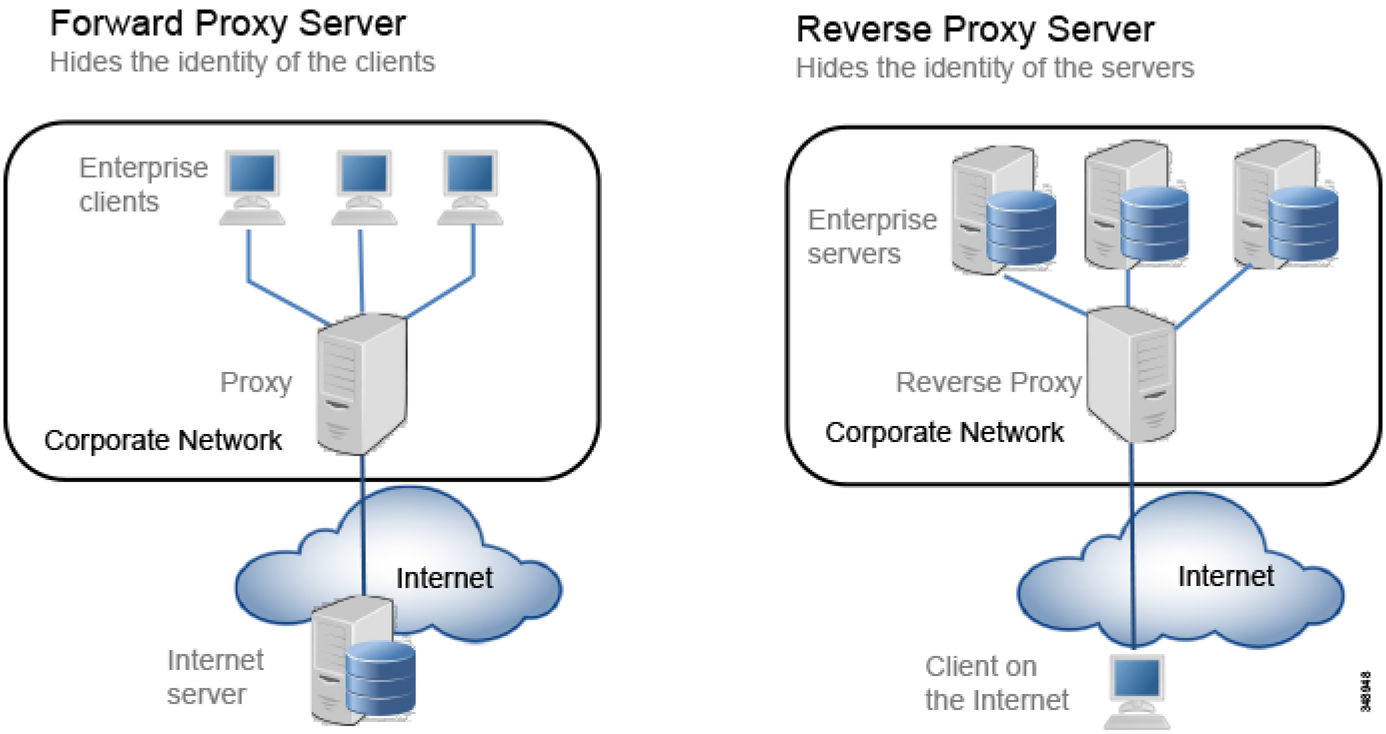
Expressway-C has an HTTPs reverse proxy server feature to manage the provisioning process. A reverse proxy is the opposite of the most common forward proxy server, also referred to as the *proxy server*.

As shown in Figure 4-9, while a forward proxy server provides services information for on-premises clients by hiding client details when connecting to Internet servers, a reverse proxy server provides information for off-premises clients by hiding on-premises server information. Clients in the corporate network, connecting through a forward proxy to an Internet server, know the identity of the server they are connecting to, but the servers do not know the identity of the clients.

On the other side, clients on the Internet connecting through a reverse proxy do not know the identity of the on-premises servers because they are connecting through the reverse proxy server, but the on-premises servers know the identity of the clients they are connecting to. This information is then returned to the client as though it originated from the on-premises servers themselves.

Expressway-C has a reverse proxy feature that provides provisioning, registration, and service details to the clients on the Internet, on behalf of collaboration application servers such as Cisco Unified CM, IM and Presence, and Unity Connection.

Figure 4-9 Forward Proxy vs. Reverse Proxy Server



Also consider that for services like Visual Voicemail, Jabber Update Server, custom HTML tabs and icons, directory photo host, Expressway-C will allow these connections if these services are specified under the *HTTP allow list*, which is a type of access list for HTTP services.

Provisioning and registration are multi-step processes that involve the client, Expressway-C, Expressway-E, Unified CM, and IM and Presence server.

The following is an overview of the major steps involved when a client registers through the Collaboration Edge.

1. Provisioning starts with the **get_edge_config** request issued from the client. For example:

   https://expressway_e.ent-pa.com:8443/ZeW50LXBhLmNvbQ/get_edge_config?service_name=_cisco-uds&service_name=_cuplogin

   Along with the request, the client sends the credentials of the user (For example, username "user1", password "user1"). The query is sent to Expressway-E, which forwards it to Expressway-C.

2. Expressway-C performs a UDS query to Unified CM to determine the home cluster for user1. This is essential for multi-cluster scenarios:

   GET cucm.ent-pa.com:8443/cucm-uds/clusterUser?username=user1

3. Once the home cluster is found, a response is sent to Expressway-C. This response includes all servers in the cluster.

4. Expressway-C asks the home cluster for provisioning information by making the following queries for user1 on behalf of the client:

   GET /cucm-uds/user/user1/devices retrieves the devices association list.

   GET /cucm-uds/servers retrieves the list of servers for the cluster.

   GET /cucm-uds/user/user1 retrieves the user and line configuration for user1.

   In response to the queries, the TFTP servers are also returned.

Subsequent queries, such as **http://us_cucm1.ent-pa.com:6970/SPDefault.cnf.xml**, are TFTP queries over HTTP. Thus, the provisioning process is done by queries to the UDS and to the TFTP server. As a result of these queries, provisioning information is forwarded to the client, and the client is able to start the registration process.

The registration process consists of two actions:

1. IM and Presence login, which is achieved via XCP router functionality on Expressway-C. The XCP router queries the IM and Presence clusters configured on Expressway-C in order to find the IM and Presence cluster where the user is configured, and the Jabber client is able to login for IM and Presence services.

2. Unified CM registration using SIP REGISTER messages, which are proxied by the Expressway SIP Proxy function.

Business-to-Business Communications

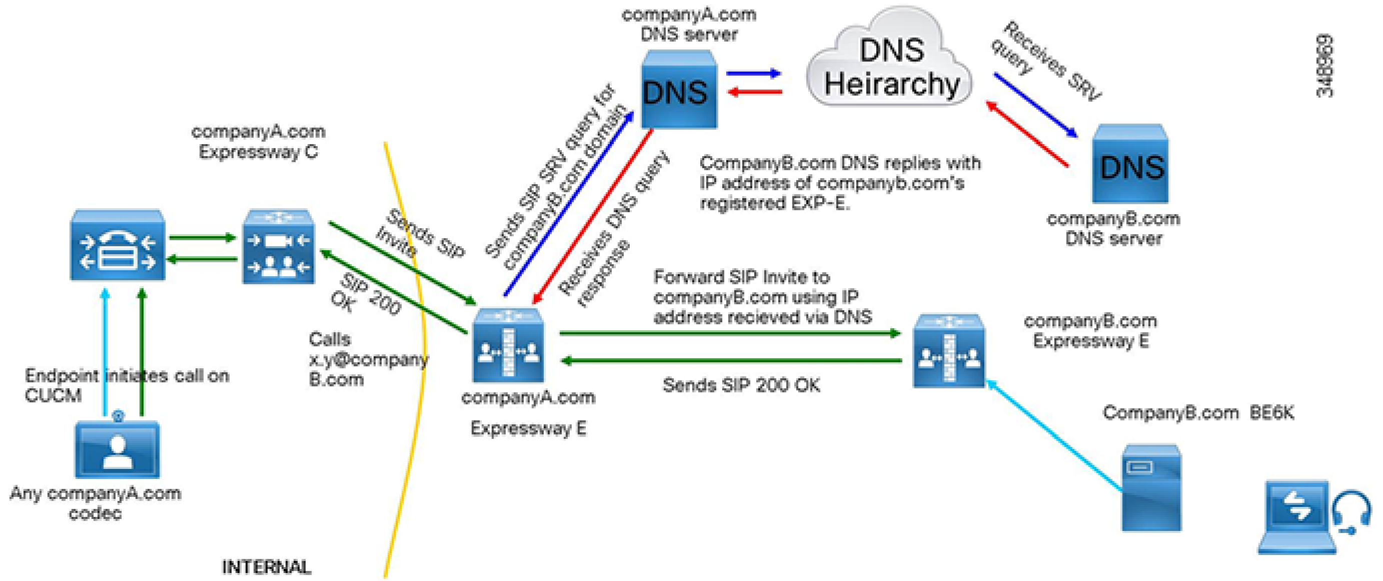
Business-to-business communications require the ability to look up the domains of remote organizations for the purpose of URI routing. This is done by creating a DNS zone on Expressway-E. This zone should be configured with the default settings. Both SIP and H.323 are set by default. This allows Expressway-E to automatically re-initiate a DNS query using the other protocol not used by the initiating call, thereby giving the call the best chance of success. Expressway-C and Expressway-E use the protocol that was used to initiate the call, and they automatically try the other protocol when SIP-to-H.323 gateway interworking is enabled on the Expressway.

SIP-to-H.323 interworking should be set to **On** for Expressway-E. If a call is receive as an H.323 call, this allows Expressway-E to interwork the call to SIP and use native SIP for the rest of the call legs to Unified CM. Likewise, an outbound call to an H.323 system will remain a SIP call until it reaches Expressway-E, where it will be interworked to H.323.

In order to receive business-to-business communications over the Internet, External SIP and H.323 DNS records are required. These records allow other organization to resolve the domain of the URI to the Expressway-E that is offering that call service. Cisco' s validated design included the SIP and SIPS SRV records and the H.323cs SRV record for business-to-business communications. The H.323ls SRV record is not necessary for Expressway-E because this record is used by an endpoint to find its gatekeeper for registration.

Figure 4-10 shows the DNS process for resolving the domain of the URI, and Example 4-1 shows an SRV lookup example.

Figure 4-10 URI Dialing with DNS



Example 4-1 SRV Record Examples for the Domain ent-pa.com

```
>nslookup
set type=srv
_sips._tcp.ent-pa.com

Non-authoritative answer:
_sips._tcp.ent-pa.comSRV service location
    priority= 1
    weight  = 10
    port    = 5061
    srv hostname= expe.ent-pa.com.
```

For more information on configuring a DNS zone on Expressway-E, refer to the *Cisco Expressway Basic Configuration Deployment Guide*.

IP-based Dialing for Business-to-Business Calls

IP-based dialing is a feature well known and used in most scenarios, when dealing with H.323 endpoints. The Cisco Collaboration Architecture uses SIP URIs and does not need IP-based dialing. However, when interacting with endpoints in other organizations that are capable of making and receiving calls using IP addresses only, the Cisco Collaboration Architecture allows IP-based dialing for both inbound and outbound calls.

Outbound Calls

Outbound IP dialing is supported on Expressway-E and Expressway-C, but it does not have full native support on Cisco Unified Communications Manager. However, it is possible to set up Unified CM to have IP-based dialing, as described below.

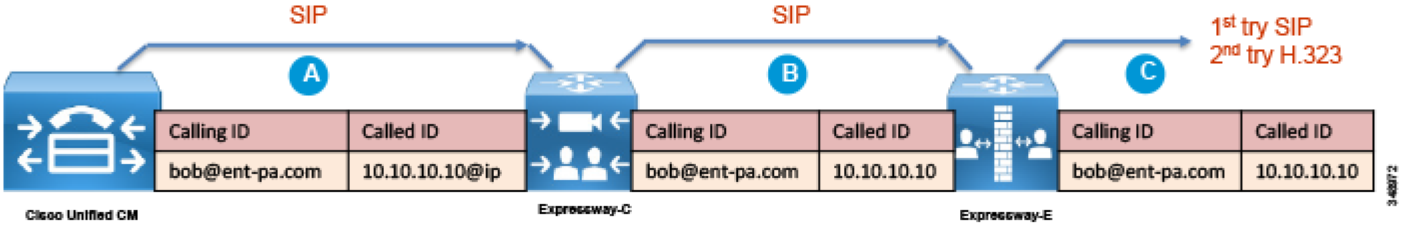
Instead of dialing the IP address alone, users on Cisco Unified CM can dial a SIP URI-based IP address as shown in this example: 10.10.10.10@ip, where "@ip" is literal and could be replaced with "external", "offsite" or other meaningful terms.

Unified CM will match a SIP route pattern configured to route the "ip" fictional domain to Expressway-C. Expressway-C strips off the domain "@ip" and sends the call to the Expressway-E, which is also configured for IP address dialing.

Calls to unknown IP addresses on Expressway -E should be set to **Direct**. Since IP-based address dialing is mostly configured in H.323 endpoints when no call control is deployed, this allows Expressway-E to send H.323 calls directly to an endpoint at a public IP address. The call will remain a SIP call until interworked on Expressway-E, as shown in Figure 4-11.

There are other options to provide for IP address dialing. One option is to replace the "." used in the IP address field with the symbol "*", as in the following example: "10*10*10*10". Cisco Unified Communications Manager will match it against a route pattern, and Expressway will replace the "*" with the "." using regular expression (regex) search rules.

Figure 4-11 Example of Outbound IP-based Dialing



Calling ID	Called ID
bob@ent-pa.com	10.10.10.10@ip

Calling ID	Called ID
bob@ent-pa.com	10.10.10.10

Calling ID	Called ID
bob@ent-pa.com	10.10.10.10

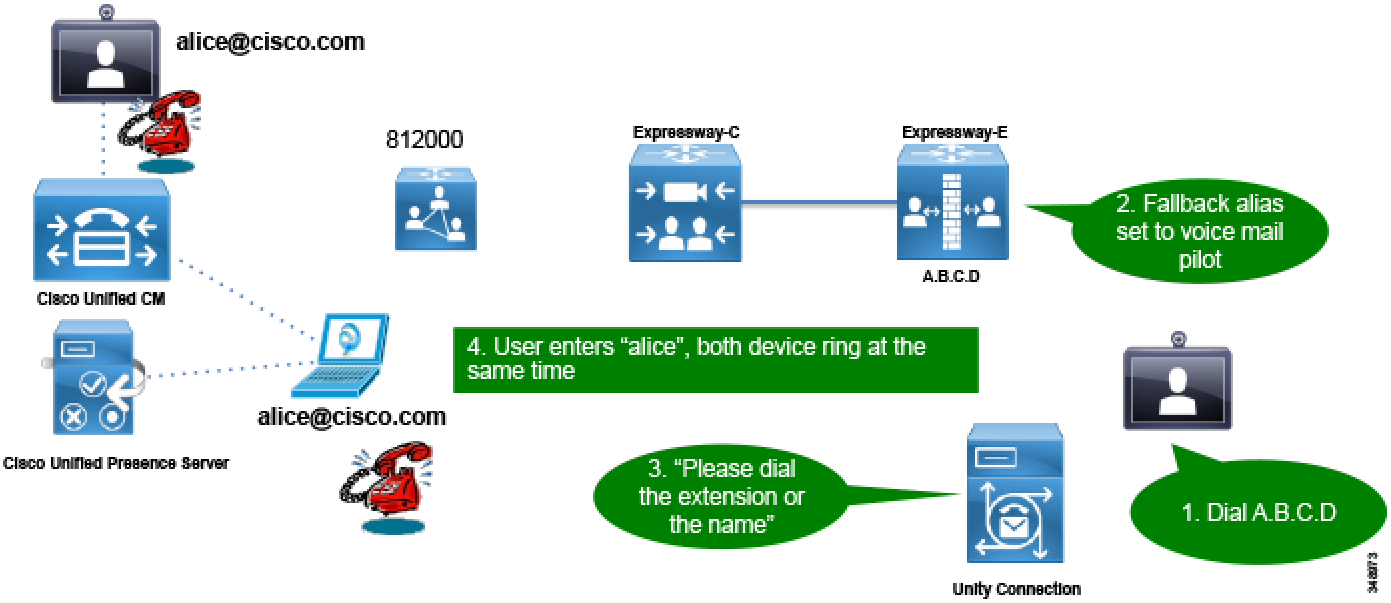
Inbound Calls

IP-based inbound calls make use of a *fallback alias* configured in Expressway-E. When a user on the Internet dials the IP address of the Expressway-E external LAN interface, Expressway-E receives the call and sends the call to the alias configured in the fallback alias setting. As an example, if the fallback alias is configured to send the call to conference number 80044123 or to the conference alias meet@ent-pa.com, the inbound call will be sent to the TelePresence Server in charge of such conferences.

If the static mapping between the IP address and the fallback alias is too limited, it is possible to set the fallback alias to the pilot number of Cisco Unity Connection. In this way it is possible to use the Unity Connection auto-attendant feature to specify the final destination through DTMF, or by speech recognition if Unity Connection is enabled to support this feature.

If Unity Connection is used as an auto-attendant feature for external endpoints dialing the IP address of the Expressway-E, remember to set the **Rerouting Calling Search Space** on the Unified CM trunk configuration for Unity Connection. Figure 4-12 shows the setup.

Figure 4-12 Example of Inbound IP-based Dialing


alice@cisco.com
812000
Expressway-C
Expressway-E
A.B.C.D
2. Fallback alias set to voice mail pilot
Cisco Unified CM
4. User enters "alice", both device ring at the same time
Cisco Unified Presence Server
alice@cisco.com
3. "Please dial the extension or the name"
1. Dial A.B.C.D
Unity Connection
348073

Deployment of External XMPP Federation through Expressway-C and Expressway-E

XMPP federation utilizes the same type of traversal connection – Unified Communications traversal – as does mobile and remote access. XMPP federation can be deployed as a standalone service. It can also be deployed on the same Expressway-C and Expressway-E pair with mobile and remote access, utilizing the same Unified Communications traversal link.

Perform the following typical tasks to deploy instant messaging and presence federation:

1. Validated email addresses for federation.

   XMPP federation through Expressway does not support translation of email addresses to XMPP addresses. Translation of email addresses to Jabber IDs is a feature of the IM and Presence server federation model. This feature is typically used to improve user experience and simplify communication for XMPP federation when the email URI convention and JID URI convention are different. When deploying XMPP federation through Expressway, the same goals of improved user experience and simplified communications apply. We recommends setting the IM and Presence domain to the same domain as the email domain. We also recommend using the LDAP sAMaccount name for UserID, email address convention, and Jabber ID. In the overall collaboration architecture, we recommend having a comprehensive and consistent strategy for URI convention that is repeatable and scalable.

2. Ensure that the IM and Presence service is operational and has XMPP federation turned off.

   XMPP federation on the IM and Presence server must be turned off so that it does not interfere with the federation configured on the Expressway.

3. Complete server certificate requirements.

   Plan ahead when setting up certificates for Expressway-C and Expressway-E. If you plan to use chat node aliases as a part of XMPP federation, the chat nodes alias FQDN must be included in the subject alternate name (SAN) field of the certificate. Doing this ahead of time avoids having to generate new certificates and possibly incurring greater expense for the public certificates on your Expressway-Es.

4. Configure the local domains for XMPP federation on Expressway-C.

5. Configure Expressway-E for XMPP federation and security.

   This step enables federation and the level of security desired for external federation. Authentication is required and is set up via the dialback secret. Securing the communications via TLS is the recommended configuration. Authorization of which foreign domains and external chat node aliases are allowed or denied, is configured in this section as well.

6. Configure how XMPP servers for federated domains and chat node aliases are located using either DNS lookups or static routes.

   The Expressway series supports federation via DNS SRV records and federation via static routes. Static routes define a path to reach external domains without having to do a DNS query. Public XMPP SRV records are used to resolve external domains that support federation. These records are required for other organizations to reach your organization when deploying an open federation model.

7. Ensure that the correct firewall ports are open.

8. Check the status of XMPP federation.

Deployment of Cisco Unified Border Element for PSTN Voice Connection Through a SIP Trunk

Cisco Unified Border Element is the recommended session border controller for PSTN centralized access. It is deployed as a demarcation point between the enterprise network and the Telecom carrier network. It gives access to the IP PSTN through its external interface and to the enterprise network through its internal interface. It enables centralized PSTN service and therefore has to be deployed where the enterprise network connects to the Telecom carrier's network.

Because all remote sites leverage central PSTN connectivity, Cisco Unified Border Element has to be highly redundant. If the PSTN central service is not available, only those offices with local PSTN access would be able to make external calls. Therefore, we recommend deploying Cisco Unified Border Element in pairs to provide redundancy.

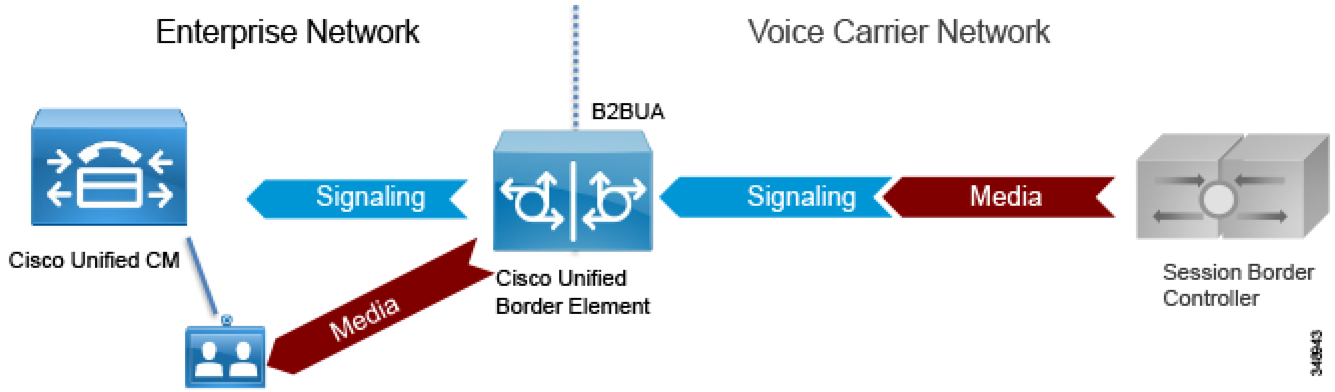
Unified Border Element is a Cisco IOS feature set supported on the Cisco IOS Integrated Services Router (ISR) and Aggregation Services Router (ASR) platforms. For information on how to choose the correct platform, see the Sizing chapter.

Cisco Unified Border Element is a session border control that terminates sessions from Unified CM and re-originates them toward the Telecom carrier network, and vice-versa. Note that in contrast to Expressway-E, which is exposed on the Internet, Cisco Unified Border Element is deployed between private networks: the corporate network and the carrier's network. From the carrier’s perspective, the traffic to the centralized PSTN originates from the Cisco Unified Border Element external interface. From the enterprise's perspective, the traffic from the carrier originates from the internal Cisco Unified Border Element interface. In this sense, the Cisco Unified Border Element performs topology hiding.

Deployment of Cisco Unified Border Element is different from that of the Expressway. While the former gives access to the carrier network - a private, controlled and secured network - the latter gives access to the Internet. For this reason, deployment of Cisco Unified Border Element does not require a DMZ.

For this Preferred Architecture, as shown in Figure 4-13, the Unified Border Element has a WAN interface on the Telecom carrier network and a LAN interface on the enterprise network.

Figure 4-13 IP PSTN Architecture

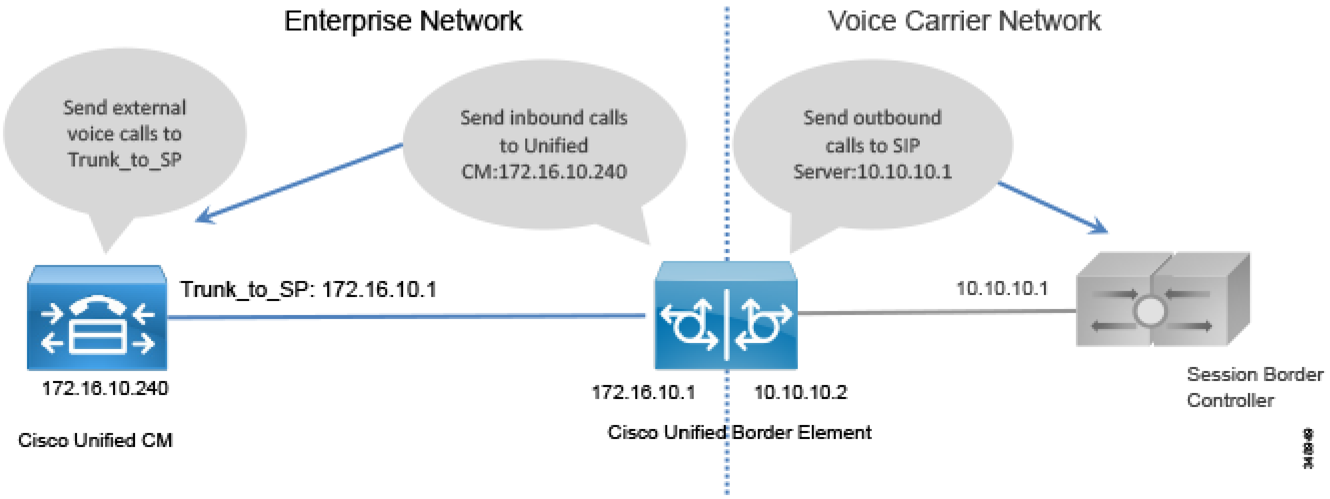


Cisco Unified Border Element performs the following functions:

- Topology hiding, as shown in Figure 4-14, including address and port translations. All traffic from Unified CM is sent to the Unified Border Element internal interface, and all traffic from the Telecom carrier soft-switch is sent to the Unified Border Element external interface. There is no direct connection between them. Figure 4-14 details the trunking configuration on Cisco Unified CM and the voice routes on the Unified Border Element.
- Delayed Offer to Early Offer conversion, and vice versa
- Media interworking — In-band and out-of-band DTMF support, DTMF conversion, fax passthrough and T.38 fax relay, volume and gain control
- Call admission control (CAC) — CAC can be performed by Unified Border Element based on resource consumption such as CPU, memory, and call arrival spike detection. CAC can be implemented at interface level or globally. While CAC configured on Unified CM is location-based, CAC configured on the Unified Border Element is resource-based. Resources-based CAC is recommended to avoid over-subscription of the Unified Border Element and for security reasons (see the section on Security for Collaboration Edge).
- Security capabilities, including RTP-to-SRTP interworking, SIP malformed packet detection, non-dialog RTP packet drops, SIP listening port configuration, digest authentication, simultaneous call limits, call rate limits, toll fraud protection, and a number of signaling and media encryption options
- Mid-call supplementary services, including hold, transfer, and conference
- PPI/PAI/Privacy and RPID — Identity Header Interworking with Telecom carriers
- Simultaneous connectivity to SIP trunks from multiple Telecom carriers

- Conversion of multicast music on hold (MoH) to unicast MoH
- Billing statistics and call detail record (CDR) collection

Figure 4-14 Trunking Considerations for Cisco Unified Border Element


Enterprise Network
Voice Carrier Network
Send external voice calls to Trunk_to_SP
Send inbound calls to Unified CM:172.16.10.240
Send outbound calls to SIP Server:10.10.10.1
Trunk_to_SP: 172.16.10.1
10.10.10.1
172.16.10.240
Cisco Unified CM
172.16.10.1
10.10.10.2
Cisco Unified Border Element
Session Border Controller
348849

PSTN Gateways

Legacy PSTN gateways are deployed in a distributed architecture, where each site has its own PSTN connection. We recommend using Cisco Unified Border Element for centralized PSTN access, but PSTN gateways can still be used as a backup for those sites heavily relying on external calls to run their daily business.

In this case the number of concurrent ISDN channels can be much lower than the number of concurrent calls to the centralized PSTN because they are used just in backup scenarios. As an example, if the normal situation allows for 30 concurrent calls to the centralized PSTN, it would be possible to size the backup ISDN gateway to support only two BRI channels, since they would be used in backup scenarios only.

Cisco voice gateways support:

- DTMF relay capabilities
- Supplementary services support — Supplementary services are basic telephony functions such as hold, transfer, and conferencing.
- Fax passthrough and T.38 fax relay

PSTN gateways support many protocols (SCCP, MGCP, H.323, SIP). SIP is the recommended protocol because it aligns with the overall Cisco collaboration solution and is the protocol of choice for new voice and video products.

Voice gateway functionality is enabled on any Cisco ISR with appropriate PVDMs and service modules or cards.

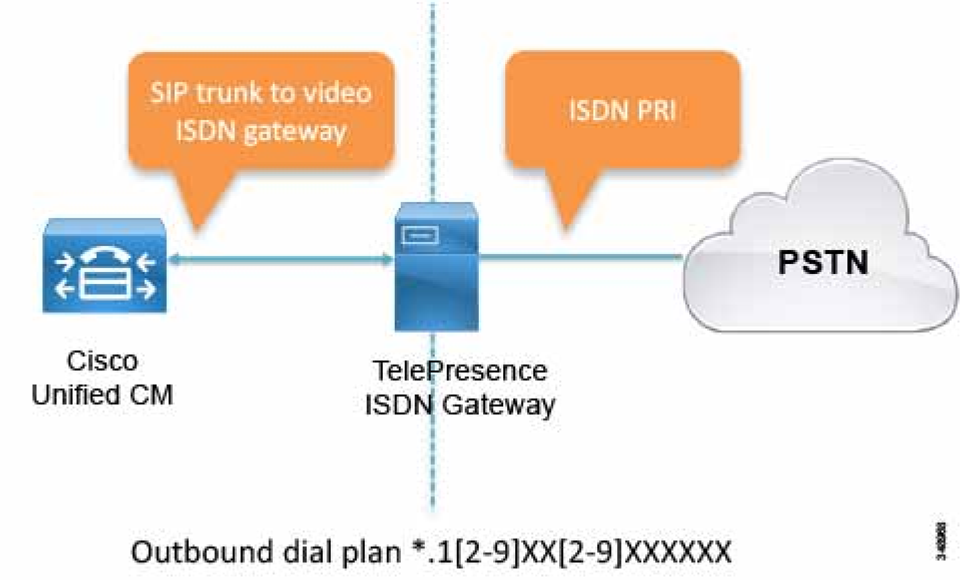
Video ISDN Gateways

The Preferred Architecture for Enterprise Collaboration incorporates the standard deployment for Cisco TelePresence ISDN Gateways with ISDN PRI trunks, and SIP trunks to Unified CM. The Cisco TelePresence ISDN Gateways include the MSE 8321 gateway and the GW 3241 gateway. The TelePresence ISDN Gateway is an optional item and is required only if there is a need to provide the ability to send and receive calls from legacy ISDN videoconferencing systems.

Requirements and Recommendations

- Use SIP instead of H.323 as the IP protocol for communicating with Unified CM.
- Make the dial plan as simple as possible on the ISDN gateway, and perform all dial string manipulations on Unified CM.
- Leave the dial plan setup for last. The ISDN gateways block all calls by default until the dial plan settings are configured. This secures the gateway until it is ready to be used.
- Use of a ***** in front of the ISDN number as a prefix to route calls to the TelePresence ISDN gateway. The * allows video ISDN calls to be differentiated from voice PSTN calls and does not conflict with existing international numbering plans. This also minimizes changes to the user experience when dialing. The Unified CM dial plan removes the * and routes the remaining digits through to the TelePresence ISDN gateway. (See Figure 4-15.)

Figure 4-15 Video ISDN Gateway Dial Plan


SIP trunk to video ISDN gateway
ISDN PRI
PSTN
Cisco Unified CM
TelePresence ISDN Gateway
Outbound dial plan *.1[2-9]XX[2-9]XXXXXX
348088

Limitations and Restrictions for ISDN Video Gateways

- Only the hold and resume features are supported during ISDN video calls.
- Video call transfer is not supported.
- ISDN calls into CMR Cloud are not supported.

Limitations and Restrictions for Mobile and Remote Access

The following limitations and restrictions apply to mobile and remote access (MRA) connectivity:

- CTI is not supported.
- Jabber desktop phone control is not supported.
- MRA does not support peer-to-peer file transfers using the IM and Presence Service or Jabber. MRA supports only managed file transfer (MFT) with IM and Presence Service 10.5.2 (and later) and Jabber 10.6 (and later) clients.
- Jabber mobile features dial-via-office reverse (DVO-R), dual-mode handoff, and session persistency are not supported.
- One-Button-To-Push for TelePresence Conductor-based TelePresence is not supported.
- TelePresence Conductor endpoint management capabilities are not supported.

Limitation and Restrictions for Cisco Unified Border Element

- For Cisco ISR, the transport protocol can be TCP and UDP only. For Cisco ASR the transport protocol can be TLS as well.
- Transcoding, DTMF interworking, IVR, SIP-to-TLS and RTP-to-SRTP conversions, and fax and modem features are preserved in failover scenarios. For Cisco ISRs, no DSP-related functions are preserved.

High Availability for Collaboration Edge

High availability is a critical aspect of designing and deploying collaboration systems. Collaboration Edge allows for redundancy, load-sharing, and call license sharing.

High Availability for Expressway-C and Expressway-E

We recommend deploying Expressway-C and Expressway-E in clusters. Each cluster can have up to six Expressway nodes and a maximum of N+2 physical redundancy. All nodes are active in the cluster. For details about cluster configuration, refer to the *Cisco Expressway Cluster Creation and Maintenance Deployment Guide*.

Expressway clusters provide configuration redundancy.The first node configured in the cluster is the *publisher*, while all other nodes are *subscribers*. Configuration is done in the publisher and automatically replicated to the other nodes.

Expressway clusters provide call license sharing and resilience. All rich media sessions and TURN licenses are shared equally across nodes in the cluster. Call licenses are contributed by the licenses configured on each node.

Expressway-C and Expressway-E deployed as virtual machines support VMware VMotion. VMware VMotion enables the live migration of running virtual machines from one physical server to another. When moving the virtual machine, Expressway-C and Expressway-E servers will maintain actives calls when handling signaling only or when handling both signaling and media. This provides high availability for the Expressway nodes as well as call resilience across Cisco Unified Computing System (UCS) hosts.

The following rules apply to Expressway clustering:

- Expressway-C and Expressway-E node types cannot be mixed in the same cluster.
- All nodes in a cluster must have identical configurations for zones, authentication, and call policy.
- Configuration changes should be made only on the master node, and this will overwrite the configuration on the other peers in the cluster when replication occurs.
- If a node becomes unavailable, the licenses it contributed to the cluster will become unavailable after 2 weeks.
- Deploy an equal number of nodes in Expressway-C and Expressway-E clusters.
- Deploy the same OVA template throughout the cluster.
- All nodes in a cluster need to be within 30 ms maximum round-trip time to all other cluster nodes. Clustering over the WAN is thus not recommended due to latency constraints.
- You must use the same cluster preshared key for all nodes within the same cluster.
- H.323 must be enabled on all nodes in a cluster for database replication. At the same time, if you also want to block H.323 calls coming from the Internet, you can configure Expressway-E with firewall rules to drop H.323 traffic on the external LAN interface.
- If mobile and remote access and business-to-business communications are enabled on the same Expressway-C and Expressway-E pairs, the SIP port number used on the SIP trunk between Unified CM and Expressway-C needs to be changed from the default 5060 or 5061.
- A DNS SRV record must be available for the cluster and must contain A or AAAA records for each node of the cluster.

Since Expressway-C is deployed in the internal network and Expressway-E in the DMZ, Expressway-C has to be connected to Expressway-E through a Unified Communications *traversal zone* for mobile and remote access. Business-to-business calls require a separate traversal zone, which retains the name of traversal client zone for Expressway-C and traversal server zone for Expressway-E. The traversal server, traversal client, and Unified Communications traversal zones include all the nodes of Expressway-C and Expressway-E, so that if one of the nodes is not reachable, another node of the cluster will be reached instead.

As shown in Figure 4-16, Expressway-C connects to all servers of the Cisco Unified CM, IM and Presence, and Unity Connection clusters, so high availability and redundancy are preserved across the entire connection path.

Figure 4-16 Expressway Services Connection

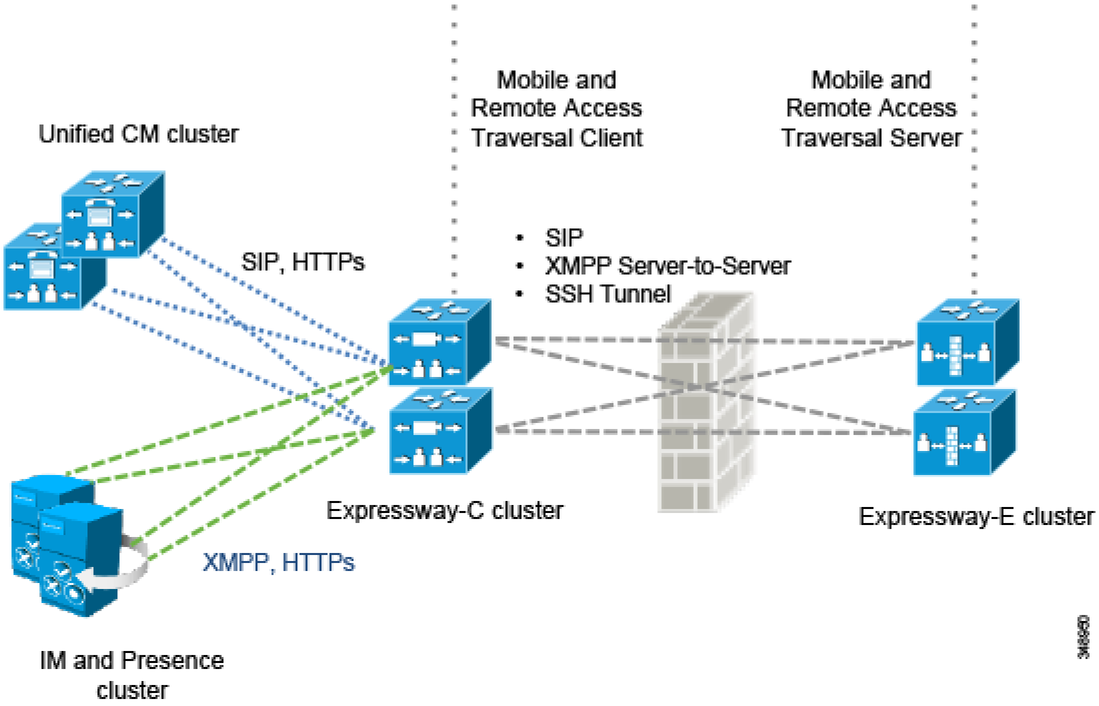


Figure 4-16 shows the high availability built into the Unified Communications traversal zone and into mobile and remote access. However, the following description applies to both Unified Communications traversal zones and standard (client and server) traversal zones.

The traversal client zone configured on Expressway-C should contain the fully qualified domain names of all of the cluster nodes of the corresponding Expressway-E cluster. Likewise, the traversal server zone should connect to all Expressway-C cluster nodes. This is achieved by including in the subject alternative names of the Expressway-C certificate the FQDNs of the Expressway-C cluster nodes and by setting the **TLS verify subject name** equal to the FQDN of the Expressway-C cluster. This creates a mesh configuration of cluster nodes across the traversal zone and provides continuous and high availability of the traversal zone until the last cluster node is unavailable.

Expressway-C connects to Unified CM via trunks for routing inbound and outbound business-to-business calls. Unified CM also trunks to Expressway-C. For high availability, the fully qualified domain names of each Expressway-C cluster node should be listed in the trunk configuration on Unified CM. If Unified CM is clustered, the fully qualified domain name (FQDN) of each member of the cluster should be listed in the neighbor zone profile of Expressway-C.

A meshed trunk configuration is created here as well. Unified CM will check the status of the nodes in the trunk configuration via a SIP OPTIONS Ping. If a node is not available, Unified CM will take that node out of service and will not route calls to it. Expressway-C will also check the status of the trunk from Unified CM via a SIP OPTIONS Ping. Calls will be routed only to nodes that are shown as active and available. This provides high availability for both sides of the trunk configuration.

DNS SRV records can add to availability of Expressway-E for inbound business-to-business traffic. For high availability all nodes in the cluster should be listed with the same priority in the SRV record. This allows all nodes to be returned in the DNS query. A DNS SRV record helps to minimize the time spent by a client on lookups since a DNS response can contain all of the nodes listed in the SRV record. The far-end server or far-end endpoint will typically cache the DNS response and will try all nodes returned in the DNS query until a response is received. This provides the best chance for a successful call.

In addition, Expressway clusters support rich media license sharing across clusters. If a node is lost from the cluster, its call licenses will continue to be shared for the next 2 weeks. Any one Expressway cannot process any more rich media licenses than its physical capacity, even though it can carry more licenses than its physical capacity.

High Availability for Cisco Unified Border Element

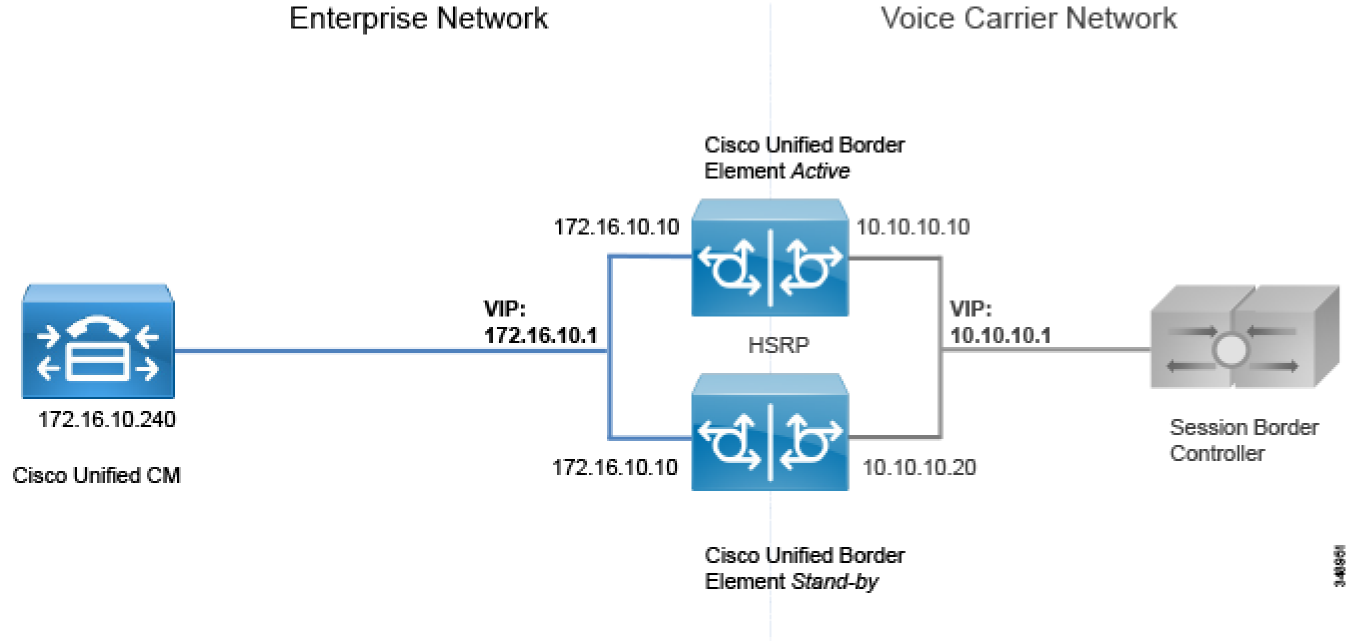
High availability for Cisco Unified Border Element can be achieved in more than one way. For the Preferred Architecture, box-to-box redundancy with call preservation is recommended because it provides both signaling and media call preservation if the Unified Border Element fails.

Unified Border Element servers are deployed in pairs, following the active/standby model. If the active Unified Border Element goes down, the standby Unified Border Element is engaged and all active sessions are transferred. This provides high availability for both signaling and media. (See Figure 4-17.)

Hot Standby Routing Protocol (HSRP) technology provides high network availability by routing IP traffic from hosts on networks without relying on the availability of any single router. HSRP is used in a group of routers for selecting an active router and a standby router. HSRP monitors both the inside and outside interfaces; if any interface goes down, the whole device is considered down, and the standby device becomes active and takes over the responsibilities of the active router.

Box-to-box redundancy uses the HSRP protocol to form an HSRP active/standby pair of routers. The active and standby servers share the same virtual IP address and continually exchange status messages. Unified Border Element session information is shared across the active/standby pair of routers, as seen in Figure 4-17, where 172.16.0.1 and 10.10.10.10 are the virtual IP addresses of the Cisco Unified Border Element pairs. This enables the standby router to immediately take over all Unified Border Element call processing responsibilities if the active router goes out of service for planned or unplanned reasons.

Figure 4-17 Cisco Unified Border Element Box-to-Box Redundancy



High Availability for Voice Gateways

PSTN gateways directly connect via physical interfaces to the PSTN network. If a gateway goes down, all communications with the PSTN are cleared. Mechanisms such as HSRP would not be of any benefit in this case, as they would in the case of PSTN access through IP trunks to a Telecom carrier. Unlike the Unified Border Element, a TDM-based PSTN gateway deployment is by nature distributed, although there are cases where a centralized PSTN with gateway interconnection is deployed. Also, a PSTN voice gateway manages a smaller amount of calls than a Unified Border Element does. Due to the nature of PSTN, media preservation is not possible in this scenario.

However, it is possible to provide signaling resilience by configuring multiple gateways in the same Unified CM route group, so that load balancing of calls will occur. If one of the gateways in the group goes down, all calls will be dropped, but new calls will be established using one of the remaining available gateways.

Security for Collaboration Edge

This section explains how to implement security in the Collaboration Edge.

Security for Expressway-C and Expressway-E

Security on Expressway-C and Expressway-E can be further partitioned into network level and application level. Network level security includes feature such as firewall rules and intrusion protection, while application level security includes authorization, authentication, and encryption.

Network Level Protection

Network level protection on Expressway-C and Expressway-E consists of two main components: firewall rules and intrusion protection.

Firewall rules enable the ability to:

- Specify the source IP address subnet from which to allow or deny traffic.
- Choose whether to drop or reject denied traffic.
- Configure well known services such as SSH and HTTP/HTTPS, or specify customized rules based on transport protocols and port ranges.
- Configure different rules for the LAN 1 and LAN 2 interfaces on Expressway-E.

The Automated Intrusion Protection feature should be used to detect and block malicious traffic and to help protect the Expressway from dictionary-based attempts to breach login security.

Automated Intrusion Protection works by parsing the system log files to detect repeated failures to access specific service categories, such as SIP, SSH, and web/HTTPS. When the number of failures within a specified time reaches the configured threshold, the source host IP address (the intruder) and destination port are blocked for a specified period of time. The host address is automatically unblocked after that time period so as not to lock out any genuine hosts that might have been temporarily misconfigured.

Mobile and Remote Access

TLS, SRTP, HTTPS, and XMPP are the only configuration options between the client on the Internet and Expressway-C for mobile and remote access.

The connection between Unified CM and Expressway-C may be encrypted and authenticated, depending on the configuration. If Unified CM is in mixed-mode, we recommend end-to-end encryption of media and signaling.

Security certificates are needed on these connections. Certificates provide the identity of servers and clients and must be deployed on Expressway-C, Expressway-E, Unified CM, and the Unified CM IM and Presence Service. The recommended configuration is to use a certificate authority (CA) to sign certificates.

CAs can be private or public. Private CA deployments have the benefit of being cost-effective, but these certificates are valid only inside the organization. Public CAs increase the security and are trusted by every organization; thus, they are commonly used for communications between different companies.

If Expressway-C and Expressway-E are used only for mobile and remote access, the company can choose to deploy a private CA or to rely on a public one. However, if Expressway-C and Expressway-E are also used for business-to-business communications, a certificate signed by public CA has to be deployed on Expressway-E. In this case, Expressway-E certificates have to be signed by a public CA such as VeriSign/Symantec, GeoTrust, GoDaddy, or others. If Expressway-E receives a business-to-business call from an entity whose certificate is signed by a CA that is present in the Trusted CA certificate list, the call is accepted. If the CA that has signed the certificate is not in the list, the call will be rejected. It is therefore important to pre-populate Expressway-E with a trust list of major CA certificates if the Expressway is enabled for business-to-business too.

For cost reduction, Expressway-C certificates may be signed by an internal CA not recognized outside the company. In this case, it is important that the internal CA certificate be included in the Trusted CA certificate list of Expressway-E in order for Expressway-C and Expressway-E to establish a connection. Table 4-3 summarizes the public and private approach for certificate deployment.

Table 4-3 Public and Private Certification Authority And Certificates

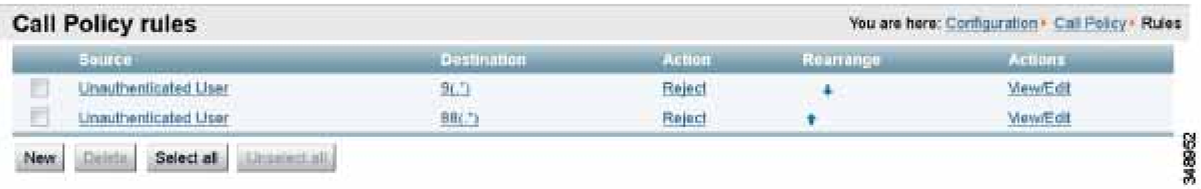
	Unified CM	IM and Presence Service	Expressway-C	Expressway-E
Certificate signed by	Internal CA	Internal CA	Internal CA	Public CA
Trust List includes	Internal CA certificate	Internal CA certificate	Internal CA and public CA certificates	Internal CA and public CA certificates

Business-to-Business Communications

Securing business-to-business communications include authentication, encryption, and authorization. Business-to-business communications use an authenticated traversal link by default. The traversal link can also benefit from the use of a Public Key Infrastructure (PKI) verified by a mutually authenticated transport layer security (MTLS) connection between Expressway-C and Expressway-E. If the business-to-business traversal link is deployed on the same Expressway-C and Expressway-E infrastructure as mobile and remote access, make sure that the traversal zone uses the FQDNs of the cluster nodes of Expressway-C and Expressway-E. This makes it straightforward to use certificates for each server to validate the offered certificate against its certificate trust for the traversal connection.

Inbound calls can be differentiated by whether they are authenticated or unauthenticated. This differentiation can be used to authorize access to protected resources. Unknown remote business-to-business calls should be treated as unauthenticated and restricted from access to protected resources such as IP voice and video gateways. This is accomplished by configuring Call Processing Language (CPL) rules with regular expressions to block access to prefixes used for gateway access. (See Figure 4-18.)

Figure 4-18 CPL Rules for Unauthenticated Callers



Call Policy rules

You are here: Configuration ▸ Call Policy ▸ Rules

	Source	Destination	Action	Rearrange	Actions
□	Unauthenticated User	9(.*)	Reject	↓	View/Edit
□	Unauthenticated User	88(.*)	Reject	↑	View/Edit

Signaling and media encryption is important for business-to-business calls, but it needs to be deployed carefully so as not to restrict or limit the ability to receive calls. There is a variety of older SIP and H.323 systems that you may be communicating with that do not support signaling or media encryption.

Requirements and Recommendations

- Set media encryption on the traversal client side (Expressway-C) of the business-to-business traversal zone to **best effort**. This means encryption for calls will always be tried first and DMZ traffic will be encrypted. It also allows fallback to unencrypted calls. If strong encryption policy is required, set the media encryption to **force encrypted**.
- Use TLS for signaling encryption for the SIP trunk between Unified CM and Expressway-C.

If mobile and remote access (MRA) is also deployed in the organization, both MRA to business-to-business calls can potentially go out unencrypted. Best-effort media encryption means that the outbound calls will be tried with encryption first.

Certificate settings for mobile and remote access and business-to-business call scenarios:

- General requirements for certificates are that the fully qualified DNS name (FQDN) of Expressway-C and Expressway-E must match the hostname in the certificate.
- While business-to-business communication does not have any other requirements for certificates, MRA has more. For a detailed explanation on how to set up certificates on Expressway for mobile and remote access, refer to the *Unified Communications Mobile and Remote Access via Cisco Expressway Deployment Guide*.

Security for Cisco Unified Border Element

Unlike Internet connections, PSTN connectivity over IP trunks is delivered through a private network offered by the Telecom carrier. As such, it is a controlled network. Security deployed for the Internet Edge is thus different from that deployed for IP PSTN access. Between the Cisco Unified Border Element and the carrier, there are no firewalls; however, in specific cases, companies and Telecom providers require the use of an enterprise DMZ.

Between the carrier and the enterprise network, the traffic is sent unencrypted. Depending on the corporate policies, internal enterprise traffic can be encrypted or not. In such cases, the Unified Border Element is able to perform TLS-to-TCP and SRTP-to-RTP conversion. Usage of the internal CA to sign Unified Border Element certificates is recommended when multiple gateways are deployed.

Because the Unified Border Element is deployed without a firewall, it is protected at various layers. As an example, it is possible to create an access control list to allow only the Telecom carrier's session border control to initiate calls from the PSTN side, and to allow only Unified CM to initiate calls from the internal network side.

The Unified Border Element is also protected against toll fraud and telephony denial of service (TDoS) attacks. Large packet arrival rates can also be mitigated through call admission control mechanisms based on CPU, memory, bandwidth utilization, and call arrival spike detection.

Security for Voice Gateways

PSTN gateways have an interface on the customer network and a second interface on the PSTN. They are deployed inside the corporate network and they are not reachable from the Internet. The PSTN is inherently secure; thus, there are no specific tools to use to protect the gateway, unless the gateway is deployed on a router that also gives access to the Internet. In this case Cisco IOS features on the gateway can be used to perform firewall and intrusion protection. In all other cases, no specific tools are required to protect the gateway (for example, denial of service (DoS) prevention and so on).

However, it is always possible to encrypt media from the endpoint to the gateway. In such cases the gateway will use TLS and SRTP. Use of CA-signed certificates is recommended in this case.

Security for Video ISDN Gateways

Video ISDN gateways have an IP interface on the customer network and another interface on the PSTN. Two common security threats that need to be guarded against with gateways are toll fraud from IP-to-ISDN and ISDN hairpinning of calls.

Basic CPL rules on Expressway-E can be used to block access to ISDN gateway resources from the Internet. Calling search spaces should be used to block access from Unified CM registered devices.

Even though the entire dial plan and permissions are configured on Cisco Unified Communications Manager, it is good security practice to use CPL rules on Expressway-E to block fraudulent attempts to access voice and video gateways before they are routed to Unified Communications Manager.

Scaling the Collaboration Edge Solution

The number of Collaboration Edge clusters deployed does not depend on the number of call control clusters, but rather on the number of connection points to the Internet. A customer with multiple Unified CM and IM and Presence clusters, and multiple TelePresence Conductor clusters, will have a single Internet Edge if that customer has a single Internet breakout point. The same environment might have multiple PSTN hop-offs if the Telecom carrier offers more than one connection point to the PSTN network. The same considerations apply for video ISDN access.

Scaling the Internet Edge Solution

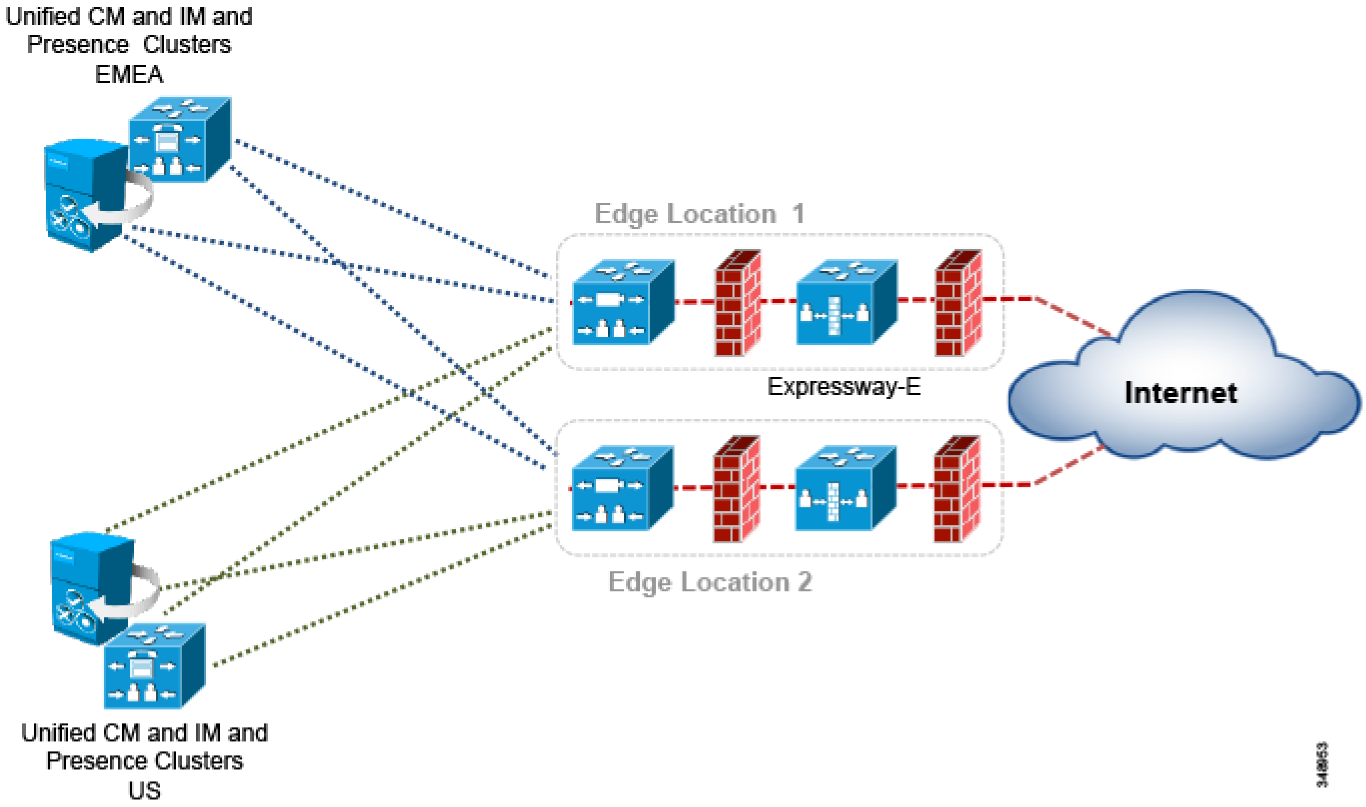
When multiple Internet Edges are deployed, it is important to set routing rules properly in order to send collaboration traffic to the nearest Internet Edge.

Mobile and Remote Access

If multiple Unified CM and IM and Presence clusters are deployed, every Expressway-C must discover all Unified CM clusters. If Expressway-C discovers only some of the clusters, it will be able to proxy registration only for those users belonging to the clusters that have already been discovered.

If a registration request is made from a client belonging to a Unified CM and IM an Presence cluster that has not been discovered by Expressway-C, that client will not be able to log in. This is the reason why it is important for each Expressway-C to discovers all Unified CM and IM and Presence clusters if users are enabled for mobility, as shown in Figure 4-19.

Figure 4-19 Service Discovery of Multiple Unified CM and IM and Presence Clusters


Unified CM and IM and Presence Clusters EMEA
Edge Location 1
Expressway-E
Internet
Edge Location 2
Unified CM and IM and Presence Clusters US
348953

When two or more Internet Edges are deployed, it is important to understand how to split the load between them. If the Internet Edges are deployed in the same datacenter or in the same area, load balancing can occur at the DNS SRV level. As an example, if the enterprise network includes three Internet Edges used for mobile and remote access, each one consisting of a cluster of two Expressway-E and Expressway-C nodes, the _collab-edge._tls.ent-pa.com will include all six Expressway-E records at the same priority and weight. This distributes the registrations and calls equally across the various Expressway-E and Expressway-C clusters.

Once a mobile-and-remote-access connected endpoint is registered through a specific Expressway cluster pair, it will stay connected until the client is disconnected or it has been switched off.

However, if the Expressway clusters are deployed across geographical regions, some intelligent mechanisms on top of the DNS SRV priority and weight record are needed to ensure that the endpoint uses the nearest Expressway-E cluster.

As an example, if an enterprise has two Expressway cluster, one in the United States (US) and the other in Europe (EMEA), it is desirable for users located in the US to be directed to the Expressway-E cluster in the US while users in Europe are directed to the Expressway-E cluster in Europe. This is facilitated by implementing GeoDNS services. GeoDNS services are cost effective and easy to configure. To show how GeoDNS services work, the example below uses an Amazon Route 53 Geo DNS server. There are many GeoDNS services available in the market, including Amazon Route 53, Edge Director, GeoScaling, Max Mind GeoIP2, and others.

With GeoDNS it is possible to route traffic based on different policies such as location (IP address routing), latency (minimum latency), and others. Amazon Route 53 allows routing by both latency and geographical location. We have chosen to configure latency-based routing, but the configuration steps are identical for geographical location routing based on IP addresses.

With latency-based routing, a client in the same site might access different datacenters over time if latency on the Internet changes. However, this does not happen instantly as soon as latency changes, since it is measured as a mean value over a period of time. Spikes due to instant congestion of the Internet are thus absorbed by the mean value.

In our scenario, two Internet Edge Expressway clusters are deployed, one in the US and one in Europe, each composed of two Expressway-C and Expressway-E servers. If the measured latency between the endpoint and the European Edge is less than the latency between the endpoint and the US Edge, the endpoint will be directed to the European Edge for registration.

Although some GeoDNS providers support GeoDNS services on SRV records, many others allow CNAME or A-records only. The recommendation is to implement GeoDNS services on SRV records. The following example shows how to configure the Geo DNS if only CNAME is supported for Geo DNS services.

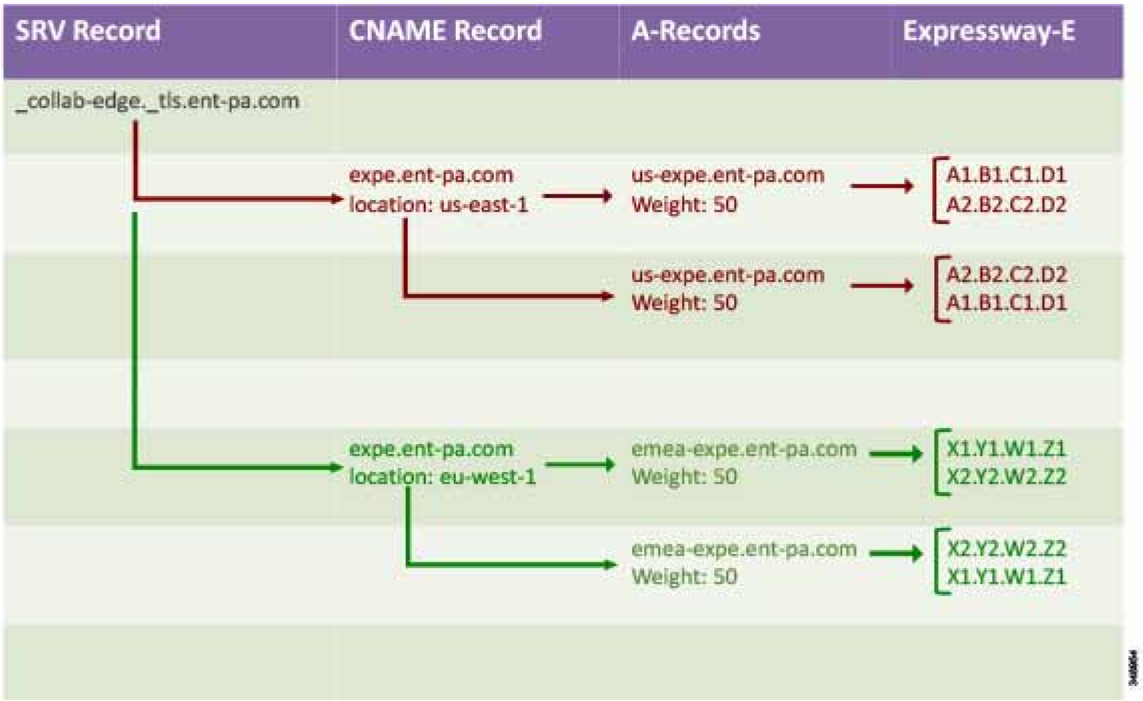
Following this scenario, a single SRV record _collab-edge._tls.ent-pa.com is configured for mobile and remote access. This record resolves into expe.ent-pa.com, a CNAME record, which is an alias that resolves into the real A-records for that resource. There are two records for expe.ent-pa.com; the first resolving into us-expe.ent-pa.com (DNS name for the US Edge) and the second resolves into emea-expe.ent-pa.com (DNS name for the EMEA edge). A-records us.expe.ent-pa.com and emea-expe.ent-pa.com resolve to the IP addresses of Expressway-E server nodes for US and Europe.

While _collab-edge._tls.ent-pa.com is configured for standard routing, expe.ent-pa.com records have a routing policy set to "latency." As a result, the locations of the Expressway-E clusters have to be specified. If the latency between the client and emea-expe.ent-pa.com is less than the latency between the client and us-expe.ent-pa.com, the registration request will be sent to the European Expressway-E. If for any reason latency changes over time and goes above the latency to the US, the us-expe.ent-pa.com will be selected instead.

Both emea-expe.ent-pa.com and us-expe.ent-pa.com are A-records, and the Jabber client or TelePresence Conductor system performs a subsequent query to emea-expe.ent-pa.com or us-expe.ent-pa.com, based on the answer of the _collab-edge._tls.ent-pa.com SRV record. However, since standard A-records cannot set priority and weight like SRV records can, another load-balancing and redundancy mechanism is needed in order to specify which server of the Expressway-E cluster the client has to connect to. This can be done by using a round-robin mechanism. As an example, two emea-expe.ent-pa.com records are created, each one with the routing policy set to "weighted." Specifying the same weight for the two records assures that an equal load-balancing process takes place between the servers of the cluster. The first record resolves into multiple Expressway-E servers of the same cluster (in this case, two servers). The second record resolves into the same set of servers, but in reverse order.

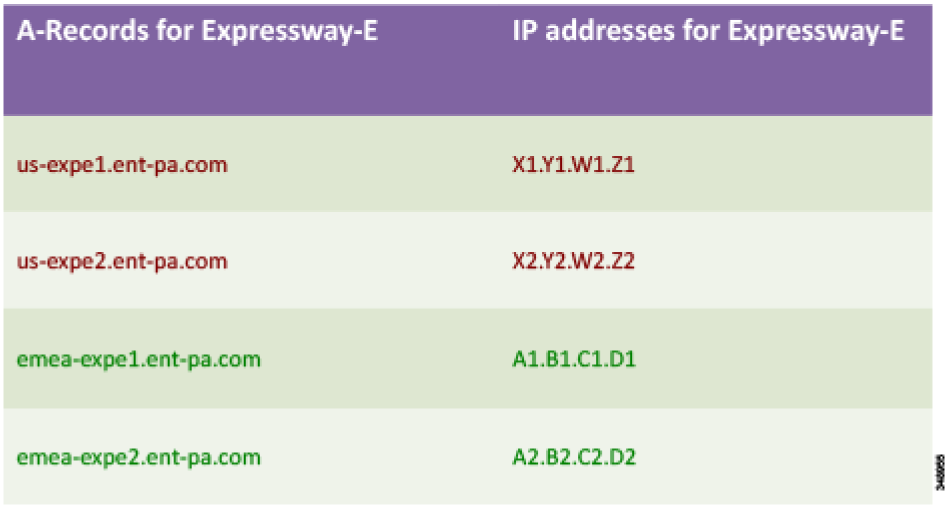
Figure 4-20 shows the DNS record structure for GeoDNS with latency-based routing between regional Expressway-E clusters and round-robin inside the same cluster. As shown in the figure, both records emea-expe.ent-pa.com resolve into the same set of Expressway-E nodes, but in different orders. This provides both redundancy and load balancing.

Figure 4-20 Route 53 DNS Record Structure for Latency-Based Routing



For each Expressway cluster node, an A-record has to be created, as shown in Figure 4-21.

Figure 4-21 DNS A-Records for Expressway Nodes



A-Records for Expressway-E	IP addresses for Expressway-E
us-expe1.ent-pa.com	X1.Y1.W1.Z1
us-expe2.ent-pa.com	X2.Y2.W2.Z2
emea-expe1.ent-pa.com	A1.B1.C1.D1
emea-expe2.ent-pa.com	A2.B2.C2.D2

Each new Expressway location deployment will require a new CNAME record and as many A-records as the number of nodes in the Expressway cluster. In addition, an A-record for each individual Expressway-E node is also needed.

Business-to-Business Communications

Scalability for business-to-business communications can be addressed by adding multiple Expressway-C and Expressway-E clusters, either in the same physical location or geographically dispersed.

When multiple Expressway-C and Expressway-E pairs are deployed, Unified CM can direct an outbound call to the edge server that is nearest to the calling endpoint, thus minimizing internal WAN traffic. Additionally, when utilizing multiple edge clusters, the Expressway-Cs should form a meshed trunk configuration with the Unified CM clusters. This adds more scalability and resiliency by allowing additional outbound traversal paths if the geographically located traversal is full or not available.

For large deployments it might be preferable to host business-to-business communications on Expressway-C and Expressway-E pairs separate from mobile and remote access. This allows the server resources to be dedicated to external Internet communications.

Considerations for Inbound Calls

DNS SRV records are used to determine which Expressway-E clusters are authorized for the SIP and H.323 ent-pa.com domain. SRV records with the same weight and priority are used to balance calls across Expressway-E cluster nodes.

When scaling inbound calls across multiple geographically dispersed Expressway-E clusters, load balancing traffic becomes the primary consideration. Expressway-C and Expressway-E do not support load balancing of SIP or H.323 traffic. Therefore load balancing of the response to the DNS query becomes an important means of scaling the solution.

Much like with the mobile and remote access service, GeoDNS is used to direct different DNS responses to the same queries. Different metrics such as network latency and geographical location should be used to provide the correct Expressway-E cluster in the DNS response. Depending on the Internet service provider providing the GeoDNS service, status monitoring of the Expressway-E servers should also be included. This allows for a more efficient DNS response, for example, that does not include an out-of-service Expressway-E.

GeoDNS is a very good method of providing the best edge, Expressway-E, for the other server or endpoint to connect to, based on the metrics chosen by the customer. The response here is typically based on the edge physically closest to the server making the query. The mechanism is the same as described in the previous section, except that the SRV records are different. As an example, a SRV record for SIP TLS would be: _sips._tcp.ent-pa.com. Figure 4-20 can be used in order to set up GeoDNS service, where _collab-edge._tls.ent-pa.com is replaced by _sips._tcp.ent-pa.com

An alternative solution is designed to return the edge that is closest to the destination endpoint or device. This requires finding or knowing where the destination endpoint is located and then returning the appropriate edge. The benefit of this solution is to minimize the use of bandwidth on the customer network by delivering the shortest internal path to the endpoint.

This can be achieved by configuring Expressway-E to direct the call to the Expressway-E in another region if the called endpoint belongs to the other region.

As an example, consider two Expressway-C and Expressway-E clusters in EMEA, and another two Expressway-C and Expressway-E clusters in APJC. The Unified CM inbound calling search space on the Expressway-C trunk in EMEA will contain the partition of the EMEA phones but not the partition of the APJC phones. Analogously the inbound calling search space on the Expressway-C trunk in APJC will contain the partition of the APJC phones but not the partition of the EMEA phones. If a user on the Internet in EMEA calls a corporate endpoint in APJC, the call will be sent by DNS to the EMEA Expressway-E cluster, the default for business-to-business calls. The EMEA Expressway-E and Expressway-C will try to send the call to the destination, but the inbound calling search space of the

Expressway-C trunk will block the call. The EMEA Expressway-E will then forward the call to the APJC Expressway-E. This time the call will be delivered to the destination because the inbound calling search space of APJC Expressway-C contains the APJC endpoints partition.

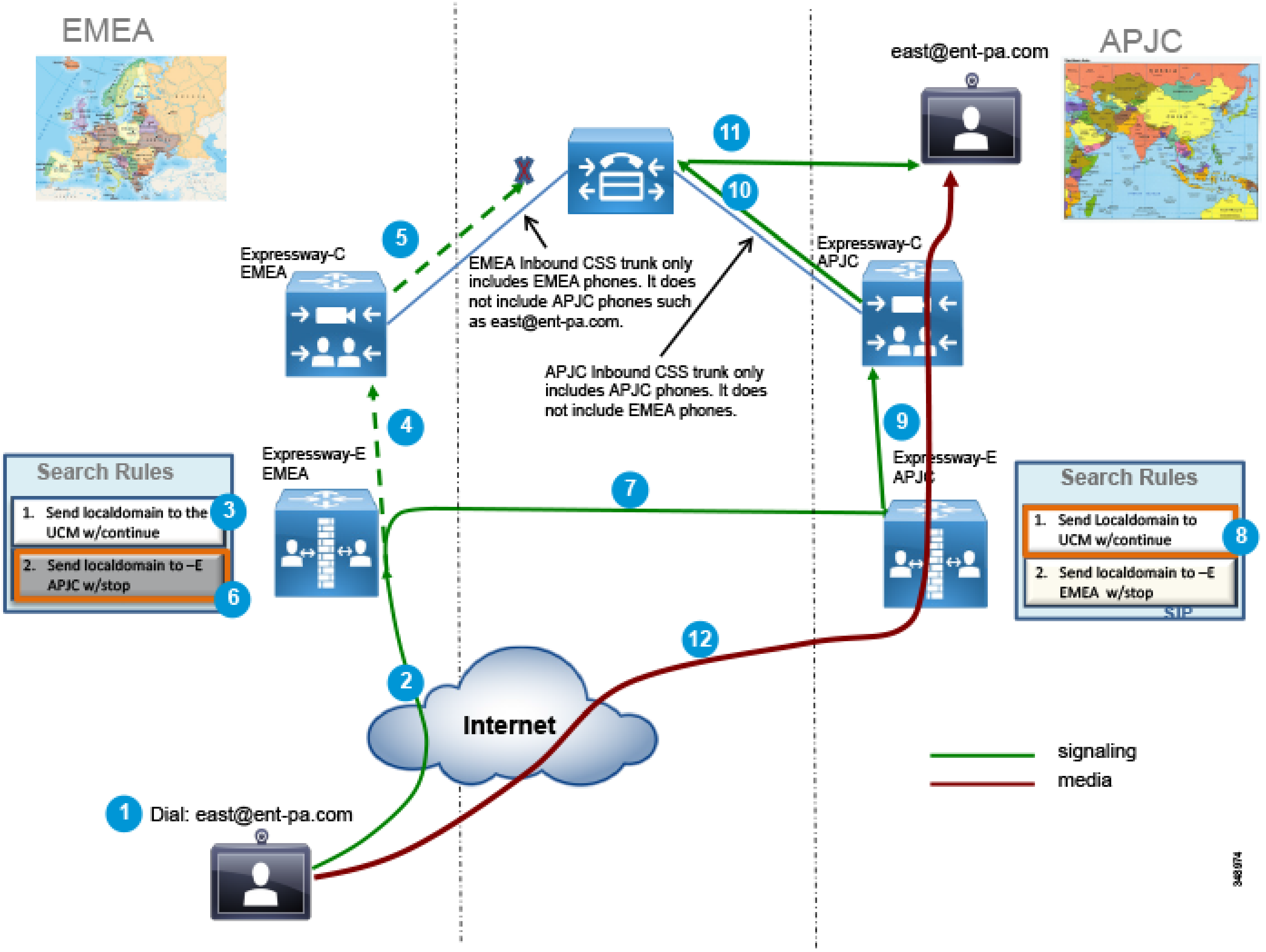
In order to allow the Expressway-E in EMEA to remove itself from the signaling and media path, it is important to make sure that there is no TCP-to-TLS or RTP-to-SRTP conversion on Expressway-E EMEA clusters, and to make sure that the call signaling optimization parameter is set to **on** in all Expressway-C and Expressway-E.

Because this is not a deterministic process, in the case of three or more Expressway edges the searching mechanism would require too much time. Therefore, this configuration is recommended for no more than two Expressway edges.

To scale to more that two edges, a different architecture called Directory Expressway can be deployed. Directory Expressway architecture is not part of the Preferred Architecture.

Figure 4-22 shows the Expressway edge design that enables selection of the edge closest to the destination endpoint.

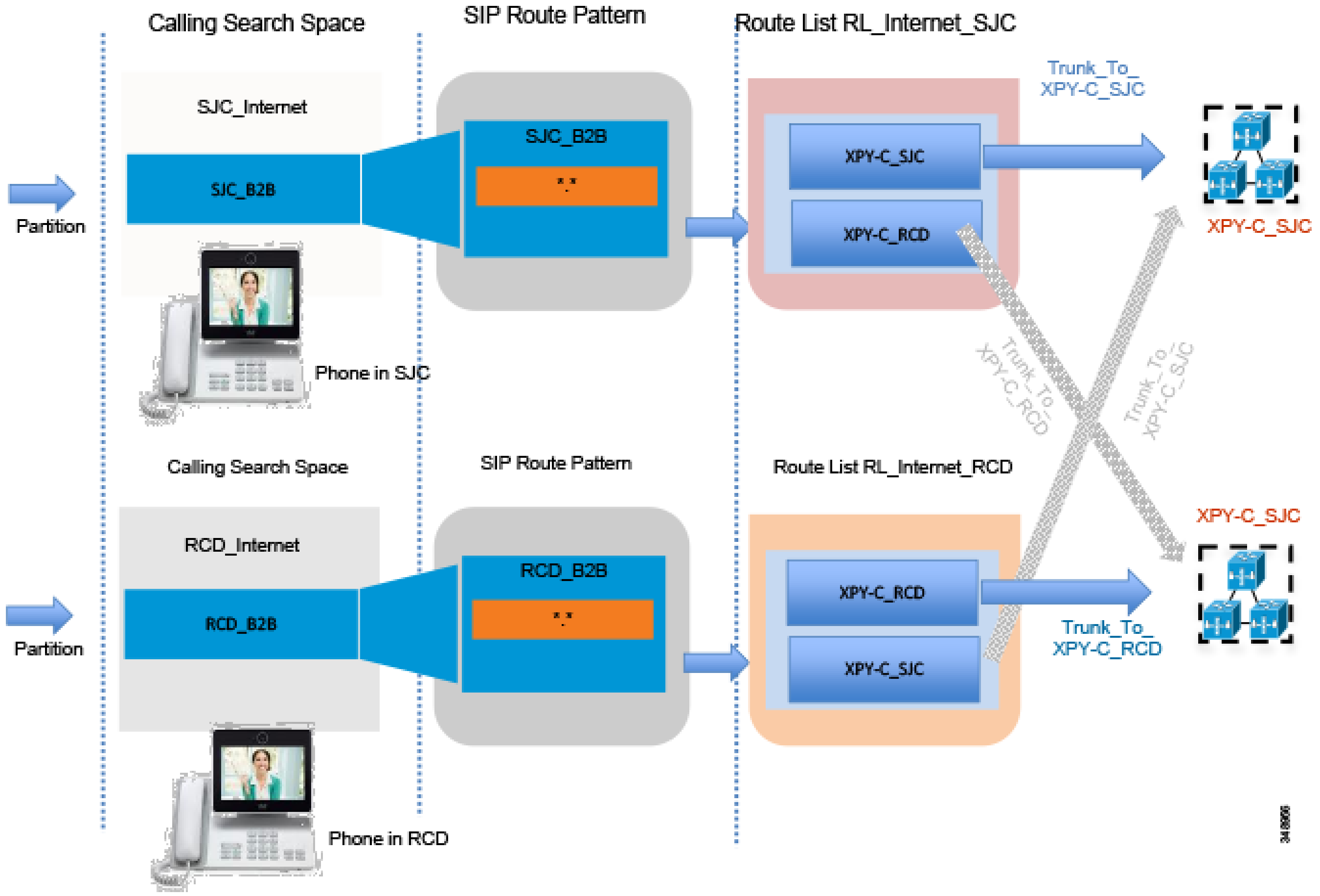
Figure 4-22 Selection of the Expressway Cluster Closest to the Destination


EMEA
APJC
east@ent-pa.com
11
10
5
Expressway-C
EMEA
EMEA Inbound CSS trunk only includes EMEA phones. It does not include APJC phones such as east@ent-pa.com.
Expressway-C
APJC
APJC Inbound CSS trunk only includes APJC phones. It does not include EMEA phones.
4
9
Search Rules
1. Send localdomain to the UCM w/continue
3
2. Send localdomain to –E APJC w/stop
6
Expressway-E
EMEA
7
Expressway-E
APJC
Search Rules
1. Send Localdomain to UCM w/continue
8
2. Send localdomain to –E EMEA w/stop
SIP
12
2
Internet
signaling
media
1
Dial: east@ent-pa.com
348974

Considerations for Outbound Calls

Outbound calls should be directed to the Expressway-C that is nearest to the calling endpoint. This can be achieved by using Cisco Unified CM mechanisms such as calling search spaces and partitions. Figure 4-23 shows the Unified CM configuration.

Figure 4-23 Partitions and Calling Search Spaces Configured in Unified CM



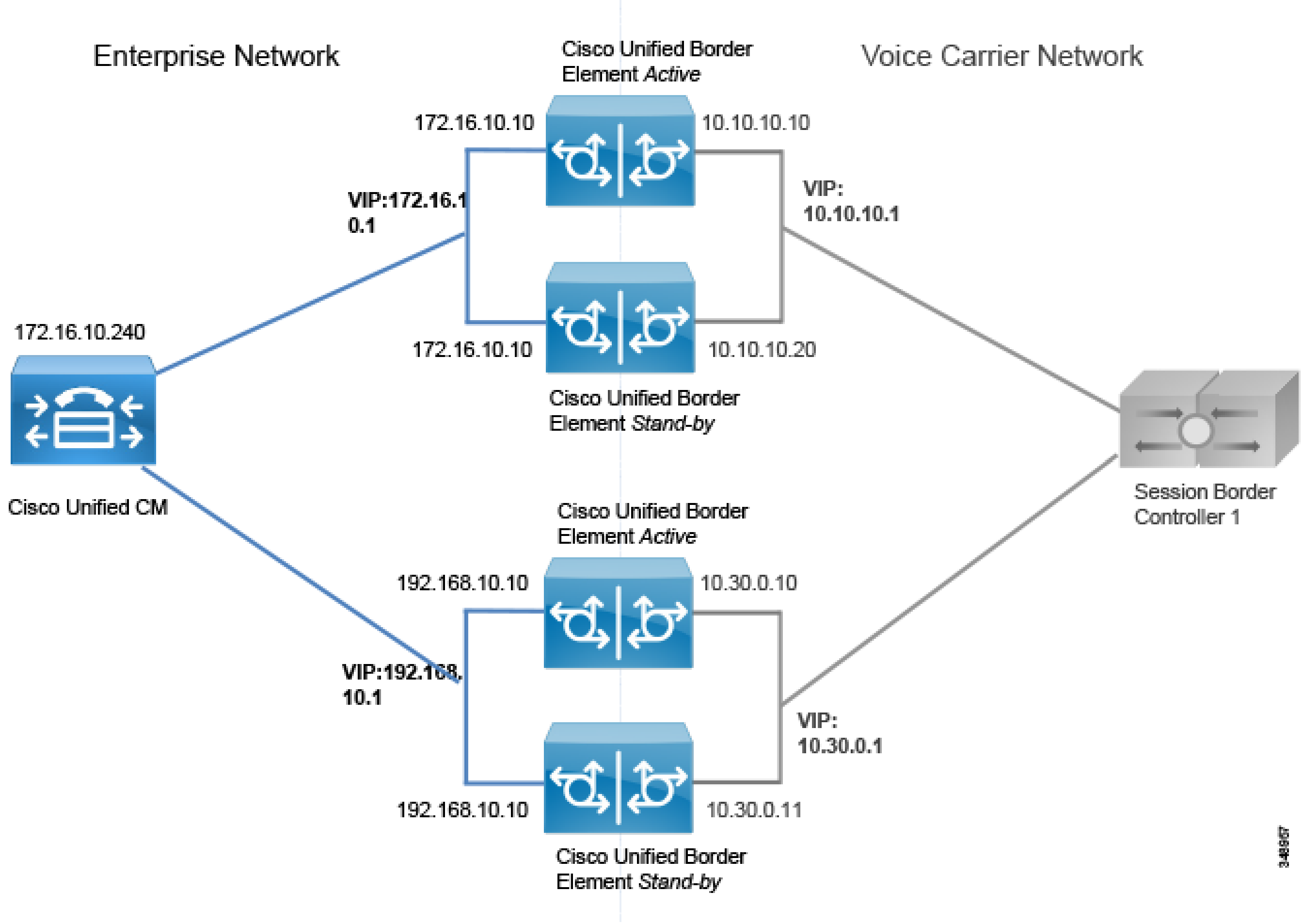
The Unified CM Local Route Group feature helps scale this solution when multiple sites access two or more Expressway-C clusters. This mechanism is also applied on ISDN gateways and Cisco Unified Border Element, and it is further described in the next section. A full description of the configuration is documented in the next two sections, since it also applies to Cisco Unified Border Element and voice gateways.

Scaling the Cisco Unified Border Element

For session capacities per platform, see the Sizing chapter.

If more than one datacenter is deployed, Cisco Unified Border Element can be deployed in each datacenter. This might happen for many reasons; for example, if it is required for a disaster recovery architecture, as shown in Figure 4-24.

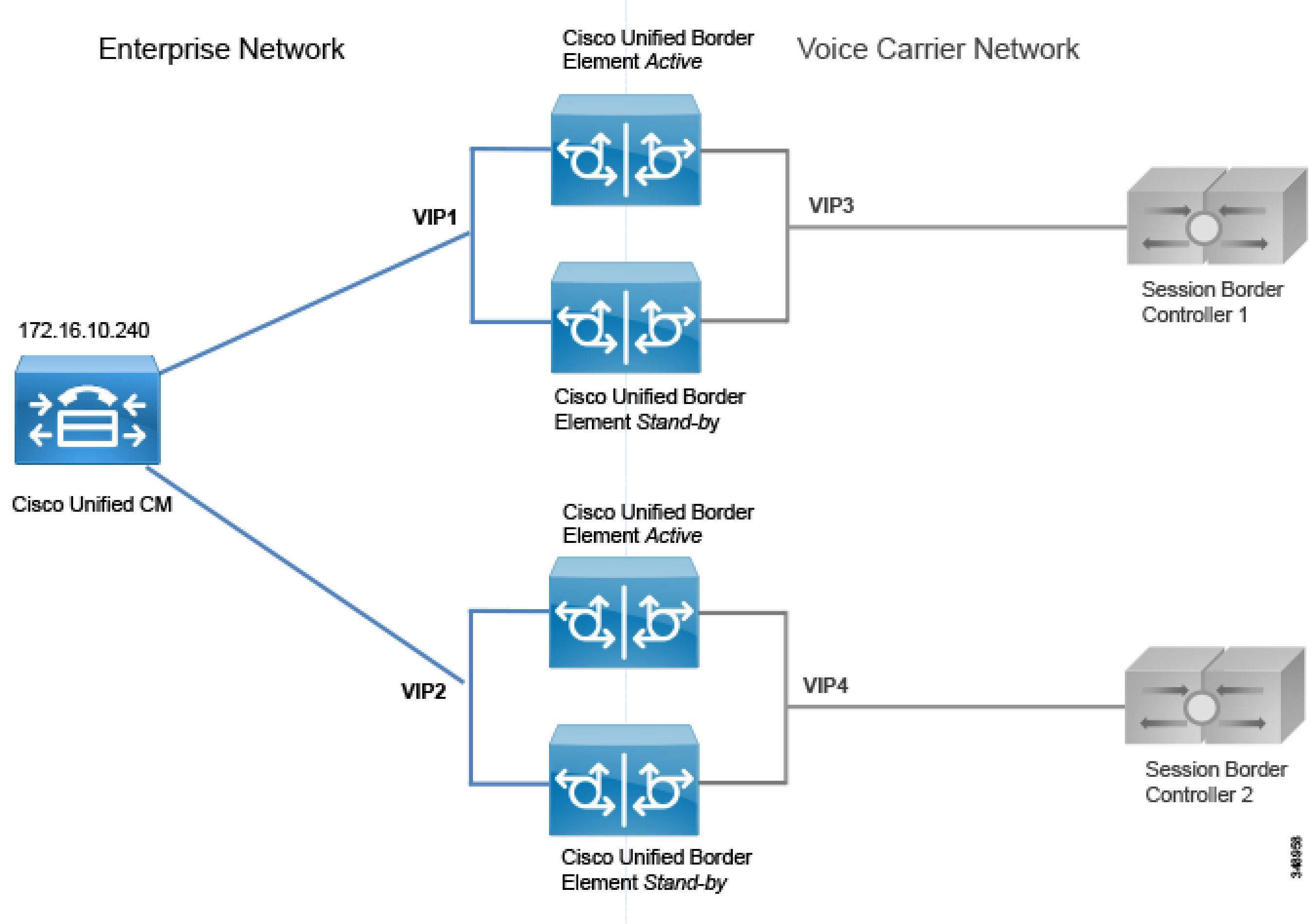
Figure 4-24 Multiple Cisco Unified Border Elements



All trunks to the Unified Border Element (usually two or three) can be inside the same route group. This would provide load balancing between datacenters. If the active router in the datacenter breaks, active calls will be preserved. If a datacenter becomes unreachable, call requests will be sent to the remaining datacenters. In this case, active calls would be dropped and users would have to reestablish them manually.

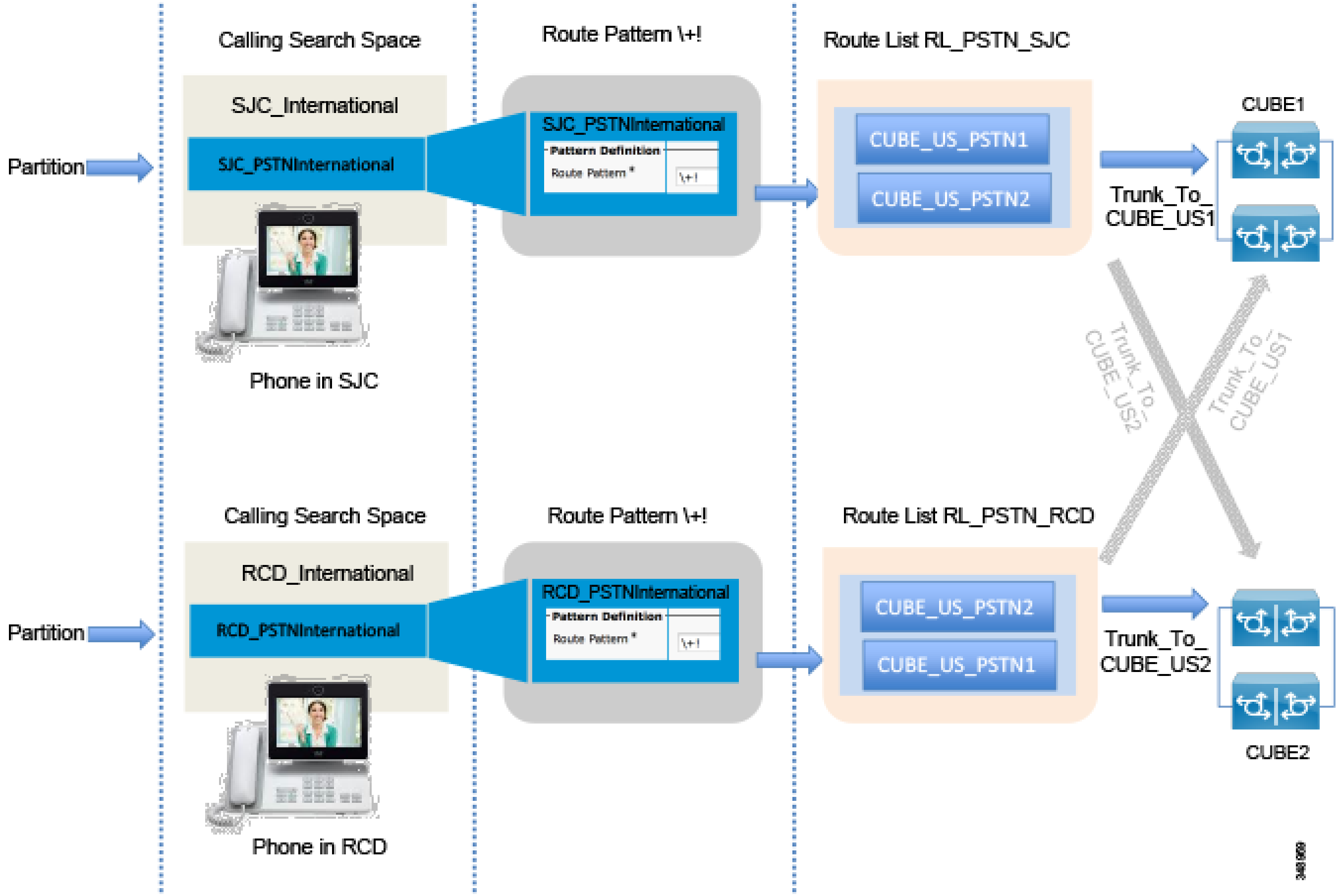
As shown in Figure 4-25, if the enterprise voice network is spread across a wide area, more than one session border controller (SBC) from the Telecom carrier is used. For each SBC, a Cisco Unified Border Element might be deployed, based on the carrier's recommendations.

Figure 4-25 Multiple Cisco Unified Border Element Connected to Different SBCs


Enterprise Network
Voice Carrier Network
Cisco Unified Border Element *Active*
VIP1
VIP3
Session Border Controller 1
172.16.10.240
Cisco Unified Border Element *Stand-by*
Cisco Unified CM
Cisco Unified Border Element *Active*
VIP2
VIP4
Session Border Controller 2
Cisco Unified Border Element *Stand-by*
348958

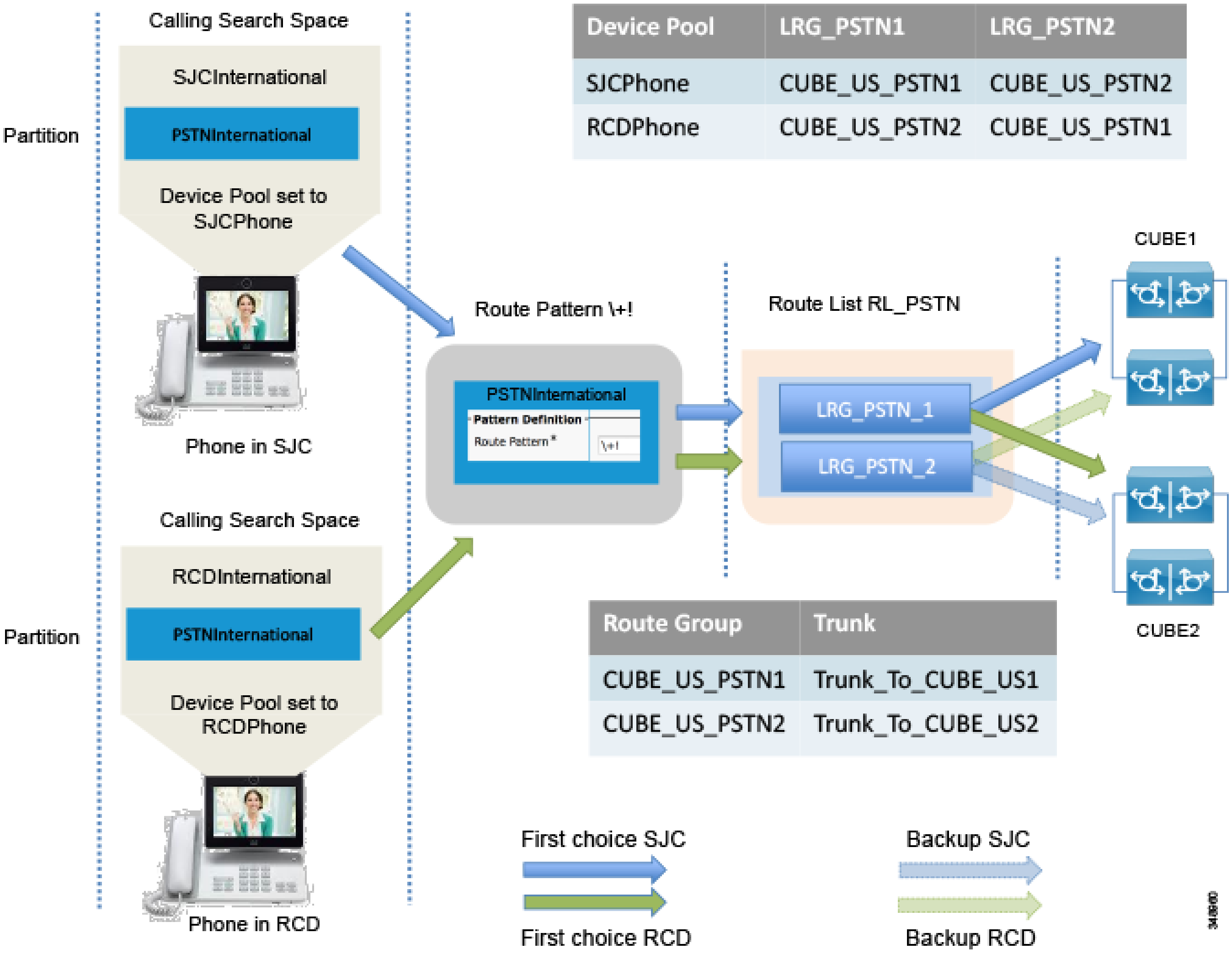
As an example, assume that another Unified Border Element is needed in the US besides the one already deployed. A new trunk called Trunk_to_CUBE_US2 is added. Figure 4-26 shows the configuration based on standard 1:1 mapping between calling search space and route pattern. This configuration has some limitations because, as the number of Unified Border Elements increases, it has a big impact on Unified CM resources. It is shown in Figure 4-26 in order to contrast this approach with the Local Route Group approach shown in Figure 4-27.

Figure 4-26 Unified CM Configuration for Cisco Unified Border Element Connection



The same route pattern, \+!, is repeated for every physical destination, and it resides in different partitions. The original partition PSTNInternational needs to be split into two, SJC_PSTNInternational and RCD_PSTNInternational, and the route pattern \+! has to be deleted and moved into the two newly created partitions. This approach works if the number of sites is not high, no more than two or three. A much better approach is to use the Local Route Group concept, as shown in Figure 4-27.

Figure 4-27 Unified CM Configuration for Cisco Unified Border Element Connection by Using the Local Route Group Approach



Device Pool	LRG_PSTN1	LRG_PSTN2
SJCPhone	CUBE_US_PSTN1	CUBE_US_PSTN2
RCDPhone	CUBE_US_PSTN2	CUBE_US_PSTN1

Route Group	Trunk
CUBE_US_PSTN1	Trunk_To_CUBE_US1
CUBE_US_PSTN2	Trunk_To_CUBE_US2

In this case, the device pool SCJPhone has LRG_PSTN1 set equal to the route group CUBE_US_PSTN1, while device pool RCDPhone has LRG_PSTN1 set equal to the route group CUBE_US_PSTN2. LRG_PSTN2 is set equal to CUBE_US_PSTN2 for SJC phones and equal to CUBE_US_PSTN1 for RCD phones. This approach is recommended because new partitions and route patterns are not required, and this approach is much more scalable than the approach shown in Figure 4-26.

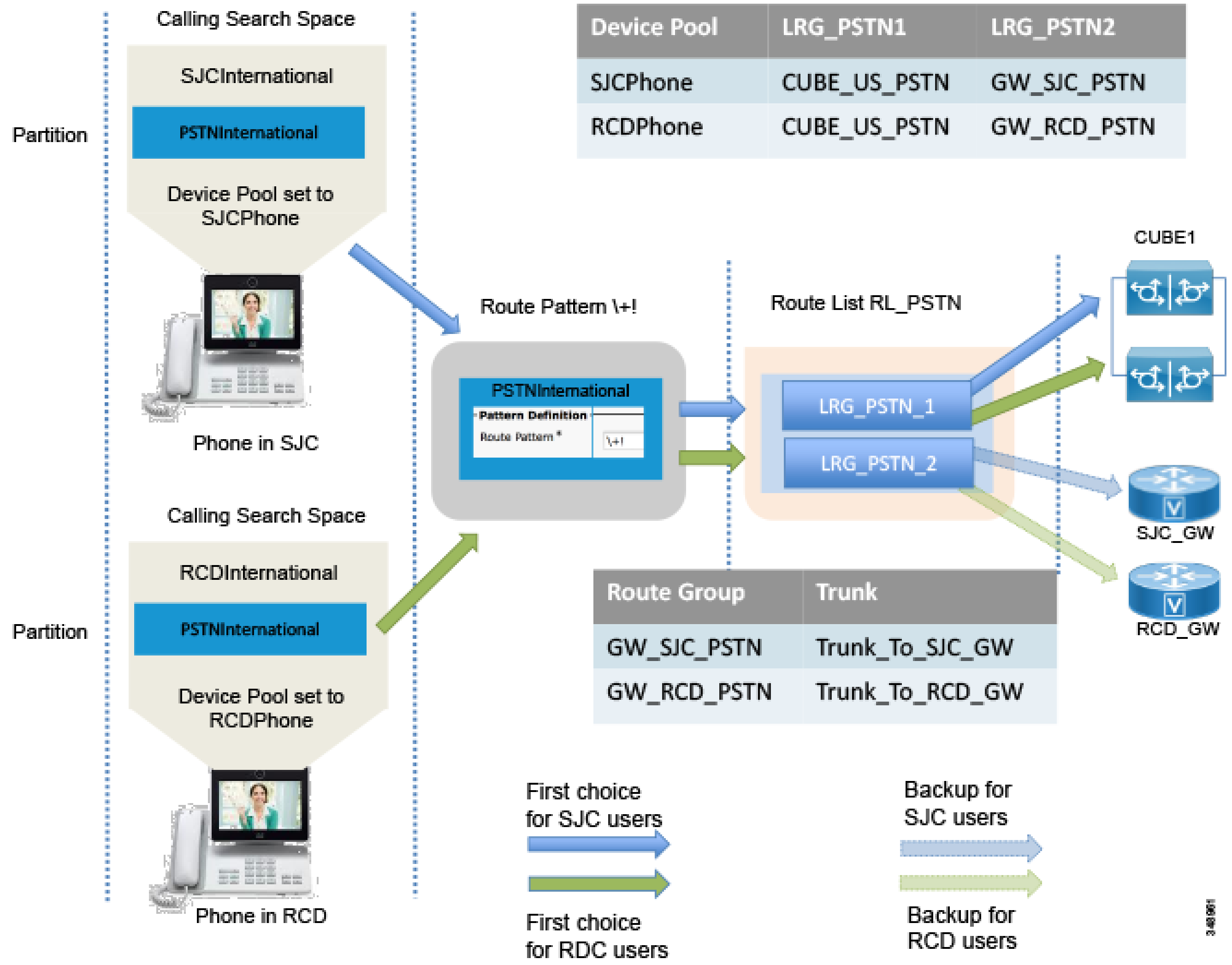
Scaling the PSTN Solution

Distributed gateways providing local PSTN access are deployed in branch offices and used as backup services.

If the number of branches is high, the route group and route list configuration construct within Unified CM does not scale well. For this deployment we recommend using the Local Route Group feature, so that route patterns to the PSTN do not have to be replicated for each site.

The configuration presented in the previous section is easily adapted to cover this scenario. What is needed is to assign the device profile LRG_PSTN1 to route group CUBE_US_PSTN, and to assign LRG_PSTN2 to the route group corresponding to the local gateway for that device pool, as shown in Figure 4-28.

Figure 4-28 Configuration for Centralized PSTN Access with Local ISDN Gateways



Device Pool	LRG_PSTN1	LRG_PSTN2
SJCPhone	CUBE_US_PSTN	GW_SJC_PSTN
RCDPhone	CUBE_US_PSTN	GW_RCD_PSTN

Route Group	Trunk
GW_SJC_PSTN	Trunk_To_SJC_GW
GW_RCD_PSTN	Trunk_To_RCD_GW

Scaling the Video ISDN Solution

Scaling the video ISDN solution is very similar to scaling the PSTN solution. If more than a few video ISDN gateways are required, using local route groups in Unified CM is the recommended method for routing calls out the PSTN. Having geographically dispersed ISDN gateways does aid in reducing toll charges for both inbound and outbound calls.

Collaboration Edge Deployment Process

This section summarizes the Collaboration Edge deployment process. Each component of Collaboration Edge is treated separately since each deployment may not require all access methods. As an example, a company might have only PSTN. Another company might use PSTN as a local backup for IP PSTN at specific local sites, have a Internet Edge deployment, and use ISDN video gateways to call those users who are not enabled for business-to-business Internet calls.

The Collaboration Edge components should be deployed in the following order:

- Deploy Expressway-C and Expressway-E
- Deploy Cisco Unified Border Element
- Deploy Cisco Voice Gateways
- Deploy Cisco ISDN Video Gateways

Deploy Expressway-C and Expressway-E

This section provides an overview of the tasks required to install and deploy Expressway-C and Expressway-E. The task should be performed in the following order:

1. Download and deploy Expressway-C and Expressway-E OVA templates and install the Expressway software. If the appliance model is used (Cisco Expressway CE500 or CE1000), there is no need to download and install OVA templates and the Expressway software.
2. Configure network interfaces and settings, including DNS and NTP, and system host name and domain name. Expressway-E has two LAN interfaces. If the external interface IP address is to be translated statically, the IP address of the translated interface has to be configured. Expressway-E will use the public IP address in payload references.
3. Configure clustering.

Deploy Mobile and Remote Access

1. Enable mobile and remote access by setting the Unified Communications mode to **Mobile and remote access**.
2. Select the domains for which mobile and remote access is enabled. Turn on **SIP registration and provisioning on Unified CM**, **IM and Presence service on Unified CM**, and **XMPP federation if inter-company federation**.
3. Upload the CA certificate to Expressway-C and Expressway-E. This is needed to discover Unified CM and IM and Presence clusters if **TLS verify mode** is **on** (recommended). This way Expressway-C will verify the identity of the cluster servers by checking the certificate.
4. Discover Unified CM and IM and Presence servers by configuring the publisher for each cluster.

5. Install certificates on both Expressway-C and Expressway-E. Both Expressway node types are able to generate a Certificate Signing Request (CSR) which is then signed by a CA. If an internal CA is use, the CSRs have to be signed by it. If Expressway-C and Expressway-E are also used for business-to-business communications, a public CA has to sign the certificate of Expressway-E, as mentioned previously. The signed certificates then need to be uploaded on Expressway-C and Expressway-E.
6. Configure a Unified Communication traversal zone between Expressway-C and Expressway-E, and allow for proxy registration to Cisco Unified CM.
7. To ensure that everything has been set up properly, check the Unified Communications status.

Note

- This configuration enables mobile and remote access. Business-to-business requires an additional configuration.
- The configuration above is done entirely on Expressway-C and Expressway-E.
- These steps are required for TCP/RTP connection to Unified CM (TLS/SRTP is not shown).

Deploy Business-to-Business Communications

This section provides an overview of the additional steps necessary to setup business-to-business communications.

1. Configure the basic Layer 3 configuration, including NTP, DNS, and system name, on both Expressway-C and Expressway-E.
2. Set up NAT configuration on Expressway-E, including IP routes necessary for routing traffic.
3. Ensure that the external firewall is set to block all traffic to Expressway-E before placing it in the DMZ.
4. Configure an administrative access policy, including local and/or remote authentication for both Expressway-C and Expressway-E.
5. Configure DNS A records in the appropriate DNS servers to be able to resolve the FQDN of each server.
6. Set up local authentication credentials in Expressway-E for the purpose of authenticating the traversal client connection coming from Expressway-C.
7. Set up the traversal server zone on Expressway-E for SIP only.
8. Set interworking on Expressway-E to **On**. This allows Expressway-E to send and receive H.323 calls and interwork them to SIP at the edge of the network, thus maintaining a single protocol inside the enterprise.
9. Set up the traversal client zone on Expressway-C for SIP only.
10. Use the FQDN of Expressway-E to enable the traversal link to allow for the possible use of PKI.
11. Configure the external DNS zone for outbound domain resolution for business-to-business communications.
12. Place basic CPL rules in place on Expressway-E to restrict access to protected resources such as video, voice, and IP PSTN gateways.
13. Set up domains for which Expressway-C and Expressway-E will have authorization.
14. Set up the dial plan on Expressway-C and Expressway-E with presearch transforms, search rules, DNS search rules, and external IP address routing.

15. Configure the SIP neighbor zone to Unified CM on Expressway-C.
16. Configure the SIP trunk on Unified CM to communicate with Expressway-C.

Deploy Cisco Unified Border Element

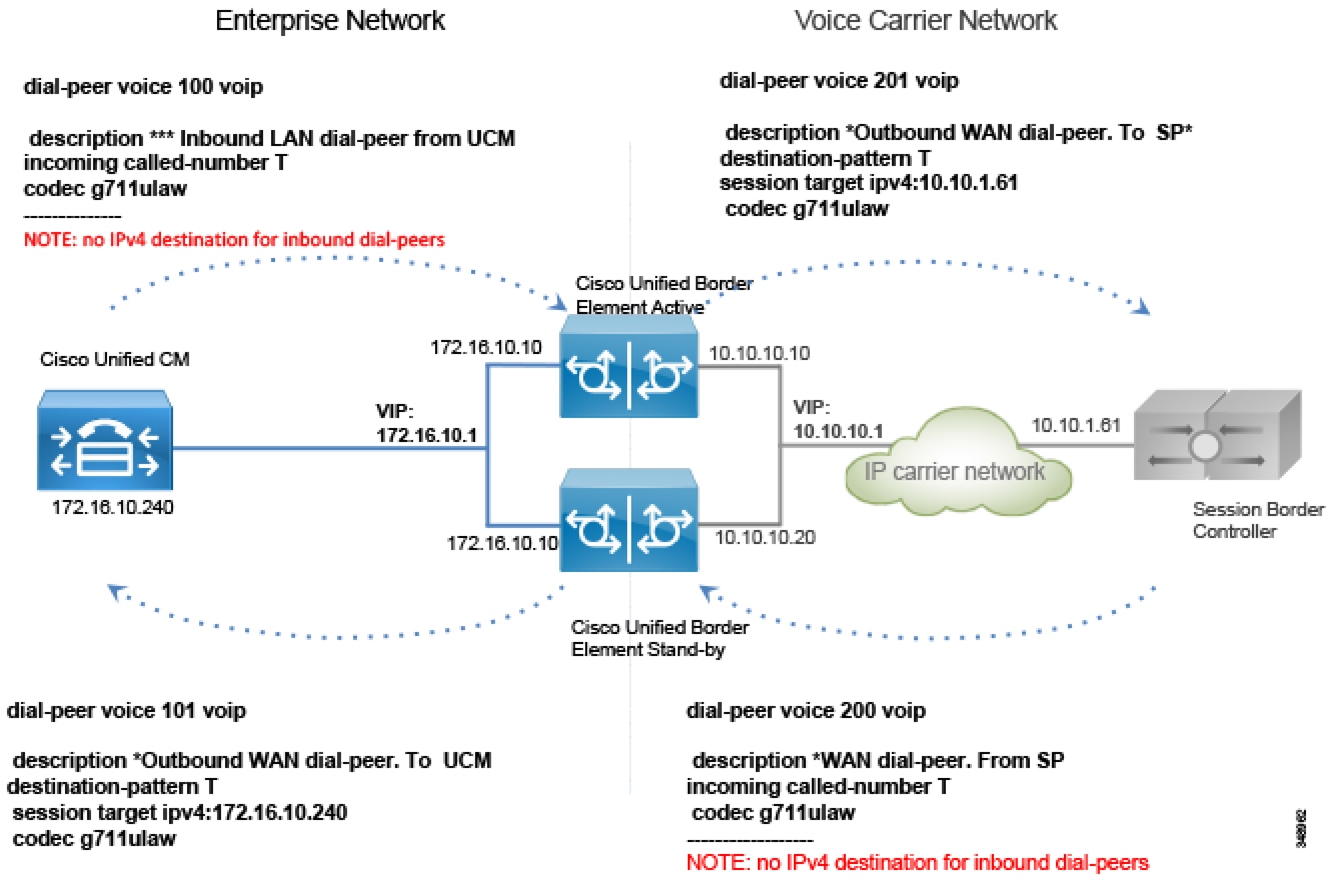
This section provides an overview of the process for deploying Cisco Unified Border Elements with box-to-box redundancy. Box-to-box redundancy has to be configured on both Unified Border Element routers, and the configuration is the same on both. It is possible to copy and paste the configuration from the active to the standby Unified Border Element.

1. Configure the network settings: the two Ethernet interfaces (one toward the LAN and the other facing the WAN) on both active and standby Unified Border Elements, as well as IP routing.
2. Enable the Unified Border Element on both routers for SIP-to-SIP calls, fax relay or passthrough, calling ID treatments as privacy headers, and enforcement of Early Offer. We recommend enabling this feature on Unified Border Element because Unified CM is configured for Best Effort Early Offer only. Although in new deployments only Early Offer will be sent from endpoints, there might be some cases involving old Cisco devices where Delayed Offer is sent instead. Even though these cases are not covered in this document, it is good practice to enforce Early Offer on Cisco Unified Border Element.
3. Enable box-to-box redundancy, and configure HSRP globally and on both the LAN and WAN interfaces for the active and standby routers.
4. Configure the voice codecs preference (in case the voice codec can be negotiated and it is not enforced by Unified CM or the Telecom carrier soft switch).
5. Configure music on hold.
6. Configure the dial-peers. Dial-peers are associated with call legs and can be matched inbound or outbound. As an example, an inbound call from Unified CM will be matched by an inbound dial-peer (corresponding to the inbound call leg). Another call leg will be generated by Cisco Unified Border Element (CUBE) toward the session border controller (SBC) of the Telecom carrier, and thus will be matched against another dial-peer. although the same dial-peer can match inbound or outbound calls, we recommend having each dial-peer match a specific call leg. Following this recommendation, there will be 4 distinct dial-peers: inbound dial-peer from Unified CM to CUBE, outbound dial-peer from CUBE to SBC, inbound dial-peer from SBC to CUBE, and outbound dial-peer from CUBE to Unified CM. Dial-peers can be matched against calling or called numbers or patterns. A dial-peer can force a single codec or can negotiate the list of codecs configured in step 4. The **incoming called-number** command makes a dial-peer inbound only.

   Inbound dial-peers do not have an associated target, while outbound dial-peers have Unified CM or the carrier's SBC defined as targets.

   Since calls to external destinations match generic patterns, dial-peer configuration on Unified Border Elements might lead to errors. As an example, in Figure 4-29 an outbound call matches both dial-peer 201 and 101, and therefore the routing does not work properly.

Figure 4-29 Inbound and Outbound Dial-Peer Configuration on Cisco Unified Border Element



```
dial-peer voice 100 voip

 description *** Inbound LAN dial-peer from UCM
incoming called-number T
codec g711ulaw
--------------
```

NOTE: no IPv4 destination for inbound dial-peers

```
dial-peer voice 201 voip

 description *Outbound WAN dial-peer. To  SP*
destination-pattern T
session target ipv4:10.10.1.61
 codec g711ulaw
```

```
dial-peer voice 101 voip

 description *Outbound WAN dial-peer. To  UCM
destination-pattern T
 session target ipv4:172.16.10.240
 codec g711ulaw
```

```
dial-peer voice 200 voip

 description *WAN dial-peer. From SP
incoming called-number T
 codec g711ulaw
-----------------
```

NOTE: no IPv4 destination for inbound dial-peers

The variable T in Figure 4-30 indicates any numeric string of any length, since calls from Unified CM might be sent to any destination in the world. A closest match might help, but when the Unified Border Element is centralized and provides the service for multiple locations, it might not be practical to list all the possible destinations in the "destination pattern" configuration. To overcome this limitation, and in order to simplify the routing process and make it more responsive, the following additional configuration is implemented:

a. Server groups in outbound dial-peers — If a server group is set as the destination in a dial-peer, and a round-robin algorithm is selected, the Unified Border Element will share the load between multiple servers:

```
voice class server-group 1
    ipv4 172.16.10.240
    ipv4 172.16.10.241
    ipv4 172.16.10.242
    ipv4 172.16.10.243
    ipv4 172.16.10.244
    hunt-scheme round-robin
```

b. SIP Out-of-Dialog OPTIONS Ping — It is possible to configure many parameters, such as the ping interval when a server is up and running, and the interval when it is down (set to 30 and 60 seconds respectively in this example):

```
voice class sip-options-keepalive 171
    transport tcp
    sip-profile 100
    down-interval 30
    up-interval 60
    retry 5
    description Target Unified CM
```

This way, the outbound dial-peer to Unified CM will be as follows:

```
dial-peer voice 101 voip
    description *Outbound WAN dial-peer. ToUnified CM
    destination-pattern T
    session protocol sipv2
    session server-group 1
    voice-class sip options-keepalive profile 171
    codec g711ulaw
```

c. The outbound call leg to the Telecom carrier will be matched by an outbound dial-peer:

```
dial-peer voice 201 voip
description *Outbound WAN dial-peer. To  SP*
destination-pattern T
session target ipv4:10.10.1.61
 codec g711ulaw
```

d. A leading "*" is sent by Unified CM on outbound calls (inbound calls from Unified Border Element’s perspective), which enables the router to distinguish the direction of the call. This character must be eliminated before the call goes out to the IP PSTN. Further, according to the configured dial plan, the calling number has to be normalized with the "+". Rule 2 prefixes a "+" and is applied to the calling number, while rule 1 replaces the leading "*" with "+". The rules are applied to the called number. Two rules might be created for this, one for the called number and one for the calling number. However, since the called number always matches the first rule, and the calling number always matches the second rule, it is possible to use a single voice translation rule. This is configured on the inbound dial-peer.

The outbound call leg (dial-peer) is bound to the inbound dial-peer via the **dpg** command, so that if any call is received with a leading "*", it is sent to the dial-peer facing the SBC and not to the one going to Cisco Unified CM:

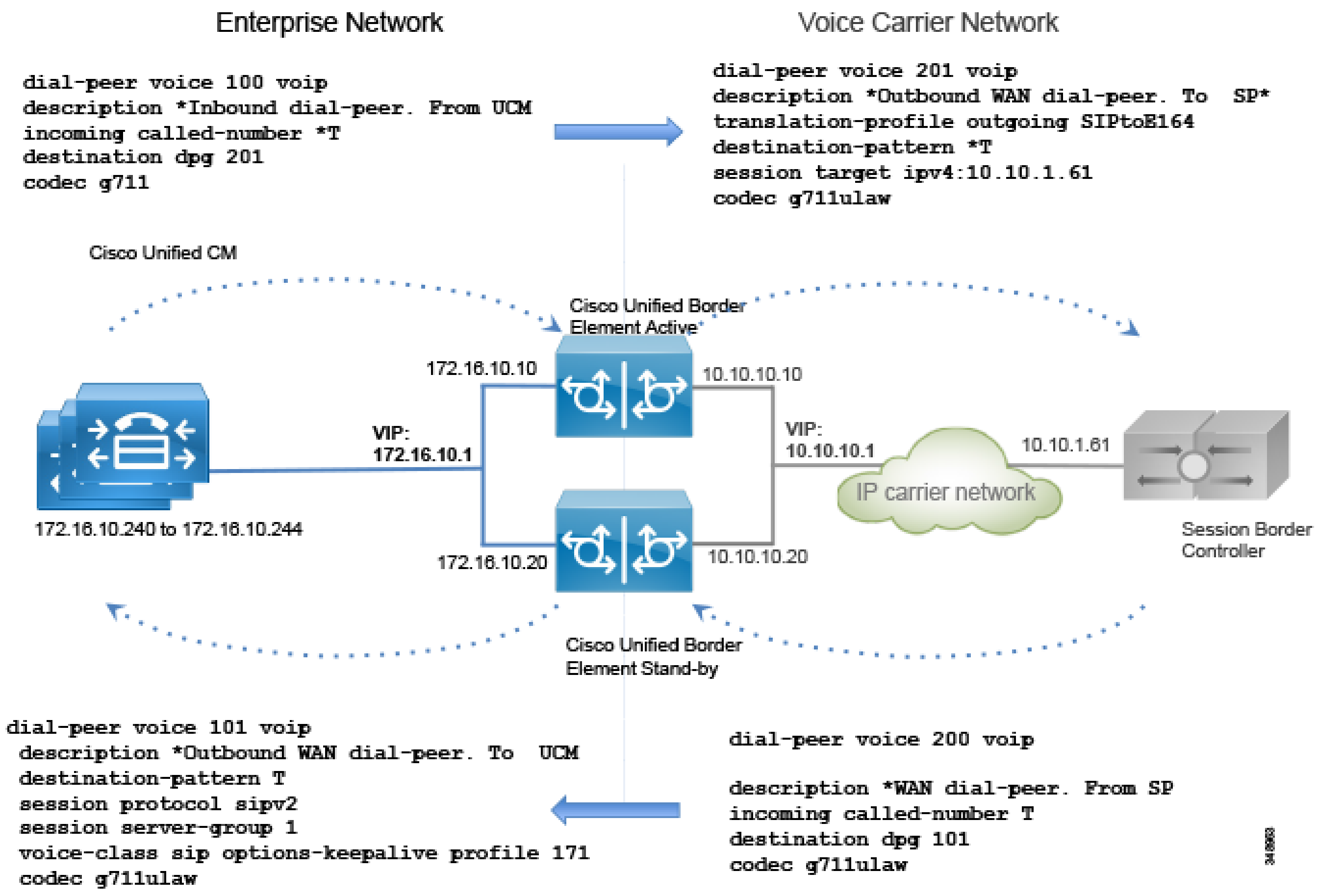
```
voice class dpg 201
    dial-peer 201

voice translation-rule 2
    rule 1 /^\*/ /+/
    rule 2 // /+/
voice translation-profile SIPtoE164
    translate called 2
    translate calling 2
dial-peer voice 100 voip
    translation-profile outgoing SIPtoE164
    incoming called-number *T
destination dpg 201
    codec g711
```

Dial-peer 200 needs also to be bound to dial-peer 101:

```
voice class dpg 101
    dial-peer 101

dial-peer voice 200 voip
    description *WAN dial-peer. From SP
    incoming called-number T
    destination dpg 101
    codec g711ulaw
```

Figure 4-30 shows this configuration.

Figure 4-30 Dial-Peer Configuration for Cisco Unified Border Element



```
dial-peer voice 100 voip
description *Inbound dial-peer. From UCM
incoming called-number *T
destination dpg 201
codec g711
```

```
dial-peer voice 201 voip
description *Outbound WAN dial-peer. To  SP*
translation-profile outgoing SIPtoE164
destination-pattern *T
session target ipv4:10.10.1.61
codec g711ulaw
```

```
dial-peer voice 101 voip
 description *Outbound WAN dial-peer. To  UCM
 destination-pattern T
 session protocol sipv2
 session server-group 1
 voice-class sip options-keepalive profile 171
 codec g711ulaw
```

```
dial-peer voice 200 voip

description *WAN dial-peer. From SP
incoming called-number T
destination dpg 101
codec g711ulaw
```

If the call leg comes from Unified CM, it will hit the Unified Border Element with a leading "*", thus matching dial-peer 100. The call is then sent to dial-peer 200 by using the outbound dial-peer group as an inbound dial-peer destination. Dial-peer 200 removes the leading "*" and sends the call to the PSTN. Note that without this feature, dial-peer 201 would also be matched, resulting in routing errors.

If the call leg comes from the SBC, it might match dial-peers 201, 101, and 200. But since the "incoming called-number" takes precedence over the "destination pattern," dial-peer 200 will be matched; and since dial-peer 200 is linked to the dial-peer 101, the call is correctly routed to the destination.

7. Configure transcoding if required. Remember that transcoding requires dedicated hardware resources (DSPs).

Perform the following configuration tasks on Unified CM:

1. Configure a Best Effort Early Offer trunk for each Unified Border Element, as specified in the Call Control chapter.
2. Configure route group CUBE_US_PSTN and add the Unified Border Element trunk as a member.
3. Configure local route group LRG_PSTN1.
4. Configure a route list that includes the default local route group and the route group LRG_PSTN1.
5. For each device pool set LRG_PSTN1 to CUBE_US_PSTN.

Deploy Cisco Voice Gateways

PSTN interfaces are available across a wide range of routers, such as Cisco ISR G2/G3 and ASR routers. PSTN interfaces include analog, BRI, and PRI ISDN voice cards. Analog interfaces are used mostly to connect fax machines and analog telephones.

Perform the following tasks to configure a PSTN gateway with ISDN voice interfaces:

1. Configure network settings and routing on the router.
2. Activate the ISDN interface.
3. Set the ISDN parameters for user side, switch-type, framing, and linecode, based on the Telecom carrier's requirements.
4. Configure the dial-peers.

The dial-peer logic is the same as for the IP PSTN and Unified Border Element, but in this case besides the "voip" dial-peers, a voice gateway also has "pots" dial-peers toward the PSTN.

If there are analog devices such as fax machines, they can be connected to the router through an analog interface.

If the router is used only for analog fax interconnection and with the PSTN interfaces attached to another router, T.38 fax-relay can be configured since it provides for better resiliency, especially if the path to the PSTN gateway traverses the WAN.

The dial-peer configuration is different from IP PSTN and the Unified Border Element configuration. Since the gateway is deployed within a specific location and serves phones for that location, the pattern destination is well known, as for example +14085554XXX.

On the other side, an incoming PSTN call has an address that is composed of plan, type, and number. Plan and type are not supported in SIP, and based on the Telecom carrier, the call can reach the gateway with a different plan and type. As an example, for a call to E.164 destination 4961007739764 on a trunk in Germany in the same area code 6100, the called party number in the outgoing ISDN SETUP message could be sent as (plan/type/number) ISDN/national/61007739764, ISDN/subscriber/7739764, or unknown/unknown/061007739764.

Based on the plan/type, the number changes, and thus the dial-peers might not match. For this reason it is necessary to force the plan/type to unknown/unknown. This way the full E164 number will be released to the destination. Dial-peer structure is described in detail in the Call Control chapter and is referenced here for consistency.

For outbound dial-peers, this rule transforms any calling party number to plan “unknown” and type “unknown”, and it transforms the called party number with the leading "*" to the +E.164 number.

```
voice translation-rule 1
        rule 1 /^\*/ // type any unknown plan any unknown
        rule 2 // // type any unknown plan any unknown
voice translation-profile ISDNunknown
        translate called 1
translate calling 1
dial-peer voice 1 pots
        translation-profile outgoing ISDNunknown
```

For inbound dial-peers, if the calling party information has a 10-digit number with type “national” (and does not include the country code "1" for US), the call will be transformed correctly to the +E.164 number, prepending "+1". If it is "unknown" the following rules will not be matched.

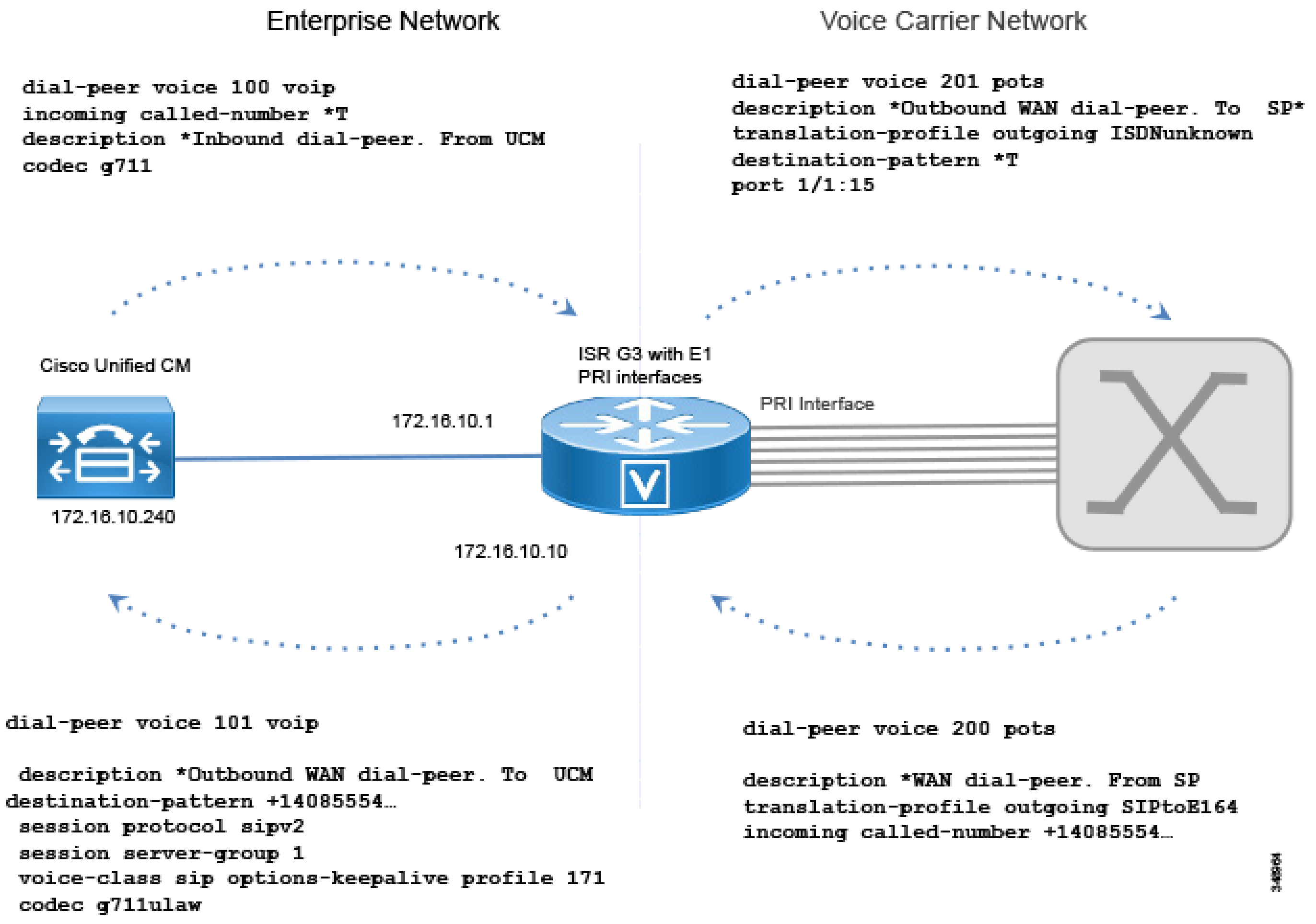
If the called number comes from an international destination, and thus it has the country code and is in the E.164 format, then rule 2 will add the leading "+".

However, since ISDN setup is hop-by-hop, we are not expecting to see many calls with type "national" since the latest switch might force it to "national". In any case, these rules normalize the calling and called party numbers correctly.

```
voice translation-rule 3
    rule 1 /^\(.+\)$/ /+1\1/ type national unknown plan any unknown
    rule 2 /^\(.+\)$/ /+\1/ type international unknown plan any unknown
voice translation-profile ISDNtoE164
    translate called 3
    translate calling 3

dial-peer voice 1 pots
    translation-profile incoming ISDNtoE164
```

Figure 4-31 shows a dial-peer configuration for G.711 and the E1 PRI interface.

Figure 4-31 Dial-Peer Configuration for Voice Gateways



```
dial-peer voice 100 voip
incoming called-number *T
description *Inbound dial-peer. From UCM
codec g711
```

```
dial-peer voice 201 pots
description *Outbound WAN dial-peer. To  SP*
translation-profile outgoing ISDNunknown
destination-pattern *T
port 1/1:15
```

```
dial-peer voice 101 voip

 description *Outbound WAN dial-peer. To  UCM
destination-pattern +14085554…
 session protocol sipv2
 session server-group 1
 voice-class sip options-keepalive profile 171
 codec g711ulaw
```

```
dial-peer voice 200 pots

description *WAN dial-peer. From SP
translation-profile outgoing SIPtoE164
incoming called-number +14085554…
```

Perform the following configuration tasks on Unified CM:

1. Configure a Best Effort Early Offer trunk for each gateway (Trunk_to_SiteID_GW, SiteID is a variable that identifies the location).
2. Configure route group LRG_PSTN1 and include the gateway trunk as member.
3. Configure local route group LRG_PSTN1.
4. Configure a route list that includes the default local route group and LRG_PSTN1.
5. For each device pool, set LRG_PSTN1 to Trunk_to_SiteID_GW. This configuration assumes, as recommended, that for each site there is a device pool SiteIDPhone.

By using the local route group configuration, it is easy to reconfigure PSTN access. As an example, it is possible to use the Unified Border Element for centralized access to the PSTN and to use the local PSTN connection as backup. In this case, the device pool would specify the Unified Border Element route group as LRG_PSTN1, and LRG_PSTN2 will include the trunk to the local gateway (Trunk_to_SiteID_GW).

Deploy Cisco ISDN Video Gateways

Deployment of a Cisco TelePresence ISDN GW 3241 or a Cisco TelePresence ISDN MSE 8321 is a fairly straightforward process:

1. Log into the web interface.
2. Allocate port licenses. Each port license activates a PRI interface. Port licenses are configured on the supervisor MSE 8050 for the 8321 ISDN gateway, and port licenses are configured on-box for the ISDN GW 3241.
3. Set up the ISDN interface. This is accomplished under **Settings** > **ISDN**. These settings are the typical settings received from the service provider for the type and form of ISDN being delivered.
4. Configure the ISDN ports. This is accomplished under **Settings** > **ISDN ports**. Directory number, channel range, and channel search order are all configured here. The **Enabled** box must be checked here for each port to enable the ISDN port for use.
5. Configure call control. This is accomplished under **Settings** > **SIP**. These are the SIP settings on the ISDN gateway that include the hostname and SIP domain used for Unified CM.
6. Configure the dial plan. This is accomplished on two tabs under the dial plan heading: **IP to ISDN** and **ISDN to IP**. Ensure that the incoming ISDN number range is translated correctly to the IP number range on the **ISDN to IP** dial plan tab.

For more information on the gateway installation and initial configuration, refer to the Cisco TelePresence ISDN Gateway installation and upgrade guides, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-isdn-gateway/products-installation-guides-list.html

CHAPTER 5

Core Applications

Revised: November 20, 2015

This chapter describes the core applications included in the Preferred Architecture for Enterprise Collaboration. While many additional applications from Cisco and our Ecosystem partners are available, this chapter focuses on a subset of core applications that are necessary for most collaboration environments. The Preferred Architecture is built with all of the available applications in mind to simplify the deployment of these applications and avoid unnecessary configuration changes.

The main section of this chapter explains how to implement Unified Messaging with Cisco Unity Connection. This section contains a description of the core architecture as well as details about the deployment process.

The second section of this chapter describes Tools for Application Deployment, namely: Cisco Prime Collaboration Deployment (PCD) and Cisco Prime License Manager (PLM). There is also a list of Additional Applications at the end of this chapter.

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:	Revision Date
TelePresence Management Suite (TMS) information has been moved to the Conferencing chapter.	Role of Cisco TMS, page 3-6 5. Deploy TelePresence Management Suite, page 3-38	November 20, 2015

Prerequisites

Before deploying the core applications for the Preferred Architecture, ensure that:

- Cisco Unified Communications Manager (Unified CM) is deployed and functioning.
- Microsoft Active Directory is installed, and the integration for each application is understood.
- The Call Control chapter of this document is understood and implemented.

Unified Messaging Application

A core application of the Preferred Architecture is Cisco Unity Connection, which provides unified messaging. (See the section on Unified Messaging with Cisco Unity Connection.)

Tools Used in a Collaboration Deployment

These software tools are useful to administrators in deploying the Enterprise Collaboration Preferred Architecture:

- Cisco Prime License Manager (PLM)
- Cisco Prime Collaboration Deployment (PCD)

Key Benefits

- Unified messaging available on multiple end-user platforms
- Eases deployment of new infrastructure components
- A single tool to manage licenses for various products

Unified Messaging with Cisco Unity Connection

Cisco Unity Connection enables unified messaging for the Cisco Preferred Architecture for Enterprise Collaboration. This section provides the information and instructions for deploying Unity Connection for voice messaging and unified messaging along with features such as single inbox and visual voicemail. This section also covers networking between two Unity Connection clusters.

Core Components

The core architecture contains these elements:

- Cisco Unified Communications Manager (Unified CM)
- Cisco Unity Connection
- Microsoft Exchange
- Microsoft Active Directory

Key Benefits

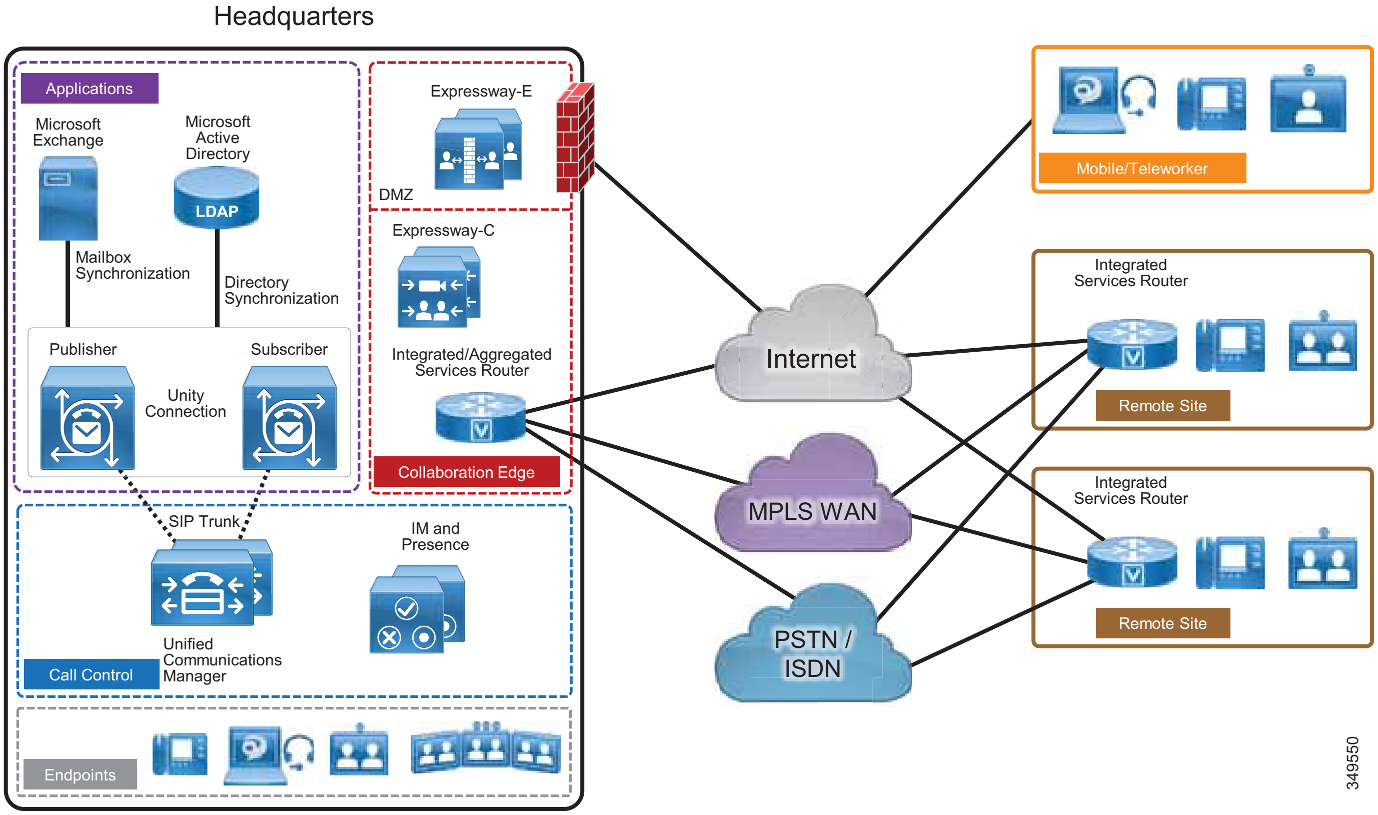
- Users can access the voicemail system and retrieve their voice messages by using:
  - Cisco Unified IP Phones, TelePresence endpoints, Jabber, and mobile devices
  - Web interface with PCs or Mac
  - Email client applications such as Microsoft Outlook
- Visual voicemail provides secure access to a visual display of voice messages on a Jabber client, listed with sender name, date, and message duration.

Core Architecture

Centralized Messaging and Centralized Call Processing

As shown in Figure 5-1, with centralized messaging Unity Connection is located in the same site as the Unified CM cluster. Remote branch sites located over the WAN from the central site rely on the centralized Unity Connection for unified messaging services. Unity Connection integrates with Unified CM using SIP for call control and RTP for the media path. Each Unity Connection cluster consists of two server nodes providing high availability and redundancy.

Figure 5-1 Architecture Overview



At the remote branch site, Cisco Unified Survivable Remote Site Telephony (SRST) is installed as a backup call agent, which is integrated with the central Unity Connection server. In the event of an IP WAN outage, all the phones at the remote branch register with SRST, which is preconfigured to send all the unanswered and busy calls to the central Unity Connection server via the PSTN.

Role of Unified CM

Unified CM provides call control capabilities and forwards calls to Unity Connection in the event that a called phone is either busy or unanswered. If a user presses the message button on the phone or dials the voicemail pilot number from an outside network, then Unified CM routes the call to Unity Connection.

Role of Unity Connection

In a centralized messaging deployment, Unity Connection provides users with the ability to store and retrieve voicemails. Typically calls forwarded to Unity Connection are direct calls or are due to a called extension that is either busy or unanswered. Message Waiting Indicator (MWI) is displayed on the endpoint for any new messages stored for the user. With each call, the following call information is typically passed between the phone system and Unity Connection:

- The extension of the called party
- The extension of the calling party (for internal calls) or the phone number of the calling party (if it is an external call and the phone system supports caller ID)
- The reason for the forward (the extension is busy, does not answer, or is set to forward all calls)

If the call is forwarded because the called party did not answer the call, Unity Connection plays the called user’s standard greeting. If the call was forwarded because the called phone was busy, Unity Connection plays the called user’s busy greeting.

Unity Connection handles direct calls differently than forwarded calls. When Unity Connection receives a call, it first attempts to determine whether the caller is a user. It does this by identifying whether the caller ID matches a user’s primary or alternate extension. If Unity Connection finds a match, it assumes that a user is calling and it asks for that user’s voicemail PIN. If Unity Connection determines that the caller ID is not associated with a user, then the call is sent to the opening greeting. An opening greeting is the main greeting that outside callers hear when they reach the Unity Connection auto-attendant.

Role of Microsoft Exchange

Unity Connection is integrated with Microsoft Exchange to enable the Single Inbox feature. Single Inbox in Unity Connection enables unified messaging and synchronizes voice messages between Unity Connection and Microsoft Exchange.This enables users to retrieve voicemail using their email client.

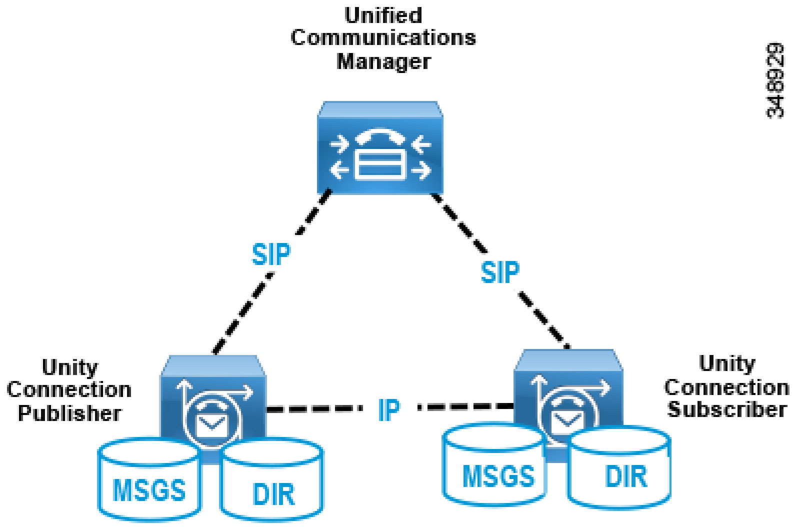
This chapter focuses on Unified Messaging with Microsoft Exchange. Unity Connection can also be integrated with IBM Lotus Sametime instant messaging application, allowing users to play their voice messages using Lotus Sametime. For more information on this topic, refer to the Unity Connection documentation available at

http://www.cisco.com/en/US/products/ps6509/index.html

High Availability for Unified Messaging

Figure 5-2 shows Unity Connection in an active/active pair, allowing the Unity Connection servers to be installed in the same or separate buildings to provide high availability and redundancy. Both servers in the active/active pair are running Unity Connection, both accept calls and HTTP requests, and both servers store user information and messages. In the event that only one server in the clustered pair is active, Unity Connection preserves the complete end-user functionality, including voice calls and HTTP requests. However, Unity Connection port capacity for calls will be reduced by half, to that of a single server.

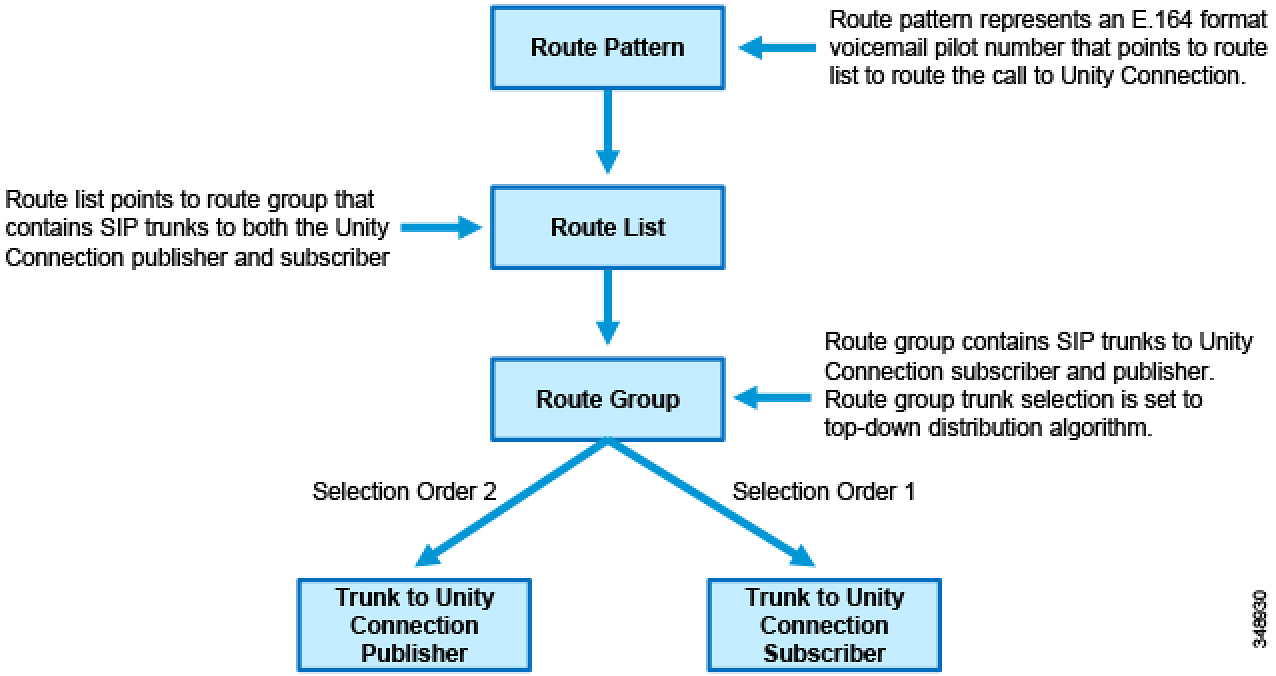
Figure 5-2 Unity Connection Cluster


Unified Communications Manager
348929
SIP
SIP
Unity Connection Publisher
IP
Unity Connection Subscriber
MSGS
DIR
MSGS
DIR

All user client and administrator sessions (for example, IMAP and Cisco Personal Communications Assistant) and administration traffic (for example, Cisco Unity Connection Administration, the Bulk Administration Tool, and backup operations) connect to the Unity Connection publisher server. If the publisher server stops functioning, the user client and administrator sessions can connect to the Unity Connection subscriber server.

This topology requires two separate Unified CM SIP trunks pointing to each Unity Connection server node in the cluster. This configuration provides both high availability and redundancy. Unified CM should be configured to route all calls to the Unity Connection subscriber node first. If the subscriber server is unavailable or all the ports of the subscriber are busy, then calls are routed to the publisher node. Given the SIP integration between Unified CM and Unity Connection, trunk selection is achieved via Unified CM route pattern, route list, and route group constructs (see Figure 5-3). Both trunks are part of the same route group and assigned to the same route list, and the trunks within the route group are ordered using a top-down trunk distribution algorithm. This approach allows Unified CM to control the preference of the Unity Connection server node selection during both normal and failover operation.

Figure 5-3 Unity Connection SIP Trunk Selection


Route Pattern
Route pattern represents an E.164 format voicemail pilot number that points to route list to route the call to Unity Connection.
Route list points to route group that contains SIP trunks to both the Unity Connection publisher and subscriber
Route List
Route Group
Route group contains SIP trunks to Unity Connection subscriber and publisher. Route group trunk selection is set to top-down distribution algorithm.
Selection Order 2
Selection Order 1
Trunk to Unity Connection Publisher
Trunk to Unity Connection Subscriber
348930

Unity Connection supports using Single Inbox with Microsoft Exchange 2010 or Exchange 2013 Database Availability Groups (DAGs) for high availability. The DAGs are deployed according to Microsoft recommendations. Unity Connection also supports connecting to a client access server (CAS) array for high availability. This section does not cover Microsoft Exchange high availability deployment. For more information about Exchange high availability deployments, refer to the Microsoft Exchange product information available at http://www.microsoft.com/.

Licensing Requirements

The licenses for Unity Connection are managed by the Cisco Prime License Manager (PLM). To use the licensed features on Unity Connection, the valid licenses for the features must be installed on the PLM server and Unity Connection must communicate with the PLM server to obtain the license. The PLM server provides centralized, simplified, and enterprise-wide management of user-based licensing.

Unified Messaging Requirements

- Unity Connection supports Microsoft Exchange 2003, 2007, 2010, and 2013 Server for Single Inbox. Unity Connection also supports interoperability with the Microsoft Business Productivity Online Suite (BPOS) Dedicated Services and Microsoft Office 365 cloud-based exchange server.
- Exchange servers and Active Directory domain controllers/global catalog servers (DC/GCs) can be installed in any hardware virtualization environment supported by Microsoft. Refer to Microsoft Exchange product information available at http://www.microsoft.com/ for more information about supported hardware platforms.
- The Microsoft Exchange message store can be stored in any storage area network configuration supported by Microsoft. Refer to Microsoft Exchange product information available at http://www.microsoft.com/ for more information about supported storage area network.
- For every 50 voice messaging ports on each server, 7 Mbps of bandwidth is required between Unity Connection and Microsoft Exchange for message synchronization.
- The default Unity Connection configuration is sufficient for a maximum of 2,000 users and 80 milliseconds of round-trip latency between Unity Connection and the Exchange servers. For more than 2,000 users and/or more than 80 milliseconds of latency, you can change the default configuration. For more information, see the information on latency in the *Design Guide for Cisco Unity Connection*, available at

  http://www.cisco.com/c/en/us/support/unified-communications/unity-connection/products-implementation-design-guides-list.html

Scaling Unity Connection

A Unity Connection cluster consists of a maximum of two nodes, one publisher and one subscriber in an active/active deployment. Under normal operation, call processing load balancing does not occur in an active/active deployment. Unified CM is configured to route all calls to the Unity Connection subscriber server first. If all ports are busy or if the subscriber server is unavailable, then calls are routed to the publisher. When sizing Unity Connection, consider the following parameters:

- Total number of current and future users.
- Required Voice messaging storage capacity.
- Number of voicemail ports supported with each platform.

For more information on Unity Connection scaling, see the Sizing chapter.

Cisco Unity Connection Deployment Process

Prerequisites

Before deploying the unified messaging architecture, ensure that:

- Unified CM is installed and configured for call control (see the Call Control chapter).
- Microsoft Exchange is installed and configured as an email server. For more information about supported exchange versions, refer to the section on Unified Messaging Requirements.

Deployment Overview

For the purposes of this Preferred Architecture, we assume a centralized messaging deployment model serving three sites in the US: SJC, RCD, and RTP. The deployment of centralized messaging starts with the Unity Connection cluster installation followed by further provisioning and configuration. To deploy centralized unified messaging with Cisco Unity Connection, perform the following tasks in the order listed here:

1. Provision the Unity Connection Cluster
2. Configure Unified CM for Unity Connection Integration
3. Unity Connection Base Configuration
4. Enable Single Inbox
5. Enable Visual Voicemail
6. Voice Mail in SRST Mode
7. HTTPS Internetworking of Two Unity Connection Clusters

Note Only non-default and other configuration field values are specified in this document. If a field configuration value is not mentioned, then the default value should be assumed.

1. Provision the Unity Connection Cluster

When clustering Unity Connection server nodes, one server is designated as the publisher server in the server pair while the other server is designated as the subscriber server.

Publisher

In Unity Connection only two servers are supported in a cluster for active/active high availability. The publisher server is the first to be installed, and it publishes the database and message store, replicating this information to the other subscriber server in the cluster.

Subscriber

Once the software is installed, the subscriber server node subscribes to the publisher to obtain a copy of the database and message store.

Unity Connection Mailbox Stores

During installation, Unity Connection automatically creates:

- A directory database for system configuration information (user data, templates, classes of service, and so forth).
- A mailbox store database for information on voice messages (who each message was sent to, when it was sent, the location of the WAV file on the hard disk, and so forth).
- An operating system directory for voice message WAV files.

Prerequisite for Unity Connection Cluster Deployment When the Servers Are to Be Installed in the Same Building

- For inbound and outbound calls to Unity Connection, the TCP and UDP ports of the firewall must be open as listed in the chapter on *IP Communications Required by Cisco Unity Connection* in the *Security Guide for Cisco Unity Connection*.
- For a cluster with two virtual machines, both must have the same virtual platform overlay.
- The servers must not be separated by a firewall.
- Both Unity Connection servers must be in the same time zone.
- Both Unity Connection server nodes must integrate to the same phone system.
- Both Unity Connection servers must have the same enabled features and configurations.

Prerequisite for Unity Connection Cluster Deployment When the Servers Are to Be Installed in Separate Buildings

- For inbound and outbound calls to Unity Connection, the TCP and UDP ports of the firewall must be open as listed in the chapter on *IP Communications Required by Cisco Unity Connection* in the *Security Guide for Cisco Unity Connection*.
- For a cluster with two virtual machines, both must have the same virtual platform overlay.
- Both Unity Connection server nodes must integrate to the same phone system.
- Both Unity Connection servers must have the same enabled features and configurations.
- Depending on the number of voice messaging ports on each Unity Connection server node, the connectivity between the server nodes must have the following guaranteed bandwidth with no steady-state congestion:
  - For every 50 voice messaging ports on each server, 7 Mbps of bandwidth is required.
  - Maximum round-trip latency must be no more than 150 milliseconds (ms).

To Deploy Unity Connection Cluster

- Determine which VMware Open Virtual Archive (OVA) template you want to deploy for the Unity Connection node based on the maximum number of ports and the maximum number of users. Refer the section on Scaling Unity Connection.
- Add both the Unity Connection nodes as host A records in the enterprise domain name service (DNS) server. For example, set the publisher Unity Connection hostname as US-CUC1.ent-pa.com and the subscriber hostname as US-CUC2.ent-pa.com.
- Determine the network parameters required for the installation:
  - Time zone for the server
  - Host name, IP address, network mask, and default gateway. Ensure that the hostname and IP address match the previous DNS configuration.
  - DNS IP addresses
  - Network Time Protocol (NTP) server IP addresses
- Download the above noted OVA file from the Cisco website.
- Deploy the Unity Connection publisher server node using the VMware vSphere Client.
- After installing the Unity Connection publisher, add the subscriber details in the cluster configuration of the primary server.
- Deploy the Unity Connection subscriber server node using the VMware vSphere Client.

Note Optionally, the Unity Connection cluster can be deployed automatically using Cisco Prime Collaboration Deployment. For details, see the section on Cisco Prime Collaboration Deployment (PCD).

2. Configure Unified CM for Unity Connection Integration

Before Unity Connection communicates with Unified CM, certain tasks must be performed on Unified CM. Unity Connection communicates to Unified CM over a SIP trunk. This section provides an overview of the tasks required to integrate Unified CM with Unity Connection.

SIP Trunk Security Profile

As far as media and signaling encryption is concerned, this guide assumes they are not used and instead non-secure SIP trunks are implemented between Unified CM and Unity Connection server nodes. Create a new SIP Trunk Security Profile for Unity Connection with device security mode set to **Non Secure**. Table 5-2 lists the SIP trunk security profile settings.

Table 5-2 SIP Trunk Security Profile Settings

Parameter	Value	Comments
Name	Unity Connection SIP Trunk Security Profile	Enter the name of the security profile.
Description	Unity Connection SIP Trunk Security Profile	Enter the description for profile.
Device Security Mode	Non Secure	Security mode for SIP trunk.
Accept out-of-dialog refer	Checked	Ensures that Unified CM accepts incoming non-INVITE, out-of-dialog refer messages that come via the SIP trunk.
Accept unsolicited notification	Checked	Ensures that Unified CM accepts incoming non-INVITE, unsolicited notification messages that come via the SIP trunk. This parameter must be checked to accept MWI messages from Unity Connection.
Accept replaces header	Checked	Ensures that Unified CM accepts new SIP dialogs, which replace existing SIP dialogs. This allows "REFER w/replaces" to be passed, which is used for Cisco Unity Connection initiated supervised transfers.

SIP Profile

Configure a SIP profile for the SIP trunk to Unity Connection. Copy the standard SIP profile and rename it to **Unity Connection SIP Profile**. Select the checkbox **Use Fully Qualified Domain Name in SIP Requests** to prevent the IP address of the Unified CM server from showing up in SIP calling party information sent by Unified CM. Ensure that the checkbox **Enable OPTIONS Ping to monitor destination status for Trunks with Service Type "None (Default)"** is checked so that the system tracks the status of connectivity to the Unity Connection node.

When the OPTIONS Ping is enabled, each node running the trunk's SIP daemon will periodically send an OPTIONS Request to each of the trunk's destination IP addresses to determine its reachability and will send calls only to reachable nodes. A destination address is considered to be "out of service" if it fails to respond to an OPTIONS Request, if it sends a Service Unavailable (503) response or Request Timeout (408) response, or if a TCP connection cannot be established. The overall trunk state is considered to be "in service" when at least one node receives a response (other than a 408 or 503) from a least one destination address. SIP trunk nodes can send OPTIONS Requests to the trunk's configured destination IP addresses or to the resolved IP addresses of the trunk's DNS SRV entry. Enabling SIP OPTIONS Ping is recommended for all SIP trunks because it allows Unified CM to track the trunk state dynamically rather than determining trunk destination state on a per-node, per-call, and time-out basis.

SIP Trunk

Create two separate SIP trunks, one for each Unity Connection server node in the cluster. Table 5-3 lists the SIP trunk settings.

Table 5-3 Parameter Settings for SIP Trunk to Unity Connection Server

Parameter	Value	Description
Name	US_CUC1_SIP_Trunk	Enter the unique name for SIP trunk to Unity Connection.
Description	Unity Connection Publisher	Enter the description for the SIP trunk.
Device Pool	Trunks_and_Apps	Enter the device pool for Unity Connection. (See the Call Control chapter.)
Run On All Active Unified CM Nodes	Checked	This ensures that outbound calls using the SIP trunk do not require intra-cluster control signaling between Unified CM call processing subscribers.
Call Routing Information – Inbound Calls		
Calling Search Space (CSS)	VoiceMail (Refer to the Call Control chapter for more about CSS configuration.)	CSS assigned contains all the on-net destinations such as DIDs, non-DID numbers, and URI partitions. If the CSS does not include all these partitions, then the MWI Unsolicited Notify messages from Unity Connection will not reach user phones.
Redirecting Diversion Header Delivery - Inbound	Checked	This ensures that the redirecting Information Element, the first redirecting number, and the call forward reason are sent and accepted as a part of incoming messages. Unity Connection uses the first redirecting number to answer the call.
Call Routing Information – Outbound Calls		
Calling and Connected Party Info Format	Deliver URI and DN in connected party, if available	This option determines whether Unified CM inserts a directory number, a directory URI, or a blended address that includes both the directory number and directory URI, in the SIP identity headers for outgoing SIP messages.

Table 5-3 Parameter Settings for SIP Trunk to Unity Connection Server (continued)

Parameter	Value	Description
Redirecting Diversion Header Delivery - Outbound	Checked	This ensures that the redirecting Information Element, the first redirecting number, and the call forward reason are sent and accepted as a part of outgoing messages. Unity Connection uses the first redirecting number to answer the call.
SIP Destination Information		
Destination Address	10.195.100.20	Enter the IP address of Unity Connection server.
SIP Trunk Security Profile	Unity Connection SIP Trunk Security Profile	See Table 5-2.
SIP Profile	Unity Connection SIP Profile	See the SIP Profile section.

Route Group

Create a separate route group RG_CUC for the Unity Connection cluster. The route group contains the SIP trunks to the Unity Connection subscriber and publisher nodes. Ensure that the SIP trunk that connects to the subscriber node (US_CUC2_SIP_Trunk) appears first in the list, followed by the publisher node (US_CUC1_SIP_Trunk). The route group distribution algorithm should be set to the **Top Down** trunk selection method. A route group configured with the **Top Down** distribution algorithm ensures that the calls are always sent to the Unity Connection subscriber server node (US-CUC2) first. If the Unity Connection subscriber server node is busy or unavailable, then the calls are sent to the publisher server node (US-CUC1).

Route List

Create a separate route list RL_CUC for the Unity Connection cluster. The route list should contain only the Unity Connection route group (RG_CUC) created previously. Ensure that the options **Enable this Route List** and **Run on all Active Unified CM Nodes** are selected.

Route Pattern

Create a separate route pattern for the voicemail pilot number pointing to the Unity Connection route list created above. This number must match the voicemail pilot number. Table 5-4 shows the route pattern configuration example.

Table 5-4 Unity Connection Pilot Number-Route Pattern Example

Parameter	Value
Route Pattern	+14085554999
Route Partition	DN
Gateway/Route List	RL_CUC
Call Classification	OnNet
Provide Outside Dial Tone	Unchecked

Voice Mail Pilot

The voicemail pilot number designates the directory number that users dial to access voice messages. Unified CM automatically dials the voicemail pilot number when a user presses the Messages button on an IP endpoint. A single voicemail pilot number is created for all three sites. Table 5-5 shows the voicemail pilot configuration example.

Table 5-5 Voicemail Pilot Example

Parameter	Value
Voice Mail Pilot number	+14085554999
Calling Search Space	DN
Description	VM Pilot
Make this the default Voice Mail Pilot for the system	Checked

Voicemail users located at remote sites can check their messages from the PSTN by dialing the voicemail access number from their own DID range. A separate translation pattern is created to translate the voicemail PSTN access number to the voicemail pilot number. Table 6 shows the translation pattern configuration for the voicemail pilot.

Table 5-6 Voicemail Pilot Translation Pattern Example

Parameter	Value
Translation Pattern	+19195551999
Partition	DN
Use Originators Calling Search Space	Checked
Route Option	Route this pattern
Called Party Transformations	
Called Party Transform Mask	+14085554999

Additional translation patterns would be created for other remote sites.

Voicemail Profile

A voicemail profile is assigned to each user's phone line on all endpoint devices and Extension Mobility profiles. The profile enables users to press the Messages button on an endpoint for one-touch access to the voicemail system. If Unity Connection is integrated with a single phone system, we recommend using the default voicemail profile. During the initial provisioning of a line on an endpoint device, the default voicemail profile (None) is assigned to the directory number. For the users who do not require voicemail access, no voicemail profile is assigned to their endpoint lines. Table 5-7 shows the settings for the voicemail profile configuration example.

Table 5-7 Voicemail Profile Example

Parameter	Value
Voice Mail Profile Name	Default
Description	VM Profile
Voice Mail Pilot	+14085554999/DN
Voice Mail Mask	Blank
Make this the default Voice Mail Profile for the System	Checked

3. Unity Connection Base Configuration

Service Activation

- After Unity Connection installation is complete, login to Cisco Unified Serviceability and activate the **DirSync** service on the publisher server node.
- Under Unified Serviceability, **Navigate** to **Tools** –> **Control Centre-Feature Services**. Verify that the Cisco DirSync service is started on publisher server node.
- Under Unity Connection Serviceability, **Navigate** to **Tools** –> **Service Management**. Verify the status of services on the publisher and subscriber Unity Connection server nodes. Table 5-8 shows the services status for this deployment.

Table 5-8 Unity Connection Services Status

Services	Unity Connection Publisher (Primary)	Unity Connection Subscriber (Secondary)
Status Only Services (Can be deactivated from OS command line interface)		
All the Services in this category	Yes	Yes
Critical Services		
Connection Conversation Manager	Yes	Yes
Connection Mailbox Sync	Yes	No
Connection Message Transfer Agent	Yes	No
Connection Mixer	Yes	Yes
Connection Notifier	Yes	No
Base Services		
All the Services in this category	Yes	Yes
Optional Services		
Connection Branch Sync Service	No	No
Connection Digital Networking Replication Agent	No	No
All other remaining services in this category (including Connection Jetty and Connection REST Service)	Yes	Yes

Database Replication

After activating services on both publisher and subscriber Unity Connection server nodes, confirm that the subscriber node can connect to the publisher node. Also check the database replication status using the OS Command line interface (CLI) command **show perf query class "Number of Replicates Created and State of Replication"** on both the nodes

Unified CM Integration

Each Unity Connection cluster is integrated with the co-located Unified CM cluster. This provides a simple integration model with each Unity Connection cluster dedicated to a Unified CM cluster. While SIP trunks are configured on the Unified CM for interconnectivity into the Unity Connection cluster, voicemail ports are used for capacity and licensing purposes on the Unity Connection system. This section discusses design considerations, capacity planning, and configuration settings of the voicemail ports.

Voicemail Port Audio Codec Configuration

In Unity Connection, a call in any audio codec format that is supported by Unity Connection SIP signaling will always be transcoded to PCM linear. From PCM linear, the recording is encoded in the system-level recording audio codec system-wide setting in Unity Connection Administration. G.711 mu-law is the default.

In this section, we refer to the audio codec that is negotiated between the calling device and Unity Connection as the *line codec*, and the audio codec that is set as the system-level recording audio codec as the *recording codec*.

Supported line codecs (advertised codecs):

- G.711 mu-law
- G.711 a-law
- G.722
- G.729
- iLBC

Supported recording codecs (system-level recording audio codecs):

- PCM linear
- G.711 mu-law (default)
- G.711 a-law
- G.729a
- G.726
- GSM 6.10

Because transcoding is inherent in every connection, there is little difference in system impact when the line codec differs from the recording codec. For example, using G.729a as the line codec and G.711 mu-law as the recording codec does not place a significant additional load on the Unity Connection server for transcoding. However, the iLBC or G.722 codecs require more computation to transcode, and therefore they place a significant additional load on the Unity Connection server. Consequently, a Unity Connection server can support only half as many G.722 or iLBC connections as it can G.711 mu-law connections.

For this example topology, the system recording codec is left at default (G.711 mu-law). The supported line codes are set to G.729 and G.711 mu-law. Using this default configuration, the users located at the same site of Unity Connection will use G711 mu-law. For the users located over the WAN from the centralized Unity Connection servers, the selected line codec will be G.729.

Use of the G.722 or iLBC codec as line codecs or advertised codecs reduces the number of voice ports that can be provisioned on the Cisco Unity Connection server. For more information on the number of voice ports supported for each platform overlay when using G.722 or iLBC codecs, refer to the documentation on Virtualization for Cisco Unity Connection.

Phone System Settings

Phone system integration enables communication between Unity Connection and Unified CM. We recommend using default **PhoneSystem** if Unity Connection is integrated with single Unified CM cluster. Table 5-9 shows the Phone System settings.

Table 5-9 Phone System Settings

Parameter	Value	Description
Phone System Name	PhoneSystem	PhoneSystem
Default TRAP Phone System	Checked	Phone system enables TRAP connections so that administrators and users without voicemail boxes can record and playback through the phone in Unity Connection web applications.
Call Loop Detection by Using Extension		
Enable for Forwarded Message Notification Calls (by Using Extension)	Checked	Unity Connection uses the extension to detect and reject new-message notifications that are sent to a device (such as a mobile phone) and that the device forwards back to Unity Connection because the device did not answer. If the call loop is not detected and rejected, the call creates a new voice message for the user and triggers Unity Connection to send a new-message notification call to the device.
Outgoing Call Restrictions		
Enable outgoing calls	Checked	Unity Connection places outgoing calls (for example, setting MWIs) as needed through the phone system.

Port Group Settings

A port group is used to control the SIP communications between the Unified CM and Unity Connection clusters. The port group allows the system to restrict and specify which Unified CM servers the Unity Connection server will accept SIP messages from, and the order and preference that the Unity Connection servers will use to route outbound calls to the Unified CM servers. The Unity Connection servers are configured to mirror the Unified CM SIP routing design for Unity Connection, hence outbound routing should be configured on Unity Connection servers to prefer the first available Unified CM subscriber node. Table 5-10 provides the port group settings.

Table 5-10 Port Group Settings

Parameter	Value	Description
Display Name	PhoneSystem-1	Descriptive name for the Phone System
Integration Method	SIP	The method of integration that is used to connect Unity Connection and Unified CM
Session Initiation Protocol (SIP) Settings		
Register with SIP Server	Checked	This ensures that Cisco Unity Connection is registered with the SIP server.
SIP Servers		
Order 0	10.195.100.21	The SIP server configured for Order 0 will have higher preference. Enter the IP address of the primary Unified CM call processing node.
Order 1	10.195.100.20	The SIP server configured for Order 1 will have lower preference. Enter the IP address of the secondary Unified CM call processing node.
Port	5060	Enter the TCP port of the Unified CM server that Unity Connection uses.

Voice Messaging Port Sizing Considerations

Each Unity Connection server in a cluster must have voice messaging ports designated for the following dial-in function in case either server has an outage:

- Answer Calls

Further, each Unity Connection server must have voice messaging ports designated for the following dial-out functions:

- Sending message waiting indications (MWIs)
- Performing message notifications
- Allowing telephone record and playback (TRAP) connections

We recommend reserving 20% of the total number of voicemail ports on the system for message notification, dial out MWI, and TRAP to reduce the possibility of call blocking on the ports for answering calls versus ports dialing out.

Alternatively, the answer and dial-out port selection can be done using previous voicemail traffic reports. Use the Port Usage Analyzer Tool to collect traffic for the last one or two weeks, then make adjustments based on actual port traffic.

Port Settings

As discussed in the previous section, ports will be either incoming or outgoing ports. Table 5-11 shows a voicemail port allocation configuration example, and Table 5-12 provides the configuration template for answer port configuration.

Table 5-11 Voicemail Port Allocation Configuration Example

Cisco Unity Connection Server	Port Range	Function
US-CUC1	1-80	Answer
US-CUC2	1-80	Answer
US-CUC1	81-100	Dial-Out
US-CUC2	81-100	Dial-Out

Table 5-12 Voicemail Answer Port Configuration Example

Parameter	Value	Description
Enabled	Checked	Check the box to enable the phone system port.
Phone System Port		
Port Name	Auto Created	Unity Connection Automatically creates the port name.
Phone System	PhoneSystem	Choose the appropriate Phone System.
Port Group	PhoneSystem-1	Choose the appropriate Port Group.
Server	US-CUC2/US-CUC1	Choose the Cisco Unity Connection subscriber node first, and similarly add ports for the Unity Connection publisher node.
Phone behavior		
Answer Call	Checked	This setting designates the port for answering the call.
Perform Message Notification	Unchecked	This setting designates the port for notifying users of messages.
Send MWI Requests	Unchecked	This setting designates the port for sending MWI on and off requests.
Allow TRAP Connections	Unchecked	This setting designates the port for Telephony Recording and Playback (TRAP) connections.

The configuration shown in the Table 5-12 should also be used to create voicemail dial out ports. However, in the case of dial out ports, uncheck the Answer Call parameter and check the Perform Message Notification, Send MWI Requests, and Allow TRAP Connection parameters instead.

Active Directory Integration

Unity Connection supports Microsoft Active Directory synchronization and authentication for Unity Connection web applications, such as Cisco Personal Communications Assistant (PCA) for end users, that rely on authentication against Active Directory. Likewise IMAP email applications that are used to access Unity Connection voice messages are authenticated against the Active Directory. For telephone user interface or voice user interface access to Unity Connection voice messages, numeric passwords (PINs) are still authenticated against the Unity Connection database.

The administrator account must be created in the Active Directory that Unity Connection will use to access the sub-tree specified in the user search base. We recommend using an account dedicated to Unity Connection, with minimum permissions set to "read" all user objects in the search base and with a password set to never expire.

Ensure that the Unified CM Mail ID field is synchronized with the Active Directory mail field. During the integration process, this causes values in the LDAP mail field to appear in the Corporate Email Address field in Unity Connection. Unity Connection uses Corporate Email Address in the Unified Messaging account to enable Single Inbox.

Unity Connection integrates with Active Directory to enable importing of user information. Integrating Unity Connection with an Active Directory provides several benefits:

- User creation — Unity Connection users are created by importing data from the Active Directory.
- Data synchronization — Unity Connection is configured to automatically synchronize user data in the Unity Connection database with data in the Active Directory.
- Single set of credentials — Configure Unity Connection to authenticate user names and passwords for Unity Connection web applications against the Active Directory, so that users do not have to maintain multiple application passwords.

Refer the Call Control chapter for Active Directory settings.

Unity Connection Partitions and CSS

All the users for this deployment are configured in the default calling search space (US-CUC1 Search Space), which contains the default partition (US-CUC1 partition).

Restriction Tables

Unity Connection uses restriction tables to prevent the voicemail system from calling unauthorized telephone numbers. These rules are normally configured to explicitly match either allowed or blocked numbers. For this deployment, the Unity Connection system is not using restriction rules for call blocking from the voicemail system but instead is using the SIP trunk incoming calling search space (CSS) to prevent unauthorized calling from Unity Connection. The SIP trunk CSS is set to allow Unity Connection to dial only on-net destinations. Table 5-13 lists the Default Transfer restriction table settings.

Table 5-13 Restriction Table in Unity Connection

Order	Blocked	Pattern
0	Uncheck the check box	+*
1	Uncheck the check box	9+*
2	Uncheck the check box	91???????*
3	Uncheck the check box	9011???????*
4	Uncheck the check box	9???????????*
5	Uncheck the check box	900
6	Uncheck the check box	*

Unity Connection contains four additional restriction tables for Default Fax, Default Outdial, Default System Transfer, and User-defined and Automatically-Added Alternate Extensions. These restriction tables can also be disabled using the settings mentioned in Table 5-13.

Class of Service

Class of service (CoS) defines limits and features for users of Unity Connection voice mail. Class of service is typically defined in a User Template, which is then applied to the user's account when it is created. For this deployment, the default Voice Mail User COS is associated with all users.

User Provisioning

Import the users into Unity Connection by using the user template from the Active Directory server. The user template contains settings that are common to a group of users. Users inherit the common settings from the user template when their account is created. Separate user templates should be created for each site in the local time zone. Table 5-14 provides the user template settings.

Table 5-14 Voicemail User Template

Section	Field	Value
Basics	Alias	SJC_User_Template
	Display Name	SJC_User_Template
	Display Name Generation	First name, then last name
	Phone System	PhoneSystem
	Class of Service	Voice Mail User COS
	Set for Self-enrollment at Next Login	Checked
	List in Directory	Checked
	Time Zone	(GMT-8:00) America/Los_Angeles
	Language	English(United States)
	Generate SMTP Proxy Address from the Corporate Email Address	Checked
Password Settings - VM	User Must Change at Next Sign-In	Checked
	Does Not Expire	Checked
	Authentication Rule	Recommended Voice Mail Authentication Rule
Change Password-Voicemail	PIN	30071982

Basing new user settings on a template minimizes the number of settings to be modified on individual user accounts, making the job of adding users quicker and less prone to error.

Note that any subsequent user template changes (after the creation of user accounts using the template) are not applied to existing user accounts; that is, the common settings are picked up from the template at user account creation time only. An individual user's settings can be changed after the template has been used to create a Unity Connection account without affecting the template or other users.

The web application password should not be changed here because Unity Connection is integrated with LDAP and user authenticates from Active Directory. You have to give these PINs and passwords to users so that they can sign in to the Unity Connection system telephone user interface (TUI) and to the Cisco Personal Communications Assistant (PCA).

Select the options **Allow Users to Use the Messaging Assistant** and **Allow Users to Use the Web Inbox and RSS Feeds** under **Voice Mail User COS class of Service** to allow users to access their web inbox using Cisco PCA.

Import the users from LDAP using the template created above.

Unity Connection User Self Enrollment

End users must enroll as Unity Connection users. The Unity Connection administrator should provide an ID (usually the user's desk phone extension) and a temporary PIN (set during User Provisioning) for each user. The first-time enrollment conversation is a set of prerecorded prompts that guide users to do the following tasks:

- Record user name.
- Record a greeting that outside callers hear when the user does not answer the phone.
- Change user PIN.
- Choose whether to be listed in the directory. (When the user is listed in the directory, callers who do not know the user's extension can reach the user by spelling or saying user's name.)

Unity Connection users can dial the voicemail pilot number from an IP endpoint within the organization or from the outside network for the self-enrollment process. If the user is calling from an extension number that is unknown to Unity Connection, either from within your organization or from outside, the user must press * (star key) when Unity Connection answers to continue the self-enrollment process. If the user hangs up before enrollment finishes, the first-time enrollment conversation plays again the next time the user signs in to Unity Connection.

4. Enable Single Inbox

Single Inbox, one of the unified messaging features in Unity Connection, synchronizes voice messages in Unity Connection and Microsoft Exchange mailboxes. When a user is enabled for a Single Inbox, all Unity Connection voice messages that are sent to the user, including those sent from Unity Connection ViewMail for Microsoft Outlook, are first stored in Unity Connection and immediately replicated to the user's Exchange mailbox. This section explains configuration tasks required for integrating Unity Connection with Microsoft Exchange 2013 and 2010 to enable Single Inbox.

Perquisites for Enabling Single Inbox with Unity Connection

- Before enabling the Single Inbox feature, ensure that Microsoft Exchange is configured and users can send and receive emails.
- Microsoft Active Directory is required for Unified Messaging service account authentication.
- Unity Connection users are imported and configured for basic voice messaging. See the section on User Provisioning.

Unity Connection Certificate Management

When you install Cisco Unity Connection, local self-signed certificates are automatically created and installed to secure communication between Cisco PCA and Unity Connection, and between IMAP email clients and Unity Connection. This means that all the network traffic (including usernames, passwords, other text data, and voice messages) between Cisco PCA and Unity Connection is automatically encrypted, and the network traffic between IMAP email clients and Unity Connection is automatically encrypted, if you enable encryption in the IMAP clients.

The other option is to use the certificate issued by the certificate authority (CA).In this case self-signed certificates are replaced with certificates issued and signed by a trusted CA.For more information on this process, refer to the section on Cisco Unified CM and IM and Presence Certificate Management.

Confirm the Exchange Authentication and SSL Settings for Unity Connection

Confirm that the Exchange server is configured for the desired web-based authentication mode (Basic, Digest, or NT LAN Manager) and web-based protocol (HTTPS or HTTP).The authentication mode must match on both Exchange and Unity Connection for them to communicate.

If you select the option to validate certificates signed by an external CA for Exchange servers and Active Directory domain controllers, obtain and install the external CA signed certificate on both the Exchange and domain controller servers.

Configure SMTP Proxy Addresses in Unity Connection

When Single Inbox is configured, Unity Connection uses SMTP proxy addresses to map the sender of a message that is sent from Unity Connection ViewMail for Microsoft Outlook to the appropriate Unity Connection user, and to map recipients to Unity Connection users.

For example, suppose an email client is configured to access Unity Connection with the email address aross@ent-pa.com. This user records a voice message in ViewMail for Outlook and sends it to user ahall@ent-pa.com. Unity Connection then searches the list of SMTP proxy addresses for aross@ent-pa.com and ahall@ent-pa.com. If these addresses are defined as SMTP proxy addresses for the Unity Connection users ahall and aross respectively, Unity Connection delivers the message as a voice message from the Unity Connection user aross to the Unity Connection user ahall.

The SMTP proxy address for the user is automatically created when you import the users via the user template. In the user template, select the **Generate SMTP Proxy Address from the Corporate Email Address** option for creating the SMTP proxy address. Refer to the section on User Provisioning for more information.

Create Unified Messaging Services Account in Active Directory and Grant Permissions for Unity Connection

Single Inbox requires an Active Directory account (called the Unified Messaging Services account), and the account must have the rights necessary for Unity Connection to perform operations on behalf of users. Unity Connection accesses Exchange mailboxes using the Unified Messaging Services account. When creating the Unified Messaging Services account, follow these guidelines:

- Do not create an Exchange mailbox for the account.
- Do not add the account to any administrator group.
- Do not disable the account, otherwise Unity Connection cannot use it to access Exchange mailboxes

Sign in to a server on which the Exchange Management Shell is installed and assign the **ApplicationImpersonation Management** role to the Unified Messaging Services account for Unity Connection using the following command:

new-ManagementRoleAssignment -Name: *RoleName* **-Role:ApplicationImpersonation -User:'** *Account* **'**

Where:

- *RoleName* is the name that you want to give the assignment; for example, Unity ConnectionUMServicesAcct. The name that you enter for RoleName appears when you run the command **get-ManagementRoleAssignment**.
- *Account* is the name of the Unified Messaging Services account in domain\alias format.

SMTP Smart Host

Unity Connection relays the message to the user email address using SMTP Smart Host. When a Unity Connection user receives a new message, Unity Connection can send a text notification to an email address. With this type of notification, you can configure Unity Connection to include a link to Cisco PCA in the body of the email message. Under the user configuration, navigate to the **Edit Notification Device** page for the user and select the option to **Include a Link to the Cisco Unity Connection Web Inbox in Message Text**. Table 5-15 lists the SMTP Smart Host configuration.

Table 5-15 SMTP Smart Host Details (System Settings > SMTP Configuration > Smart Host)

Parameter	Value
SmartHost	US-EXCH1.ent-pa.com

Unified Messaging Service

In Unity Connection Administration, expand **Unified Messaging**, then select **Unified Messaging Services**.

- Unified Messaging Services define the type of Microsoft Exchange and authentication method that Unity Connection will use to communicate with Microsoft Exchange.
- Configure Unified Messaging Services to communicate with a specific Exchange server using an FQDN.
- Configure the Unity Connection Unified Messaging Services for the same Web-based Authentication Mode (Basic, Digest, or NT LAN Manager) and Web-Based Protocol (HTTPS or HTTP) that is configured on Microsoft Exchange.
- Enter the Active Directory account credentials created in the section Create Unified Messaging Services Account in Active Directory and Grant Permissions for Unity Connection.
- Select the options to **Access Exchange Calendar and Contacts** and **Synchronize Connection and Exchange Mailboxes (Single Inbox)** to enable Unified Messaging features.
- Self-signed certificates cannot be validated. If you want the Unity Connection server to validate SSL certificates from Exchange, then use the public certificates from a certification authority (CA) instead of self-signed certificates. Refer to the section on Unity Connection Certificate Management for details.

Unified Messaging Account

In Unity Connection Administration, expand **Users** then select **Users**. On the Edit User Basics page, in the Edit menu, select **Unified Messaging Account**s.

- When you create a user account, Unity Connection does not automatically create a unified messaging account for that user. A unified messaging account can be created for one user or multiple users. Use the Bulk Administration Tool (BAT) to create the unified messaging account for large number of users.
- Unified messaging requires that you enter the Exchange email address for each Unity Connection user. On the Unified Messaging Account page, select **Use Corporate Email Address: None Specified** to cause Unity Connection to use the corporate email address specified on the Edit User Basics page as the Exchange email address.
- In the Active Directory integration, the Unified CM Mail ID field is synchronized with the Active Directory mail field. This causes values in the LDAP mail field to appear in the Corporate Email Address field in Unity Connection.

For more information on creating unified messaging accounts for multiple users with the Bulk Administration Tool, refer to the *System Administration Guide for Unity Connection*.

Voice Mail User COS

Edit the Voice Mail User Class of Service (**Class of Service** –> **Voice Mail User COS**) to enable the user for Single Inbox. In the **Licensed Features** select the option to **Allow Users to Access Voicemail Using an IMAP Client and/or Single Inbox**. Also select the option to **Allow IMAP Users to Access Message Bodies**.

Install ViewMail for Outlook on User Workstations

Cisco ViewMail for Microsoft Outlook provides a visual interface from which users can send, listen to, and manage their Unity Connection voice messages from within Outlook. Download Unity Connection ViewMail for Microsoft Outlook from the Cisco website and install it on each user workstation. After installing ViewMail, open the ViewMail settings or Options tab and associate an email account with a Unity Connection server. Enter the user information and Unity Connection server details.

When using another email client to access Unity Connection voice messages in Exchange, or in cases when ViewMail for Outlook is not installed, note the following:

- The email client treats Unity Connection voice messages like emails with .wav file attachments.
- When a user replies to or forwards a Unity Connection voice message, the reply or forward is treated like an email, even if the user attaches a .wav file. Message routing is handled by Exchange, not by Unity Connection, so the message is never sent to the Unity Connection mailbox for the recipient.

5. Enable Visual Voicemail

Visual Voicemail provides access to Unity Connection directly from the voicemail tab on Jabber clients. Users can view a list of voice messages and play messages from Jabber. Users can also delete voice messages.

Unity Connection Configuration

- Ensure that the Unity Connection users are imported and configured for basic voice messaging. Refer to the section on User Provisioning.
- Ensure that the Unity Connection **Connection Jetty** service and **Connection REST Service** are up and running. Both services are activated during Service Activation under the **Optional Services** category.
- Ensure that **Class of Service** is enabled for voicemail access from the IMAP client. Refer the section on Voice Mail User COS.
- Edit the Unity Connection Voice Mail Class of Service (CoS) to allow users to use web inboxes. Under the **Features** tab, select the option to **Allow Users to Use Unified Client to Access Voicemail**.
- Select the following options under the API settings (**System Settings** > **Advanced**):
  - Allow Access to Secure Message Recordings through CUMI
  - Display Message Header Information of Secure Messages through CUMI
  - Allow Message Attachments through CUMI

Unified CM Configuration

Add a **Voicemail** UC service for each Unity Connection server node. Table 5-16 shows the voicemail UC service configuration.

Table 5-16 Voicemail Service Settings (User Management > User Settings > UC Service)

Parameter	Value	Comments
Product Type	Unity Connection	Enter the product name of the voicemail system.
Name	us-cuc1	Enter the name of the voicemail service. Choose the display name that will help to distinguish between publisher and subscriber voicemail services.
Description	us-cuc1	Enter the display name that will help to distinguish between publisher and subscriber voicemail services.
Host Name/IP address	us-cuc1.ent.pa.com	Enter the address of the voicemail service in either IP address or FQDN format.
Port	443	Enter the port to connect with the voicemail service.
Protocol	HTTPS	Select the protocol to route voice messages securely.

Apply the **Voicemail** UC service created previously to the **Standard** Service Profile (**User Management** –> **User Settings** –> **Service Profile**). Ensure that the Voicemail UC service created for Unity Connection publisher (us-cuc1.ent.pa.com) is set to the primary profile and the Unity Connection subscriber (us-cuc2.ent.pa.com) is set to the secondary profile. To synchronize credentials for the voicemail service, select **Unified CM - IM and Presence** from the **Credentials source for voicemail service** drop-down list.

6. Voice Mail in SRST Mode

With the centralized messaging deployment model, during a WAN outage the branch site's Survivable Remote Site Telephony (SRST) routes the unanswered and busy calls to the central Unity Connection. Incoming calls that reach a busy signal, calls that are unanswered, and calls made by pressing the message button are forwarded to Unity Connection. This configuration allows phone message buttons to remain active. To enable this functionality, configure POTS dial peer access to Unity Connection through PRI.

When calls are routed over the PSTN to Unity Connection, Redirected Dialed Number Information Service (RDNIS) is critical. Incorrect RDNIS information can affect calls to voicemail that are rerouted over the PSTN. If the RDNIS information is not correct, the call will not reach the voicemail box of the dialed user but will instead receive the automated 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 whether it provides guaranteed RDNIS delivery end-to-end for your circuits.

Unified CM Configuration

Ensure that the settings mentioned in Table 5-17 are enabled in Unified CM configuration for the SIP trunk to the central site PSTN gateway.

Table 5-17 Settings for the SIP Trunk to the PSTN gateway for Voicemail in SRST Mode

Parameter	Value	Comments
Call Routing Information – Inbound Calls		
Redirecting Diversion Header Delivery - Inbound	Checked	This ensures that the redirecting Information Element, the first redirecting number, and the call forward reason are sent and accepted as a part of incoming messages. Unity Connection uses the first redirecting number to answer the call.
Call Routing Information – Outbound Calls		
Redirecting Diversion Header Delivery - Outbound	Checked	This ensures that the redirecting Information Element, the first redirecting number, and the call forward reason are sent and accepted as a part of outgoing messages. Unity Connection uses the first redirecting number to answer the call

Branch SRST Router Configuration

Configure the following command on the branch site SRST router to enable voicemail access over PRI.

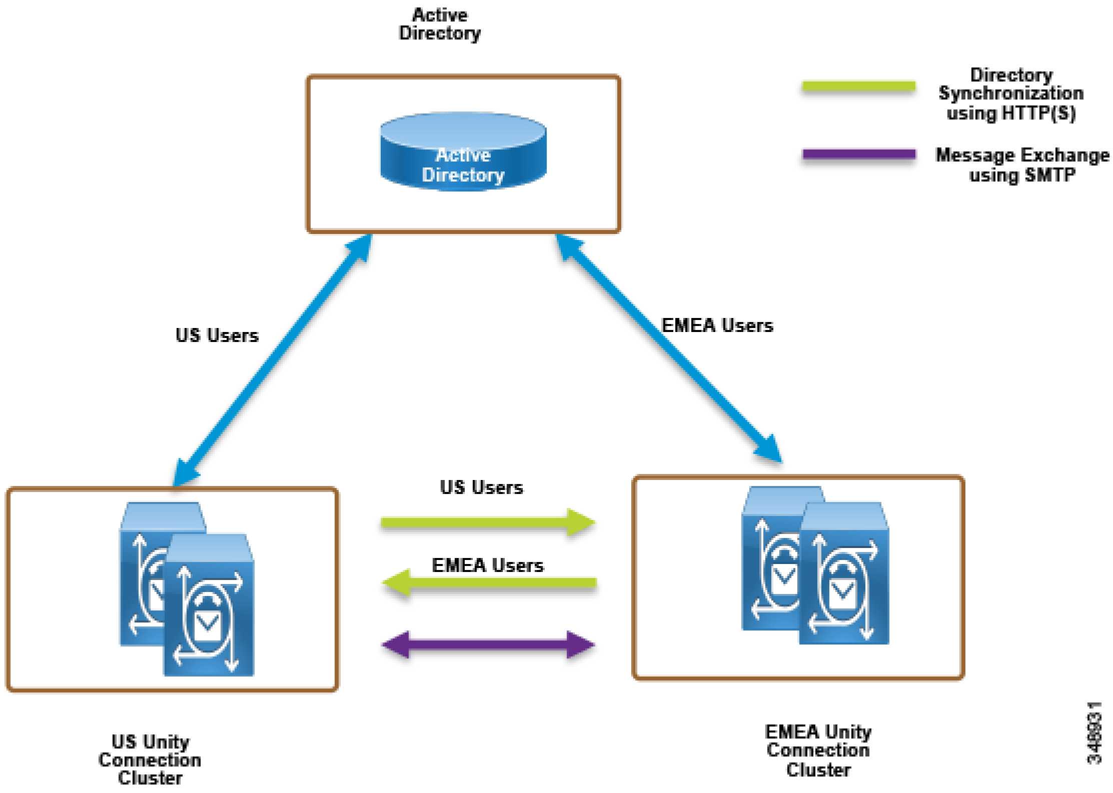
```
!
!
dial-peer voice 10 pots
destination-pattern +14085554999
direct-inward-dial
port 1/0:15
!
!
voice register pool 1
call-forward b2bua busy +14085554999
call-forward b2bua noan +14085554999 timeout 12
!
!
```

7. HTTPS Internetworking of Two Unity Connection Clusters

Figure 5-4 shows HTTPS internetworking of two Unity Connection clusters. HTTPS networking connects multiple Unity Connection clusters so that they can share directory information and exchange of voice messages. You can join two or more Unity Connection servers or clusters to form a well-connected network, referred to as a Unity Connection site. The servers that are joined to the sites are referred to as *locations*. Within a site, each location uses HTTPS protocol to exchange directory information and SMTP protocol to exchange voice messages with each other.

Within a site, Unity Connection locations automatically exchange directory information, so that a user in one location can dial out to or address messages to a user in any other system by name or extension, provided that the target user is reachable in the search scope of the originating user. The networked systems function as though they share a single directory.

Figure 5-4 HTTPS Internetworking of Two Unity Connection Clusters


Active Directory
Active Directory
Directory Synchronization using HTTP(S)
Message Exchange using SMTP
US Users
EMEA Users
US Users
EMEA Users
US Unity Connection Cluster
EMEA Unity Connection Cluster
348931

In HTTPS networking, Unity Connection clusters are joined together using a hub-and-spoke topology. In this topology, all the directory information among the spokes is shared through the hub that connects the spokes. The number of Unity Connection locations that can be connected in an HTTPS network and the maximum number of users in HTTPS networking depend on the deployed OVA template. For more information on the maximum number of supported locations and maximum directory size, refer to the information on *directory object limits* in the *System Requirements for Cisco Unity Connection*.

In HTTPS networking, the directory replication is accomplished by means of a Feeder service and a Reader service running on each location in the network. The Reader service periodically polls the remote location for any directory changes since the last poll interval. The Feeder service checks the change tracking database for directory changes and responds to poll requests with the necessary information.

In the HTTPS networking, when the publisher server of a cluster location is up and running, it is responsible for the synchronization of directory information. However, if the publisher server is down, the subscriber server takes the role of synchronizing directory information.

Depending upon the server of a cluster (publisher or subscriber) with which the directory synchronization is being performed, the directory synchronization can be either of the following types:

- Standard — Specifies that the directory synchronization is done by the publisher server with the connected locations.
- Alert — Specifies that the publisher server is unreachable and the subscriber server is responsible for providing directory information to the connected locations. However, the subscriber server has the directory information stored that was last synchronized with the publisher server when it was running.

In the event of a publisher failure, directory synchronization occurs in the Alert mode. During the Alert mode, the connected nodes in the HTTPS network have limited access to directory synchronization with the subscriber. The limited access means that the connected nodes can fetch only the directory information that was last synchronized with the publisher when it was running. When the publisher comes up, the nodes that are directly connected to the publisher synchronize the updated directory information through the publisher. Therefore, the key benefit of the Alert mode is that the connected nodes remain synchronized with the subscriber server even when the publisher is down.

The clusters that are networked together are directly accessible through TCP/IP port 25 (SMTP).In addition, both locations must be able to route to each other via HTTP on port 8081 or HTTPS on port 8444.

For the purposes of this deployment documentation, we assume there is HTTPS networking between US and EMEA Unity Connection clusters. Table 5-18 shows the server node information of both clusters that are joined using HTTPS networking.

Table 5-18 Unity Connection Cluster Details for HTTPS Networking

	US Unity Connection Cluster		EMEA Unity Connection Cluster	
Server	**Hostname**	**IP address**	**Hostname**	**IP address**
Publisher	US-CUC1	10.195.100.30	EMEA-CUC1	10.195.99.30
Subscriber	US-CUC2	10.195.100.31	EMEA-CUC2	10.195.99.31

To set up HTTPS networking between two Unity Connection clusters, perform the following tasks described in this section.

Check the Display Name and SMTP Domain of Each Unity Connection Server

- The Unity Connection server that you join to an HTTPS network must have a unique display name and SMTP domain.
- Before enabling HTTPS networking, verify the display name and SMTP domain of the Unity Connection publisher server in the **Networking** –> **Locations** settings.

Create the HTTPS Network Between Unity Connection Clusters

- To create an HTTPS network of Unity Connection servers, start by linking two clusters together by creating an HTTPS link and then ensuring that the subscribers of each cluster are added for the SMTP Access.
- On each Unity Connection publisher, add a new HTTPS link. Table 5-19 shows the HTTPS Link settings.

Table 5-19 HTTPS Link Settings (Networking > HTTP(s) Links)

Parameter	Value	Comments
Link to Cisco Unity Connection Remote Location		
Publisher (IP address/FQDN/Hostname)	emea-cuc1.ent-pa.com	Enter the IP address, fully qualified domain name (FQDN), or hostname of the remote Unity Connection publisher node.
Username	Name of admin user	Enter the Username of an administrator at the location specified in the above publisher field. The administrator user account must be assigned the System Administrator role.
Password	Password of the admin user	Enter the password for the administrator specified in the Username field.
Transfer Protocol		
Use Secure Socket Layer (SSL)	Checked	This option enables SSL to encrypt directory synchronization traffic between the various HTTPS locations.
Accept Self-Signed Certificates	Check the check box only if a self-signed certificate is used	Check this check box to allow the local node in a network to use a self-signed certificate to negotiate SSL with this location. Uncheck this check box to require the local node in a network to use a certificate signed by a certificate authority (CA).

Configure SMTP Access for Cluster Subscriber Servers

In an HTTPS network that includes a Unity Connection cluster server pair, you can join only the publisher server of the pair to the network. In order for all locations in the network to communicate directly with the cluster subscriber server node when the subscriber is the primary server, all network locations should be configured to allow SMTP connections from the subscriber server.

In this example we are adding the EMEA subscriber to the SMTP configuration of the US publisher, as well as adding the US subscriber to the EMEA publisher SMTP configuration.

- In the US cluster on the US publisher, add the EMEA subscriber to the SMTP configuration (System Settings). In the **Edit** menu, select **Search IP Address Access List**. On the New IP Address page, enter the IP address of an EMEA subscriber server (10.195.99.31 in this example). Ensure that the **Allow Connection** option is selected.
- Repeat the above steps on the EMEA cluster publisher, emea-cuc1.ent-pa.com, to add the US cluster subscriber IP address.

Replication Between the Locations

After creating the HTTPS network, verify that the complete database is replicated between the two locations added to network. When initial replication begins, it can take a few minutes to a few hours for the data to be fully replicated between all locations, depending on the size of your directory.

Open the **HTTP(S) Link** created in the above step, and check the following values:

- Time of Last Synchronization

  Indicates the time stamp of the last time the local reader service attempted to poll the remote location feeder service for directory changes on the remote locations, regardless of whether a response was received.

- Time of Last Failure

  Indicates the time stamp of the last time the local reader service encountered an error while attempting to poll the remote location feeder service. If the value of this field is 0, or if the Time of Last Synchronization value is later than the Time of Last Error value, replication is likely to be progressing without problems.

- Object Count

  Indicates the number of users that the local Unity Connection location has synchronized from the remote location.

Add Remote Location Partition to Local Unity Connection CSS

When you initially set up a network between locations, users that are provisioned on the US cluster will not able to send voice messages to users on the EMEA cluster because the users in each location are in separate partitions and separate user search spaces that do not contain the partitions of users in the other locations.

- Edit the us-cuc1 calling search space (CSS) configured for the US Unity Connection server to include the EMEA location Unity Connection server partition emea-cuc1.
- Edit the emea-cuc1 CSS configured for the EMEA Unity Connection server to include the US location Unity Connection server partition us-cuc1.

Related Documentation

- *Voice Messaging* chapter of the *Cisco Collaboration System SRND*

  http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab11/collab11/vmessage.html

- *Design Guide for Cisco Unity Connection*

  http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/connection/11x/design/guide/11xcucdgx.html

- *HTTPS Networking Guide for Cisco Unity Connection*

  http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/connection/11x/https_networking/guide/11xcuchttpsnetx.html

- *Unified Messaging Guide for Cisco Unity Connection*

  http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/connection/11x/unified_messaging/guide/11xcucumgx.html

Tools for Application Deployment

In addition to applications such as Unity Connection, described in this chapter, there are two useful tools to help administrators deploy the Enterprise Collaboration Preferred Architecture:

- Cisco Prime Collaboration Deployment — Assists the administrator by automating many of the steps necessary to install Unified CM clusters with IM and Presence Servers and Unity Connection clusters.
- Cisco Prime License Manager — Is deployed as a standalone tool to provide the administrator with a single management point for the various licenses used in a deployment.

Cisco Prime Collaboration Deployment (PCD)

Prime Collaboration Deployment (PCD) assists the administrator with the tasks of deploying new clusters of Unified CM and IM and Presence servers and Unity Connection. The automation greatly assists administrators by configuring all common settings in the nodes of the cluster.

PCD is a standalone application that must be installed on its own virtual machine prior to beginning any other installation tasks. To install a new cluster using PCD:

1. Deploy the host hardware and configure the ESXi.
2. Download the necessary OVA template and Cisco ISO images for the target release.
3. Deploy the recommended OVA template for your enterprise:
    a. Create the virtual machines for each node on the ESXi hosts (one virtual machine for each node to be installed).
    b. Configure the network settings on the new virtual machines.
4. Add the ESXi host to the PCD user interface.
5. Create a new task within PCD for the type of cluster being created. Define the nodes to be installed and their associated virtual machines.

Cisco PCD will then complete the process of installing the application on the virtual machines and will provide an email notification to the administrator when the task is complete.

Cisco Prime License Manager (PLM)

Cisco Prime License Manager (PLM) provides simplified, enterprise-wide management of user-based licensing, including license fulfillment. Cisco Prime License Manager handles licensing fulfillment, supports allocation and reconciliation of licenses across supported products, and provides enterprise-level reporting of usage and entitlement.

A single virtual machine is required for an enterprise, and the application should be backed up through VMware tools. As a standalone instance of PLM, all nodes of the Preferred Architecture could be effectively managed. Since this is not an end-user facing application and does not have any real-time usage impacts, no clustering is required. Since one of the features of PLM is the ability to support e-Fulfillment of additional licenses, this server must have access to the Internet to pull new license files.

In cases where PLM does not have access to the Internet, the system administrator can create a request using the "Generate License Request…" option. The administrator submits the license request on the Cisco License Registration website and receives an applicable license file by email. The administrator then uploads this file to the system using the "Fulfill licenses from File…" option.

Within the Preferred Architecture, the following products are supported by PLM:

- Cisco Unified CM
- Cisco Unity Connection

Key features of PLM that benefit the administrator:

- License usage history
- e-Fulfillment of new licenses

License Usage History

The ability for an administrator to track usage of collaboration portfolio licenses over time gives that administrator better ability to plan for additional licenses when needed. In addition, this license usage tool assists the administrator with remaining in compliance with all license usage rules.

An application is allowed 60 days of non-compliance, during which administrators can make changes if there are insufficient licenses or if the PLM node has lost communication with the application node. After 60 days of non-compliance, the Unified CM application(s) will no longer allow administrative changes; however, the application(s) will continue to function (call control) with no loss of service. After 60 days of non-compliance, the Unity Connection application(s) will allow administrative changes, but the application(s) will not continue to function (users will not have access to voice messaging).

e-Fulfillment of New Licenses

When the administrator needs to procure additional licenses, the e-Fulfillment tool within PLM simplifies the number of steps required, and it imports the licenses into the appropriate product for use.

Additional Applications

There are many additional applications provided by Cisco and ecosystem partners that enhance a collaboration environment. Table 5-20 is not intended to be all-inclusive, but it lists some frequently referenced applications for customer deployments.

Table 5-20 Additional Cisco Applications for the Preferred Architecture

Application Name	Functions	Integration Method
Contact Center Enterprise (CCE)	Provides internal and external customer collaboration technologies, including agent login, Interactive Voice Response (IVR) for call vectoring, outbound connection methods, and multi-channel agent interactions.	Enterprise contact centers operate on a dedicated Unified CM cluster that is trunked to the enterprise Unified CM cluster.
Contact Center Express (CCX)	Provides dial-by-name and a subset of Contact Center ideal for small contact centers or internal use.	Communicates through JTAPI to Unified CM.
TelePresence Content Server (TCS)	Provides video, audio, and content recording functionality that can be included in scheduled calls through a check-box in TMS or dialed, allowing any endpoint to easily be a recording station.	Integrates with Unified CM via a SIP trunk and enables recording for devices registered to Unified CM.
Show and Share	Provides an internal stored video content portal.	TCS automatically uploads content to Show and Share. No other integration to call control is required.

Table 5-20 Additional Cisco Applications for the Preferred Architecture (continued)

Application Name	Functions	Integration Method
Prime Collaboration Provisioning	Provides an administrative portal for "Day 2" operations.	Standalone software that communicates through SSH and HTTPS interfaces of infrastructure devices and endpoints.
Prime Collaboration Assurance	Provides quality and fault detection services for collaboration deployment administrator	Standalone software that communicates through SSH and HTTPS interfaces of infrastructure devices and endpoints
Prime Collaboration Analytics	Provides up to one year of usage data for usage and fault trend analysis by the collaboration deployment administrator.	Deployed with Prime Collaboration Assurance and utilizes data collected by that application.
Attendant Console	Gives corporate operators or receptionists a desktop application to handle incoming calls.	Standard version installs on the end user's Windows computer and connects to Unified CM. Advanced version runs on a dedicated server, and the end users log into the application.
MediaSense	Provides recording for both full-time and selective recording scenarios in Unified CM.	Recording Profiles are configured in Unified CM, and MediaSense is connected to Unified CM and Cisco Unified Border Element through SIP trunks.
Jabber Guest	Provides click-to-connect functionality for business-to-consumer (B2C) collaboration.	Requires a dedicated Expressway-C and Expressway-E pair, using a distinct domain from the enterprise Expressway-C and Expressway-E implementation used for Mobile and Remote Access and business-to-business video calls. Unified CM has SIP trunks to this dedicated Expressway pair.

CHAPTER 6

Bandwidth Management

Revised: November 20, 2015

This chapter describes the bandwidth management strategy for the Cisco Preferred Architecture (PA) for Enterprise Collaboration.

Certain requirements might put your deployment outside the PA design guidelines and recommendations, in which case you might have to use other documentation such as the Cisco Collaboration SRND and related product documentation for a more customized architecture.

The first part of this chapter provides an architectural overview and introduces some fundamental design concepts, while the second part covers deployment procedures. The Architecture section discusses topics such as identification and classification, queuing and scheduling, provisioning and admission control, using the hypothetical customer topology presented in the examples throughout this document. The next portion of this chapter is the Bandwidth Management Deployment section. The deployment examples in that section help you to understand the implementation of certain design decisions more clearly than an abstract discussion of concepts can. The order of the topics in the Bandwidth Management Deployment section follows the recommended order of configuration.

Note This chapter is a new addition to the 11.0 release of this document. We recommend that you read this entire chapter before attempting to implement bandwidth management for the Preferred Architecture for Enterprise Collaboration.

Core Components

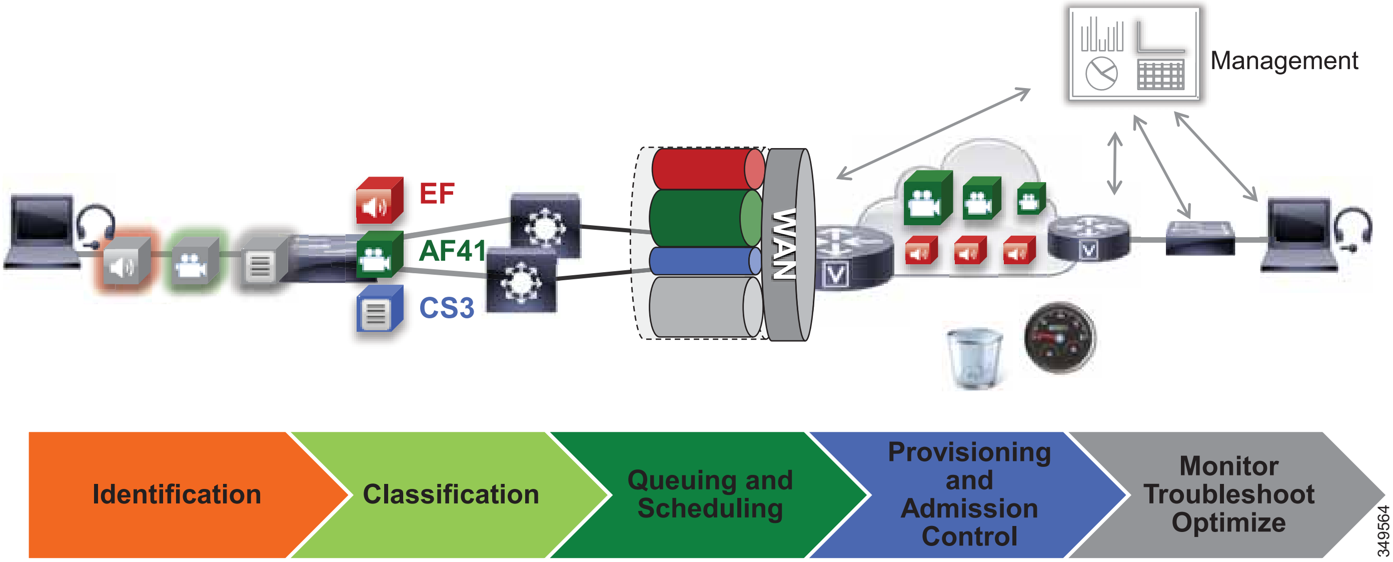
The Quality of Service (QoS) architecture contains these key components:

- Cisco Unified Communications Manager
- Cisco endpoints
- Cisco Expressway
- Cisco Unity Connection
- Cisco TelePresence Server
- Cisco TelePresence Conductor
- Network infrastructure:
  - Cisco routers
  - Cisco switches

Figure 6-1 illustrates the design approach to QoS used in the Cisco PA for Enterprise Collaboration. This approach consists of the following phases:

- **Identification and classification** — Refers to concepts of trust and techniques for identifying media and call signaling for endpoints and applications. It also includes the process of mapping the identified traffic to the correct DSCP to provide the media and signaling with the correct per-hop behavior end-to-end across the network.
- **Queuing and scheduling** — Consists of general WAN queuing and scheduling, the various types of queues, and recommendations for ensuring that collaboration media and signaling are correctly queued on egress to the WAN.
- **Provisioning and admission control** — Refers to provisioning the bandwidth in the network and determining the maximum bit rate that groups of endpoints will utilize. This is also where call admission control can be implemented in areas of the network where it is required.
- **Monitoring, troubleshooting, and optimization** — Ensures the proper operation and management of voice and video across the network. Cisco Prime Collaboration offers a suite of tools to perform these functions. Monitoring, troubleshooting and optimization are not covered in the Preferred Architectures but are part of the overall approach.

Figure 6-1 Architecture for Bandwidth Management



Recommended Deployment

- Identify media and SIP signaling traffic from the endpoints.
- Classify and mark traffic at the access switch edge.
  - Mark all audio with Expedited Forwarding class EF (includes all audio of voice-only and video calls).
  - Mark all critical desktop and room system video with an Assured Forwarding class of AF41.
  - Mark all Jabber, mobile and remote access (MRA), and edge video with an Assured Forwarding class of AF42.
  - Mark all call signaling with CS3.
  - Configure QoS on all media originating and terminating applications and MCUs across the solution.
- Apply simplified WAN edge policies for identifying, classifying, marking, and queuing collaboration traffic:
  - WAN edge ingress re-marking policy
  - WAN edge egress queuing and scheduling policy
- Group video endpoints into classes according to maximum video bit rate, to limit bandwidth consumption based on endpoint type and usage within the solution.
- Deploy Enhanced Locations Call Admission Control and limit video calling only in areas of the network where bandwidth resources are restricted.

Key Benefits

This deployment of bandwidth management provides the following benefits:

- Provides prescriptive recommendations to simplify deployment with a simplified QoS architecture
- Makes more efficient use of network resources
- Supports mobile and multi-media Collaboration devices
- Takes into account "unmanaged" network segments (Internet)
- Is "future-proof" because it facilitates introduction of new services, features, and endpoints

Architecture

Bandwidth management is about ensuring the best possible user experience end-to-end for all voice and video endpoints, clients, and applications in the Collaboration solution. This chapter provides a holistic approach to bandwidth management, incorporating an end-to-end Quality of Service (QoS) architecture with call admission control and video rate adaptation and resiliency mechanisms to ensure the best possible user experience for deploying pervasive video over managed and unmanaged networks.

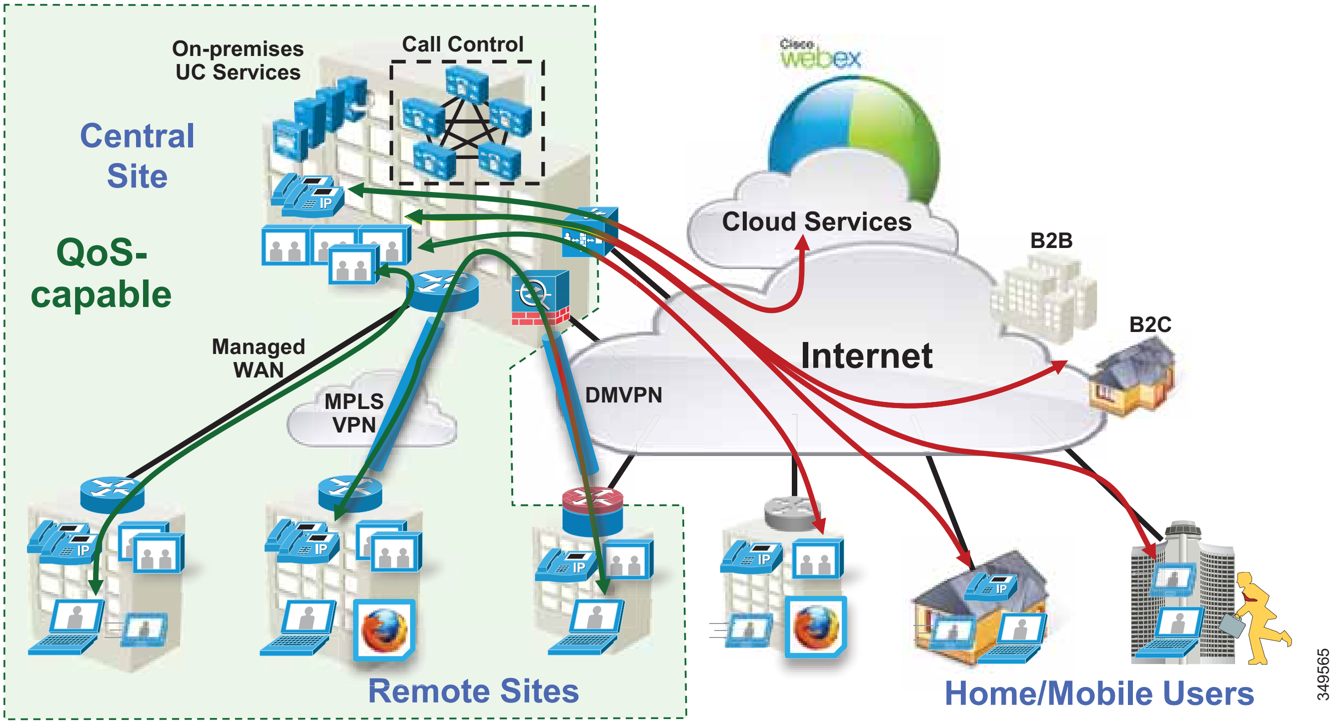
This section starts with a discussion of collaboration media, the differences between audio and video, and the impact this has on the network. Next this section outlines an end-to-end QoS architecture for collaboration that includes: identification and classification of collaboration media and SIP signaling for endpoints, clients, and applications; WAN queuing and scheduling; and bandwidth provisioning and admission control. The next section on Bandwidth Management Deployment explains the steps involved in implementing this architecture in both the collaboration and network architecture.

Note The Network Infrastructure chapter of the Collaboration SRND lays the foundation for QoS in the LAN and WAN. If you are unfamiliar with the concepts of QoS, it is important to read that chapter to fully understand the concepts discussed therein. This chapter assumes an understanding of QoS.

Introduction

In this Preferred Architecture, usage of the Internet and cloud-based services such as WebEx are an important aspect of the solution, which means that some of the Collaboration infrastructure is located outside of the managed enterprise network and located in the cloud. The enterprise office connectivity options also range from remote sites and mobile users connected over managed leased lines directly connected to MPLS or other technologies, to being connected over the Internet through technologies such as Dynamic Multipoint VPN (DMVPN), for example. Figure 6-2 illustrates the convergence of a traditional on-premises Collaboration solution in a managed (capable of QoS) network with cloud services and sites located over an unmanaged (not capable of QoS) network such as the Internet. On-premises remote sites are connected over this managed network, where administrators can prioritize collaboration media and signaling with QoS, while other remote sites and branches connect into the enterprise over the Internet, where collaboration media and signaling cannot be prioritized or prioritized only outbound from the site. Many different types of mobile users and teleworkers also connect over the Internet into the on-premises solution. So the incorporation of the Internet as a source for connecting the enterprise with remote sites, home and mobile users, as well as other businesses and consumers, has an important impact on bandwidth management and user experience.

Figure 6-2 Managed versus Unmanaged Network



This section presents a strategy for leveraging smart media techniques in Cisco video endpoints, building an end-to-end QoS architecture, and using the latest design and deployment recommendations and best practices for managing bandwidth to achieve the best user experience possible based on the network resources available and the various types of networks that collaboration media traverse.

Collaboration Media

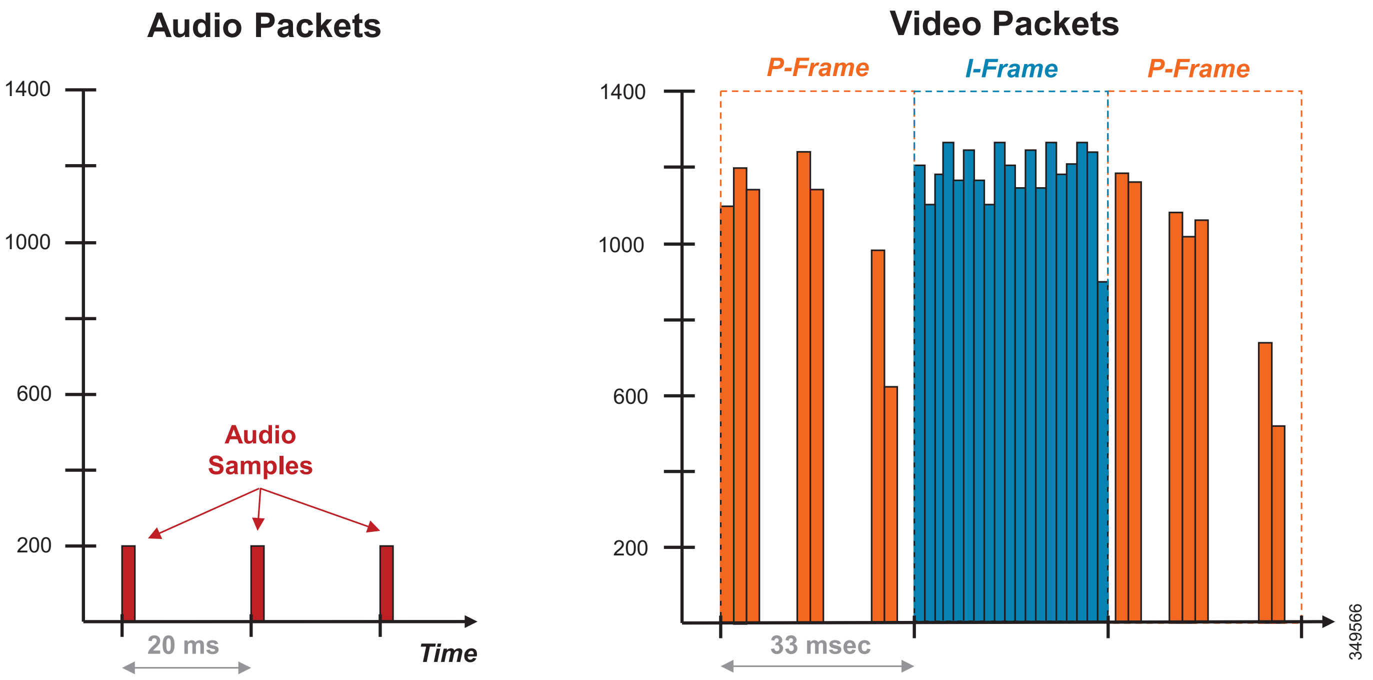
This section covers the characteristics of audio and video streams in real-time media as well as the smart media techniques that Cisco video endpoints employ to ensure high fidelity video in the event of packet loss, delay, and jitter.

Audio versus Video

Voice and video are often thought of as quite similar, and although they are both real-time protocol (RTP) applications, the similarities stop there. Voice is generally considered well behaved because each packet is a fixed size and fixed rate. Video frames are spread over multiple packets that travel as a group. Because one lost packet can ruin a P-frame, and one bad P-frame can cause a persistent artifact, video generally has a tighter loss requirement than audio. Video is asymmetrical. Voice can also be asymmetrical but typically is not. Even on mute, an IP phone will send and receive the same size flow.

Video increases the average real-time packet size and has the capacity to quickly alter the traffic profile of networks. Without planning, this could be detrimental to network performance. Figure 6-3 shows the difference between a series of audio packets and video packets sent during a specific time interval.

Figure 6-3 Audio versus Video



As illustrated in Figure 6-3, the audio packets are the same size, sent at exactly the same time intervals, and represent a very smooth stream. Video, on the other hand, sends a larger group of packets over fixed intervals and can vary greatly from frame to frame. Figure 6-3 shows the difference in the number of packets and packet sizes for an I-Frame compared to P-frames. This translates to a stream of media that is very bursty in nature when compared to audio. This burstiness is illustrated in Figure 6-4, which shows the bandwidth profile over time of an HD video stream. Note the large bursts when I-Frames are sent.

Figure 6-4 Bandwidth Usage: High-Definition Video Call

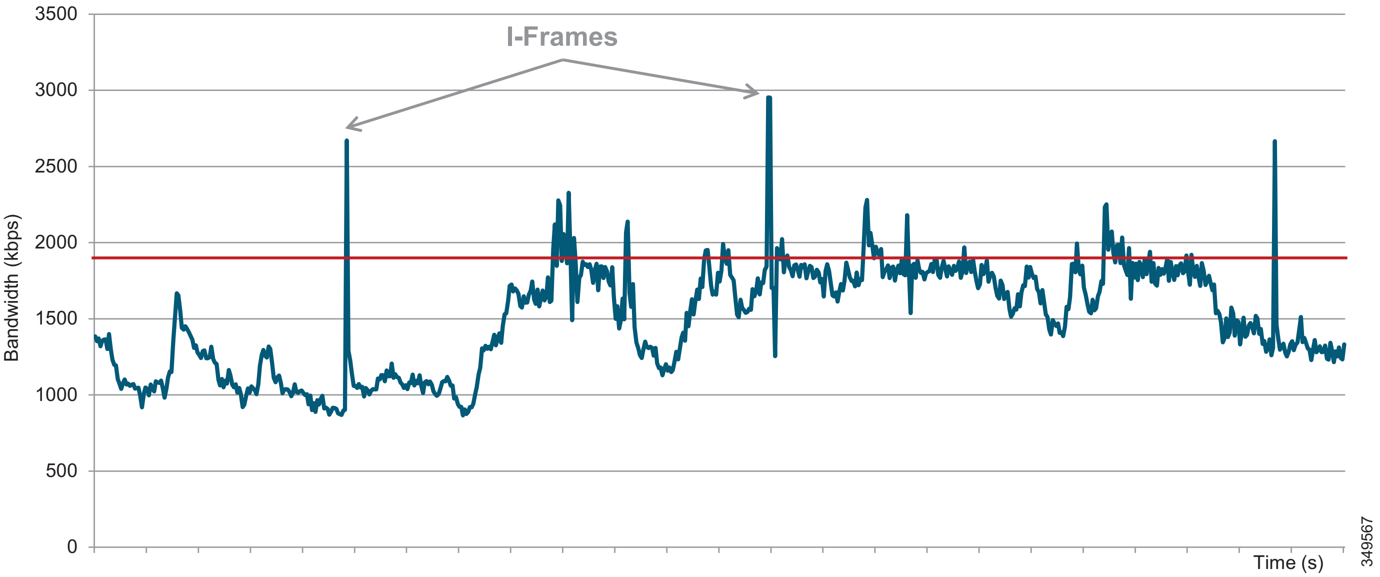
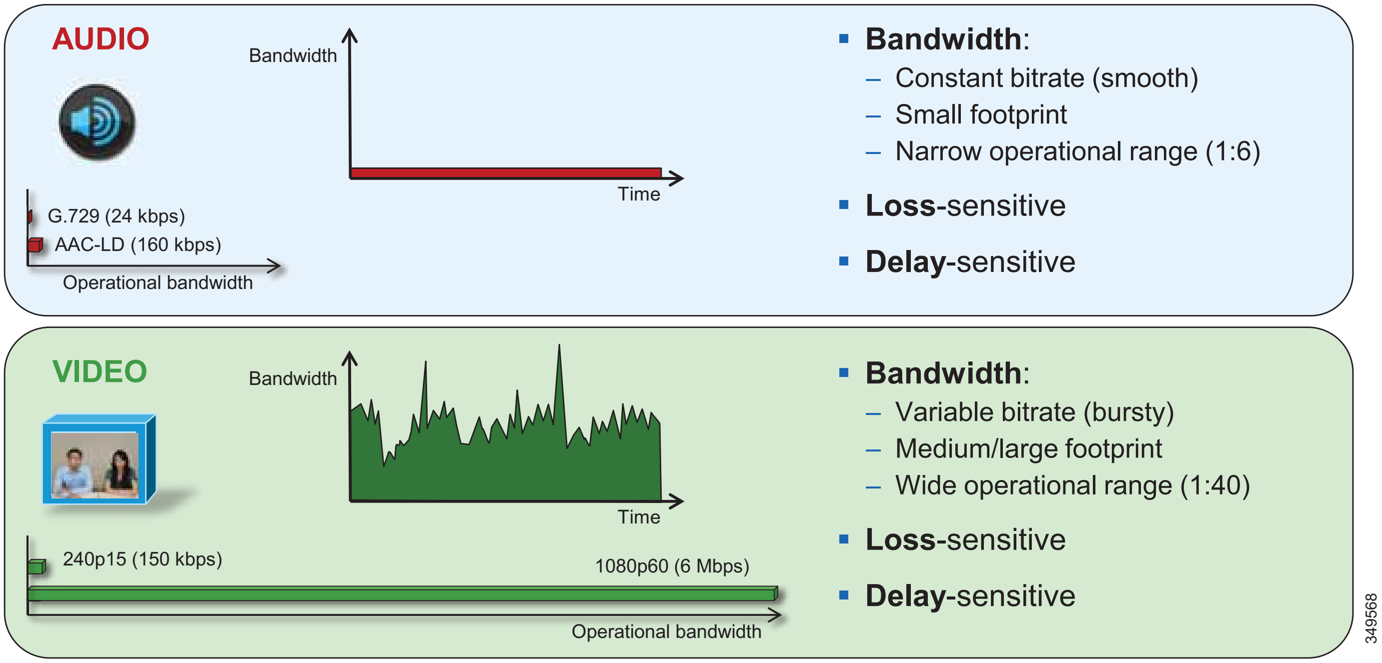


Figure 6-4 shows an HD video call at 720p at 30 fps and 1920 kbps (1792 kbps video + 128 kbps audio). The red line indicates the average bit rate for the duration of the call.

While audio and video are both transported over UDP and are sensitive to loss and delay, they are quite different in their network requirements and profile. As shown in Figure 6-5, audio is a constant bit rate, has a smaller footprint compared to video, and has a narrower operational range of 1:6 ratio when comparing the lowest bit rate audio codec to one of the highest bit rate codecs. Video, on the other hand, has a variable bit rate (is bursty), has a medium to large footprint when compared to audio, and has a wide operational range of 1:40 (250p at 15 fps up to 1080p at 60 fps).

Figure 6-5 Video Traffic Requirements and Profiles



The main point here is that audio and video, while similar in transport and sensitivity to loss and delay, are quite different in the methods employed to manage their bandwidth requirements in the network. Also, while video is pertinent to a full collaboration experience, audio is critical. For example, during a video call, if video is lost or distorted due to a network outage or some other network related event, communication can continue provided that audio is not lost during that outage. This is a critical concept in bandwidth management in the PA.

"Smart" Media Techniques (Media Resilience and Rate Adaptation)

When deploying video pervasively across an organization, administrators will inevitably encounter insufficient bandwidth to handle the load of video required during the busy hour in some bottleneck areas of the Wide Area Network (WAN). In light of this it is important to prioritize video correctly, to ensure that audio is not affected by any video packet loss that may occur, and to ensure that certain types of video can leverage video rate adaptation to manage the amount of bandwidth used during times of congestion. The media resilience and rate adaptation techniques allow for an optimized video experience in the face of congestion and packet loss over managed and unmanaged networks, but that is not all. These techniques, when used as a strategy coupled with QoS, can offer the ability for an organization to deploy video pervasively by allowing the endpoints to reduce their bit rate and thus their bandwidth utilization during congestion and packet loss, while also allowing the endpoints to increase their bit rate and thus bandwidth utilization during more idle times of the day outside of the busy hour, thereby maximizing video quality.

Every Cisco video endpoint employs a number of smart media techniques to avoid network congestion, recover from packet loss, and optimize network resources. The following smart media techniques are some of the techniques employed by Cisco video endpoints:

- Media resilience techniques
  - Encoder pacing
  - Gradual Decoder Refresh (GDR)
  - Long-Term Reference Frame (LTRF) with Repair
  - Forward Error Correction (FEC)
- Rate adaptation

Media Resilience Techniques

Table 6-1 summarizes the media resilience techniques supported on various Cisco video endpoints.

Table 6-1 Media Resilience Support in Cisco Collaboration Video Endpoints

Video Endpoint or Bridge	Encoder Pacing	Rate Adaptation	FEC	LTRF Repair
Cisco IP Phone 8800 Series	Yes	No	No	No
Cisco Jabber	Yes	Yes	Yes	Yes
Cisco DX Series	Yes	Yes	No	No
Cisco TelePresence MX Series	Yes	Yes	Yes	Yes
Cisco TelePresence SX Series	Yes	Yes	Yes	Yes
Cisco TelePresence IX Series	No	No	No	No
Cisco TelePresence Server	Yes	Yes	Yes	Yes

Encoder Pacing

Encoder pacing is a simple technique used to spread the packets as evenly as possible in order to smooth out the peaks of the bursts of bandwidth.

Gradual Decoder Refresh (GDR)

GDR is a method of gradually refreshing the picture over a number of frames, giving a smoother, less bursty bit stream.

Long Term Reference Frame (LTRF)

A Long Term Reference Frame (LTRF) is a reference frame stored in the encoder and decoder, which allows the video endpoints to recover more efficiently from packet loss with less bandwidth utilization over the network path in order to resend lost frames.

Forward Error Correction (FEC)

Forward error correction (FEC) provides redundancy to the transmitted information by using a predetermined algorithm. The redundancy allows the receiver to detect and correct a limited number of errors occurring anywhere in the message, without the need to ask the sender for additional data. FEC gives the receiver an ability to correct errors without needing a reverse channel (such as RTCP) to request retransmission of data, but this advantage is at the cost of a fixed higher forward channel bandwidth (more packets sent). FEC protects the most important data (typically the repair P-frames) to make sure the receiver is receiving those frames. The endpoints do not use FEC on bandwidths lower than 768 kbps, and there must also be at least 1.5% packet loss before FEC is introduced. Endpoints typically monitor the effectiveness of FEC; and if FEC is not efficient, they make a decision not to do FEC.

Rate Adaptation

Rate adaptation or dynamic bit rate adjustments adapt the call rate to the variable bandwidth available, down-speeding or up-speeding the video bit rate based on the packet loss condition. An endpoint will reduce bit rate when it receives messages from the receiver indicating there is packet loss; and once the packet loss has decreased, up-speeding of the bit rate will occur.

Opportunistic Video and Prioritized Audio

Opportunistic Video and Prioritized Audio is a concept and a QoS strategy combined. It consists of defining a number of video endpoints that opportunistically utilize available video bandwidth. During the busy hour these endpoints rate adapt or throttle down their bit rate to accommodate limited bandwidth availability, all the while not impacting prioritized video. During the idle hour they utilize available bandwidth to optimize video quality by increasing the video bit rate. Prioritized audio for both audio-only and audio of video calls ensures that all audio is prioritized in the network and is thus not impacted by any loss that can occur in the video queues. Prioritizing voice from all types of collaboration media ensures that even during times of extreme congestion when video is experiencing packet loss and adjusting to that loss, the audio streams are not experiencing packet loss and are allowing the user to carry on an uninterrupted audio experience. Prioritizing audio from both voice-only and video calls is a paradigm shift from the previous historic model where audio and video of video calls were always marked with the same QoS. Opportunistic video with prioritized audio maintains an acceptable video experience while simultaneously ensuring that voice media for voice-only and video calls is not compromised. This of course applies to the managed network, since an unmanaged network such as the Internet is not QoS enabled and thus provides no guarantees with regard to packet loss. Nevertheless, the media resiliency and rate adaptation mechanisms also attempt to ensure that media over unmanaged networks has the best possible quality in the face of packet loss, delay, and jitter.

QoS Architecture for Collaboration

Quality of Service (QoS) ensures reliable, high-quality voice and video by reducing delay, packet loss, and jitter for media endpoints and applications. QoS provides a foundational network infrastructure technology, which is required to support the transparent convergence of voice, video, and data networks. With the increasing amount of interactive applications (particularly voice, video, and immersive applications), real-time services are often required from the network. Because these resources are finite, they must be managed efficiently and effectively. If the number of flows contending for such priority resources were not limited, then as these resources become oversubscribed, the quality of all real-time traffic flows would degrade, eventually to the point of futility. "Smart" media techniques, QoS, and admission control ensure that real-time applications and their related media do not oversubscribe the network and the bandwidth provisioned for those applications. These smart media techniques coupled with QoS and, where needed, admission control, are a powerful set of tools used to protect real-time media from non-real-time network traffic and to protect the network from over-subscription and the potential loss of quality of experience for end users of voice and video applications.

Identification and Classification

QoS Trust and Enforcement

The enforcement of QoS is crucial to any real-time audio, video, or immersive video experience. Without the proper QoS treatment (classification, prioritization, and queuing) through the network, real-time media can potentially incur excessive delay or packet loss, which compromises the quality of the real-time media flow. In the QoS enforcement paradigm, the issue of trust and the trust boundary is equally important. Trust is the permitting or the "trusting" of QoS marking (Layer 2 CoS or Layer 3 IP DSCP) of the traffic by the endpoint or device, to allow the traffic to continue through the network. The trust boundary is the place in the network where the trust occurs. It can occur at any place in the network; however, we recommend enforcing trust at the network edge such as the LAN access ingress or the WAN edge, or both, as is feasible and applicable.

There are three main categories of trust:

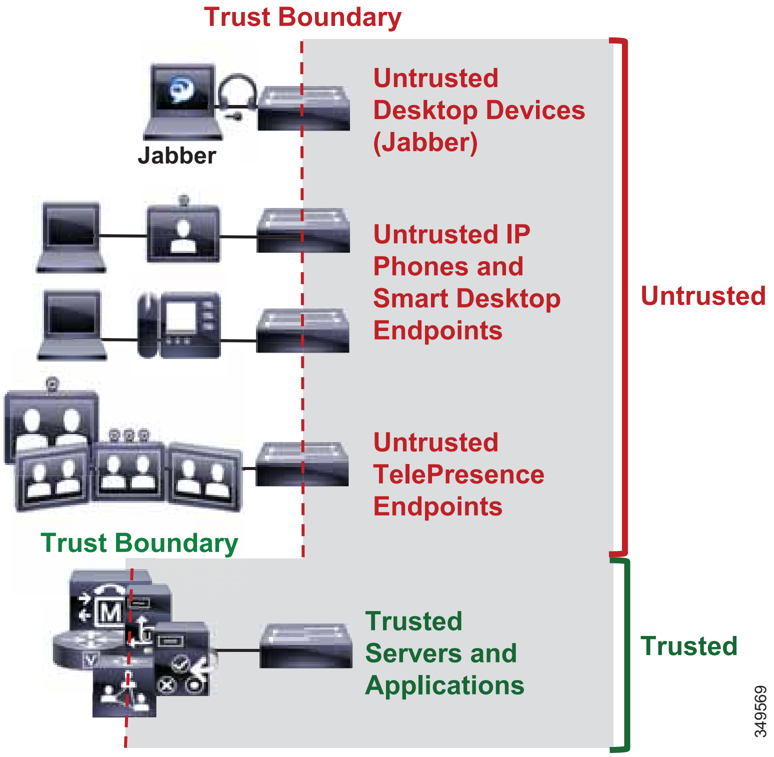
- **Untrusted** — These devices include unsecure PCs, Macs, or hand-held mobile devices running Jabber clients, IP phones, and smart desktop endpoints.
- **Trusted** — These devices can include secure PCs and servers, video conferencing endpoints, analog and video conferencing gateways, PSTN gateways, Cisco Unified Border Element, trusted application servers, and other similar devices.
- **Conditionally trusted** — These devices typically include endpoints that support Cisco Discovery Protocol (CDP). Although Cisco TelePresence, IP Phone and Smart Desktop phones support CDP, they are not conditionally trusted in the PA.

In the PA, trusted and untrusted switch ports are used but conditionally trusted ports are not used. Conditional trust is not recommended in the PA for the following reasons:

- Complexity across a variety of switches — Enabling conditional trust across a variety of switch types can become complex. Some of the older switch types do not trust by default, while newer switches do trust by default. Furthermore, the commands for enabling trust and the process of trust enforcement are different across platforms.
- Even more important is the lack of Layer 3 DSCP re-marking on PC ports of IP phones and smart desktop endpoints. The endpoints re-mark only Layer 2 CoS. Because of this and the inability to correctly re-mark PC traffic at the DSCP level, using access lists to re-mark IP phones and smart desktop endpoints is a preferred method in the PA.
- A single ACL that maps directly to all switch ports is easier to manage than specifying only a limited number of ports for trust.

Figure 6-6 illustrates the types of trust used in the PA, and which devices are trusted and untrusted.

Figure 6-6 Trust Boundaries in the Preferred Architecture


Trust Boundary
Jabber
Untrusted Desktop Devices (Jabber)
Untrusted IP Phones and Smart Desktop Endpoints
Untrusted
Untrusted TelePresence Endpoints
Trust Boundary
Trusted Servers and Applications
Trusted
349569

Classification and Marking

This section discusses classification and marking for endpoints.

All Cisco endpoints derive their DSCP marking from Unified CM. Unified CM houses the QoS configuration for endpoints in two places, in the Service Parameters for the CallManager service (**Clusterwide Parameters (System - QoS)**) and in the SIP Profile (applicable only to SIP devices). The SIP Profile configuration of QoS settings overrides the Service Parameter configuration. This allows Unified CM administrators to set different QoS policies for groups of endpoints. Unified CM passes this QoS configuration to the endpoints in a configuration file over TFTP during endpoint registration. This configuration file contains the QoS parameters as well as a number of other endpoint specific parameters. For QoS purposes there are two categories of video endpoints: TelePresence endpoints (any endpoint with TelePresence in the phone type name) and all other non-TelePresence video endpoints (referred to as "UC Video endpoints").

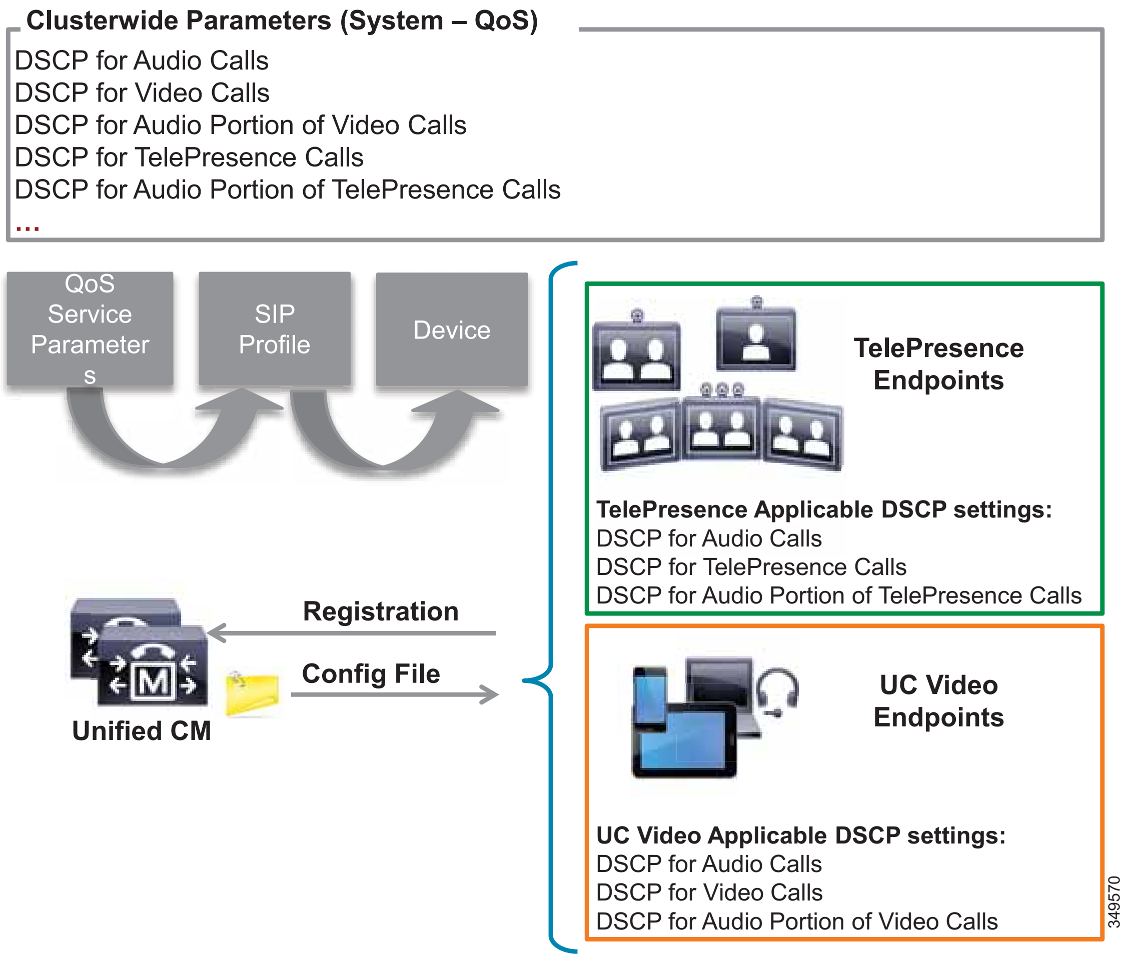
Table 6-2 shows the Preferred Architecture endpoints and their classification.

Table 6-2 PA Video Endpoints

Endpoint	TelePresence Endpoint	UC Video Endpoint
Cisco IP Phone 8800 Series		X
Cisco Jabber		X
Cisco DX Series		X
Cisco TelePresence MX Series	X	
Cisco TelePresence SX Series	X	
Cisco TelePresence IX Series	X	

Figure 6-7 illustrates how the two categories of Cisco video endpoints derive DSCP. These categories apply only to QoS and call admission control (CAC).

Figure 6-7 How Cisco Endpoints Derive DSCP


Clusterwide Parameters (System – QoS)
DSCP for Audio Calls
DSCP for Video Calls
DSCP for Audio Portion of Video Calls
DSCP for TelePresence Calls
DSCP for Audio Portion of TelePresence Calls
...
QoS Service Parameters
SIP Profile
Device
Registration
Config File
Unified CM
TelePresence Endpoints
TelePresence Applicable DSCP settings:
DSCP for Audio Calls
DSCP for TelePresence Calls
DSCP for Audio Portion of TelePresence Calls
UC Video Endpoints
UC Video Applicable DSCP settings:
DSCP for Audio Calls
DSCP for Video Calls
DSCP for Audio Portion of Video Calls
349570

The configuration file is populated with the QoS parameters from the CallManager service parameters or the SIP Profile, when configured, and sent to the endpoint upon registration. The endpoint then uses the correct DSCP parameters for each type of media stream, depending on which category of endpoint it is. Table 6-3 lists the DSCP parameters, the type of endpoint, and the type of call flow determining the DSCP marking of the stream.

Table 6-3 DSCP for Basic Call Flows

DSCP Parameter	TelePresence Endpoint	UC Video Endpoint	Call Flow
DSCP for Audio Calls	X	X	Voice-only
DSCP for Video Calls		X	Video – Audio and video stream of a video call, unless the endpoint supports the **DSCP for Audio Portion of Video Calls** parameter (see Table 6-4)
DSCP for Audio Portion of Video Calls		X	Audio stream of a video call – Applicable only to endpoints that support the parameter
DSCP for TelePresence Calls	X		Immersive video – Audio and video stream of an immersive video call, unless the endpoint supports the **DSCP for Audio Portion of TelePresence Calls** parameter (see Table 6-4)
DSCP for Audio Portion of TelePresence Calls	X		Audio stream of a video call – Applicable only to endpoints that support the parameter

Table 6-4 Endpoint Support for DSCP for Audio Portion of Video and TelePresence Calls

Video Endpoint	DSCP for Audio Portion of Video Call	DSCP for Audio Portion of TelePresence Call
IP Phone 8845 and 8865 Series	Yes	No
DX Series	Yes	No
TelePresence IX Series	No	Yes
TelePresence SX and MX Series (CE 8.0 or later firmware)	No	Yes

Trusted Core Devices and Applications

Like endpoints, devices and applications in the collaboration portfolio source and terminate media and signaling streams. These trusted applications require the appropriate configuration on the application itself as well as the switch to which the application is connected in order to transparently pass the QoS marking of the media and signaling.

Core trusted devices and applications:

- Cisco Unified Communications Manager and IM and Presence Service
- Cisco Expressway
- Cisco Unity Connection
- Cisco TelePresence Server
- Cisco TelePresence Conductor
- Cisco IOS SIP Gateway and Cisco Unified Border Element

It is important to ensure that DSCP Trust is enabled on the switch ports to which these endpoints and application servers are connected. QoS DSCP trust is typically enabled by default on all newer Cisco switches; however, it is important to verify each switch platform to determine if this QoS trust is enabled, since some platforms do not trust DSCP by default.

Endpoints and Clients

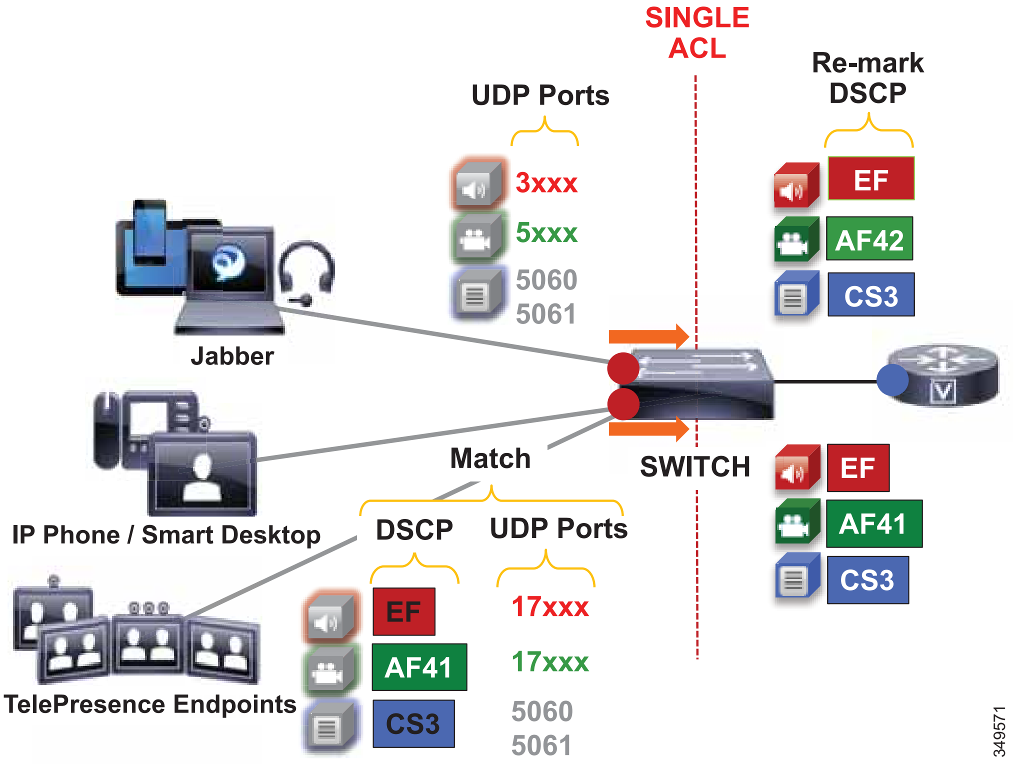
For the endpoints, the DSCP marking of packets on ingress into the switch needs to be re-marked using network access control lists (ACLs) to ensure that collaboration media and SIP signaling are marked appropriately and that other PC data traffic is marked to DSCP of Best Effort (DSCP 0) or as is appropriate based on the QoS data policies set forth by the enterprise.

The method used here consists of mapping identifiable media and signaling streams based on specific protocol ports, such as UDP and TCP ports, then making use of network access lists to remark QoS of the signaling and media streams based on those protocol port ranges. This method applies to all Cisco Jabber clients (Cisco Jabber for Windows, Cisco Jabber for Mac OS, Cisco Jabber for iPhone, Cisco Jabber for iPad, and Cisco Jabber for Android) because they all behave similarly when allocating media and signaling port ranges. Unlike Cisco Jabber clients, endpoints such as IP phones, 8800 Series IP and video phones, and DX Series with PC ports require an additional measure of matching on DSCP as well as UDP ports. The reason for this is that the IP phones and video endpoints use the same UDP port range for both audio and video, and thus in order to differentiate audio and video, matching on both UDP port range and DSCP allow for the proper identification of media traffic.

The concept is simple. An access list is used in the network access layer equipment (switch) to identify the media and signaling streams based on UDP port ranges and DSCP matching, and then it is set to re-mark them to the appropriate DSCP values. Although this technique is easy to implement and can be widely deployed, it is however not a 100% secure method and this point should be noted. This PA assumes that other security measures will be implemented to ensure the correct access to the network as well as any securing of the operating systems (OS) on the PCs and Macs used for Jabber to impede user tampering of OS related QoS settings.

Figure 6-8 illustrates the use of network access control lists (ACLs) to map identifiable media and signaling streams to DSCP for Jabber clients.

Figure 6-8 Endpoint Marking



Example 6-1 Switch ACL-Based QoS Policy for Untrusted Endpoints in Figure 6-8:

- Jabber clients
  - Match UDP Port Range 3xxx –> Re-mark to DSCP EF
  - Match UDP Port Range 5xxx –> Re-mark to DSCP AF42
  - Match TCP Port 5060 or 5061 –> Re-mark to DSCP CS3
- IP phones and video endpoints
  - Match UDP Port Range 17xxx with DSCP EF –> Re-mark to DSCP EF
  - Match UDP Port Range 17xxx with DSCP AF41 –> Re-mark to DSCP AF41
  - Match TCP Port 5060 or 5061 –> Re-mark to DSCP CS3
- Generic matching
  - Matches the rest of the traffic and sets DSCP to 0 (Best Effort or BE) using a default class-map

Endpoints send and receive other data and signaling such as ICMP, DHCP, TFTP, BFCP, LDAP, XMPP, FECC, CTI, and so forth. The QoS values for this traffic should follow the enterprise best practices for each type of traffic. Without doing this step, all other traffic apart from media and SIP signaling will be set to a DSCP of BE (DSCP 0) by the class-default in this configuration. We recommend either passing through the traffic marking by matching on DSCP and then re-marking the DSCP to the same value, or else using the TCP and UDP ports for each protocol that the endpoints use for communications.

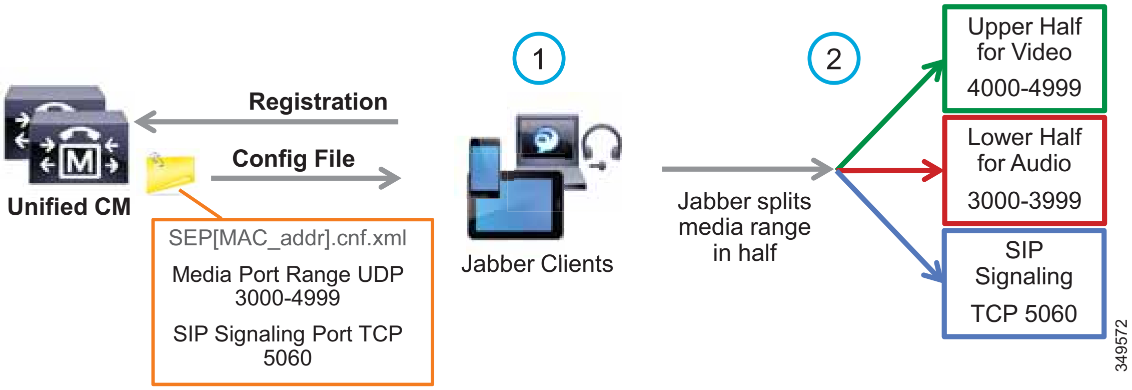
QoS for Cisco Jabber Clients

As discussed, this method involves classifying media and signaling by identifying the various streams from the Jabber client based on IP address, protocol, and/or protocol port range. Once identified, the signaling and media streams can be classified and re-marked with a corresponding DSCP. The protocol port ranges are configured in Unified CM and are passed to the endpoint to use during device registration. The network can then be configured via access control lists (ACLs) to classify traffic based on IP address, protocol, and protocol port range, and then to re-mark the classified traffic with the appropriate DSCP as discussed in the preceding section.

Cisco Jabber provides identifiable media streams based on UDP protocol port ranges and identifiable signaling streams based on TCP protocol port ranges. In Unified CM, the signaling port for endpoints is configured in the SIP Security Profile, while the media port range is configured in the SIP Profile of the Unified CM administration pages.

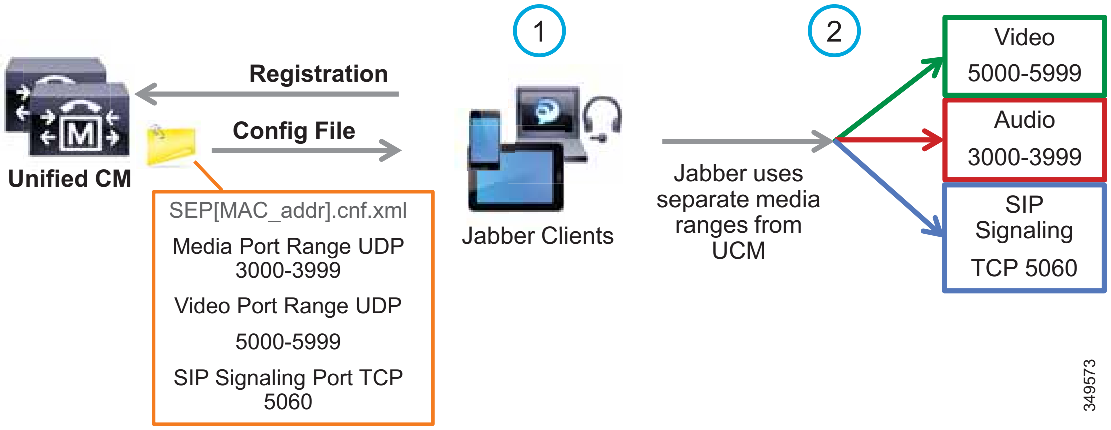
For the media port range, all endpoints and clients use the SIP profile parameter **Media Port Ranges** to derive the UDP ports used for media. By default, media port ranges are configured with **Common Port Range for Audio and Video**. When Jabber clients receive this port range in their configuration file, they split the port range in half and use the lower half for the audio streams of both voice and video calls and the upper half for the video streams of video calls. This is illustrated in Figure 6-9.

Figure 6-9 Media and Signaling Port Range – Common



Jabber can also use the **Media Port Ranges** > **Separate Port Range for Audio and Video** configuration. In this configuration the Unified CM administrator can specify a non-contiguous audio and video port range, as illustrated in Figure 6-10.

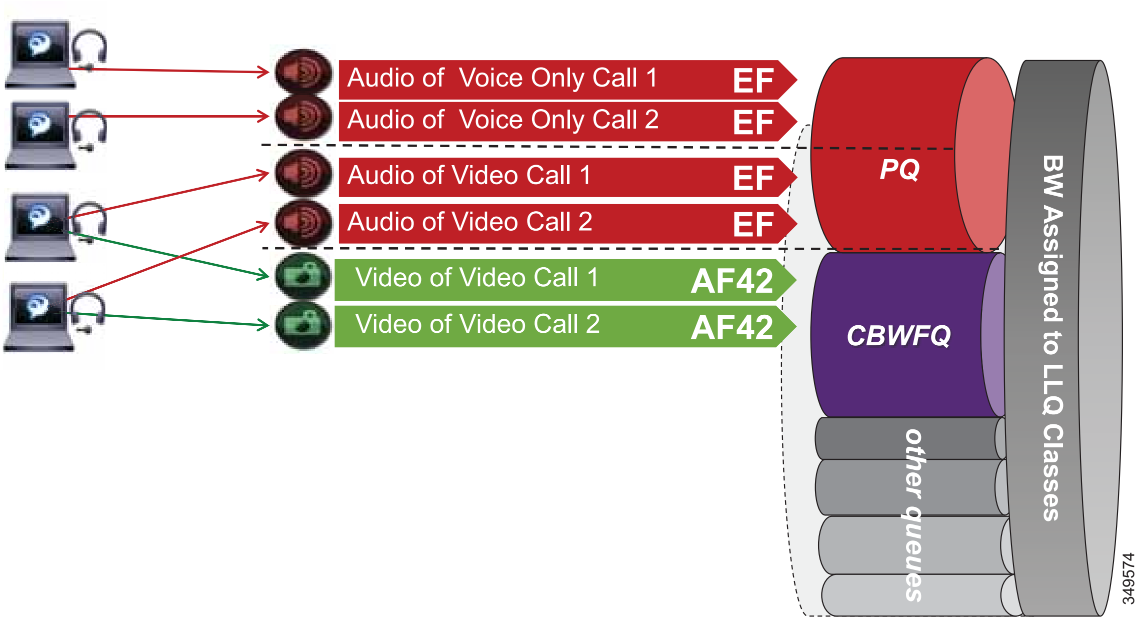
Figure 6-10 Media and Signaling Port Range – Separate



Caution **Security Alert**: If you use identifiable media streams for QoS classification at the network level, the trust model does *not* extend to the application itself. Apart from prioritizing streams from the intended application, other applications *could* potentially be configured to use the same identification criteria (media port range) and marking, and therefore achieve network prioritization. Because this unintended traffic would not be accounted for in call admission control or in the provisioning of the network, severe overall impact to real-time conversations could occur. It is for this reason that it is a good practice to define restricted port ranges whenever possible to identify the media streams.

When utilizing this technique, it is important to ensure that the audio portion of these video calls that will be re-marked to the audio traffic class (EF), and the video portions that will be re-marked to the video traffic class (AF4), are provisioned in the network accordingly. Figure 6-11 is an example of placing audio traffic into a Priority Queue (PQ) and video traffic into a Class Based Weighted Fair Queue (CBWFQ). The combination of PQ and CBWFQ is often referred to as low latency queuing (LLQ). Note that, because it is not possible to use port ranges in Cisco Jabber endpoints to differentiate the audio portion of voice-only calls from the audio portion of video calls, all audio using this technique will be re-marked to EF. It is important to provision the PQ adequately to support voice-only calls and the audio portion of video calls. An example of such provisioning is illustrated in Figure 6-11. For more information on the design and deployment recommendations for provisioning queuing and scheduling in the network, see the WAN Queuing and Scheduling section.

Figure 6-11 Provisioning Jabber QoS in the Network



According to RFC 3551, when RTCP is enabled on the endpoints, it uses the next higher odd port. For example, a device that establishes an RTP stream on port 3500 would send RTCP for that same stream on port 3501. This function of RTCP is also true with all Jabber clients. RTCP is common in most call flows and is typically used for statistical information about the streams and to synchronize audio and video in video calls to ensure proper lip-sync. In most cases, video and RTCP can be enabled or disabled on the endpoint itself or in the common phone profile settings.

Utilizing the Network for Classification and Marking

Based on the identifiable media and signaling streams created by the endpoints, common network QoS tools can be used to create traffic classes and to re-mark packets according to those classes.

These QoS mechanisms can be applied at different layers, such as the access layer (access switch), which is closest to the endpoint and the router level in the distribution, core, or services WAN edge. Regardless of where classification and re-marking occurs, we recommend using DSCP to ensure end-to-end per-hop behaviors.

As previously mentioned, Cisco Unified CM allows the port range utilized by SIP endpoints to be configured in the SIP Profile. As a general rule, a port range of a minimum of 100 ports (for example, 3000 to 3099) is sufficient for most scenarios. A smaller range could be configured as long as there are enough ports for the various audio, video, and associated RTCP ports (RTCP runs over the odd ports in the range), as well as ensuring against port conflict with other applications on the operating system of the device that may be using these ports since this could cause port collisions.

Access Layer (Layer 2 Definitions)

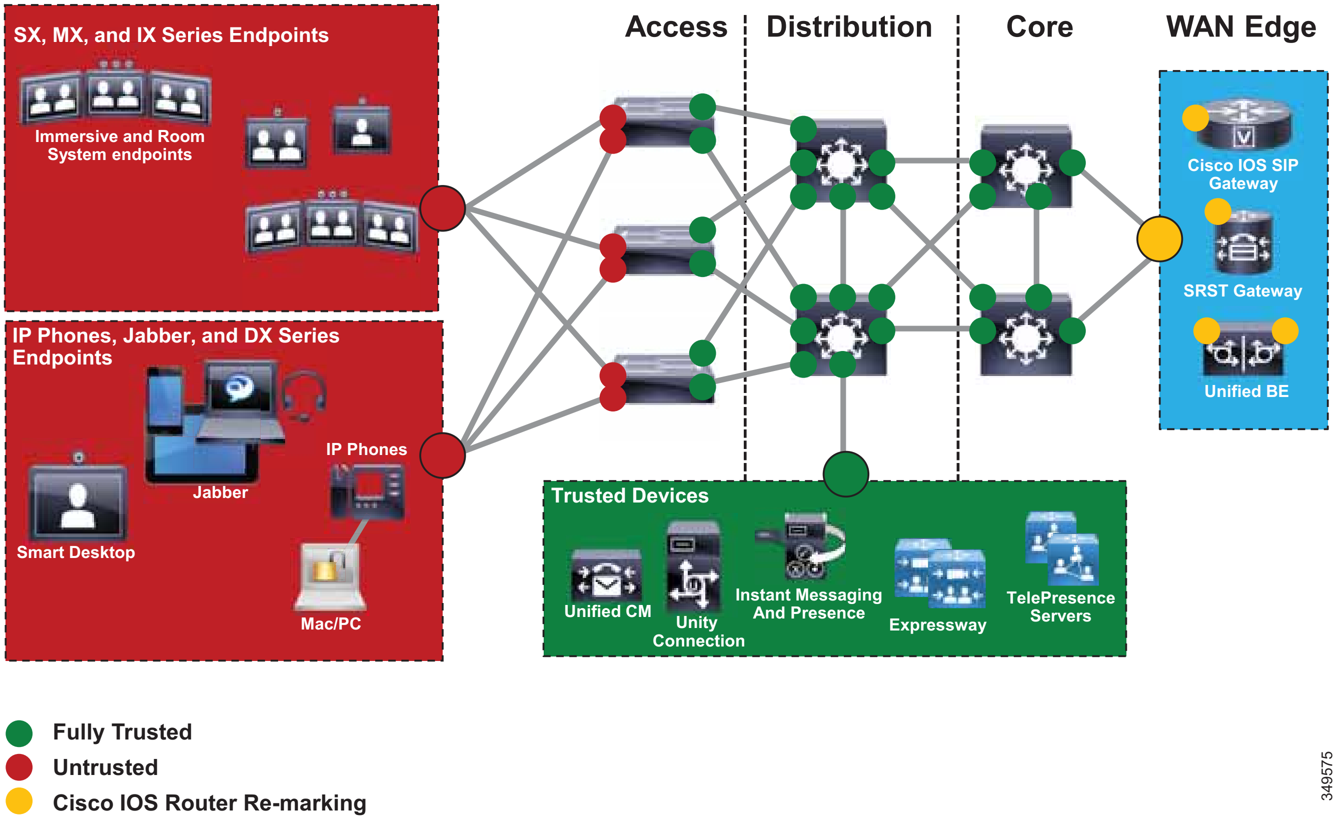
When utilizing the access layer to classify traffic, the classification occurs at the ingress of traffic into the network, thus allowing the flows to be identified as they enter. In environments where QoS policies are applied not only in the WAN but also within the LAN, all upstream components can rely on traffic markings when processing. Classification at the ingress allows different methods to be utilized based on different types of endpoints.

Configuring QoS policies in the access layer of the network could result in a significant amount of devices that require configuration, which can create additional operational overhead. The QoS policy configurations should be standardized across the various switches of the access layer through templates. You can use configuration deployment tools to relieve the burden of manual configuration. The PA simplifies this process by using a single group of ACLs that can be used across the various switching platforms.

Distribution/Core/Services WAN Edge (Layer 3 Definitions)

Another location where QoS marking can take place is at the Layer 3 routed boundary. In a campus network Layer 3 could be in the access, distribution, core, or services WAN edge layers. The recommendation is to classify and re-mark at the access layer, then trust through the distribution and core of the network, and finally re-classify and re-mark at the WAN edge if and when needed. For smaller networks such as branch offices where there are no Layer 3 switching components deployed, QoS marking can be applied at the WAN edge router. At Layer 3, QoS policies are applied to the Layer 3 routing interfaces. In most campus networks these would be VLAN interfaces, but they could also be Fast Ethernet or Gigabit Ethernet interfaces. Figure 6-12 illustrates the areas of the network where the various types of trust are applied in relation to the places in the network – access, distribution, core, or WAN edge.

Figure 6-12 Trust and Enforcement – Places in the Network



Endpoint Identification and Classification Considerations and Recommendations

Summary of design and deployment considerations and recommendations:

- Use DSCP markings whenever possible because these are IP layer end-to-end, more granular, and more extensible than Layer 2 markings.
- Mark as close to the endpoint as possible, preferably at the LAN switch level.
- When trying to minimize the number of media ports used by the Cisco Jabber client, a minimum range of 100 ports is recommended. This is to ensure that there are enough ports for all of the streams, such as RTCP, RTP for audio and video, BFCP, and RTP for secondary video for desktop sharing sessions, as well as to avoid any overlap with other applications on the same computer.
- Ensure a QoS policy includes other pertinent collaboration traffic to be re-marked, otherwise a value of 0 (Best Effort, or BE) will be placed on all remaining traffic.

WAN Queuing and Scheduling

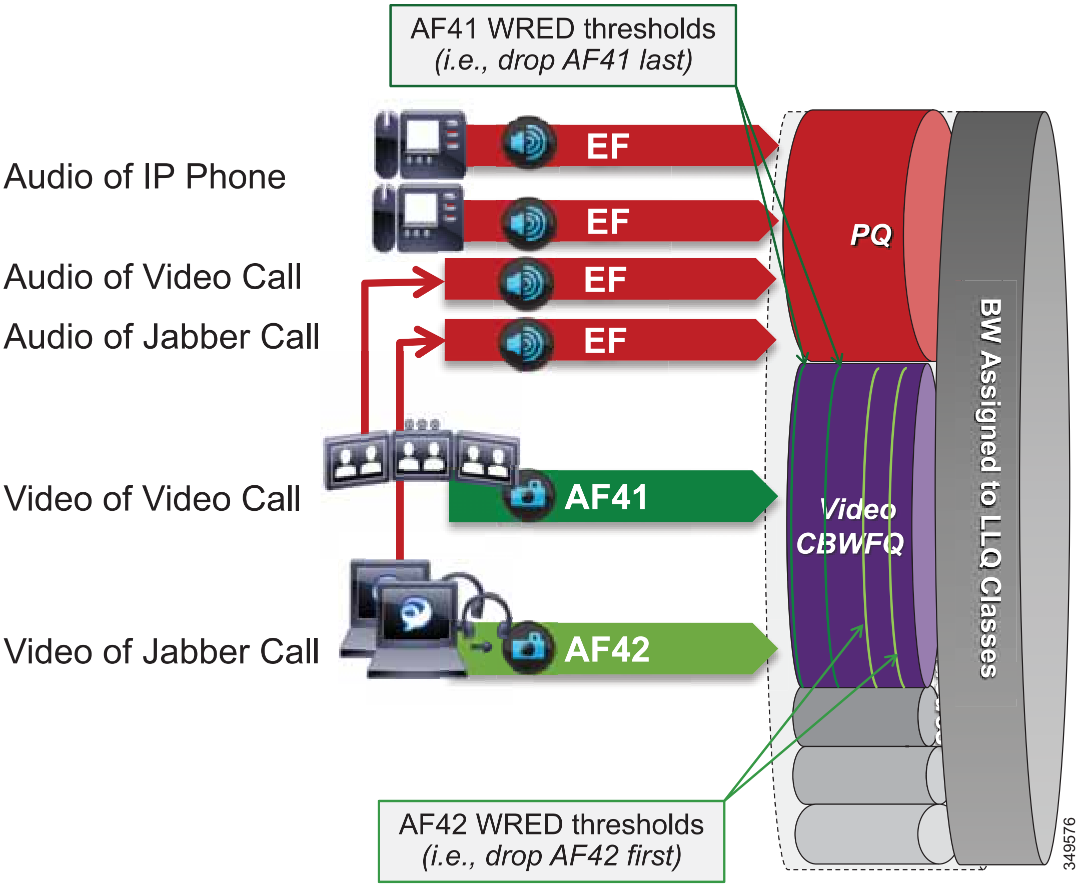
As discussed in the Identification and Classification section, Unified CM has the ability to differentiate the video endpoint types as well as the their media streams. This provides the network administrator the ability to treat the video from different endpoints differently in the network. The recommended approach in the PA is to use AF42 DSCP markings for Jabber clients and AF41 for all smart desktop, room system, and immersive video endpoints. These values are in line with RFC 4594.

PA Queuing and Scheduling Approach (Single Video Queue)

A single rate-based queue with multiple DSCPs with differing drop probabilities is used in the PA for managing multiple types of video across an integrated collaboration media and data network. In this approach to scheduling video traffic in the WAN, the single video queue is configured with two AF4 drop probabilities using AF41 and AF42, where AF42 has a higher drop precedence or probability than AF41. The premise behind a single video queue with this service class with hierarchical drop precedence is that, when one class of video is not using the bandwidth within the queue, the rest of the queue bandwidth is available for the other class of video. This approach dedicates bandwidth reserved for video to be available only to the video queue during times of congestion. Other approaches such as dual rate-based video queues sub-optimally allocate excess video bandwidth from one queue to all queues on the interface equally.

Although different strategies for optimized video bandwidth utilization can be designed based on this single video queue with hierarchical DSCP drop probabilities, the PA approach is illustrated in Figure 6-13.

Figure 6-13 Single Video Queue Approach



In Figure 6-13 the audio of a voice call is marked as EF and placed into a Priority Queue (PQ) with a strict policer on how much bandwidth the PQ can allocate to this traffic. Video calls are separated into two classes, AF41 for prioritized video and AF42 for opportunistic or Jabber video. Using a CBWFQ with Weighted Random Early Detection (WRED), the administrator can adjust the drop precedence of AF42 over AF41, thus ensuring that during times of congestion when the queue is filling up, AF42 packets are dropped from the queue at a higher probability than AF41. See the *WAN Quality of Service* section in the Network Infrastructure chapter of the Collaboration SRND for more details on the function of WRED.

The above example illustrates how an administrator using a single CBWFQ with DSCP-based WRED for all video can protect one type of video (TelePresence video) from packet loss during periods of congestion over another type of video (desktop). With this "single video queue approach," unlike the "dual video queue approach," when one type of video is not using bandwidth in the queue, the other type of video gains full access to the entire queue bandwidth if and when needed. This is a significant improvement when deploying pervasive video.

Achieving this holistically across the entire solution depends on a number of conditions. Below is a list of conditions required to achieve marking all audio to a DSCP of EF:

- The customer equipment (CE) or service provider (SP) owned WAN equipment must support AF4 QoS containing both AF41 and AF42 QoS markings as well as Weighted Random Early Detection (WRED).
- Enhanced Locations Call Admission Control (ELCAC) can be implemented in conjunction with marking all audio as EF. ELCAC relies on the correct DSCP setting in order to ensure the protection of the queues that voice and video CAC pools represent. Changing the DSCP of audio streams of the video calls requires updating how ELCAC deducts bandwidth for video calls. This can be done by setting the service parameter **Deduct Audio Bandwidth from Audio Pool for Video Call**, under the Call Admission Control section of the CallManager service, to **True**. This parameter can be set to true or false:
  - **True** (recommended) — Cisco Unified CM splits the audio and video bandwidth allocations for video calls into separate pools. The bandwidth allocation for the audio portion of a video call is deducted from the audio pool, while the video portion of a video call is deducted from the video pool.
  - **False** (default)— Cisco Unified CM applies the legacy behavior, which is to deduct the audio and video bandwidth allocations of video call from the video pool. This is the default setting.

Opportunistic Video

When video is deployed pervasively across the organization, bandwidth constraints typically determine the video resolution that can be achieved during the busiest hour of the day based on the bandwidth available and the number of video calls during that busy hour. To address this challenge, the PA has targeted a group of endpoints whose video is treated opportunistically by the network by using a single video queue with DSCP-based WRED coupled with a strategy for identification and classification of the Jabber clients' collaboration media.

Opportunistic video is achieving the best video quality based on the WAN bandwidth resources available at any given time. To achieve this, a number of requirements must be met:

- Select a group of video endpoints to be opportunistic. In the case of the PA, Jabber clients are used as the opportunistic video endpoints.
- Ensure the WAN is configured with a single video queue using DSCP-based WRED with AF4 DSCP class servicing with drop precedence of AF41 and AF42. (While AF43 could be used, only two DSCP values are necessary in the PA.)
- Identify and classify the video of opportunistic endpoints with AF42.
- Identify and classify all other video endpoints with AF41.

Provisioning and Admission Control

Provisioning bandwidth and ensuring the correct bit rate is negotiated between various groups of endpoints are important aspects of bandwidth management. In a Unified CM environment, bit rate is negotiated through Unified CM, which uses a concept of regions to set maximum audio and maximum video bit rates for any given call flow. This section focuses on the maximum bit rate for video and TelePresence calls.

Unified CM locations (see Enhanced Locations Call Admission Control) work in conjunction with regions to define the characteristics of a call flow. Regions define the type of compression or bit rate (8 kbps or G.729, 64 kbps or G.722/G.711, and so forth) that is used between any two devices. Location links define the amount of available bandwidth for the path between devices. Each device and trunk in the system is assigned to both a region (by means of a device pool) and a location (by means of a device pool or by direct configuration on the device itself):

- Regions allow the bandwidth of video calls to be set. The audio limit on the region can result in filtering out codecs with higher bit rates. However, for video calls, the video limit constrains the quality (resolution and transmission rate) of the video.
- Locations define the amount of total bandwidth available for all calls on that link. When a call is made on a link, the regional value for that call must be subtracted from the total bandwidth allowed for that link.

Building a region matrix to manage maximum video bit rate (video resolution) for groups of devices can assist in ensuring that certain groups of devices do not over-saturate the network bandwidth. The following guidelines apply to creating a region matrix:

- Group devices into maximum video bit rate categories.
- The smaller the number of groups, the easier it is to calculate bandwidth requirements.
- Consider the default region settings to simplify the matrix and provide intra-region and inter-region defaults.
- Use a single audio codec across the entire organization, both LAN and WAN, to simplify the region matrices.

For more information about region settings, see the section on Enhanced Locations Call Admission Control.

Table 6-5 lists an example of a maximum video session bit rate region matrix for three groups of devices.

Note Table 6-5 is only an example of how to group devices and what maximum bit rate might be suggested for a general resolution between the groups of devices.

Table 6-5 Example Group Region Matrix for Three Groups of Devices

Endpoint Groupings	Video_1.5MB	Video_2.5MB	Video_20MB
Video_1.5MB	1,500 kbps	1,500 kbps	1,500 kbps
Video_2.5MB	1,500 kbps	2,500 kbps	2,500 kbps
Video_20MB	1,500 kbps	2,500 kbps	20,000 kbps

For the example in Table 6-5, the three groups are:

- Video_1.5MB

  These devices would typically be the largest group of deployed video capable endpoints and thus would benefit from the opportunistic video approach. Classified as opportunistic video, they can go up to a maximum of 1,500 kbps (720p @ 30 fps) and will rate-adapt downward based on packet loss.
- Video_2.5MB

  These devices would be room systems such as the Cisco TelePresence MX or SX Series as well as smart desktop endpoints such as the Cisco DX Series. At 2,500 kbps maximum video bit rate, these endpoints would typically be capable of 720p @ 30 fps.

- Video_20MB

  This class is for the larger Cisco TelePresence IX Series endpoints as well as TelePresence Servers and MCUs (Cisco Conductor SIP trunk) set to a maximum of 20 Mbps to allow for endpoints capable of it to run at 1080p @ 60 fps. (The IX Series, for example, requires 18 MB to do 1080p @ 60 fps on three screens). Single-screen systems use much less bandwidth, but this group would be for devices utilizing their maximum bit rate capacity.

To simplify the configuration of the regions, it is important to standardize on one audio codec to be used throughout the entire organization. The first consideration is to decide whether to have a lower bit rate codec for audio calls between sites. Historically as part of managing bandwidth, enterprises have used a lower bit rate codec such as G.729 over the WAN while using a higher bit rate, wider band codec such as G.722 for calls within the LAN or MAN. Typically when deploying video at 1 to 2.5 MB per call, audio (even at 80 kbps per call) consumes so much less bandwidth that many enterprises have moved to using a higher bit rate, better quality codec such as G.722 across the entire organization (LAN and WAN). This decision has an impact on the region matrix and whether per-site regions are required or not. The concept here is that if inter-region audio or video bit rates are to be different, then per-site regions will be required. This augments the configuration of regions to the number of sites (N) multiplied by the number of video groups (X):

Number of regions required on average = $N*X$

If audio bit rates will be the same across the WAN and LAN, then only the regions for the video groups are required (X).

Enhanced Locations Call Admission Control

The call admission control function is an important component of a Collaboration system, especially when multiple sites are connected through an IP WAN and limited bandwidth resources are available for audio and video calls.

Call Admission Control Architecture

Unified CM Enhanced Locations Call Admission Control

Cisco Unified CM provides Enhanced Locations Call Admission Control (ELCAC) to support complex WAN topologies as well as distributed deployments of Unified CM for call admission control where multiple clusters manage devices in the same physical sites using the same WAN up-links.

To support more complex WAN topologies, Unified CM implements a location-based network modeling functionality. This provides Unified CM with the ability to support multi-hop WAN connections between calling and called parties. This network modeling functionality has also been incrementally enhanced to support multi-cluster distributed Unified CM deployments. This allows each cluster to "share" locations by enabling the clusters to communicate with one another to reserve, release, and adjust allocated bandwidth for the same locations across clusters.

Network Modeling with Locations, Links, and Weights

Enhanced Locations CAC is a model-based static CAC mechanism. ELCAC involves using the administration interface in Unified CM to configure locations and links to model the "routed WAN network" in an attempt to represent how the WAN network topology routes media between groups of endpoints for end-to-end audio and video calls. Although Unified CM provides configuration and serviceability interfaces in order to model the network, it is still a "static" CAC mechanism that does not take into account network failures and network protocol rerouting. Therefore, the model needs to be updated when the WAN network topology changes or bandwidth allocations across the WAN are increased or decreased. Enhanced Locations CAC is also call oriented, and bandwidth deductions are per-call not per-stream, so asymmetric media flows where the bit-rate is higher in one direction than in the other will always deduct for the highest bit rate bi-directionally. In addition, unidirectional media flows will be deducted as if they were bidirectional media flows.

Enhanced Locations CAC incorporates the following configuration components to allow the administrator to build the network model using locations and links:

- Locations — A location represents a LAN. It could contain endpoints or simply serve as a transit location between links for WAN network modeling. For example, an MPLS provider could be represented by a location.
- Links — Links interconnect locations and are used to define bandwidth available between locations. Links logically represent the WAN link and are configured in the Location user interface (UI).
- Weights — A weight provides the relative priority of a link in forming the effective path between any pair of locations. The effective path is the path used by Unified CM for the bandwidth calculations, and it has the least cumulative weight of all possible paths. Weights are used on links to provide a "cost" for the "effective path" and are pertinent only when there is more than one path between any two locations.
- Path — A path is a sequence of links and intermediate locations connecting a pair of locations. Unified CM calculates least-cost paths (lowest cumulative weight) from each location to all other locations and builds a map of the various paths. Only one "effective path" is used between any pair of locations.
- Effective Path — The effective path is the path with the least cumulative weight and is the bandwidth accounting path that is always used between any two locations.
- Bandwidth Allocation — Is the amount of bandwidth allocated in the model for each type of traffic: audio, video, and immersive video (TelePresence).
- Location Bandwidth Manager (LBM) — Is the active service in Unified CM that assembles a network model from configured location and link data in one or more clusters. It determines the effective paths between pairs of locations, determines whether to admit calls between a pair of locations based on the availability of bandwidth for each type of call, and deducts (reserves) bandwidth for the duration of each call that is admitted.
- Location Bandwidth Manager Hub — Is a Location Bandwidth Manager (LBM) service that has been designated to participate directly in intercluster replication of fixed locations, links data, and dynamic bandwidth allocation data. LBMs assigned to an LBM hub group discover each other through their common connections and form a fully-meshed intercluster replication network. Other LBM services in a cluster with an LBM hub participate indirectly in intercluster replication through the LBM hubs in their cluster.

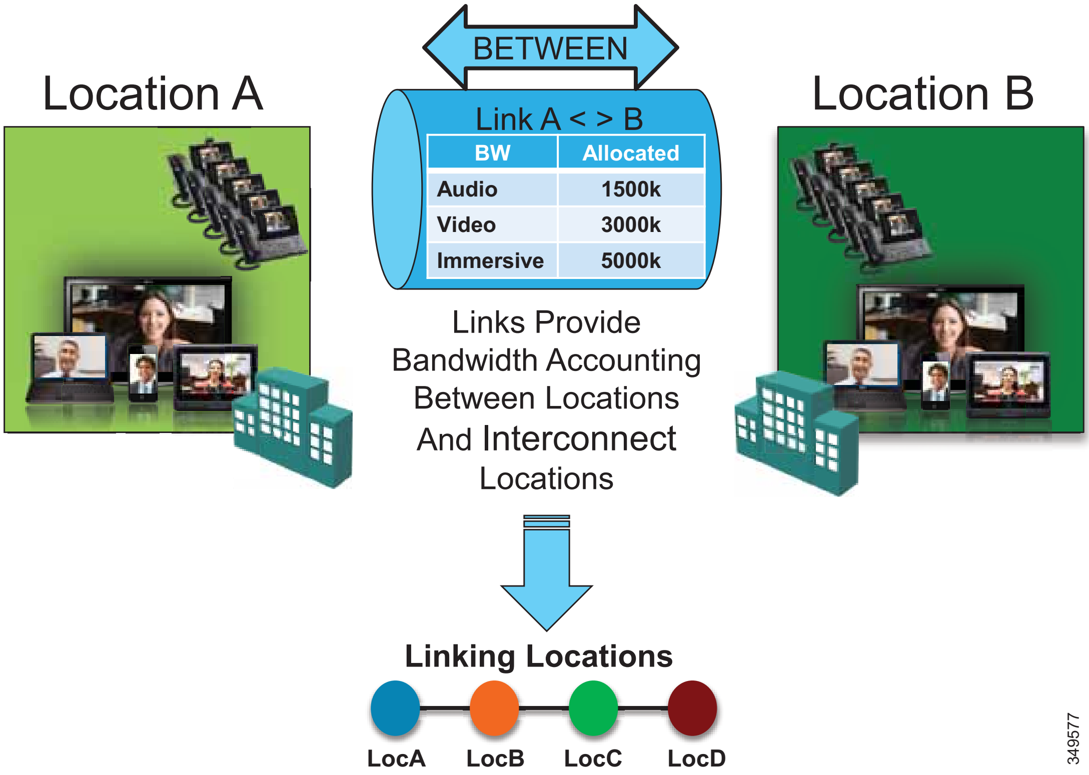
Locations, Links, and Weight

Unified CM uses the concept of locations to represent a physical site and to create an association with media devices such as endpoints, voice messaging ports, trunks, gateways, and so forth, through direct configuration on the device itself, through a device pool, or even through device mobility. Locations logically represent the Local Area Network (LAN). Unified CM also uses a configuration parameter called links to interconnect locations and to define bandwidth available between locations. Links logically represent the Wide Area Network (WAN). This section describes locations and links and how they are used (see Figure 6-14).

The location configuration itself consists of three main parts: links, intra-location bandwidth parameters, and RSVP locations settings. The intra-location bandwidth parameters are set to unlimited by default and should remain that way because there is little or no reason to limit bandwidth within a location (LAN). The RSVP locations settings are not considered here for Enhanced Location CAC because they apply only to RSVP implementations.

The link bandwidth parameters allow the administrator to characterize the provisioned bandwidth for audio, video, and immersive calls between "adjacent locations" (that is, locations that have a link configured between them). This feature offers the administrator the ability to create a string of location pairings in order to model a multi-hop WAN network.

Figure 6-14 Locations and Links



Weight is configurable on the link and provides the ability to force a specific path choice when multiple paths between two locations are available. When multiple paths are configured, only one will be selected based on the cumulative weight, and this path is referred to as the *effective path*. This weight is static and the effective path does not change dynamically. When two paths have equal weight, one path is randomly chosen; therefore it is important to ensure that only one path exists or has the least cumulative weight to ensure that it is the effective path for the Location Bandwidth Manager (LBM). This is especially important in multi-cluster environments.

When you configure a device in Unified CM, the device can be assigned to a location. A location can be configured with links to other locations in order to build a topology. The locations configured in Unified CM are virtual locations and not real, physical locations. As mentioned, Unified CM has no knowledge of the actual physical topology of the network. Therefore, any changes to the physical network must be made manually in Unified CM to map the real underlying network topology with the Unified CM locations model. If a device is moved from one physical location to another, the system administrator must either perform a manual update on its location configuration or implement the device mobility feature so that Unified CM can correctly calculate bandwidth allocations for calls to and from that device. Each device is in location **Hub_None** by default. Location Hub_None is an example location that typically serves as a hub linking two or more locations, and it is configured by default with unlimited intra-location bandwidth allocations for audio, video, and immersive bandwidth.

Unified CM allows the administrator to define separate voice, video, and immersive video bandwidth pools for each link between locations. In the PA, only voice and video bandwidth pools are used. Typically the link between locations is set to a finite number of kilobits per second (kbps) to match the provisioned amount of bandwidth available for audio and video in the WAN links between physical sites. Some WANs do not require any limitations because they are over-provisioned for the expected amount of traffic. If the bandwidth values are set to a finite number of kilobits per second (kbps), Unified CM will track all calls within the location and all calls that use the location as a transit location (a location that is in the calculation path but is not the originating or terminating location in the path).

The following devices must be configured in a location:

- Endpoints
- Conference bridges
- Gateways
- SIP trunks
- Music on hold (MoH) servers
- Annunciator (via device pool)

Table 6-6 lists the amount of bandwidth requested for various call speeds. For all audio (both audio-only calls as well as audio of a video call), Unified CM counts the media bit rates plus the IP and UDP overhead. For example, a G.711 or G.722 audio call consumes 80 kbps (64 kbps bit rate + 16 kbps for IP/UDP headers) deducted from the audio bandwidth allocation of the location and link. For a video call, Unified CM counts only the payload (no IP/UDP header overhead) for video streams, but the audio portion is calculated with IP and UDP overhead. For example, for a video call at a bit rate of 384 kbps where audio is set to use a 64 kbps bit rate, Unified CM will allocate 320 kbps from the video bandwidth allocation and take 64 kbps for the audio and add the 16 kbps for IP/UDP headers to derive 80 kbps for the audio pool deduction. For the same call where the audio is set to use a 8 kbps bit rate, Unified CM will allocate 376 kbps from the video bandwidth allocation and take 8 kbps for the audio and add the 16 kbps for IP/UDP headers to derive 24 kbps for the audio pool deduction.

***Table 6-6 Amount of Bandwidth Requested by the Locations and Links Bandwidth Deduction Algorithm in the PA Configuration*[1]**

Call Speed (Session Bit Rate)	Audio Pool Bandwidth	Video Pool Bandwidth
G.711 or G.722 audio call (64 kbps)	80 kbps	N/A
G.729 audio call (8 kbps)	24 kbps	N/A
512 kbps video call with G.729 audio (8 kbps)	24 kbps	504 kbps
512 kbps video call with G.711 or G.722 audio (64 kbps)	80 kbps	448 kbps
768 kbps video call with G.729 audio (8 kbps)	24 kbps	760 kbps
768 kbps video call with G.711 or G.722 audio (64 kbps)	80 kbps	704 kbps
1,024 kbps video call with G.729 audio (8 kbps)	24 kbps	1,016 kbps
1,024 kbps video call with G.711 or G.722 audio (64 kbps)	80 kbps	960 kbps

[1] Only 8 kbps and 64 kbps are used in these examples, but the same principle also applies to other audio bit rate codecs. The audio bit rate (payload only) is subtracted the video bit rate to get the adjusted video bit rate value.

For example, assume that the link configuration for the location Branch 1 to Hub_None allocates 256 kbps of available audio bandwidth and 1,024 kbps of available video bandwidth. In this case the path from Branch 1 to Hub_None 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 link between locations can also support different numbers of video calls, depending on the video and audio codecs being used (for example, one video call requesting 1,024 kbps of bandwidth or two video calls with each requesting 512 kbps of bandwidth).

When a call is placed from one location to the other, Unified CM deducts the appropriate amount of bandwidth from the effective path of locations and links from one location to another. When the call has completed, Unified CM returns the bandwidth to those same links over the effective path. If there is not enough bandwidth at any one of the links over the path, 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-location call is denied by call admission control, Unified CM can automatically reroute the call to the destination through the PSTN connection by means of the Automated Alternate Routing (AAR) feature (see Automated Alternate Routing, page 2-49).

Note AAR is invoked only when Enhanced Locations Call Admission Control denies the call due to a lack of network bandwidth along the effective path. In such cases, the calls are redirected to the target specified in the Call Forward No Answer field of the called device. AAR is not invoked when the IP WAN is unavailable or other connectivity issues cause the called device to become unregistered with Unified CM.

Also, AAR is applicable only to intra-cluster endpoint-to-endpoint calls. For all inter-cluster calls that fail CAC, the route groups are used to try different SIP routing paths.

Video devices can be enabled to **Retry Video Call as Audio** if a video call between devices fails CAC. This option is configured on the video endpoint or SIP trunk configuration page in Unified CM and is applicable to video endpoints or trunks placing calls. For some video endpoints, **Retry Video Call as Audio** is enabled by default and not configurable on the endpoint.

Locations, Links, and Region Settings

Location links work in conjunction with regions to define the characteristics of a call over the effective path of locations and links. Regions define the type of compression or bit rate (8 kbps or G.729, 64 kbps for G.722 or G.711, and so forth) that is used between devices, and location links define the amount of available bandwidth for the effective path between devices. You assign each device in the system to both a region (by means of a device pool) and a location (by means of a device pool or by direct configuration on the device itself).

You can configure locations in Unified CM to define:

- Physical sites (for example, a branch office) or transit sites (for example, an MPLS cloud) — A location represents a LAN. It could contain endpoints or simply serve as a transit location between links for WAN network modeling.
- Link bandwidth between adjacent locations — Links interconnect locations and are used to define bandwidth available between locations. Links logically represent the WAN link between physical sites.
  - Audio Bandwidth — The amount of bandwidth that is available in the WAN link for voice and fax calls being made from devices in the location to the configured adjacent location. Unified CM uses this bandwidth value for Enhanced Locations Call Admission Control.
  - Video Bandwidth — The amount of video bandwidth that is available in the WAN link for video calls being made from devices in the location to the configured adjacent location. Unified CM uses this bandwidth value for Enhanced Locations Call Admission Control.
  - Immersive Video Bandwidth — Not used in the PA configuration.

You can configure regions in Unified CM to define:

- The maximum audio bit rate
- The maximum session bit rate for video calls (includes audio)
- The maximum session bit rate for immersive video calls (includes audio) — Not used in the PA configuration.
- Audio codec preference lists

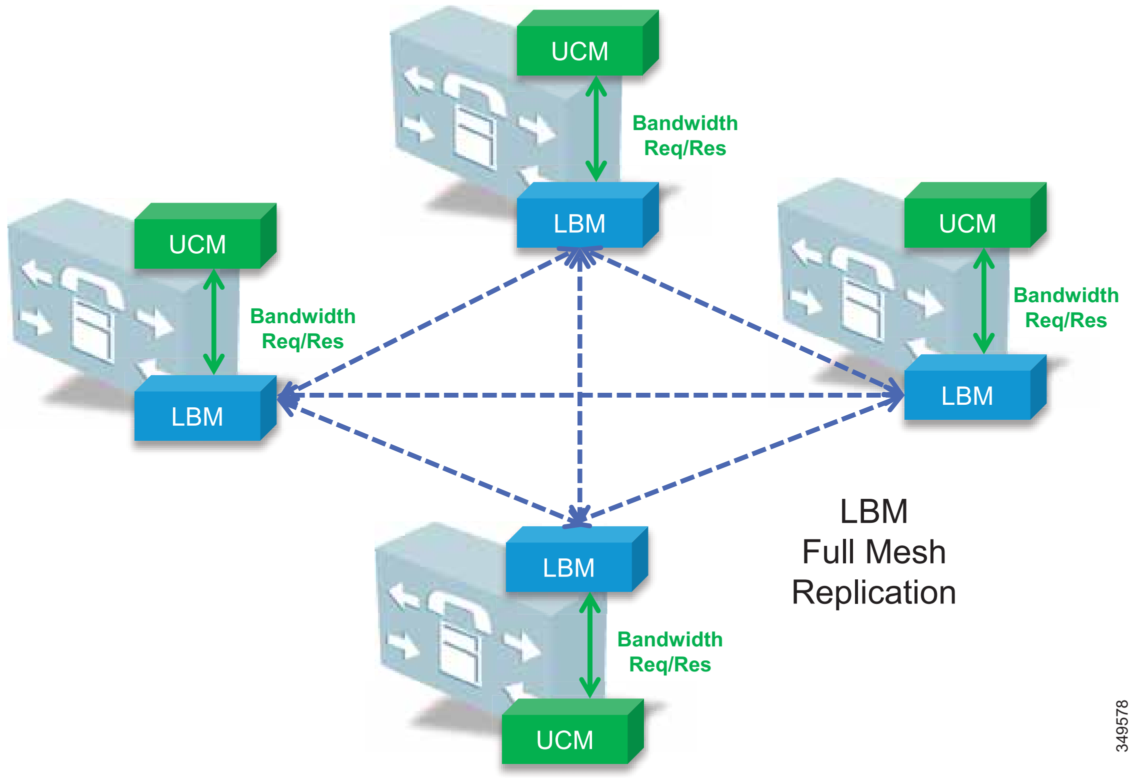
Location Bandwidth Manager

The Location Bandwidth Manager (LBM) is a Unified CM Feature Service managed from the serviceability web pages and is responsible for all of the Enhanced Locations CAC bandwidth functions. The LBM should be configured to run on each subscriber node in the cluster that is also running the Cisco CallManager service.

The LBM performs the following functions:

- Assembles topology of locations and links
- Calculates the effective paths across the topology
- Services bandwidth requests from the Cisco CallManager service (Unified CM call control)
- Replicates the bandwidth information to other LBMs (see Figure 6-15)
- Provides configured and dynamic information to serviceability
- Updates Location Real-Time Monitoring Tool (RTMT) counters

Figure 6-15 LBM Local Replication Network



By default the CallManager service communicates with the local LBM service.

Deducting All Audio from the Voice Pool

This PA utilizes a new Unified CM 11.*x* feature that allows the administrator to deduct the audio bandwidth of video calls from the voice pool. Because ELCAC relies on the correct DSCP setting in order to ensure the protection of the queues that voice and video CAC pools represent, changing how Unified CM deducts bandwidth from the video pool requires the DSCP of audio streams of the video calls to be marked the same as the audio streams of audio-only calls.

In Unified CM this feature is enabled by setting the service parameter **Deduct Audio Bandwidth from Audio Pool for Video Call** to **True** under the Call Admission Control section of the CallManager service. False is the default setting, and by default Unified CM deducts both audio and video streams of video calls from the video pool.

Multi-Cluster Considerations

This section covers the following topics:

- Intercluster ELCAC
- LBM Hub Replication Network
- Common Locations (Shared Locations) and Links
- Shadow Location
- Location and Link Management Cluster

Intercluster ELCAC

Intercluster Enhanced Locations CAC extends the concept of network modeling across multiple clusters. In intercluster Enhanced Locations CAC, each cluster manages its locally configured topology of locations and links and then propagates this local topology to other remote clusters that are part of the LBM intercluster replication network. Upon receiving a remote cluster's topology, the LBM assembles this into its own local topology and creates a global topology. Through this process the global topology is then identical across all clusters, providing each cluster a global view of enterprise network topology for end-to-end CAC. Figure 6-16 illustrates the concept of a global topology with a simplistic hub-and-spoke network topology as an example.

Figure 6-16 Example of a Global Topology for a Simple Hub-and-Spoke Network

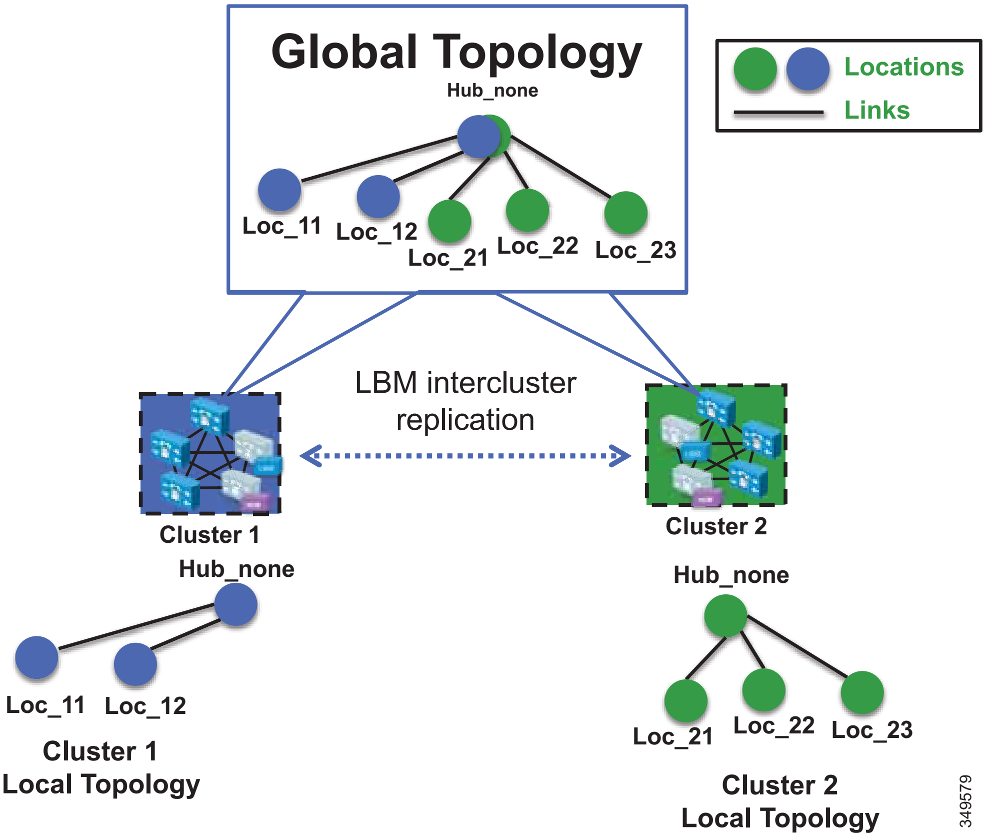


Figure 6-16 shows two clusters, Cluster 1 and Cluster 2, each with a locally configured hub-and-spoke network topology. Cluster 1 has configured Hub_None with links to Loc_11 and Loc_12, while Cluster 2 has configured Hub_None with links to Loc_21, Loc_22, and Loc_23. When intercluster Enhanced Locations CAC is enabled, Cluster 1 sends its local topology to Cluster 2, as does Cluster 2 to Cluster 1. After each cluster obtains a copy of the remote cluster's topology, each cluster overlays the remote cluster's topology over its own. The overlay is accomplished through common locations, which are locations that are configured with the same name. Because both Cluster 1 and Cluster 2 have the common location Hub_None with the same name, each cluster will overlay the other's network topology with Hub_None as a common location, thus creating a global topology where Hub_None is the hub and Loc_11, Loc_12, Loc_21, Loc_22 and Loc_23 are all spoke locations. This is an example of a simple network topology, but more complex topologies would be processed in the same way.

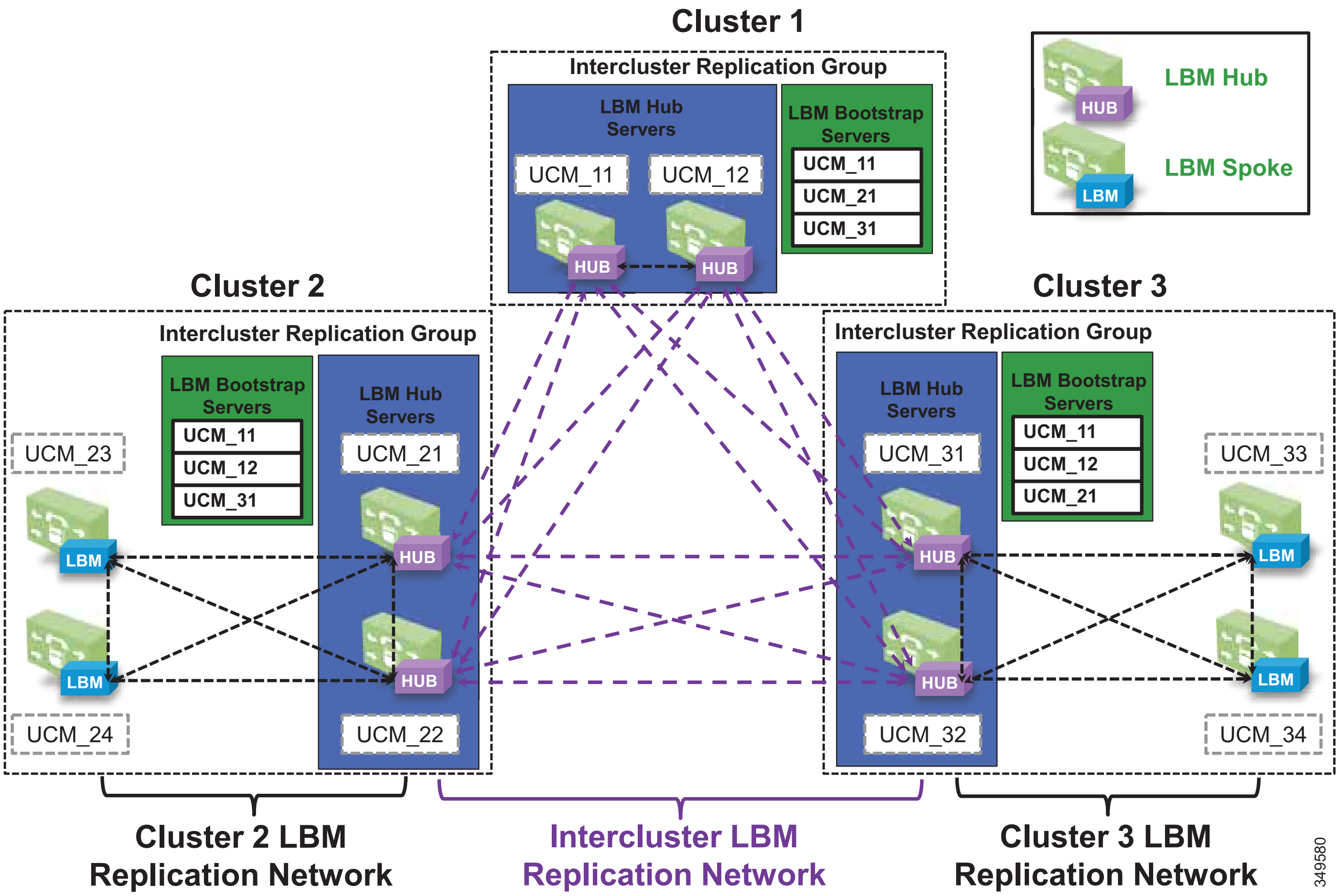
LBM Hub Replication Network

The intercluster LBM replication network is a separate replication network of designated LBMs called LBM hubs. LBM hubs create a separate full mesh with one another and replicate their local cluster's topology to other remote clusters. Each cluster effectively receives the topologies from every other remote cluster in order to create a global topology. The designated LBMs for the intercluster replication network are called *LBM hubs*. The LBMs that replicate only within a cluster are called *LBM spokes*. The LBM hubs are designated in configuration through the LBM **intercluster replication group**. The LBM role assignment for any LBM in a cluster can also be changed to a hub or spoke role in the intercluster replication group configuration. (For further information on the LBM hub group configuration, refer to the Cisco Unified Communications Manager product documentation available at http://www.cisco.com/en/US/products/sw/voicesw/ps556/tsd_products_support_series_home.html.)

In the LBM intercluster replication group, there is also a concept of bootstrap LBM. Bootstrap LBMs are LBM hubs that provide all other LBM hubs with the connectivity details required to create the full-mesh hub replication network. Bootstrap LBM is a role that any LBM hub can have. If all LBM hubs point to a single LBM hub, that single LBM hub will tell all other LBM hubs how to connect to one another. Each replication group can reference up to three bootstrap LBMs.

Once the LBM hub group is configured on each cluster, the designated LBM hubs will create the full-mesh intercluster replication network. Figure 6-17 illustrates an intercluster replication network configuration with LBM hub groups set up between three clusters (Cluster 1, Cluster 2, and Cluster 3) to form the intercluster replication network.

Figure 6-17 Example Intercluster Replication Network for Three Clusters



In Figure 6-17, two LBMs from each cluster have been designated as the LBM hubs for their cluster. This provides redundancy for the LBM hub role. These LBM hubs form the intercluster LBM replication network. The bootstrap LBMs configured in each LBM intercluster replication group are designated as UCM_11 and UCM_12. These two LBM hubs from Cluster 1 serve as points of contact or bootstrap LBMs for the entire intercluster LBM replication network. UCM_21 and UCM_31 in Cluster 2 and Cluster 3, respectively, serve as backup bootstrap LBM hubs when the primaries are not available (that is, when Cluster 1 is not available). Establishing the intercluster LBM replication network means that each LBM hub in each cluster connects to UCM_11, replicates its local topology, and gets the remote topology. They also get the connectivity information for the other clusters from UCM_11, connect to the other remote clusters, and replicate their topologies. This creates the full-mesh replication network. If UCM_11 is unavailable, the LBM hubs will connect to UCM_12. If Cluster 2 LBM hubs are unavailable, Cluster 2 and Cluster 3 LBM hubs will connect to UCM_31, and Cluster 3 LBM hubs will connect to UCM_21.

The LBM has the following roles with respect to the LBM intercluster replication network:

- LBM hubs (local LBMs)
  - Communicate directly to other remote hubs as part of the intercluster LBM replication network
- LBM spokes (local LBMs)
  - Communicate directly to local LBM hubs in the cluster and indirectly to the remote LBM hubs through the local LBM hubs
- Bootstrap LBMs
  - LBM hubs responsible for interconnecting all clusters' LBM hubs in the replication network
  - Can be any LBM hub(s) in the network
  - Can indicate up to three bootstrap LBM hubs per LBM intercluster replication group
- LBM hub replication network — Bandwidth deduction and adjustment messages
  - LBM optimizes the LBM messages by choosing a sender and receiver from each cluster

LBM hubs can also be configured to encrypt their communications. This allows intercluster ELCAC to be deployed in environments where it is critical to encrypt traffic between clusters because the links between clusters might reside over unprotected networks. For further information on configuring encrypted signaling between LBM hubs, refer to the Cisco Unified Communications Manager product documentation available at

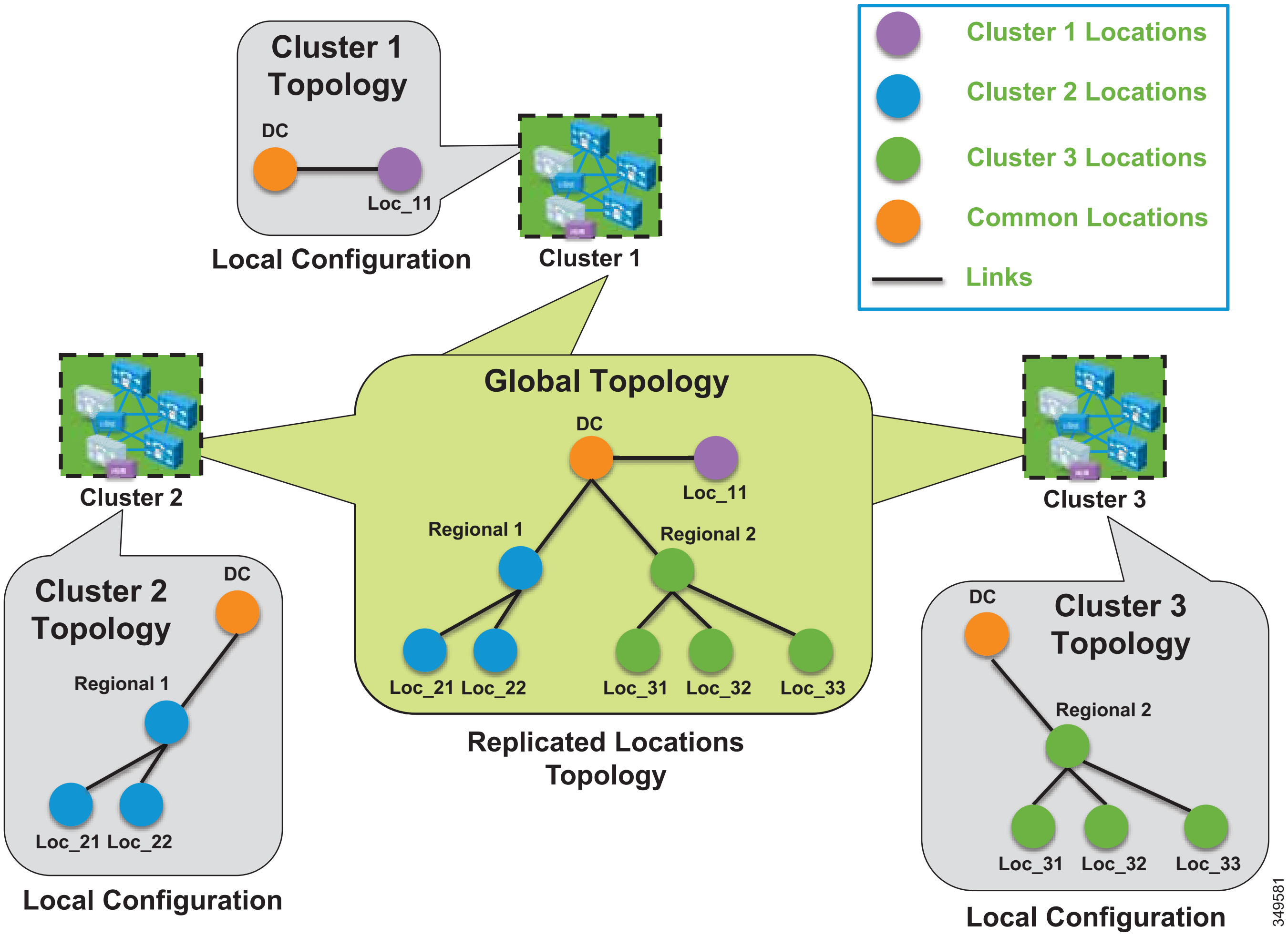
http://www.cisco.com/en/US/products/sw/voicesw/ps556/tsd_products_support_series_home.html

Common Locations (Shared Locations) and Links

Common locations are locations that are named the same across clusters. Common locations play a key role in how the LBM creates the global topology and how it associates a single location across multiple clusters. A location with the same name between two or more clusters is considered the same location and thus is a shared location across those clusters. If a location is meant to be shared between multiple clusters, it must have exactly the same name. After replication, the LBM will check for configuration discrepancies across locations and links. Any discrepancy in bandwidth value or weight between common locations and links can be seen in serviceability, and the LBM calculates the locations and link paths with the most restrictive values for bandwidth and the lowest value (least cost) for weight.

Common locations and links can be configured across clusters for a number of different reasons. You might have a number of clusters that manage devices in the same physical site and use the same WAN up-links, and therefore the same location needs to be configured on each cluster in order to associate that location to the local devices on each cluster. You might also have clusters that manage their own topology, yet those topologies interconnect at specific locations and you will have to configure those locations as common locations across each cluster so that, when the global topology is being created, the clusters have the common interconnecting locations and links on each cluster to link each remote topology together effectively. Figure 6-18 illustrates linking topologies together and shows the common topology that each cluster shares.

Figure 6-18 Using Common Locations and Links to Create a Global Topology



In Figure 6-18, Cluster 2 has devices in locations Regional 1, Loc_21, and Loc_22, but it requires configuring DC and a link from Regional 1 to DC in order to link to the rest of the global topology. Cluster 3 is similar, with devices in Regional 2 and Loc_31, Loc_32, and Loc_33, and it requires configuring DC and a link from DC to Regional 2 to map into the global topology. Cluster 1 has devices in Loc_11 only, and it requires configuring DC and a link to DC from Loc_11 to map into Cluster 2 and Cluster 3 topologies.

The key to topology mapping from cluster to cluster is to ensure that at least one cluster has a common location with another cluster so that the topologies interconnect accordingly.

Shadow Location

The shadow location is used to enable a SIP trunk to pass Enhanced Locations CAC information such as location name, among other things, required for Enhanced Locations CAC to function between clusters. In order to pass this location information across clusters, the SIP intercluster trunk (ICT) must be assigned to the shadow location. The shadow location cannot have a link to other locations, and therefore no bandwidth can be reserved between the shadow location and other locations. Any device other than a SIP ICT that is assigned to the shadow location will be treated as if it was associated to Hub_None.

Location and Link Management Cluster

In order to avoid configuration overhead and duplicated configuration across clusters that share a large number of locations, a Location and Link Management Cluster can be configured to manage all locations and links in the global topology. All other clusters uniquely configure the locations that they require for location-to-device association and do not configure links or any bandwidth values other than unlimited. The Location and Link Management Cluster is a design concept and is simply any cluster that is configured with the entire global topology of locations and links, while all other clusters in the LBM replication network are configured only with locations set to unlimited bandwidth values and without configured links. When intercluster Enhanced Locations CAC is enabled and the LBM replication network is configured, all clusters replicate their view of the network. The designated Location and Link Management Cluster has the entire global topology with locations, links, and bandwidth values; and once those values are replicated, all clusters use those values because they are the most restrictive. This design alleviates configuration overhead in deployments where a large number of common locations are required across multiple clusters.

Recommendations

Location and Link Management Cluster:

- One cluster should be chosen as the management cluster (the cluster chosen to manage administratively locations and links).
- The management cluster should be configured with the following:
  - All locations within the enterprise will be configured in this cluster.
  - All bandwidth values and weights for all locations and links will be managed in this cluster.

All other clusters in the enterprise:

- All other clusters in the enterprise should configure only the locations required for association to devices but should not configure the links between locations. This link information will come from the management cluster when intercluster Enhanced Locations CAC is enabled. By default there is always a link configured between a newly added location and hub_none. This link should be removed if hub_none is either not used or is not correct in the topology being built.
- When intercluster Enhanced Locations CAC is enabled, all of the locations and links will be replicated from the management cluster.

LBM will always use the lowest, most restrictive bandwidth and lowest weight value after replication.

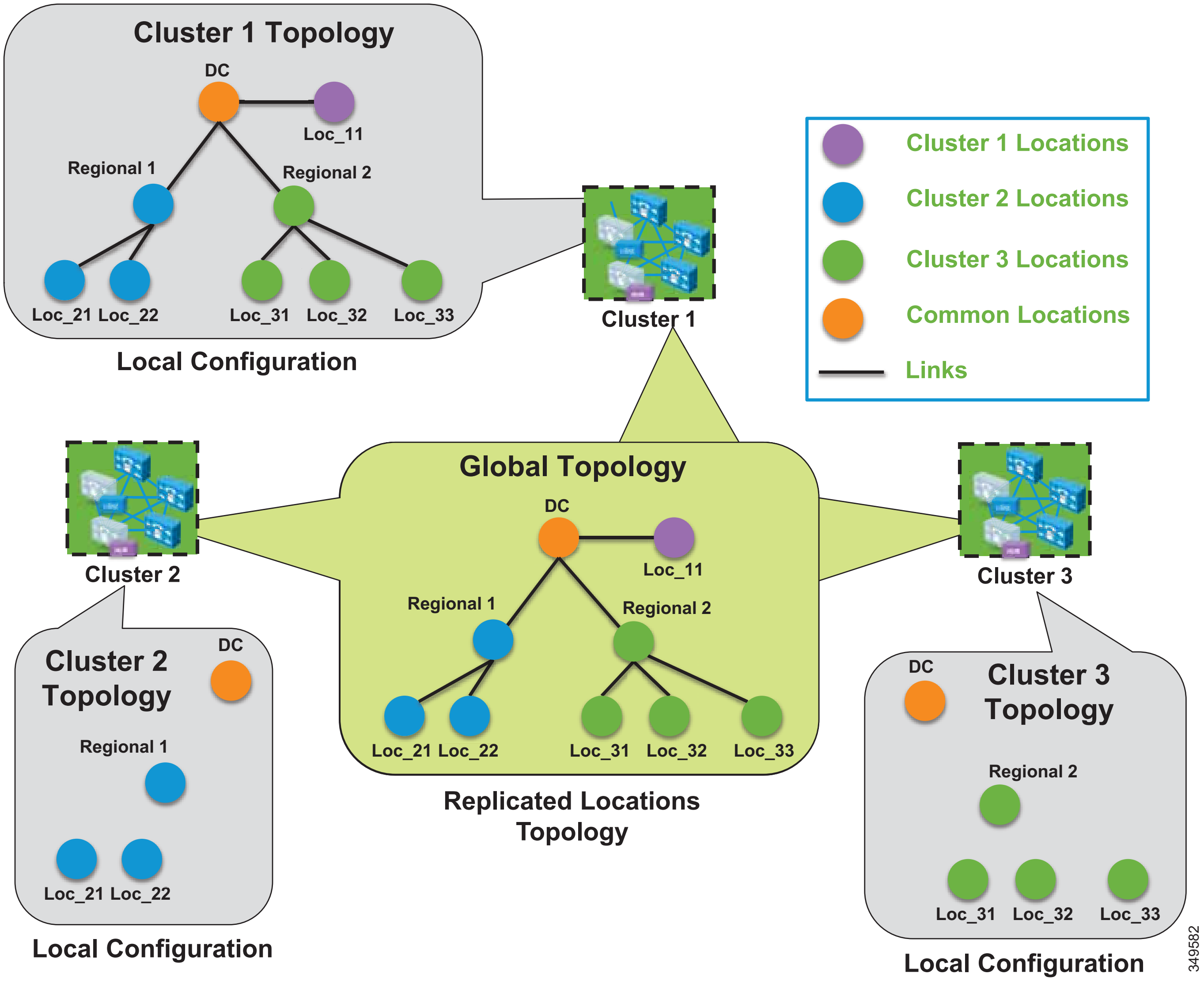
Benefits

- Manages enterprise CAC topology from a single cluster.
- Alleviates location and link configuration overhead when clusters share a large number of common locations.
- Alleviates configuration mistakes in locations and links across clusters.
- Other clusters in the enterprise require the configuration only of locations needed for location-to-device and endpoint association.
- Provides a single cluster for monitoring of the global locations topology.

Figure 6-19 illustrates a Location and Link Management Cluster for three clusters.

Note As mentioned, any cluster can act as the Location and Link Management Cluster. In Figure 6-19, Cluster 1 is the Location and Link Management Cluster.

Figure 6-19 Example of Cluster 1 as a Location and Link Management Cluster



In Figure 6-19 there are three clusters, each with devices in only a regional and remote locations. Cluster 1 has the entire global topology configured with locations and links, and intercluster LBM replication is enabled among all three clusters. None of the clusters in this example share locations, although all of the locations are common locations because Cluster 1 has configured the entire location and link topology. Note that Cluster 2 and Cluster 3 configure only the locations that they require to associate to devices and endpoints, while Cluster 1 has the entire global topology configured. After intercluster replication, all clusters will have the global topology with locations and links.

Design Considerations for Call Admission Control

This section describes how to apply the call admission control mechanisms to various IP WAN topologies. Unified CM Enhanced Locations CAC network modeling support, together with intercluster enhanced locations, can support most of the network topologies in any Unified CM deployment model. Enhanced Locations CAC is still a statically defined mechanism that does not query the network, and therefore the administrator still has to provision Unified CM accordingly whenever network changes affect admission control. This is where a network-aware mechanism such as RSVP can fill that gap and provide support for dynamic changes in the network, such as when network failures occur and media streams take different paths in the network. This is often the case in designs with load-balanced dual or multi-homed WAN up-links or unequally sized primary and backup WAN up-links.

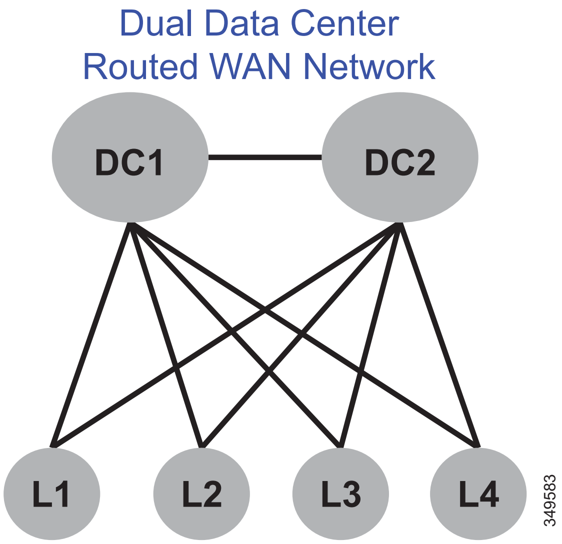
To learn how Enhanced Locations CAC functions, and for more design and deployment details of Enhanced Locations CAC, see the Enhanced Locations Call Admission Control information in the *Bandwidth Management* chapter of the Cisco Collaboration SRND.

This section explores a few typical topologies and explains how Enhanced Locations CAC can be designed to manage them.

Dual Data Center Design

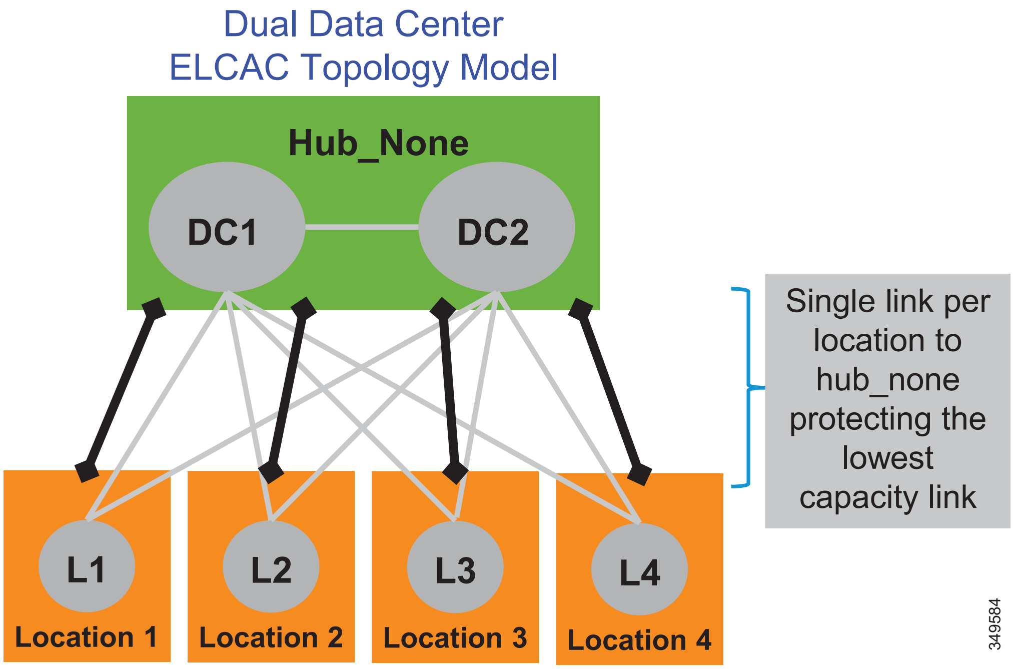
Figure 6-20 illustrates a simple dual data center WAN network design where each remote site has a single WAN up-link to each data center. The data centers are interconnected by a high-speed WAN connection that is over-provisioned for data traffic.

Figure 6-20 Dual Data Center WAN Network



Typically these WAN up-links from the remote sites to the data centers are load-balanced or in a primary/backup configuration, and there are limited ways for a static CAC mechanism to handle these scenarios. Although you could configure this multi-path topology in Enhanced Locations CAC, only one path would be calculated as the effective path and would remain statically so until the weight metric was changed. A better way to support this type of network topology is to configure the two data centers as one data center or hub location in Enhanced Locations CAC and configure a single link to each remote site location. Figure 6-21 illustrates an Enhanced Locations CAC locations and links overlay.

Figure 6-21 Enhanced Locations CAC Topology Model for Dual Data Centers



Design Recommendations

The following design recommendations for dual data centers with remote dual or more links to remote locations, apply to both load-balanced and primary/backup WAN designs:

- A single location (Hub_None) represents both data centers.
- A single link between the remote locations and Hub_None protects the remote site up-links from over-subscription during normal conditions or failure of the highest bandwidth capacity links.
- The capacity of link bandwidth allocation between the remote site and Hub_None should be equal to the lowest bandwidth capacity for the applicable Unified Communications media for a single link. For example, if each WAN up-link can support 2 Mbps of audio traffic marked EF, then the link audio bandwidth value should be no more than 2 Mbps to support a failure condition or equal-cost path routing.

MPLS Clouds

When designing for Multiprotocol Label Switching (MPLS) any-to-any connectivity type clouds in the Enhanced Locations CAC network model, a single location can serve as the MPLS cloud. This location will not have any devices associated to it, but all of the locations that have up-links to this cloud will have links configured to the location representing the cloud. In this way the MPLS cloud serves as a transit location for interconnecting multiple variable-sized bandwidth WAN up-links to other remote locations.

Design Recommendations

- The MPLS cloud should be configured as a location that does not contain any endpoints but is used as a hub to interconnect locations.
- The MPLS cloud serves as a transit location for interconnecting multiple variable-sized bandwidth WAN up-links to other remote locations.
- Remote sites with connectivity to dual MPLS clouds should treat those connections as a single link and size to the lowest capacity of the links in order to avoid over-subscription during network failure conditions.

Call Admission Control Design Recommendations for Video Deployments

Admission control and QoS are complementary and in most cases co-dependent. Current Cisco product offerings such as audio and video endpoints, voice and video gateways, voice messaging, and conferencing all support native QoS packet marking based on IP Differentiated Services Code Point (IP DSCP). Note, however, that Jabber for Windows clients specifically do not follow the same native marking ability that other clients do, because the Windows operating system requires the use of Group Policy Objects (GPO) using application, IP addresses, and UDP/TCP port ranges to mark traffic with DSCP from the operating system itself. Group Policy Objects are very similar in function to network access lists in their ability to mark traffic.

QoS is critical to admission control because without it the network has no way of prioritizing the media to ensure that admitted traffic gets the network resources it requires above that of non-admitted or other traffic classifications. Unified CM's CallManager service parameters for QoS as well as the SIP Profile settings provide five main QoS settings that are applicable to endpoint media classification. Table 6-7 shows the five main DSCP parameters along with their default and recommended values and Per Hop Behavior (PHB) equivalents.

Table 6-7 QoS Settings for Endpoint Media Classification

Cisco CallManager Service Parameters Clusterwide Parameters (System - QoS)	Default Values		Recommended Values	
	DSCP	PHB	DSCP	PHB
DSCP for Audio Calls	46	EF	46	EF
DSCP for Video Calls	34	AF41	34	AF41
DSCP for Audio Portion of Video Calls	34	AF41	46	EF
DSCP for TelePresence Calls	32	CS4	34	AF41
DSCP for Audio Portion of TelePresence Calls	32	CS4	46	EF

The **DSCP for Audio Calls** setting is used for any device that makes an audio-only call. The **DSCP for Video Calls** setting is used for the audio and video traffic of any device that is classified as "desktop." **DSCP for TelePresence Calls** is used for the audio and video traffic of any device that is classified as "immersive." The **DSCP for Audio Portion of Video Calls** and **DSCP for Audio Portion of TelePresence Calls** differentiate only the audio portion of video calls, dependent on the classified video call.

Enhanced Locations CAC Design Considerations and Recommendations

The following design recommendation applies to video solutions that employ Enhanced Locations CAC:

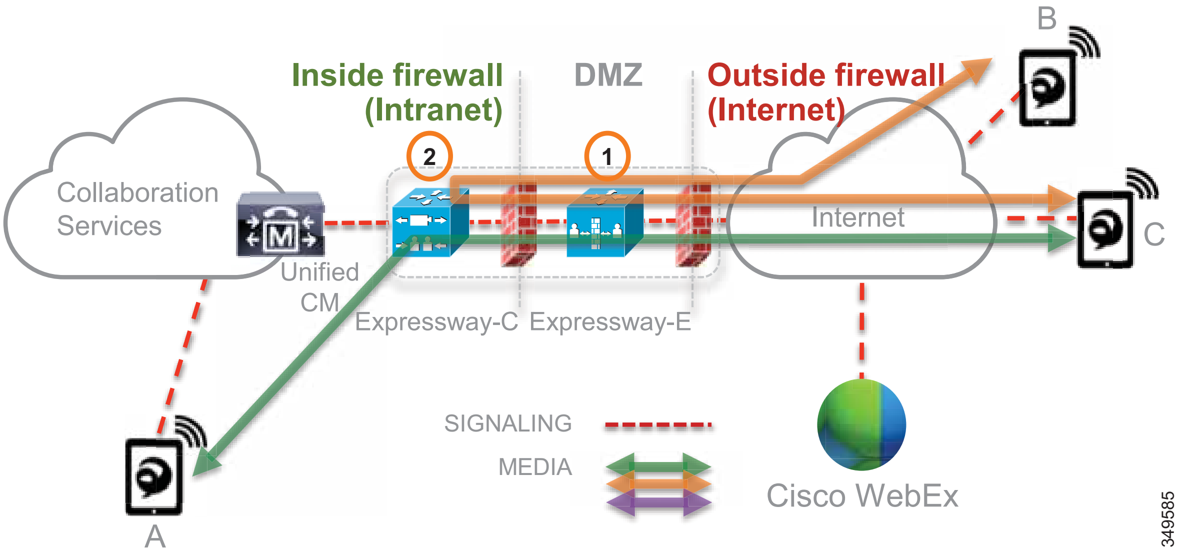
- Intercluster SIP trunks should be associated with the shadow location.

Design Recommendations for Cisco Expressway Deployments with Enhanced Locations CAC

In the Cisco Expressway mobile and remote access (MRA) solution, endpoints supporting the feature can register to Unified CM through a Cisco Expressway deployment without the use of a VPN. Cisco Expressway-C and Expressway-E servers are deployed, each with redundancy for high availability. Expressway-E is placed in the DMZ between the firewall to the Internet (outside) and the firewall to the enterprise (inside), while Expressway-C is placed inside the enterprise. Figure 6-22 illustrates this deployment. It also illustrates the following media flows:

- For Internet-based endpoints calling one another, the media is routed through Cisco Expressway E and Expressway C back out to the Internet, as is illustrated between endpoints B and C in Figure 6-22.
- For Internet-based endpoints calling internal endpoints, the media flows through Expressway-E and Expressway-C, as is illustrated between endpoints A and C in Figure 6-22.

Figure 6-22 Deployment of Cisco Expressway Mobile and Remote Access (MRA)



Enhanced Locations CAC for Cisco Expressway deployments requires the use of a feature in Unified CM called Device Mobility. Enabling Device Mobility on the endpoints allows Unified CM to know when the device is registered through Cisco Expressway or when it is registered from within the enterprise. Device Mobility also enables Unified CM to provide admission control for the device as it roams between the enterprise and the Internet. Device Mobility is able to do this by knowing that, when the endpoints register to Unified CM with the IP address of Expressway-C, Unified CM will associate the applicable Internet location. However, when the endpoint is registered with any other IP address, Unified CM will use the enterprise location that is configured directly on the device (or from the device pool directly configured on the device). It is important to note that Device Mobility does not have to be deployed across the entire enterprise for this function to work. Configuration of Device Mobility in Unified CM is required only for the Expressway IP addresses, and the feature is enabled only on the devices that require the function (that is to say, those devices registering through the Internet).

Bandwidth Management Deployment

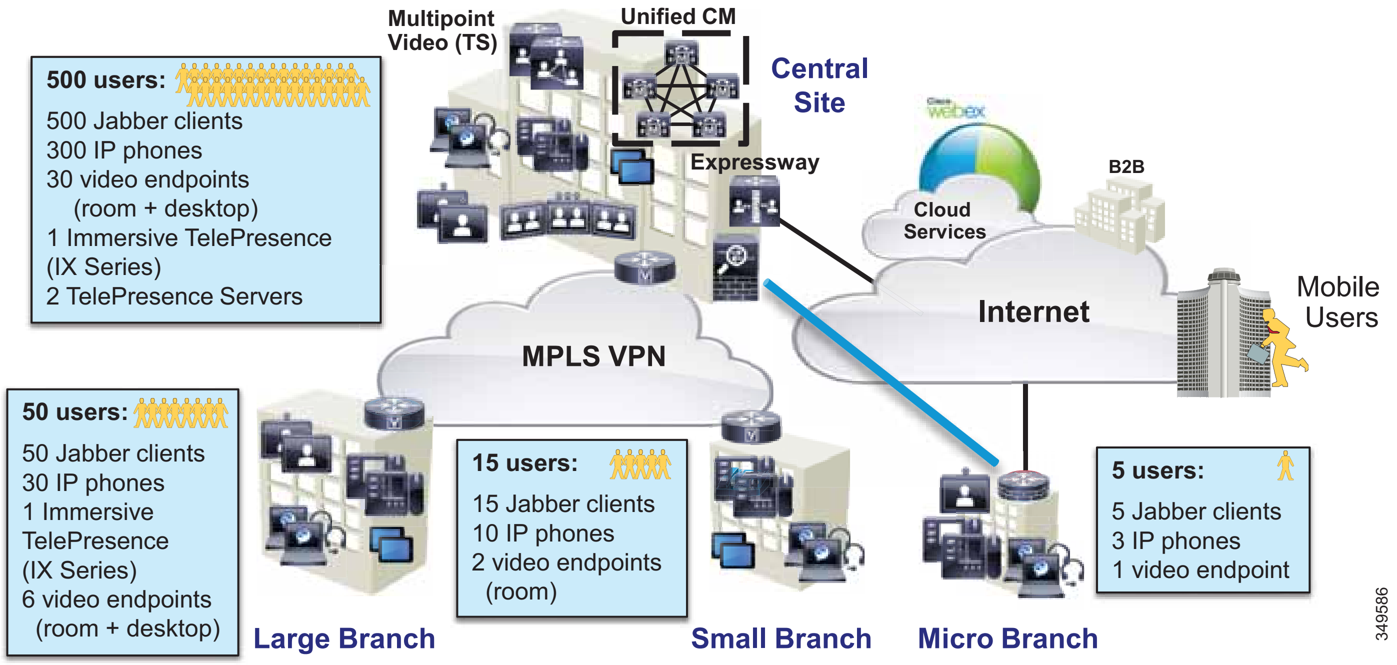
This section describes how to deploy bandwidth management for the PA. It explores all aspects discussed previously in this chapter, including identification and classification, WAN queuing and scheduling, provisioning, resource control, and bandwidth allocation guidelines for each site type.

Deployment Overview

The Preferred Architecture example in this section is a large enterprise with users across a large geographic area and with a headquarters site where the data center sits as well as multiple large, small, and micro sized branches with roughly 500, 50, 15, and 5 users in each branch type, respectively. To simplify the illustration of the network, these categories of sites (headquarters, large, small, and micro) are used as a template to size bandwidth considerations for each site that has a similar size user base and endpoint density. Figure 6-23 illustrates this with numbers of users and endpoints at each type of site. The enterprise in this example has deployed Jabber with video to ensure that users have access to a video terminal for conferencing. The TelePresence video conferencing resources are located in the data center at the Headquarters Site. IP phones are for voice-only communications. Video endpoints are Jabber clients, collaboration desktop endpoints (DX Series), and room-based endpoints (MX Series and SX Series). The Headquarters Site has an immersive TelePresence unit such as the IX Series.

The IT department is tasked with determining the bandwidth requirements for the WAN edge for each type of site. Each section below lists the requirements and illustrates a methodology for applying QoS, determining bandwidth and queuing requirements, and determining admission control requirements.

Figure 6-23 Preferred Architecture for Enterprise Collaboration



Deployment of bandwidth management for the Enterprise Collaboration Preferred Architecture involves the following major tasks:

- Identification and Classification
  - Access Layer Endpoint Identification and Classification
  - Application Server QoS
  - WAN Edge Identification and Classification
  - WAN Edge Queuing and Scheduling
- Provisioning and Admission Control
  - Enhanced Locations CAC
  - Deploy Device Mobility for Mobile and Remote Access (MRA)
  - Bandwidth Allocation Guidelines

Identification and Classification

In this phase the QoS requirements are established across the enterprise. The topics covered in this section include:

- Access Layer Endpoint Identification and Classification
  - Endpoints: Jabber
  - Endpoints: Desktop and TelePresence
- Application Server QoS
- WAN Edge Identification and Classification
- WAN Edge Queuing and Scheduling

This phase of the deployment involves the following high-level steps:

1. Configure endpoints in Unified CM with QoS for Jabber clients and desktop and telepresence endpoints.
2. Deploy an access layer policy for endpoint identification and classification for untrusted endpoints.
3. Configure application server QoS for media and SIP signaling.
4. Deploy a WAN Edge ingress marking policy for collaboration media and SIP signaling.
5. Deploy a WAN Edge egress queuing policy for collaboration media and SIP signaling.

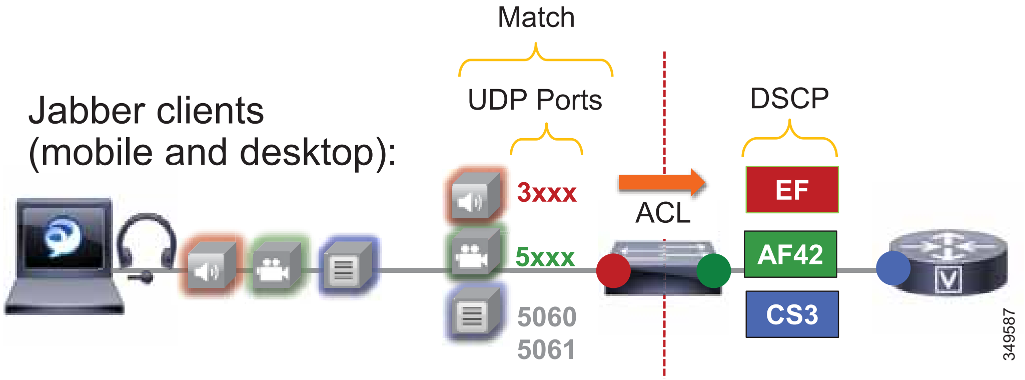
Access Layer Endpoint Identification and Classification

In this section endpoint QoS and media port ranges are configured in the network and in Unified CM.

Endpoints: Jabber

Jabber endpoints are untrusted and typically sit in the data VLAN. Specific UDP port ranges are used to re-mark signaling and media at the access layer switch. In this case Unified CM is configured with a SIP Profile specifically for all Jabber clients to use the **Separate Media and Signaling Port Range** value of 3000 to 3999 for audio and 5000 to 5999 for video. The SIP signaling port of 5060 is used for SIP signaling and 5061 for secure SIP signaling. The SIP signaling port is configured in the SIP Security Profile in Unified CM. This is illustrated in Figure 6-24.

Figure 6-24 Jabber Endpoint QoS



The administrator creates an ACL for the access switches for the data VLAN to re-mark UDP ports to the following DSCP values:

- Audio: UDP ports 3000 to 3999 marked as EF
- Video: UDP ports 5000 to 5999 marked as AF42
- Signaling: TCP ports 5060 to 5061 marked as CS3

Jabber classification summary:

- Audio streams of all Jabber calls (voice-only and video) are marked as EF
- Video streams of Jabber video calls are marked as AF42

For the Jabber endpoints, we also recommend changing the default QoS values in the Jabber SIP Profile. This is to ensure that, if for any reason the QoS is "trusted" via a wireless router or any other network component, then the correct "trusted" values are the same as they would be for the re-marked value. Therefore, the QoS parameters in the SIP Profile should be set as listed in Table 6-8, and the UDP port ranges should be set as listed in Table 6-9.

Table 6-8 QoS Parameter Settings in SIP Profile for Jabber Endpoints

QoS Service Parameter Name (SIP Profile)	Default Value	Changed Value
DSCP for Audio Calls	EF	No change
DSCP for Video Calls	AF41	AF42
DSCP for Audio Portion of Video Calls	AF41	EF
DSCP for TelePresence Calls	CS4	AF41
DSCP for Audio Portion of TelePresence Calls	CS4	EF

Table 6-9 UDP Port Settings for Jabber Endpoints

Media Port Ranges > Separate Port Range for Audio and Video	Value
Audio start port	3000
Audio stop port	3999
Video start port	5000
Video stop port	5999

The settings in Table 6-8 ensure that audio of Jabber clients is set to EF, and the video will be set to AF42 if for any reason the traffic goes through a trusted path and is not re-marked via UDP port range at the access switch. This is simply to ensure a consistent configuration across Jabber endpoints.

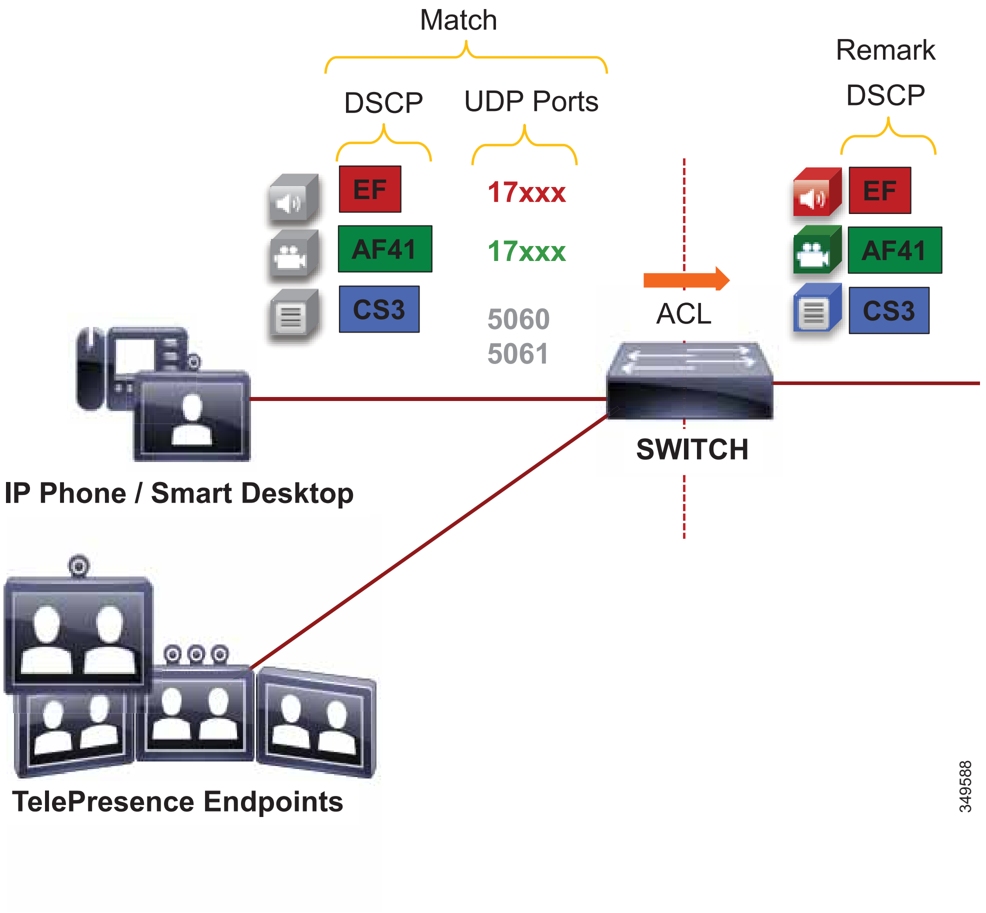
For Jabber on mobile devices, we recommend copying the **Standard SIP Profile for Mobile Device** when building a new SIP profile for these devices, because the default standard SIP profile for mobile devices includes recommended timer values for maintaining Jabber registration on Android and Apple iOS devices. These timers are required for any SIP profile assigned to dual-mode and tablet Jabber client devices.

Note Jabber for Mac, iPad, iPhone, and Android all natively mark DSCP by the OS. Jabber for Windows, however, requires Group Policy Objects to re-mark DSCP by the OS. Without Group Policy Objects, Jabber for Windows will mark all traffic with a DSCP of 0. This is why specific port ranges are used for Jabber and without matching on DSCP.

Endpoints: Desktop and TelePresence

IP phones, smart desktop, and TelePresence endpoints also rely on an access layer switch ACL to re-mark traffic. Specific UDP port ranges and DSCP are used to re-mark signaling and media at the access layer switch. In this case Unified CM is configured with a SIP Profile specifically for all IP phones, smart desktop, and TelePresence endpoints to use the common Media and Signaling Port Range value of 17000 to 17999 for audio and video. The SIP signaling port of 5060 is used for SIP signaling and 5061 for secure SIP signaling. The SIP signaling port is configured in the SIP Security Profile in Unified CM. This is illustrated in Figure 6-25.

Figure 6-25 Desktop and TelePresence Endpoint QoS



The administrator creates an ACL for the access switch ports to re-mark UDP ports to the following DSCP values:

- Audio: UDP ports 17000 to 17999 with DSCP of EF to be re-marked as EF
- Video: UDP ports 17000 to 17999 with DSCP of AF41 to be re-marked as AF41
- Signaling: TCP ports 5060 to 5061 marked as CS3

Desktop and TelePresence endpoint classification summary:

- Audio streams of all desktop and TelePresence endpoint calls (voice-only and video) are marked EF.
- Video streams of desktop and TelePresence endpoint video calls are marked AF41.

For the desktop and TelePresence endpoints, the default QoS values must be changed in the SIP Profile and set as shown in Table 6-10, and the UDP port ranges should be set as listed in Table 6-11.

Table 6-10 QoS Parameters in SIP Profile for Desktop and TelePresence Endpoints

QoS Service Parameter Name (SIP Profile)	Default Value	Changed Value
DSCP for Audio Calls	EF	No change
DSCP for Video Calls	AF41	No change
DSCP for Audio Portion of Video Calls	AF41	EF
DSCP for TelePresence Calls	CS4	AF41
DSCP for Audio Portion of TelePresence Calls	CS4	EF

Table 6-11 UDP Port Settings for Desktop and TelePresence Endpoints

Media Port Ranges > Common Port Range for Audio and Video	Value
Media start port	17000
Media stop port	17999

Example Switch ACL-Based QoS Policy for Endpoint Switch Ports

Desktop and TelePresence endpoints:

- Match UDP port range 17xxx with DSCP EF –> Re-mark to DSCP EF
- Match UDP port range 17xxx with DSCP AF41 –> Re-mark to DSCP AF41
- Match TCP ports 5060 to 5061 –> Re-mark to DSCP CS3

Jabber clients

- Match UDP port range 3xxx –> Re-mark to DSCP EF
- Match UDP port range 5xxx –> Re-mark to DSCP AF42
- Match TCP ports 5060 to 5061 –> Re-mark to DSCP CS3

Generic matching

- Matches the rest of the traffic and sets DSCP to 0 (Best Effort or BE) using a default class-map

Note The following is an example access control list based on the Cisco Common Classification Policy Language (C3PL).

```
! This section configures the ACLs to match the UDP port ranges and DSCP.
ip access-list extended QOS_VOICE
   permit udp any range 17000 17999 any dscp ef
   permit udp any range 3000 3999  any
ip access-list extended QOS_PRIORITIZED_VIDEO
   permit udp any range 17000 17999 any dscp af41
ip access-list extended QOS_JABBER_VIDEO
   permit udp any range 5000 5999 any
ip access-list extended QOS_SIGNALING
   permit tcp any any range 5060 5061
   permit tcp any range 5060 5061 any
```

```
! This section configures the classes that match on the ACLs above.
class-map match-any VOICE
  match access-group name QOS_VOICE
class-map match-any PRIORITIZED_VIDEO
  match access-group name QOS_PRIORITIZED_VIDEO
class-map match-any JABBER_VIDEO
  match access-group name QOS_JABBER_VIDEO
class-map match-any SIGNALING
  match access-group name QOS_SIGNALING

! This section configures the policy-map matching the classes configured above and sets
DSCP for voice, video, and SIP signaling on ingress. Note that the class-default sets
everything that does not match the above to a DSCP of 0 (BE).
policy-map INGRESS_MARKING
 class VOICE
   set dscp ef
 class PRIORITIZED_VIDEO
   set dscp af41
 class JABBER_VIDEO
   set dscp af42
 class SIGNALING
   set dscp cs3
class class-default
   set dscp 0

! This section applies the policy-map to the interface.
  Switch (config-if)# service-policy input INGRESS-MARKING
```

As mentioned, endpoints send and receive other data and signaling such as ICMP, DHCP, TFTP, BFCP, LDAP, XMPP, FECC, CTI, and so forth. The QoS values for this traffic should follow the enterprise's best practices for each type of traffic. Without doing this step, all other traffic apart from media and SIP signaling will be set to a DSCP of BE (DSCP 0) by the class-default in this configuration. We recommend either passing through the traffic marking by matching on DSCP and then re-marking the DSCP to the same value, or else using the TCP and UDP ports for each protocol that the endpoints use for communications.

The following example illustrates this. A class-map is created to match on a DSCP of AF21 which is transactional data, and the policy sets that data to AF21, effectively re-marking the DSCP to the same value. This is simply an example of matching on a DSCP to re-mark to the same DSCP:

```
class-map match-any TRANSACTIONAL-DATA
  match dscp af21

policy-map INGRESS_MARKING
….
 class TRANSACTIONAL-DATA
   set dscp af21
```

TCP and UDP port ranges can also be used. For more information on the TCP and UDP ports used for communication between the endpoints and Unified CM, see the *Cisco Unified Communications Manager TCP and UDP Port Usage* information in the *System Configuration Guide for Cisco Unified Communications Manager*, available at

http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-installation-and-configuration-guides-list.html

Also see the endpoints administration guides or the Jabber planning guide to determine the various protocols and ports used for other endpoint traffic. Some examples of these documents include:

- *Cisco DX Series Administration Guide*, available at

  http://www.cisco.com/c/en/us/support/collaboration-endpoints/desktop-collaboration-experience-dx600-series/products-maintenance-guides-list.html

- *Cisco Jabber Planning Guide*, available at

  http://www.cisco.com/c/en/us/support/unified-communications/jabber-android/products-installation-guides-list.html

Application Server QoS

Configure QoS on all media originating and terminating applications and MCUs across the solution. This section covers non-default configuration on all application servers in the PA. It is also equally important to ensure that the switch ports to which the application servers are connected trust the QoS set by the servers. Some switches such as the Cisco Catalyst 3850 Series trust the QoS by default, so verify the switch configuration to ensure that the switch port is trusted by default or enable QoS trust.

QoS settings for the various application servers:

- Cisco Unified CM (endpoint)

  System > **Service Parameters** > **Select Publisher** > **Select Cisco CallManager Service** > **Clusterwide Parameters (System - QOS)** > Change the QoS values from their defaults and set them as indicated in Table 6-12.

Table 6-12 QoS Parameter Settings for Unified CM Endpoints

QoS Service Parameter Name (SIP Profile)	Default Value	Changed Value
DSCP for Audio Calls	EF	No change
DSCP for Video Calls	AF41	No change
DSCP for Audio Portion of Video Calls	AF41	EF
DSCP for TelePresence Calls	CS4	AF41
DSCP for Audio Portion of TelePresence Calls	CS4	EF

- Cisco Unity Connection

  System settings > **Advanced** > **Telephony**

  – Default = Audio (46 / EF), Video (46 / EF), Signaling (24 / CS3)

  – Change Video to 34 / AF41

- Cisco TelePresence Conductor and TelePresence Server

  – Leave defaults

- Cisco Expressway

  System > **Quality of Service**

  – Sets QoS for *all* media and signaling to a single DSCP

  – Default = 0

  – Change to: 36 / AF42

WAN Edge Identification and Classification

At the WAN edge on ingress from the enterprise to the service provider, it is expected that the packets that arrive with a specific DSCP value because the collaboration traffic have been re-marked at the access layer switch. On ingress it is important to re-mark any traffic at the WAN edge that could not be re-marked at the access layer, as a failsafe in case any traffic from the access switches was trusted through the LAN. While QoS is important in the LAN, it is paramount in the WAN; and as routers assume a trust on ingress traffic, it is important to configure the correct QoS policy that aligns with the business requirements and user experience. The WAN edge re-marking is always done on the ingress interface into the router, while the queuing and scheduling is done on the egress interface. The following example walks through the WAN ingress QoS policy as well as the egress queuing policy. Figure 6-26 through Figure 6-31 illustrate the configuration and the re-marking process.

In Figure 6-26 the packets from endpoints are identified and classified with the appropriate DSCP marking via a trusted port or via an ACL. Because there are typically areas of the switched access network that either cannot be configured with the correct QoS policies or re-mark collaboration traffic to Best Effort DSCP (BE), the WAN ingress policy is a good place for a catch-all policy to readdress what the access layer might have missed before the traffic heads into the WAN.

Figure 6-26 Example Router Ingress QoS Policy Process – 1

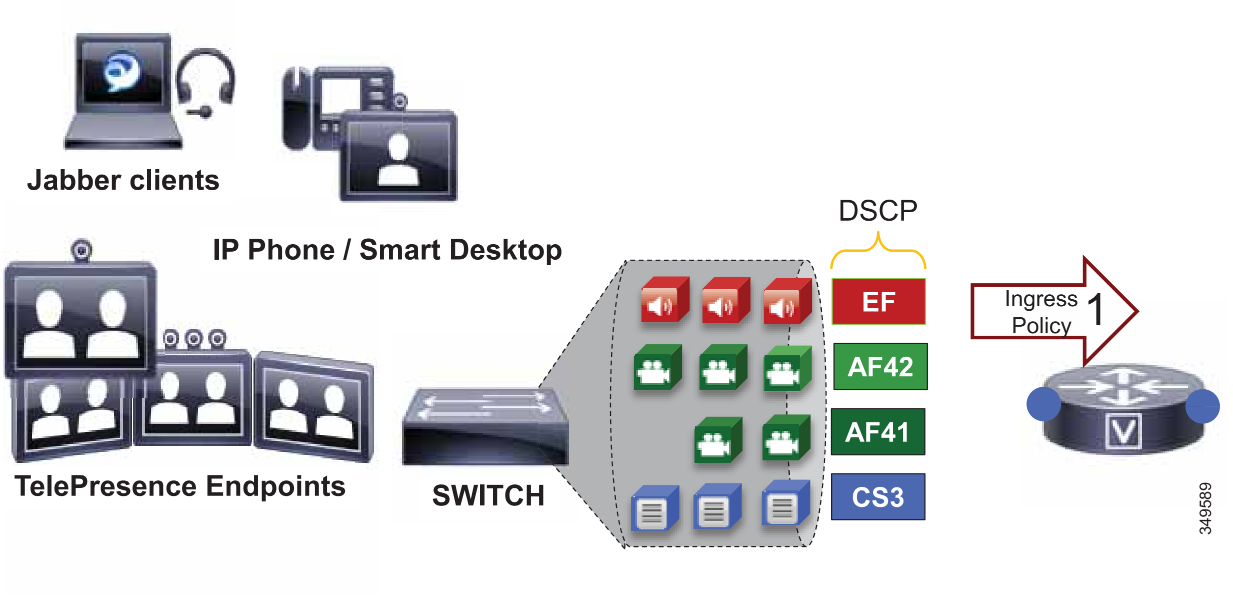
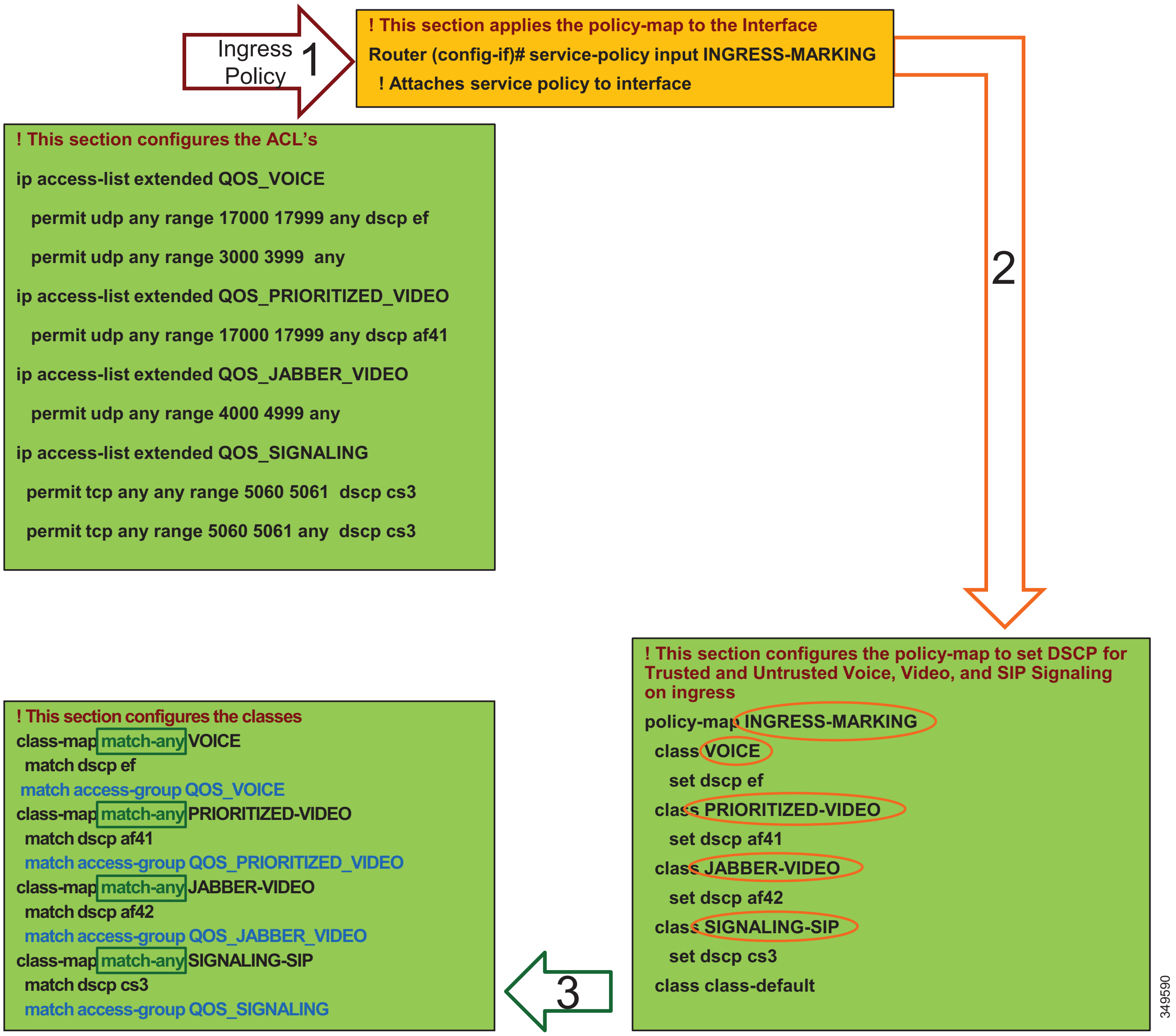


Figure 6-27 through Figure 6-29 illustrate the policy matching criteria and DSCP re-marking. The illustrations show the following steps:

1. In step 1, packets arrive at the router ingress interface, which is configured with an input service policy.
2. In step 2, the policy-map is configured with four classes of traffic to set the appropriate DSCP (voice = EF; prioritized-video = AF41; Jabber-video = AF42; signaling = CS3).
3. In step 3, each one of these classes matches a class-map of the same name configured with match-any criterion. This match-any criterion means that the process will start top-down, and the first matching criterion will be executed to set the DSCP according to each class in the policy-map statements.

Figure 6-27 Example Router Ingress QoS Policy Process – 2



```
! This section applies the policy-map to the Interface
Router (config-if)# service-policy input INGRESS-MARKING
  ! Attaches service policy to interface
```

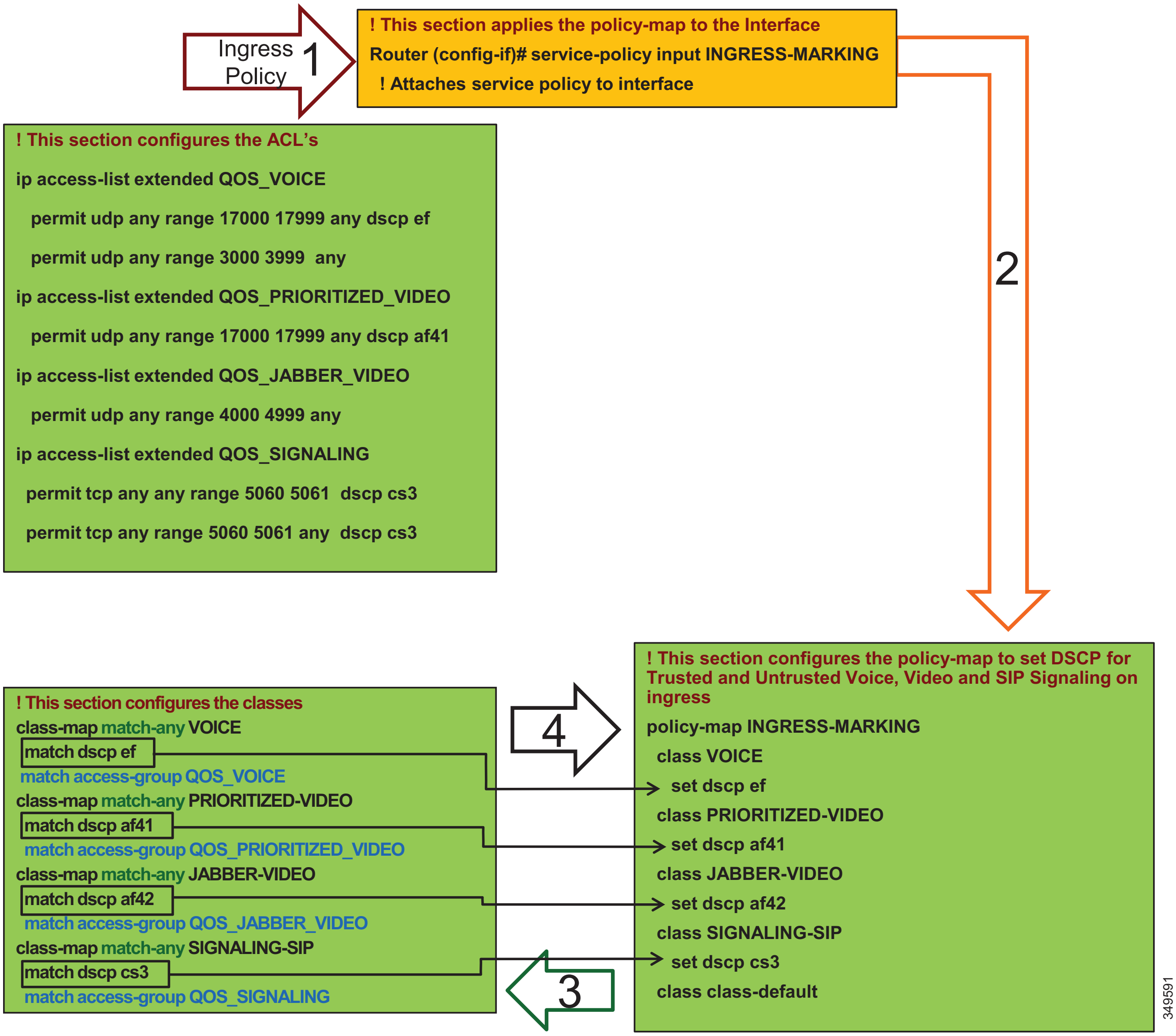
```
! This section configures the ACL's
ip access-list extended QOS_VOICE
   permit udp any range 17000 17999 any dscp ef
   permit udp any range 3000 3999  any
ip access-list extended QOS_PRIORITIZED_VIDEO
   permit udp any range 17000 17999 any dscp af41
ip access-list extended QOS_JABBER_VIDEO
   permit udp any range 4000 4999 any
ip access-list extended QOS_SIGNALING
  permit tcp any any range 5060 5061  dscp cs3
  permit tcp any range 5060 5061 any  dscp cs3
```

```
! This section configures the policy-map to set DSCP for
Trusted and Untrusted Voice, Video, and SIP Signaling
on ingress
policy-map INGRESS-MARKING
  class VOICE
     set dscp ef
  class PRIORITIZED-VIDEO
     set dscp af41
  class JABBER-VIDEO
     set dscp af42
  class SIGNALING-SIP
     set dscp cs3
  class class-default
```

```
! This section configures the classes
class-map match-any VOICE
  match dscp ef
 match access-group QOS_VOICE
class-map match-any PRIORITIZED-VIDEO
  match dscp af41
  match access-group QOS_PRIORITIZED_VIDEO
class-map match-any JABBER-VIDEO
  match dscp af42
  match access-group QOS_JABBER_VIDEO
class-map match-any SIGNALING-SIP
  match dscp cs3
  match access-group QOS_SIGNALING
```

4. In step 4, the first match statement in the class-map is **match dscp**. If the traffic matches the DSCP, then DSCP is set again to the same value that was matched and as is configured in the policy-map statements. In this case the router is simply matching on DSCP and resetting the DSCP to the same value.

Figure 6-28 Example Router Ingress QoS Policy Process – 3



```
! This section applies the policy-map to the Interface
Router (config-if)# service-policy input INGRESS-MARKING
  ! Attaches service policy to interface
```

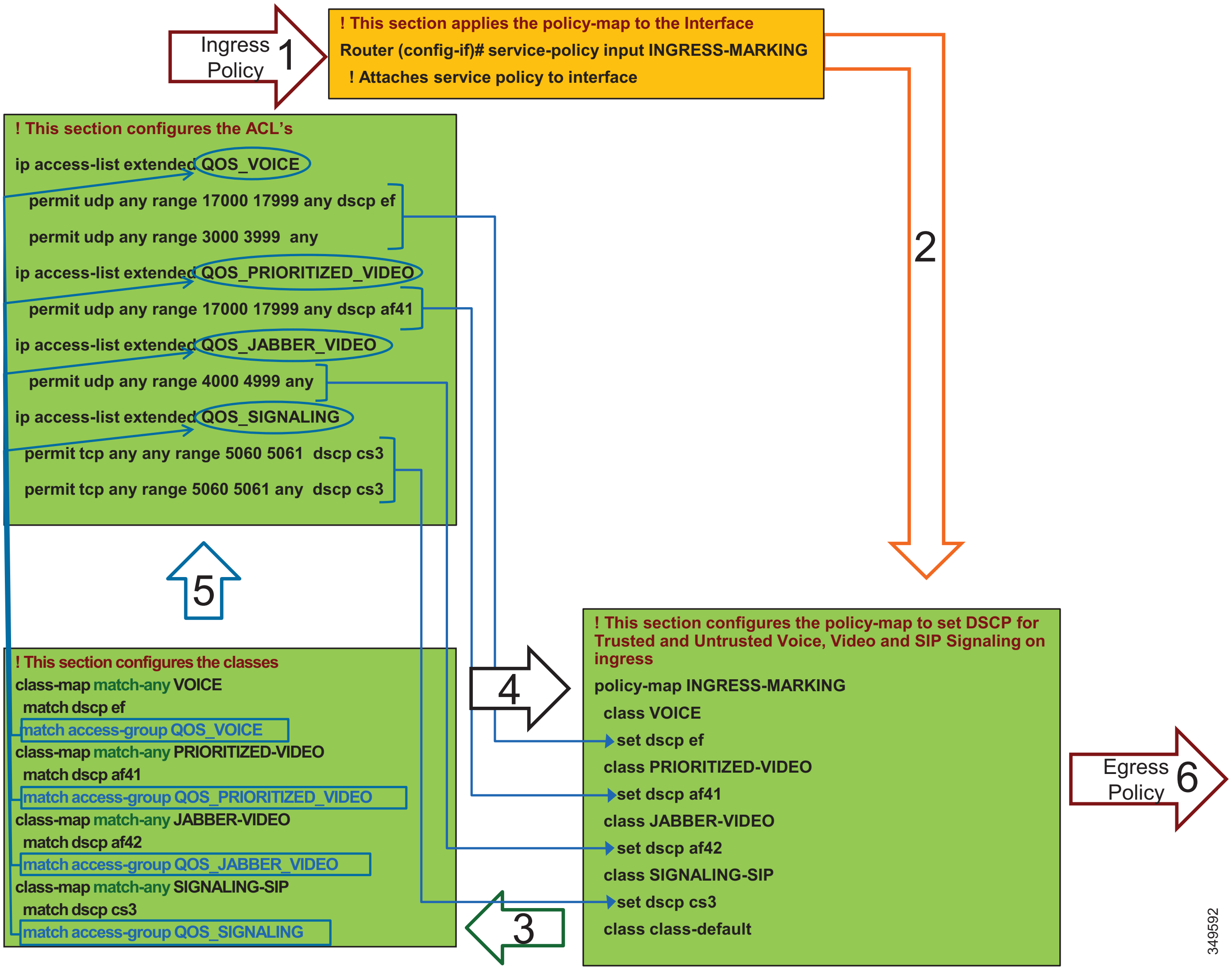
```
! This section configures the ACL's
ip access-list extended QOS_VOICE
  permit udp any range 17000 17999 any dscp ef
  permit udp any range 3000 3999  any
ip access-list extended QOS_PRIORITIZED_VIDEO
  permit udp any range 17000 17999 any dscp af41
ip access-list extended QOS_JABBER_VIDEO
  permit udp any range 4000 4999 any
ip access-list extended QOS_SIGNALING
  permit tcp any any range 5060 5061  dscp cs3
  permit tcp any range 5060 5061 any  dscp cs3
```

```
! This section configures the classes
class-map match-any VOICE
 match dscp ef
 match access-group QOS_VOICE
class-map match-any PRIORITIZED-VIDEO
 match dscp af41
 match access-group QOS_PRIORITIZED_VIDEO
class-map match-any JABBER-VIDEO
 match dscp af42
 match access-group QOS_JABBER_VIDEO
class-map match-any SIGNALING-SIP
 match dscp cs3
 match access-group QOS_SIGNALING
```

```
! This section configures the policy-map to set DSCP for
Trusted and Untrusted Voice, Video and SIP Signaling on
ingress
policy-map INGRESS-MARKING
 class VOICE
  set dscp ef
 class PRIORITIZED-VIDEO
  set dscp af41
 class JABBER-VIDEO
  set dscp af42
 class SIGNALING-SIP
  set dscp cs3
 class class-default
```

5. In step 5, if DSCP was not matched, then the next line in the class-map statement is parsed, which is the ACL that matches the UDP ports set in Unified CM for the Jabber clients in the Identification and Classification section. When the ACL criteria are met (protocol, port range, and/or DSCP), then the traffic is set as is configured in the corresponding policy-map statements.

Figure 6-29 Example Router Ingress QoS Policy Process – 4



```
! This section applies the policy-map to the Interface
Router (config-if)# service-policy input INGRESS-MARKING
 ! Attaches service policy to interface
```

```
! This section configures the ACL’s
ip access-list extended QOS_VOICE
  permit udp any range 17000 17999 any dscp ef
  permit udp any range 3000 3999  any
ip access-list extended QOS_PRIORITIZED_VIDEO
  permit udp any range 17000 17999 any dscp af41
ip access-list extended QOS_JABBER_VIDEO
  permit udp any range 4000 4999 any
ip access-list extended QOS_SIGNALING
  permit tcp any any range 5060 5061  dscp cs3
  permit tcp any range 5060 5061 any  dscp cs3
```

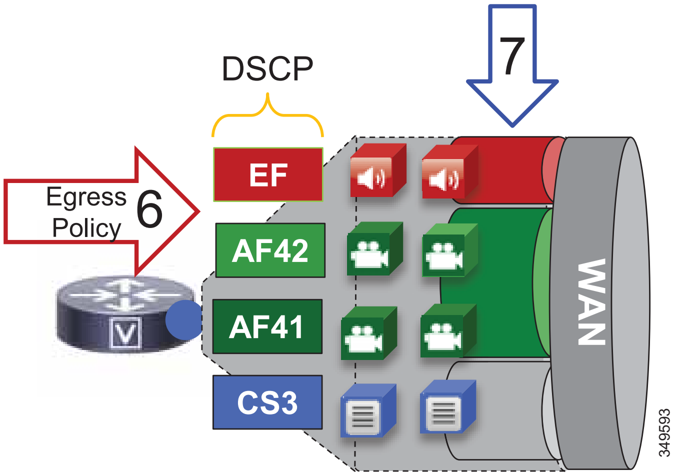
```
! This section configures the classes
class-map match-any VOICE
 match dscp ef
 match access-group QOS_VOICE
class-map match-any PRIORITIZED-VIDEO
 match dscp af41
 match access-group QOS_PRIORITIZED_VIDEO
class-map match-any JABBER-VIDEO
 match dscp af42
 match access-group QOS_JABBER_VIDEO
class-map match-any SIGNALING-SIP
 match dscp cs3
 match access-group QOS_SIGNALING
```

```
! This section configures the policy-map to set DSCP for
Trusted and Untrusted Voice, Video and SIP Signaling on
ingress
policy-map INGRESS-MARKING
  class VOICE
   set dscp ef
  class PRIORITIZED-VIDEO
   set dscp af41
  class JABBER-VIDEO
   set dscp af42
  class SIGNALING-SIP
   set dscp cs3
  class class-default
```

Note This is an example QoS ingress marking policy based on the Cisco Common Classification Policy Language (C3PL). Refer to your specific router configuration guide for any updated C3PL commands and for information on how to configure a similar policy on a Cisco router supporting C3PL.

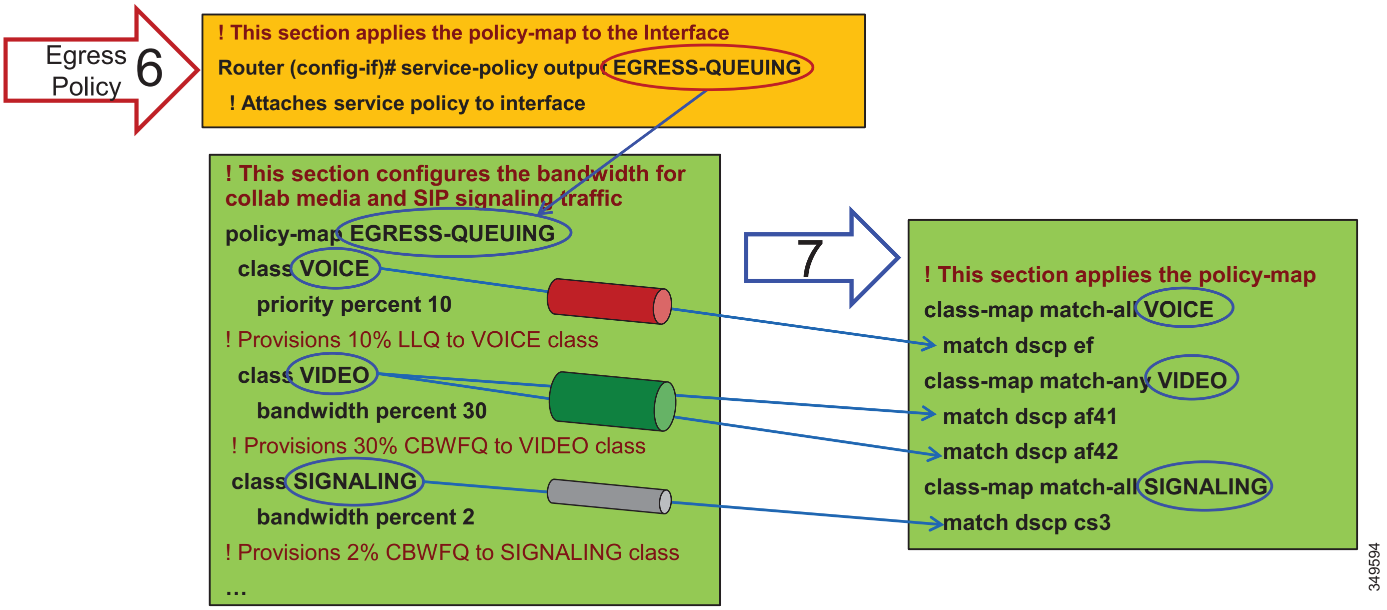
6. In step 6, the traffic goes to an outbound interface to be queued and scheduled by an output service policy that has three queues created: a Priority Queue called VOICE, a CBWFQ called VIDEO, and another CBWFQ called SIGNALING. This is illustrated in Figure 6-30 through Figure 6-31. This highlights the fact that the egress queuing policy is based only on DSCP as network marking occurring at the access switch and/or on ingress into the WAN router ingress interface. This is an example simply to illustrate the matching criteria and queues, and it does not contain the WRED functionality. For information on WRED, see the WAN Edge Queuing and Scheduling section.

Figure 6-30 Example Router Egress Queuing Policy Process – 1



7. In step 7, the traffic is matched against the class-map match statements. All traffic marked EF goes to the VOICE PQ, AF41 and AF42 traffic goes to the VIDEO CBWFQ, and CS3 traffic goes to the SIGNALING CBWFQ.

Figure 6-31 Example Router Egress Queuing Policy Process – 2



```
! This section applies the policy-map to the Interface
Router (config-if)# service-policy output EGRESS-QUEUING
 ! Attaches service policy to interface
```

```
! This section configures the bandwidth for
collab media and SIP signaling traffic
policy-map EGRESS-QUEUING
  class VOICE
    priority percent 10
! Provisions 10% LLQ to VOICE class
  class VIDEO
    bandwidth percent 30
 ! Provisions 30% CBWFQ to VIDEO class
 class SIGNALING
    bandwidth percent 2
! Provisions 2% CBWFQ to SIGNALING class
…
```

```
! This section applies the policy-map
class-map match-all VOICE
  match dscp ef
class-map match-any VIDEO
  match dscp af41
  match dscp af42
class-map match-all SIGNALING
  match dscp cs3
```

Note This is an example egress queuing policy based on the Cisco Common Classification Policy Language (C3PL). Refer to your specific router configuration guide for any updated C3PL commands and for information on how to configure a similar policy on a Cisco router supporting C3PL.

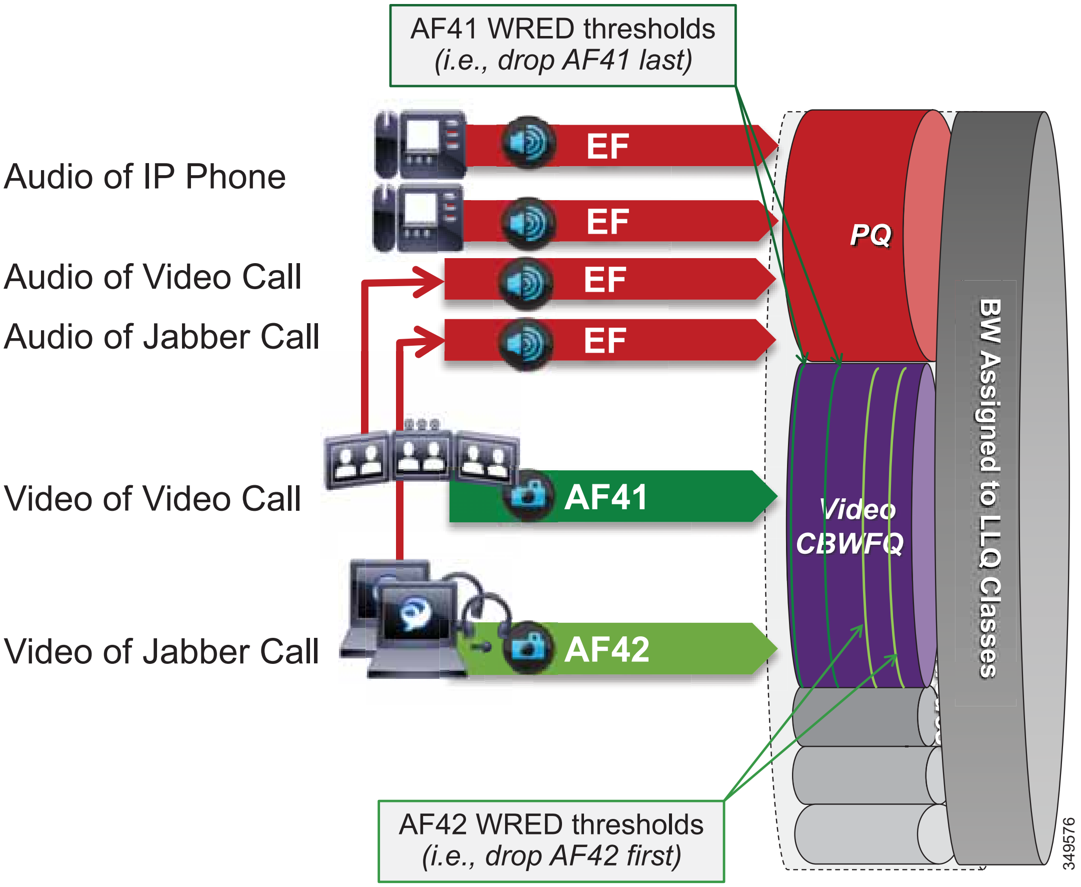
Example Configuration of Egress Queuing

```
! This section applies the policy-map classes to match media and signaling QoS.
class-map match-any VIDEO
 match dscp af41
 match dscp af42
class-map match-any VOICE
 match dscp ef
class-map match-any SIGNALING
 match dscp cs3

! This section configures the bandwidth for Collaboration media and SIP signaling traffic.
policy-map EGRESS-QUEUING
 class VOICE
  priority percent 10
 class VIDEO
  bandwidth percent 30
  fair-queue
 class SIGNALING
  bandwidth percent 2
…

! This section applies the policy-map to the interface.
Router (config-if)# service-policy output EGRESS-QUEUING
  ! Attaches service policy to interface
```

WAN Edge Queuing and Scheduling

This section covers the interface queuing. Figure 6-32 illustrates the voice PQ, video CBWFQ, and WRED thresholds used for the CBWFQ.

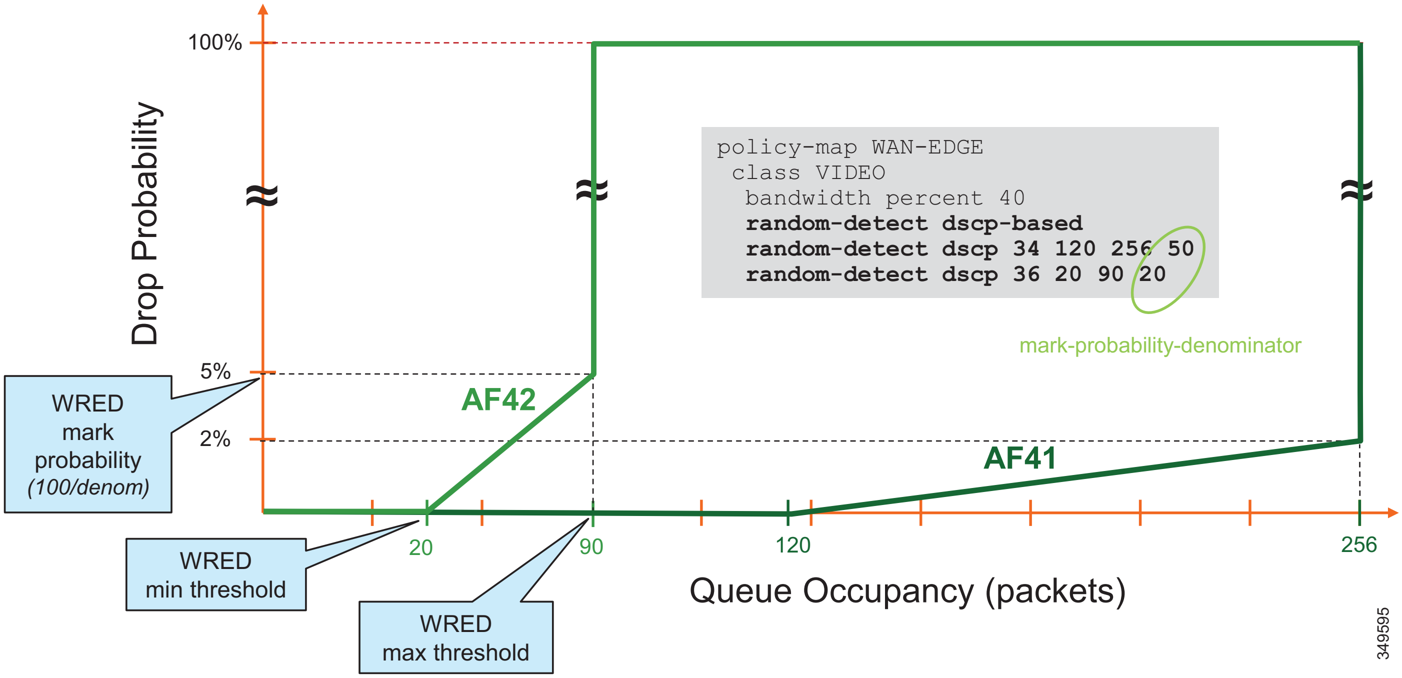
Figure 6-32 Queuing and Scheduling Collaboration Media



- All audio from all endpoints marked EF is mapped to the PQ.
- Video calls and Jabber share the same CBWFQ.
  - EF for audio streams of video calls from endpoints
  - AF41 for video streams of video calls from endpoints
  - EF for audio streams of all calls from Jabber clients
  - AF42 for video streams of video calls from Jabber clients
- WRED is configured on the video queue.
  - Minimum to maximum thresholds for AF42: approximately 10% to 30% of queue limit
  - Minimum to maximum thresholds for AF41: approximately 45% to 100% of queue limit

Weighted Random Early Detection (WRED) threshold minimum and maximum values are configured in the Video CBWFQ. To illustrate how the WRED thresholds are configured, assume that the interface had been configured with a queue depth of 256 packets. Then following the guidelines above, the WRED minimum and maximum thresholds for AF42 and AF41 would be configured as illustrated in Figure 6-33.

Figure 6-33 Threshold Example for Video CBWFQ with WRED

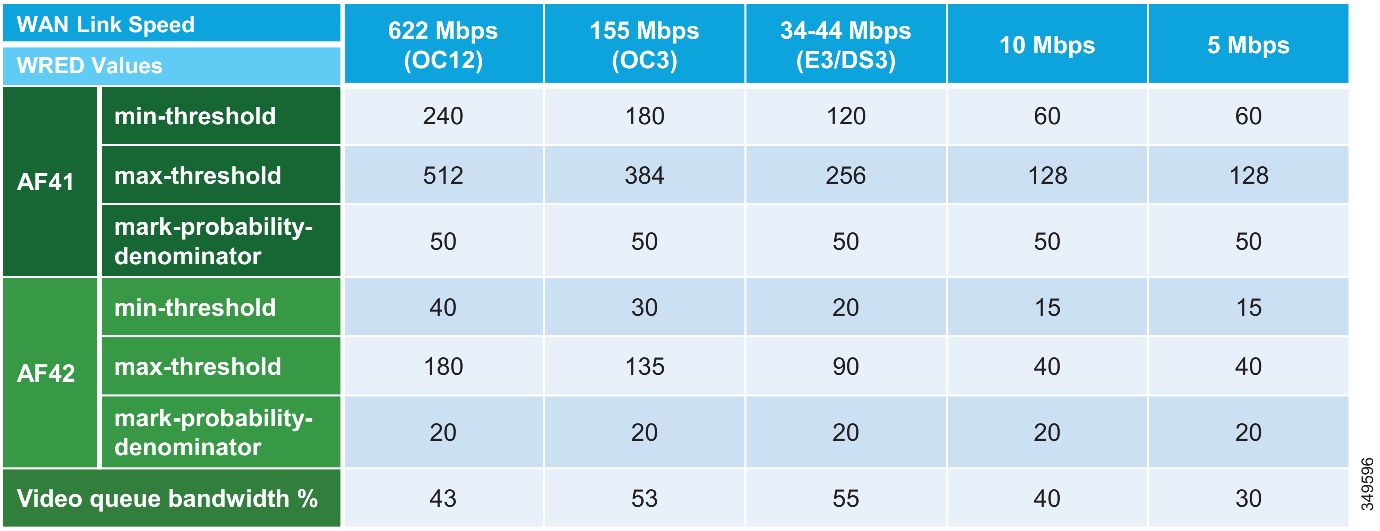


```
policy-map WAN-EDGE
 class VIDEO
  bandwidth percent 40
  random-detect dscp-based
  random-detect dscp 34 120 256 50
  random-detect dscp 36 20 90 20
```

Recommended WRED Thresholds

Figure 6-34 lists the WRED thresholds for each traffic class (AF41 and AF42) and the recommended mark probability denominators that have been tested for various link speeds. These are just examples, and testing and customization are expected based on the amount of traffic in each traffic class and the aggressiveness required in the WRED drop probability during the busy hour.

Figure 6-34 Recommended WRED Thresholds by Link Speed



WAN Link Speed		622 Mbps (OC12)	155 Mbps (OC3)	34-44 Mbps (E3/DS3)	10 Mbps	5 Mbps
WRED Values						
AF41	min-threshold	240	180	120	60	60
	max-threshold	512	384	256	128	128
	mark-probability-denominator	50	50	50	50	50
AF42	min-threshold	40	30	20	15	15
	max-threshold	180	135	90	40	40
	mark-probability-denominator	20	20	20	20	20
Video queue bandwidth %		43	53	55	40	30

The following example configuration is for WRED in the video Class-Based Weighted Fair Queue (CBWFQ) of a DS3 link (44 Mbps).

```
policy-map EGRESS-QUEUING
 class VOICE
  priority percent 10
 class VIDEO
  bandwidth percent 30
  random-detect dscp-based
  random-detect dscp 34 120 256 50
  random-detect dscp 36 20 90 20
fair-queue
 class SIGNALING
  bandwidth percent 2
```

Note The WRED values might have o be customized for the specific environment. For example, if there is a much larger amount of AF42 traffic than AF41 traffic, then it makes sense to adjust the WRED threshold variables to suit those cases. Tweaking the variables and monitoring levels of drop is always the best way to achieve the desired results.

Provisioning and Admission Control

This section addresses admission control and provisioning bandwidth to the queues for each site type. It covers the following topics:

- Enhanced Locations CAC
  - Region Configuration
  - Deploy Enhanced Locations Call Admission Control
- Deploy Device Mobility for Mobile and Remote Access (MRA)
- Bandwidth Allocation Guidelines

This phase of the deployment involves the following high-level steps:

1. Configure Enhanced Locations CAC.
2. Configure a region matrix for maximum video bit rate groups.
3. Deploy Device Mobility for mobile and remote access (MRA) endpoints.
4. Follow bandwidth allocation guidelines.

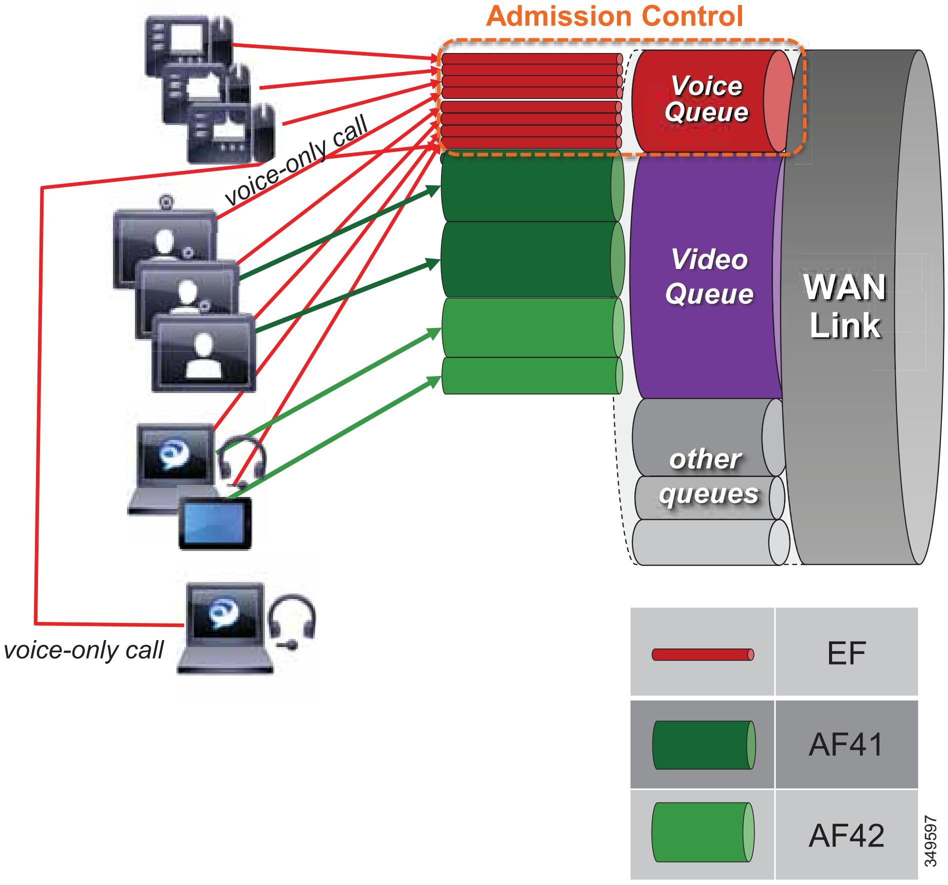
Enhanced Locations CAC

Admission control is not used in this case to manage the video bandwidth but instead to manage the audio traffic to ensure that the Priority Queue (PQ) is not over-subscribed. In this specific example the Voice pool in Enhanced Locations CAC admits the audio for both the voice-only calls and the video calls.

In Unified CM this feature is enabled by setting the service parameter **Deduct Audio Bandwidth from Audio Pool for Video Call** to **True** under the Call Admission Control section of the CallManager service. **False** is the default setting, and by default Unified CM deducts both audio and video streams of video calls from the video pool. This parameter changes that behavior and is key to the QoS alterations in the Preferred Architecture.

Figure 6-35 illustrates the various call flows, their corresponding audio and video streams, and queues to which queue they are directed.

Figure 6-35 Provisioning and Admission Control



The following conditions apply to the example in Figure 6-35:

- The Priority Queue is provisioned for all calls from endpoints and is protected by admission control (*E-LCAC voice BW pool*).
- The Video queue is over-provisioned for room-based video systems:
  - Ratios are applied to desktop video endpoint usage.
  - Jabber video calls can use any bandwidth unused by video room systems.
  - During congestion, video streams of Jabber calls are subject to WRED drops and dynamically reduce video bit rate.

Region Configuration

Group video endpoints into classes of maximum video bit rate to limit bandwidth consumption based on endpoint type and usage within the solution. Three regions are required in total (see Table 6-13), and three device pools are required per site. This applies to a configuration where a single audio codec of G.722 is used across the entire organization, both LAN and WAN. Otherwise three regions per site are also required. See the considerations for regions in the Architecture section.

Table 6-13 Example Region Matrix for Three Groups

Endpoint Groupings	Video_1.5MB	Video_2.5MB	Video_20MB
Video_1.5MB	1,500 kbps	1,500 kbps	1,500 kbps
Video_2.5MB	1,500 kbps	2,500 kbps	2,500 kbps
Video_20MB	1,500 kbps	2,500 kbps	20,000 kbps

Deploy Enhanced Locations Call Admission Control

Limit video calling based only in areas of the network where bandwidth resources are restricted beyond AF41 marked traffic; otherwise, video bandwidth in the Location links should be unlimited.

- Enable LBM services on every node where the Cisco CallManager service is enabled.
- Configure locations.
  - On the locations and links management cluster, configure all locations and links in the organization.
  - On all other clusters (subordinate to a locations and links management cluster), configure only locations and remove any links to/from the locations.
- Add locations to each device pool. The devices that must be configured in a location either directly or via a device pool include:
  - IP phones (via device pool)
  - Conference bridges (via device pool)
  - Gateways (via device pool)
  - SIP trunks (via device pool)
  - Music on hold (MoH) servers (directly)
  - Annunciator (via device pool)

Intercluster Configuration

- Configure the LBM hub group
  - Used to assign LBMs to the hub role
  - Used to define three remote hub members that replicate hub contact information for all of the hubs in the LBM hub replication network

    An LBM is a hub when it is assigned to an LBM hub group.

    An LBM is a spoke when it is not assigned to an LBM hub group.
  - Name: Cluster1_LBM_Hub_1
  - Bootstrap Servers: <*names or IP addresses of bootstrap servers*> (see the LBM Hub Replication Network section)
  - Select up to two LBMs in the cluster to serve as hubs.
- Recommendations for location configuration when intercluster ELCAC is implemented:
  - A cluster requires the location to be configured locally for location-to-device pool association.
  - Each cluster should have locations configured with the immediately neighboring locations of other clusters, so that each cluster's topology can inter-connect and create a single global topology. This does not apply to Location and Link Management Cluster deployments.
  - Discrepancies of bandwidth limits and weights on common locations and links are resolved by using the lowest bandwidth and weight values.
  - Naming locations consistently across clusters is critical. Follow the practice: "Same location, same name; different location, different name."
  - The Hub_none location should be renamed to be unique in each cluster. If Hub_none is left as default on all clusters, then it will be treated as the same location, which may or may not be desired, depending on the locations design being configured (see the Enhanced Locations Call Admission Control section).
  - Cluster-ID should be configured and must be unique on each cluster for serviceability reports to be usable.

Deploy Device Mobility for Mobile and Remote Access (MRA)

Configure Device Mobility

Figure 6-36 illustrates an overview of the device mobility configuration. Although this is a minimum configuration requirement for Device Mobility for ELCAC to function for Internet-based devices, Device Mobility can be configured to support mobility for these same endpoints within the enterprise. See the Cisco Collaboration SRND for more information on Device Mobility for devices within the enterprise.

Figure 6-36 Device Mobility Configuration and Location Association

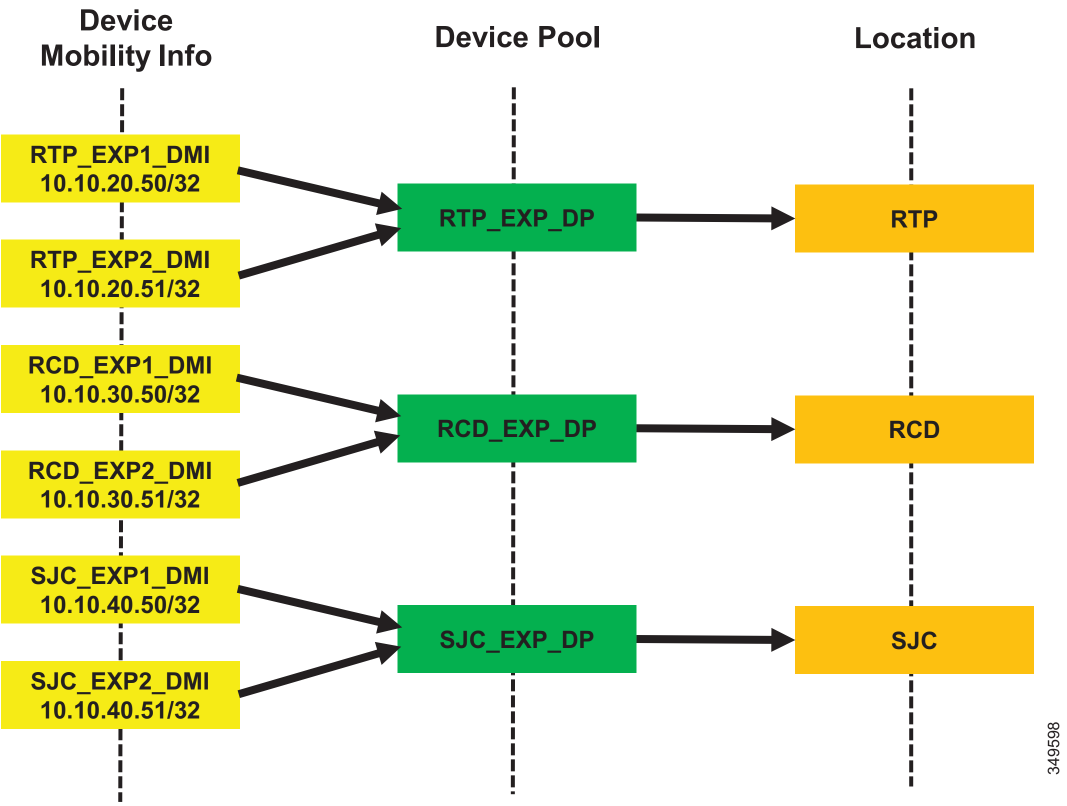
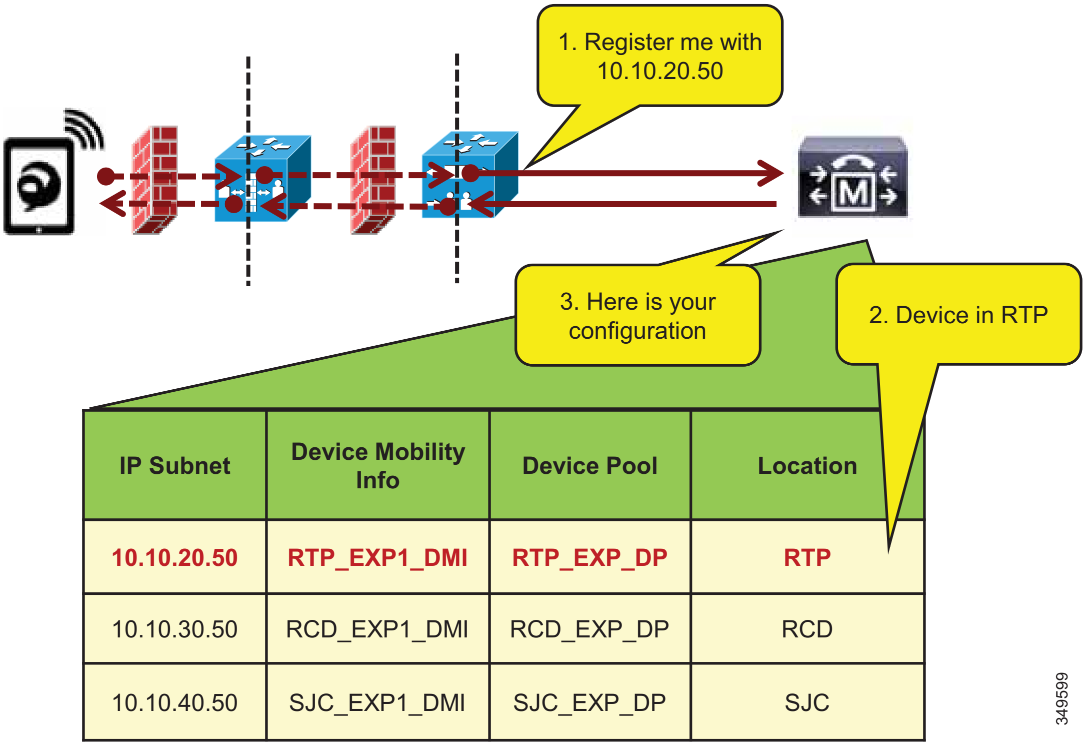

Device Mobility Info
Device Pool
Location
RTP_EXP1_DMI 10.10.20.50/32
RTP_EXP2_DMI 10.10.20.51/32
RCD_EXP1_DMI 10.10.30.50/32
RCD_EXP2_DMI 10.10.30.51/32
SJC_EXP1_DMI 10.10.40.50/32
SJC_EXP2_DMI 10.10.40.51/32
RTP_EXP_DP
RCD_EXP_DP
SJC_EXP_DP
RTP
RCD
SJC
349598

Figure 6-36 shows a simplified version of device mobility for the example deployment of ELCAC. The IP addresses of the Expressway-C servers are configured in the device mobility information. In this example there is a redundant pair of Expressway-C servers for each of the three sites: RTP, BLD, and SJC. RTP_EXP1_DMI and RTP_EXP2_DMI are configured respectively with the server IP addresses of the RTP Expressway-C servers. These two are associated to a new device pool called RTP_EXP_DP, which has the location RTP configured on it. Each site is configured similarly. With this configuration, when any device enabled for device mobility registers to Unified CM with the IP address that corresponds to the device mobility information in RTP_EXP1_DMI or RTP_EXP2_DMI, it will be associated with the RTP_EXP_DP device pool and thus with the RTP location.

With the above configuration, when an Internet-based device registers through the Expressway to Unified CM, it will register with the IP address of Expressway-C. Unified CM then uses the IP address configured in the device mobility information and associates the device pool and thus the Internet location associated to this device pool. This process is illustrated in Figure 6-37.

Figure 6-37 Association of Device Pool and Location Based on Expressway IP Address


1. Register me with 10.10.20.50
2. Device in RTP
3. Here is your configuration
349599

IP Subnet	Device Mobility Info	Device Pool	Location
10.10.20.50	RTP_EXP1_DMI	RTP_EXP_DP	RTP
10.10.30.50	RCD_EXP1_DMI	RCD_EXP_DP	RCD
10.10.40.50	SJC_EXP1_DMI	SJC_EXP_DP	SJC

In Figure 6-37. the client registers with Unified CM through the Expressway in RTP. Because the signaling is translated at the Expressway-C in RTP, the device registers with the IP address of that Expressway-C. The device pool RTP_EXP_DP is associated to the device based on this IP address. The RTP_EXP_DP pool is configured with the RTP location, and therefore that location is associated to the device. Thus, when devices register to the Expressway, they get the correct location association through device mobility. When the endpoint relocates to the enterprise, it will return to its static location configuration. Also, if the endpoint relocates to another Expressway in SJC, for example, it will get the correct location association through device mobility.

Configure Device Mobility Information (DMI) for Expressway-Cs:

- Create two DMIs per Expressway-C group (two Expressway-C nodes in a pair).
- Add the IP address of the Expressway-C node in a subnet with a mask of 32 bits (this matches the IP address exactly).
- Add the site device pool to respective DMIs. This is the device pool of the site where the Expressway pairs are located, which should contain the correct region and location.

  Example for one DMI:

  Name: SJC_EXP1_DMI

  Subnet: 10.10.40.50

  Mask: 32

  Selected Device Pool: SJC_Video_1.5MB

Enable devices for device mobility. The bulk administration tool (BAT) can and should be used to facilitate this step.

Bandwidth Allocation Guidelines

The bandwidth allocations in Figure 6-38 are unique guidelines based on this example enterprise. They provide some guidance on percentages of available bandwidth for different common classes of Collaboration traffic.

Figure 6-38 Bandwidth Allocation Guidelines

WAN Link Speed Class	622 Mbps (OC12)	155 Mbps (OC3)	34-44 Mbps (E3/DS3)	10 Mbps	5 Mbps	<2 Mbps (T1/E1)
Control (%)	1	1	1	1	2	10
Voice (%)	10	10	10	20	25	30
Video (%)	43	53	55	40	30	--
Signalling (%)	2	2	2	2	2	5
Scavenger (%)	1	1	1	1	1	1
Default (%)	43	33	31	36	40	54

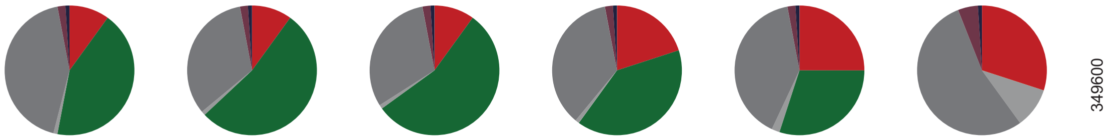
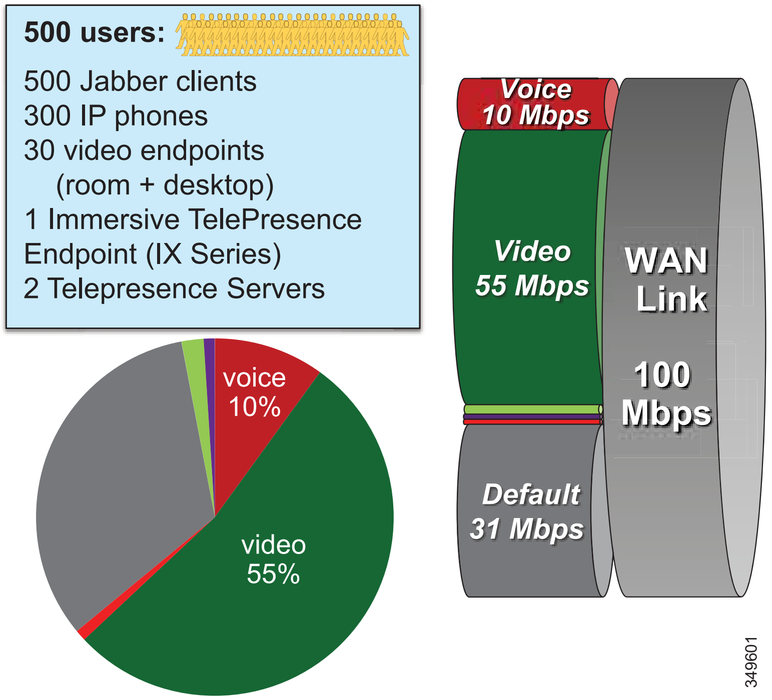


Figure 6-39 through Figure 6-42 illustrate each site (Central, Large Branch, Small Branch, Micro Branch) and the link bandwidth provisioned for each class based on the number of users and available bandwidth for each class. Keep in mind that these values are based on bandwidth calculated for Layer 3 and above. Therefore, the values do not include the Layer 2 overhead, which is dependent on the link type (Ethernet, Frame-relay, MPLS, and so forth). See the Network Infrastructure chapter of the Collaboration SRND for more information on L2 overhead. Also note that the audio portion of bandwidth for the video calls is deducted from the voice pool, so the voice queue is provisioned to include the audio bandwidth of both voice-only and video calls.

Note The calculations in the following examples use the maximum bandwidth for the number of endpoints and then multiply that value by a percentage to account for active calls. For example, for 30 video endpoints (30 calls possible) at 20% active video call rate, the calculation would be:
1.2 Mbps * 30 calls * 0.2 = 7.2 Mbps.

Figure 6-39 Central Site



Central Site Link (100 Mbps) Bandwidth Calculation

As illustrated in Figure 6-39, the Central Site has the following bandwidth requirements:

- Voice queue (PQ): 10 Mbps (L3 bandwidth)

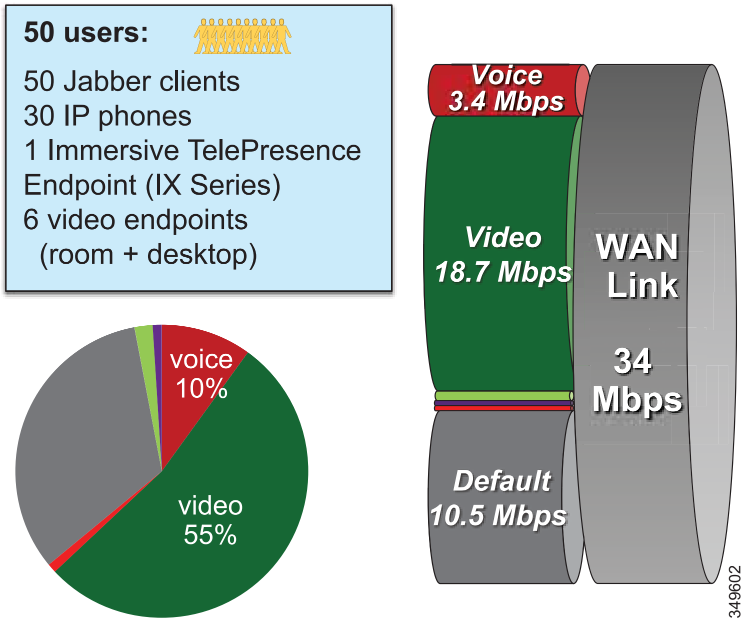
  125 calls @ G.711 or G.722
- Unified CM Location link bandwidth for the voice pool:

  125 ∗ 80 kbps = 10 Mbps
- Video queue: 55 Mbps (L3 bandwidth)
  - Immersive video endpoint (IX Series): 3 Mbps ∗ 1 call = 3 Mbps
  - Video endpoints: 1.2 Mbps ∗ 30 calls ∗ 0.2 = 7.2 Mbps
  - TelePresence Servers: 1.5 Mbps ∗ 40 calls ∗ 0.5 = 30 Mbps
  - 55 Mbps – (3 Mbps + 7.2 Mbps + 30 Mbps) = 14.8 Mbps for Jabber media

    11 Jabber video calls @ 720p, or 18 @ 576p, or 50 @ 288p

    (Plus any leftover bandwidth)

Figure 6-40 Large Branch



Large Branch Link (34 Mbps) Bandwidth Calculation

As illustrated in Figure 6-40, the Large Branch site has the following bandwidth requirements:

- Voice queue (PQ): 3.4 Mbps (L3 bandwidth)

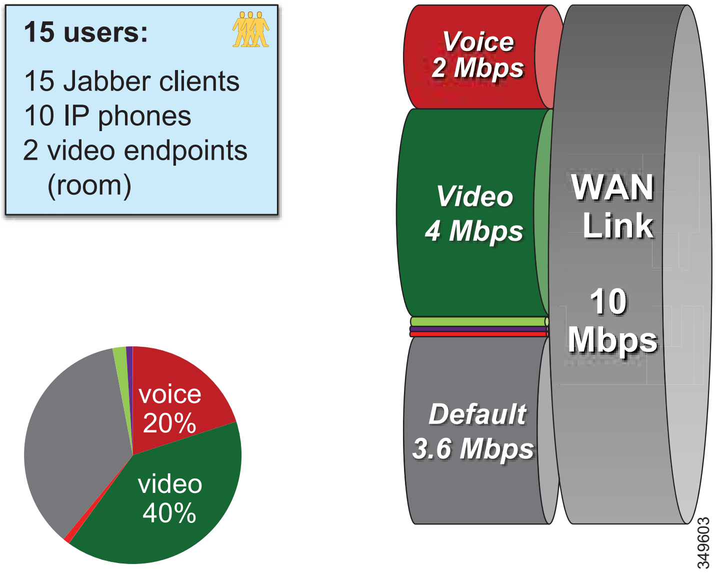
  42 calls @ G.711 or G.722
- Unified CM Location link bandwidth for the voice pool:

  42 ∗ 80 kbps = 3.360 Mbps
- Video queue: 18.7 Mbps (L3 bandwidth)
  - Immersive video endpoint (IX Series): 3 Mbps ∗ 1 call = 3 Mbps
  - Video endpoints: 1.2 Mbps ∗ 6 calls = 7.2 Mbps
  - 18.7 Mbps – (3 Mbps + 7.2 Mbps) = 8.5 Mbps for Jabber media

    6 Jabber video calls @ 720p, or 10 @ 576p, or 36 @ 288p

    (Plus any leftover bandwidth)

Figure 6-41 Small Branch



Small Branch Link (10 Mbps) Bandwidth Calculation

As illustrated in Figure 6-41, the Small Branch site has the following bandwidth requirements:

- Voice queue (PQ): 2 Mbps (L3 bandwidth)

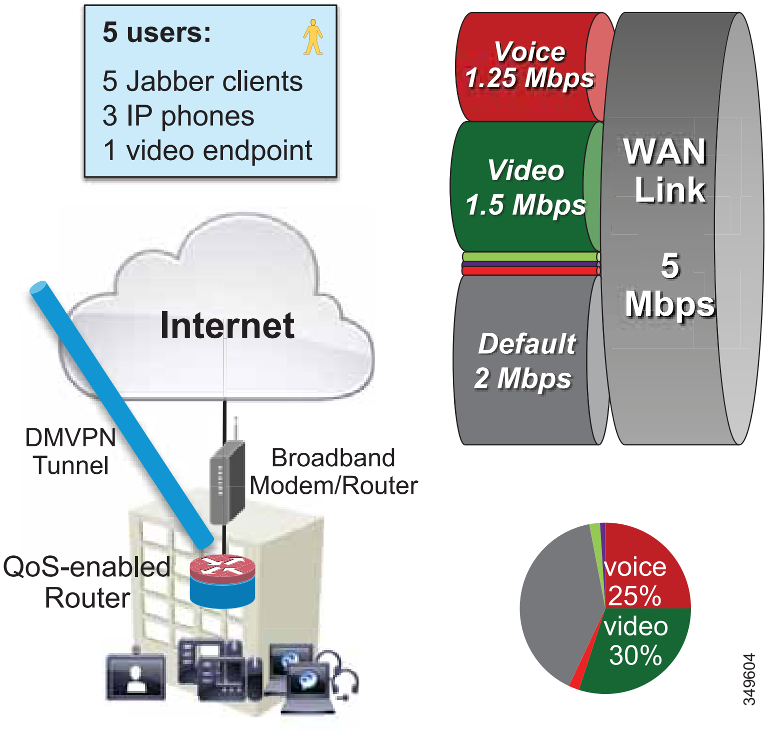
  25 calls @ G.711 or G.722
- Unified CM Location link bandwidth for the voice pool:

  25 ∗ 80 kbps = 2 Mbps
- Video queue: 4 Mbps (L3 bandwidth)
  - Video endpoints: 1.2 Mbps * 2 calls = 2.4 Mbps
  - 4 Mbps – 2.4 Mbps = 1.6 Mbps for Jabber media

    1 Jabber video call @ 720p, or 2 @ 576p, or 5 @ 288p

    (Plus any leftover bandwidth)

Figure 6-42 Micro Branch



Micro Branch Broadband Internet Connectivity (5 Mbps) Bandwidth Calculation

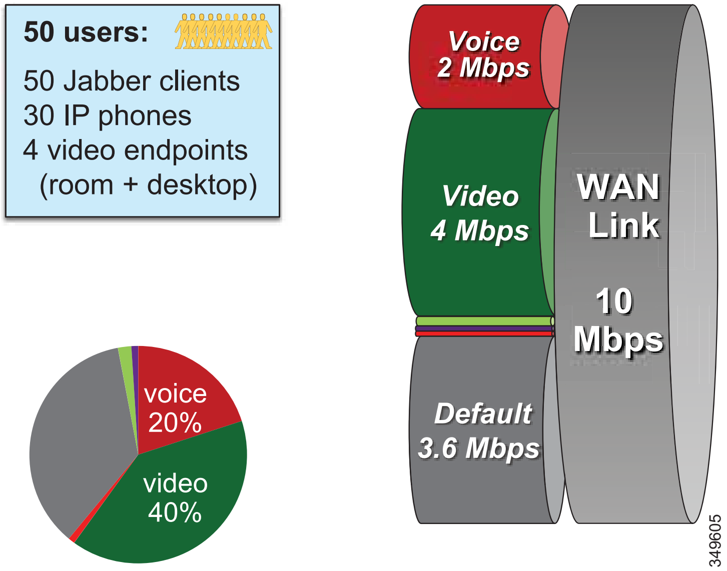
As illustrated in Figure 6-42, the Micro Branch site has the following bandwidth requirements:

- Broadband Internet connectivity + DMVPN to central site
- Configure interface of VPN router to match broadband up-link speed
- Enable QoS on VPN router to prevent bufferbloat from TCP flows
- Asymmetric download/upload broadband: consider limiting transmit bit rate on video endpoint
- Bandwidth calculation will ultimately depend on the broadband bandwidth available and should follow the same recommendations as in the Small Branch site link for provisioning.

Large Branch with Constrained WAN Link (Enhanced Locations CAC Enabled for Video)

In specific branch sites with lower-speed WAN links, over-provisioning the video queue is not feasible. ELCAC can be applied to these location links for video to ensure that video calls do not over-subscribe the link bandwidth. This template requires using site-specific region configuration to limit maximum bandwidth used by video endpoints and Jabber clients. Also keep in mind that device mobility is required if Jabber users roam across sites.

Figure 6-43 Large Branch with Constrained WAN Link (Enhanced Locations CAC Enabled for Video)



As illustrated in Figure 6-43, a Large Branch site with a constrained WAN link (10 Mbps) has the following bandwidth requirements:

- Voice queue (PQ): 2 Mbps (L3 bandwidth)

  25 calls @ G.711 or G.722
- Unified CM Location link bandwidth for the voice pool:

  25 ∗ 80 kbps = 2 Mbps
- Video queue: 4 Mbps (L3 bandwidth)
  - Possible usage:

    1 call @ 720p (1,220 kbps) + 3 calls @ 576p (810 kbps) = 3,650 kbps

    Or 2 calls @ 576p (768 kbps) + 5 calls @ 288p (320 kbps) = 3136 kbps
  - Unified CM Location link bandwidth for video calls: 3.7 Mbps (L3 bandwidth)
  - Leaves room for L2 overhead

CHAPTER 7

Sizing

Revised: November 20, 2015

Sizing the components of the Preferred Architecture for Enterprise Collaboration solution is an important part of the overall solution design.

For a given deployment, the goal of the sizing process is to determine:

- The type of platform to be used for each Cisco Collaboration product. Most products are deployed with virtualization only, but some products such as the Cisco TelePresence Server can also be deployed as an appliance or blade, depending on the requirements.
- The specifications and number of instances to be deployed for each Cisco Collaboration product. For the products that are deployed with virtualization, this corresponds to the selection of the virtual machine hardware specification defined in the Open Virtual Archive (OVA) template and the number of virtual machines. For the products that are not deployed with virtualization, this corresponds to the type and number of appliances or blades.

Sizing can be a complex exercise because of numerous parameters to take into considerations. In order to simplify the sizing exercise, this chapter provides some sizing examples with corresponding assumptions. We will refer to these sizing examples as *simplified sizing deployments*. If the requirements of your particular deployment are within those assumptions, then you can use the simplified sizing deployments in this document as a reference. If not, then the normal sizing calculations have to be performed as described in the *Sizing* chapter of the *Cisco Collaboration SRND* and product documentation available at http://www.cisco.com/go/ucsrnd.

Once the sizing is done for the products that are deployed with virtualization, determine how to place the virtual machines on Cisco Unified Computing System (UCS) servers, and consider the co-residency rules. Ultimately, this virtual machine placement process determines how many UCS servers are required for the solution.

This chapter explains sizing for all modules that are covered in this document, namely: Call Control, Conferencing, Collaboration Edge, and Core Applications. This chapter also covers Virtual Machine Placement and Platforms.

For products that are deployed as virtual machines, this document does not provide details on the virtual machine OVA template specification. For that information, refer to the documentation on *Unified Communications in a Virtualized Environment*, available at http://www.cisco.com/go/uc-virtualized.

What's New in This Chapter

Table 7-1 lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 7-1 New or Changed Information Since the Previous Release of This Document

New or Revised Topic	Described in:	Revision Date
Locations and regions	Unified CM Assumptions, page 7-3	November 20, 2015
Recommended TelePresence Server platforms and capacities	TelePresence Server Platform Sizing, page 7-8	November 20, 2015
Cisco Expressway sizing	Cisco Expressway Sizing, page 7-10	November 20, 2015

Call Control

As discussed in the Call Control chapter, the Cisco Unified Communications Manager (Unified CM) and IM and Presence Service are provided through a Unified CM cluster and an IM and Presence cluster.

A Cisco Unified CM cluster consists of one publisher node, two dedicated TFTP servers, and one or multiple call processing node pairs. The number of call processing pairs depends on the size of the deployment and is discussed later in this section. The call processing nodes are deployed in pairs for 1:1 redundancy.

IM and Presence nodes are also deployed in pairs. The number of IM and Presence pairs also depends on the size of the deployment, and this will be discussed later in this section. The IM and Presence nodes are deployed in pairs for 1:1 redundancy.

Unified CM Sizing

For Unified CM, the simplified sizing guidance covers deployments with up to 10,000 users and 10,000 devices. Unified CM supports more users and more devices under different assumptions or by adding more call processing pairs, but this is outside the scope of the simplified sizing guidance provided in this chapter. Table 7-2 describes the simplified sizing deployments. The assumptions made for those deployments are documented below this table. If the number of users or endpoints in your deployment is outside of the values in Table 7-2, or if the requirements of your specific deployment fall outside of the assumptions, do not use these simplified sizing deployments, but rather perform the normal sizing procedure documented in the *Sizing* chapter of the *Cisco Collaboration SRND* available at http://www.cisco.com/go/ucsrnd and in the product documentation available at http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/tsd-products-support-series-home.html.

Table 7-2 Unified CM Simplified Sizing Deployments

Deployment Size	Unified CM Nodes to be Deployed (7.5k-User OVA Template Used for Each Unified CM Node)
Up to 5,000 users or devices	**5 nodes:** 1 Publisher, 2 TFTP, 1 call processing pair (2 call processing subscribers)
Between 5,000 and 10,000 users or devices	**7 nodes:** 1 Publisher, 2 TFTP, 2 call processing pairs (4 call processing subscribers)

Table 7-2 sizes deployments based on the maximum number of users and devices, whichever number is greater. For example, in a deployment with 5,000 users and an average of two devices per user (for example, each user has a desk phone and a Jabber client in softphone mode), the 7-node deployment is required because there are 10,000 devices in total.

The 7.5k-user virtual machine configuration (OVA template) is used in these simplified sizing deployments in order to optimize the overall resources consumed on the UCS server. This OVA template requires a full UC performance CPU platform such as the Cisco Business Edition 7000; and it is not supported on the Business Edition 6000, for example. For more information on those OVA virtual machine configuration templates and on the platform requirements, refer to the documentation at www.cisco.com/go/uc-virtualized.

A Unified CM call processing pair deployed with the 7.5k-user OVA template could support up to 7,500 users under some conditions. But in this design, we use some assumptions that put an additional load on Unified CM; for instance, we assume that each user can be configured with a Remote Destination Profile for Single Number Reach, each user can use Extension Mobility, each endpoint can be CTI controlled, some shared lines are configured, mobile and remote access is enabled, and so forth. Therefore the capacity per Unified CM call processing pair is reduced, as shown in Table 7-2. The following description provides more information on the assumptions used in this simplified sizing model.

Unified CM Assumptions

The following assumptions apply to the two simplified sizing deployments listed in Table 7-2:

- Average of up to 4 busy hour call attempts (BHCA, the number of call attempts during the busy hour) per user.
- Average of up to 2 DNs per device.
- Up to 500 shared lines per call processing subscriber pair, each line being shared with an average of up to 3 devices.
- Jabber clients registering to Unified CM (softphone mode) must be counted against the device limit.
- Up to 3,000 partitions; 6,000 calling search spaces (CSSs); and 12,000 translation patterns per cluster.
- Per Unified CM cluster, up to 1,000 route patterns; 1,000 route lists; and 2,100 route groups. Per Unified CM call processing pair, up to 100 hunt pilots, 100 hunt lists, 50 circular/sequential line groups with an average of 5 members per line group, and 50 broadcast line groups with an average of 10 members per line group.
- Up to 500 CTI ports and 100 CTI route points per Unified CM call processing pair.
- GDPR/ILS is enabled when multiple Unified CM clusters are deployed.
- Extension Mobility (EM) — All users can use EM, but no Extension Mobility Cross Cluster (EMCC) users.

- Unified CM media resources — Unified CM software conference bridges (software CFBs) and Unified CM media termination points (MTPs) should not be used in this design. Instead, use TelePresence Servers and Cisco IOS-based MTP, respectively.
- Average of up to one remote destination or mobility identity per mobility user. For example, in a deployment with 5,000 users, there can be up to 5,000 remote destinations or mobility identities.
- Up to 40,000 users synchronized with active directory (but only up to 5,000 or 10,000 active users would place or receive calls, depending on the simplified sizing deployment selected in Table 7-2).
- Up to 1,500 concurrent active calls (conferencing and non-conferencing sessions) per Unified CM call processing pair. For example, if all calls are conference calls and if the average number of participants in a conference is 10, then this design assumes up to 150 conference calls per Unified CM call processing pair.
- Up to 15 calls per second (cps) per Unified CM call processing pair
- The Contact Source for Jabber is not based on Unified CM User Data Service (UDS) in this design, but rather Basic Directory Integration (BDI) or Enhanced Directory Integration (EDI). If Unified CM UDS is configured for the Contact Source, the maximum number of users per Unified CM call processing pair is reduced to 3,750.
- Up to 2,500 concurrent mobile and remote access endpoints per Unified CM call processing pair.

Other capacity limits that are applicable to the Cisco Collaboration solution and that are documented in the *Cisco Collaboration SRND* and product documentation, also apply. For example:

- Computer Telephony Integration (CTI) — All devices can be enabled for CTI, with up to 5 lines per device and 5 J/TAPI applications monitoring the same CTI device.
- Annunciator – 48 per Unified CM call processing pair. Music on hold (MoH) – 250 concurrent MoH sessions per call processing pair. For a larger number of annunciators or concurrent MoH sessions, deploy standalone Unified CM subscribers as MoH servers.
- Gateway – Up to 2,100 per cluster.
- Locations and regions — When adding regions, select **Use System Default** for the Audio Codec Preference List and Audio and Session Bit Rate values. Changing these values for individual regions from the default has an impact on server initialization and publisher upgrade times. Hence, with a total of 2,000 regions you can modify up to 200 regions to use non-default values. With a total of 1,000 or fewer regions, you can modify up to 500 of them to use non-default values. A maximum of 2,000 locations is supported, and they do not have usage limitations like regions do.
- Extension Mobility (EM) – Up to 250 EM users per Unified CM call processing node, or 375 per cluster across two active call processing nodes.

IM and Presence Sizing

For IM and Presence, simplified sizing guidance covers deployments with up to 15,000 users. The IM and Presence Service supports more users by adding IM and Presence node pairs, but this is outside of the simplified sizing guidance provided in this chapter. Table 7-3 describes the simplified sizing deployments. Again, if the number of users in your deployment is outside of the values in Table 7-3, do not use these simplified sizing deployments, but rather perform the normal sizing procedure documented in the *Sizing* chapter of the *Cisco Collaboration SRND* and product documentation.

Table 7-3 IM and Presence Simplified Sizing Deployments

Deployment Size	IM and Presence Nodes to be Deployed
Less than 2,000 users	One IM and Presence pair using the 2k-user OVA template
Between 2,000 and 5,000 users	One IM and Presence pair using the 5k-user OVA template
Between 5,000 and 15,000 users	One IM and Presence pair using the 15k-user OVA template

These OVA virtual machine configuration templates require a full UC performance CPU platform such as the Cisco Business Edition 7000. For more information on those OVA virtual machine configuration templates and on the platform requirements, refer to the documentation available at www.cisco.com/go/uc-virtualized.

The two IM and Presence nodes are deployed as a pair in order to provide redundancy if one of the nodes fails.

SRST Sizing

The number of phones and DNs supported on a Cisco Integrated Services Router (ISR) in Survivable Remote Site Telephony (SRST) mode depends on the platform. Table 7-4 provides capacity examples for only three platforms. For information on other SRST platforms, including information on the required amount of DRAM and flash memory, refer to the SRST documentation available at

http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cusrst/requirements/guide/srs10spc.html

Table 7-4 SRST Sizing Examples

Platform	Maximum Number of Phones	Maximum Number of DNs
Cisco 2901 Integrated Service Routers	35	200
Cisco 3925 Integrated Service Routers	730	1,000
Cisco 4451-X Integrated Service Routers	1,500	2,500

Conferencing

Sizing a deployment for conferencing is primarily an exercise in deciding how many concurrent connections are required to TelePresence Servers. Considerations include:

- Geographical location — Each region served by Unified CM should have dedicated conferencing resources. For example, there could be one central location for the US where Unified CM, TelePresence Servers, and other servers are installed, and one central location for EMEA.
- Preference for TelePresence Server platforms — Virtualized or non-virtualized
- TelePresence Server platform capacities
- TelePresence Conductor platform capacities
- Type of conferencing — Audio and/or video; scheduled and/or non-scheduled
- Conference video resolution — Higher quality conferences use more resources.
- Large conference requirements — For example, all-hands meetings

Conference resources are generally dedicated to a region in order to keep as much of the conference media on the regional network; therefore, sizing can be considered on a region-by-region basis.

Conference Port Usage Guidelines

Audio and video conference sizing depends heavily on specific details about the customer, their user base, and their conferencing habits. The guidelines in this section can be used as a basis for sizing a conferencing deployment, but user-to-port ratios will vary greatly depending on the deployment environment and the requirements of the organization.

Table 7-5 shows suggested ratios to start planning conference resource requirements. These numbers vary depending on the capabilities of deployed endpoints, availability of alternative audio conferencing such as Cisco WebEx, and users' comfort level in creating and joining conferences. As a starting point, the following formulas can be used to calculate port requirements:

- Audio ports = 50 + (<number of users> / 9)
- Video ports = 8 + (<number of users> / 15)

Table 7-5 Recommended Number of Conference Ports

Number of Users	Number of Audio Ports	Number of Video Ports
1,000	161	75
1,750	244	125
3,000	383	208
5,000	605	342
10,000	1,161	675

The numbers in Table 7-5 can be used for either scheduled or non-scheduled conferencing. It is expected that, for scheduled meetings, customers can use existing usage data to draw more definite conclusions about concurrent meeting usage.

Understanding what type of meetings a customer expects to take place will help further refine the number of ports required. The total number of ports can be calculated with the formula:

Total ports = <Average number of participants in a meeting> * <Concurrent meetings>

For example, with 3,000 users, Table 7-5 suggests 208 ports. This can, for instance, correspond to an average of 3 participants per meeting and 69 concurrent meetings, or an average of 6 participants per meeting and 34 concurrent meetings. By assessing the suggested port numbers in this manner, it is easier to determine whether the total number of ports is likely to be sufficient for the deployment.

Another important point to consider is what the maximum meeting size is likely to be. In most cases the largest meeting is an all-hands meeting type. For instance, if a customer has 1,000 users but has a requirement to join 96 systems in an all-hands TelePresence conference, this would override the 75 port suggestion.

Screen Licenses and Port Capacity

Video resolution determines the quality of users' video experience and the number of video connections that a Cisco TelePresence Server can support. For optimal experiences, we recommend enabling high definition (HD) video calls at a minimum resolution of 720p and 30 frames per second (fps). Depending on the budget and capability of an organization's endpoints and network, HD video calls might not always be possible. Table 7-6 shows TelePresence Server port capacity based on video quality, assuming the video streaming rate is 30 fps. The number of audio ports per screen license is not shown and is equal to 52, with a maximum of 200 audio ports supported per TelePresence Server.

Table 7-6 TelePresence Server Port Capacity Based on Video Quality

Screen Licenses[1]	1080p Ports[2]	720p Ports[3]	480p Ports[3]	360p Ports[3]
1	1	2	3	4
5	5	10	15	20
10	10	20	30	40
20	20	40	60	80
48	48	96	144	192

[1] The number of screen licenses that can be deployed on a TelePresence Server depends on the platform.

[2] Assumes a separate content channel sharing at a maximum of 720p resolution and 15 fps.

[3] Assumes a separate content channel sharing at a maximum of 720p resolution and 5 fps.

Note With Cisco TelePresence Conductor and TelePresence Server, a single conference resource can host multiple simultaneous conferences with different resolution limits. There is no need to dedicate a TelePresence Server to a single resolution.

As can be seen from Table 7-6, the desired video quality has a direct effect on the amount of resources consumed on a TelePresence Server and, as a result, a direct impact on the number of TelePresence Servers required for the deployment.

TelePresence Server Platform Sizing

Cisco TelePresence Server is available in several different models and platforms with differing conference support and scalability. Table 7-7 lists the recommended TelePresence Server platforms for enterprise deployments, along with some of their associated port capacities. For more details, for information on other TelePresence Server platforms, or for information on other video and data channel resolutions, refer to the *Cisco TelePresence Data Sheet*, available at

http://www.cisco.com/c/en/us/products/conferencing/telepresence-server/datasheet-listing.html

Table 7-7 TelePresence Server Platforms and Capacities

TelePresence Server Platform[1]	Cluster Support	HD 1080p Port Capacity[2]	HD 720p Port Capacity[3]	SD 480p Port Capacity[3]	SD 360p Port Capacity[3]
Multiparty Media 410v	No	27	54	81	108
Multiparty Media 820v	Yes, up to two blades can be clustered.	30 per blade. Up to 60 per cluster.	60 per blade. Up to 120 per cluster.	90 per blade. Up to 180 per cluster.	120 per blade. Up to 240 per cluster.

[1] TelePresence Servers support a maximum of 200 audio connections for any standalone deployment or cluster and with any audio codec.

[2] Assumes content sharing at 720p resolution and 15 frames per second (fps).

[3] Assumes content sharing at 720p resolution and 5 frames per second (fps).

There are other considerations to keep in mind too. For example, a TelePresence Server supports a maximum of 200 calls on any standalone server or cluster, with up to 104 calls in each conference.

TelePresence Conductor Sizing

The total number of TelePresence Servers for non-scheduled conferences is limited by the capacity of TelePresence Conductor. Table 7-8 lists TelePresence Conductor capacities.

Table 7-8 TelePresence Conductor Capacities

OVA Template	Total Number of TelePresence Servers	Total Number of Concurrent Participants Across All TelePresence Servers
Small OVA template	30	50
Large OVA template or appliance	30	2,400

Clustering provides only high availability; it does not increase the maximum number of conference bridges or concurrent calls that can be supported.

If a deployment grows beyond the capacity of a single TelePresence Conductor cluster, it is possible to create additional independent TelePresence Conductor clusters and continue to add TelePresence Servers there.

An independent TelePresence Conductor cluster should be used per regional Unified CM cluster. Using the topology example in this document (see the Call Control chapter), there would be one TelePresence Conductor cluster for the US Unified CM cluster and another one for the EMEA Unified CM Cluster.

Cisco TelePresence Management Suite (TMS)

We recommend two simplified sizing deployments for Cisco TMS, illustrated in Table 7-9. There are other possible TMS deployments, but they are not covered in this guide. For instance, the single server deployment that has all TMS, TMSPE, TMSXE, and Microsoft SQL components residing in the same virtual machine is not described here because it does not provide redundancy.

The two deployments in Table 7-9 provide high availability. The redundant node is deployed for resiliency, not for scalability. A load balancer providing a single virtual IP address for the primary and backup nodes is also required.

Table 7-9 Cisco TMS Simplified Deployments and Capacities

Deployment Model	Deployment	Cisco TMS	Cisco TMSXE	Cisco TMSPE
Regular Deployment (1 vCPU OVA template)	2 nodes total: each with TMS, TMSPE, and TMSXE Additional servers for Microsoft SQL	< 200 controlled systems (endpoints added to TMS for scheduling) < 100 concurrent participants < 50 concurrent ongoing scheduled conferences	< 50 endpoints bookable in Microsoft Exchange	< 1,000 Collaboration Meeting Rooms (CMRs)
Large Deployment (4 vCPU OVA template)	4 nodes total: 2 each with both TMS and TMSPE; and 2 with TMSXE only Additional servers for Microsoft SQL	< 5,000 controlled systems (endpoints added to TMS for scheduling) < 1,800 concurrent participants < 250 concurrent ongoing scheduled conferences	< 1,800 endpoints bookable in Microsoft Exchange	< 48,000 Collaboration Meeting Rooms (CMRs)

Other factors that influence Cisco TMS performance and scaling include:

- The number of users accessing the Cisco TMS web interface.
- Concurrency of scheduled or monitored conferences.
- Simultaneous usage of the Cisco TMS Booking API (TMSBA) by multiple extensions or custom clients. Booking throughput is shared by all scheduling interfaces, including the Cisco TMS New Conference page.

For more information on sizing Cisco TMS, refer to the *Cisco TelePresence Management Suite Installation and Upgrade Guide*, available at

http://www.cisco.com/c/en/us/support/conferencing/telepresence-management-suite-tms/products-installation-guides-list.html

Collaboration Edge

This section covers sizing of Cisco Expressway and Cisco Unified Border Element, two key components of the Collaboration Edge.

Cisco Expressway Sizing

Table 7-10 shows the maximum capacity that a single Expressway node can handle at any point of time.

The Expressway nodes are clustered together to provide redundancy and larger scalability. The cluster configurations that are recommended and that are covered in this document consist of clusters of 2, 3, or 6 nodes. Table 7-11 shows the cluster capacity for those recommended deployments. It is important to note that all of the deployment models account for redundancy. With a cluster of 2 or 3 nodes, one node can fail without impacting the cluster capacity (N+1 redundancy). With a full cluster of 6 nodes, two nodes can fail without impacting the cluster capacity (N+2 redundancy).

In order to better understand the relationship between the cluster capacity and the level of redundancy, the following example analyses the video capacity during normal operations and after a failover, using the medium OVA template:

> The maximum video call capacity per node is 100 sessions. In a 3-node cluster in a non-resilient deployment, the video call cluster capacity is 300, but it would be reduced by one-third if one node fails. In order to provide resiliency and maintain the cluster capacity if one of the three nodes fails, the recommended high-available 3-node cluster capacity is limited to 200 video sessions. During normal operations, video calls are load-balanced across the cluster, with each node handling approximately 66 video calls. If one node fails, the remaining nodes can then handle all 200 video sessions because each node can handle 100 video sessions, and therefore the cluster capacity is maintained.

Table 7-10 Expressway Node Capacity

OVA Template	Mobile and Remote Access Proxy Registrations per Node[1]	Video Calls Capacity per Node	Audio-Only Calls Capacity per Node
Virtual machine with medium OVA template or Cisco Expressway CE1100 Appliance with 1 Gb small form-factor pluggable (SFP) transceivers	2,500	100	200
Virtual machine with large OVA template or Cisco Expressway CE1100 Appliance with 10 Gb small form-factor pluggable (SFP) transceivers	2,500	500	1,000

[1] Proxy registration considerations apply only to mobile and remote access, not to business-to-business communications.

Table 7-11 Cisco Expressway Simplified Sizing Deployments and Associated Cluster Capacity

Deployment Model	Expressway Cluster Deployment	Redundancy Model	Mobile and Remote Access Proxy Registrations per Cluster[1]	Video Calls Capacity per Cluster	Audio-Only Calls Capacity per Cluster
Virtual machine with medium OVA template or Cisco Expressway CE1100 Appliance with 1 Gb SFP					
Deployment 1	2 nodes	N+1	2,500	100	200
Deployment 2	3 nodes	N+1	5,000	200	400
Deployment 3	6 nodes	N+2	10,000	400	800
Virtual machine with large OVA template or Cisco Expressway CE1100 Appliance with 10 Gb SFP					
Deployment 4	2 nodes	N+1	2,500	500	1,000
Deployment 5	3 nodes	N+1	5,000	1,000	2,000
Deployment 6	6 nodes	N+2	10,000	2,000	4,000

[1] Proxy registration considerations apply only to mobile and remote access, not to business-to-business communications.

Note The large OVA template is supported only with limited hardware. Refer to the documentation at http://www.cisco.com/go/uc-virtualized for more information.

The following assumptions are used for the Expressway simplified sizing deployments in Table 7-11:

- All video calls are encrypted. The average call rate across all the video calls is 768 kbps. For example, half of the video calls could be at 384 kbps and the other half at 1152 kbps.
- All audio calls are encrypted, and the average bandwidth across all audio calls is 64 kbps.
- For virtual machines using the medium OVA template or Cisco Expressway CE1100 Appliance with 1 Gb small form-factor pluggable (SFP) transceiver, the call rate is up to 5 calls per second (cps) per node.
- For virtual machines using the large OVA template or Cisco Expressway CE1100 Appliance with 10 Gb small form-factor pluggable (SFP) transceiver, the call rate is up to 10 calls per second (cps) per node.

The following guidelines apply when clustering Cisco Expressway:

- Expressway clusters support up to 6 nodes (cluster capacity up to 4 times the node capacity).
- Expressway-E and Expressway-C nodes cluster separately; an Expressway-E cluster consists of Expressway-E nodes only, and an Expressway-C cluster consists of Expressway-C nodes only.
- Expressway peers should be deployed in equal numbers across Expressway-E and Expressway-C clusters. For example, a three-node Expressway-E cluster should be deployed with a three-node Expressway-C cluster.
- The capacity of all nodes across and within each Expressway-E and Expressway-C cluster pair must be the same. For example, an Expressway-E node using the large OVA template must not be deployed if the nodes in the Expressway-E cluster or in the corresponding Expressway-C cluster are using the medium OVA template.
- An Expressway-E and Expressway-C cluster pair can be formed by a combination of nodes running on an appliance or running as a virtual machine, as long as the node capacity is the same across all nodes.
- Multiple Expressway-E and Expressway-C clusters may be deployed to increase capacity.

For more information on Expressway, refer to the *Cisco Expressway Administrator Guide*, available at

http://www.cisco.com/c/en/us/support/unified-communications/expressway-series/products-maintenance-guides-list.html

Cisco Expressway Sizing Example

A company has 6,000 users, and on average 1,000 users are traveling at any given time. 80% of the mobile users require mobile and remote access at any given time. In this case, Expressway has to be sized to allow for 800 concurrent registrations (80% of 1,000).

Moreover, 10% of the mobile users are in a call at the same time. 5% of these users are calling through Expressway, while the remaining 5% are calling through the cellular network, so that the number of concurrent calls to the Expressway is 40 (5% of 800).

In the corporate network, 1% of the users are on a business-to-business calls at the same time. This accounts for an additional 50 calls (1% of (6,000 – 1,000)).

In this case we need to size the cluster to support 800 concurrent registrations and 90 concurrent calls (40+50).

Table 7-10 shows that a medium OVA template supports up to 100 concurrent calls and 2,500 concurrent registrations. We can therefore deploy an Expressway-C cluster consisting of two nodes using the medium OVA template, and an Expressway-E cluster also consisting of two nodes using the medium OVA template. Each Expressway server node can manage the whole amount of 800 registrations and 90 calls at the same time, as shown by Deployment 1 in Table 7-11. Clustering is needed because, if one of the two Expressway nodes goes down, the other node can handle the whole amount of traffic. Under normal conditions, calls and registrations are load-balanced between the two nodes of the Expressway-C and Expressway-E clusters.

After some time, the business-to-business calls in this example increase from 1% to 3%. We now need to account for 190 concurrent calls (40+150) instead of 90. The maximum that a medium OVA template can handle is 100 calls, so we need to deploy a larger cluster in this case. Table 7-11 shows that Deployment 2 can account for 200 concurrent calls even in case of a server failure. Therefore, the administrator in this example decides to add another medium OVA node to the Expressway-C and Expressway-E clusters, for a total of 3 nodes per cluster.

Cisco Unified Border Element Sizing

Cisco Unified Border Element is supported on a wide range of Cisco routing platforms, including platforms such as the Cisco 2900, 3900, and 4400 Series Integrated Services Routers (ISR) and the Cisco 1000 Series Aggregation Service Routers (ASR). Cisco Unified Border Element also provides redundancy on the following platforms:

- The Cisco ISR platforms, which can provide box-to-box redundancy with both signaling and media preservation for active calls.
- The Cisco ASR platforms, which can provide box-to-box or in-box redundancy with media and signaling preservation (stateful failover) for active calls.

Table 7-12 provides capacity examples for a few platforms. For information on other platforms and for more detailed, information including required amount of DRAM and flash memory, refer to the *Cisco Unified Border Element Data Sheet* and the *Cisco Unified Border Element and Gatekeeper Ordering Guide*, both available at

http://www.cisco.com/c/en/us/products/unified-communications/unified-border-element/datasheet-listing.html

Table 7-12 Cisco Unified Border Element Capacity Examples

Platform	Maximum SIP Trunk Sessions
Cisco 2901 Integrated Service Router	100
Cisco 3925 Integrated Service Router	800
Cisco 4451-X Integrated Service Router	6,000
Cisco 1004 and 1006 Aggregation Services Routers	16,000

Cisco Unified Border Element Sizing Example

A company has 8,000 users. During the busiest hour, 10% of them are in a call at the same time. 8% of these users are calling external destinations, while the remaining users are engaged in internal calls. The Telecom carrier and the enterprise have agreed that G.711 can be used on all calls, therefore no transcoding is needed. For this deployment, 640 SIP sessions (8% of 8,000) are needed. Table 7-12 shows that a Cisco 3925 ISR can support up to 800 sessions. Thus, for this example two Cisco 3925 ISRs with Cisco Unified Border Element software are selected, one active and one standby to provide redundancy.

Core Applications

This section covers sizing for Cisco Unity Connection.

Cisco Unity Connection

As discussed in the section on the Cisco Unity Connection Deployment Process, the recommended Unity Connection deployment in this design consists of one publisher and one subscriber in active/active mode.

This guide covers three simplified sizing deployments for Unity Connection, depending on the number of users. These deployments are shown in Table 7-13. There are other possible deployments with Unity Connection, but they are not covered in this guide. Refer to the *Cisco Collaboration SRND* and product documentation for information on the other possible deployments.

Table 7-13 Cisco Unity Connection Simplified Sizing Deployments

Deployment Size	Unity Connection Nodes to be Deployed for Active/Active
1,000 users	One Unity Connection pair using 1k-user OVA template
1,000 to 5,000 users	One Unity Connection pair using 5k-user OVA template
5,000 to 10,000 users	One Unity Connection pair using 10k-user OVA template

Cisco Unity Connection Assumptions

The OVA template limits should not be exceeded. For example, with the 5k-user OVA template, there is a limit of 200 ports with G.711 or 50 ports with G.722. For more information on the OVA template limits, refer to:

- Cisco Unity Connection virtualization information at http://docwiki.cisco.com/wiki/Virtualization_for_Cisco_Unity_Connection
- Cisco Unity Connection product documentation available at http://www.cisco.com/c/en/us/support/unified-communications/unity-connection-version-10-x/model.html

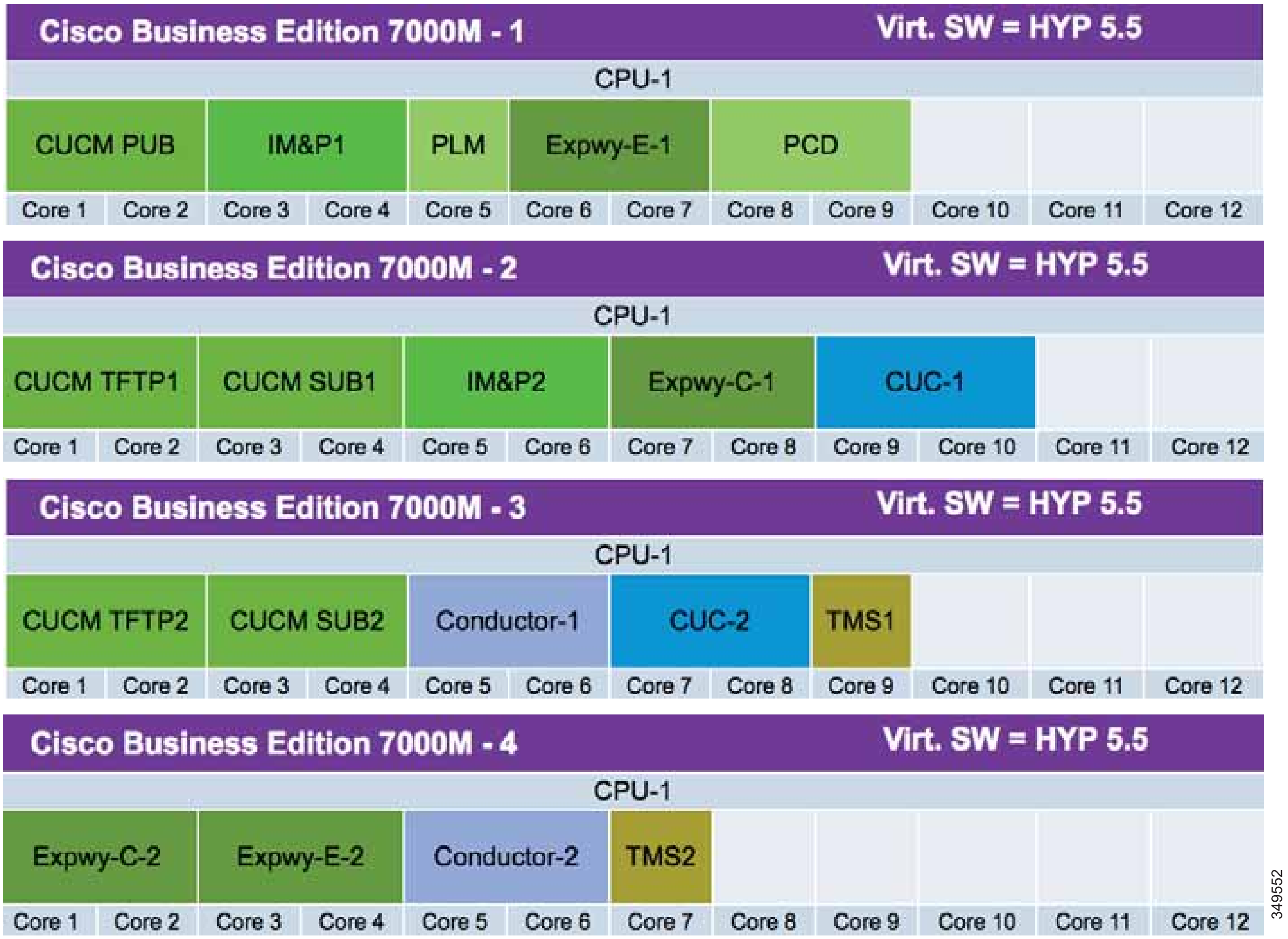
It is also important to consider the amount of storage required to store voice mail. The message storage depends on the size of the virtual disk. For example, the approximate message storage using the G.711 codec is 137k minutes with the 5k-user OVA template, which is defined with one vDisk of 200 GB. Note that with the 10k-user OVA template, different vDisk sizes are available to address different message storage requirements. For more information, refer to the *Cisco Unity Connection Supported Platforms List*.

Virtual Machine Placement and Platforms

With Cisco Collaboration products that are deployed with virtualization, after sizing the deployment, the next step is to determine how to place the virtual machines together on the Cisco Unified Computing System (UCS) servers, which will ultimately determine how many UCS servers are required for the solution. This process is performed with the Collaboration Virtual Machine Placement Tool (VMPT), which requires a cisco.com login and which is available at http://www.cisco.com/go/vmpt.

Figure 7-1 shows an example of using VMPT for a deployment with 5,000 users. This example assumes that Cisco Business Edition 7000M is deployed. It does not include the TelePresence servers, which could be deployed, for example, with the Multiparty Media 410v or Multiparty Media 820v platforms.

Figure 7-1 Virtual Machine Placement Example Using VMPT



In general, in addition to using VMPT, it is a good practice to validate the virtual machine placement by ensuring that the deployment meets all the co-residency requirements documented at

http://docwiki.cisco.com/wiki/Unified_Communications_Virtualization_Sizing_Guidelines#Application_Co-residency_Support_Policy

The main placement and co-residency rules are:

- No over-subscription — All virtual machines require a one-to-one mapping between virtual hardware and physical hardware. For example, with the CPU there must be a one-to-one mapping between virtual hardware and physical hardware, even when hyper-threading is enabled.
- VMware latency sensitivity, available with vSphere 5.5 and later versions, should be set to **High** for the Unity Connection virtual machines. If not, one spare physical core must be reserved for the ESXi scheduler on each ESXi host where Unity Connection is installed.
- Most of the applications discussed in this guide support co-residency with third-party applications, which means they can be installed on the same UCS server. However, it is important to understand that, with co-residency of third-party applications, the third-party applications must follow the same rules as Cisco collaboration applications. For example, once a third-party application is installed on the same host as a Cisco collaboration application, CPU over-subscription is not supported with that third-party application, a physical core needs to be reserved for the ESXi scheduler when deploying Unity Connection, and so forth. With Cisco Business Edition platforms, the ESXi license also dictates some of the co-residency options. For example, with the Cisco UC Virtualization Hypervisor/Foundation, there is a limit on the number of third-party applications that can be co-resident.

Redundancy Consideration

Even though the hardware platforms can be highly redundant, it is good practice to plan for hardware redundancy. For example, do not deploy the primary and backup application virtual machines on the same UCS server, as shown in the example in Figure 7-1. Instead, deploy primary and backup virtual machines on different servers to provide redundancy in case a host fails.

Platforms

For the products that are deployed with virtualization, Cisco Business Edition 7000 can be an excellent solution. It is easy to order and easy to deploy. It includes the Cisco UCS server hardware and a hypervisor license. VMware vSphere Hypervisor (ESXi) is pre-installed. Business Edition 7000 is also pre-loaded with the Cisco Collaboration software set and some of the Cisco Collaboration applications are also pre-installed.

APPENDIX A

Product List

Revised: November 20, 2015

This product list identifies the Cisco products in the Preferred Architecture for Enterprise Collaboration, along with their recommended software versions.

Table A-1 Products and Software Versions for Enterprise Collaboration 11.x Preferred Architecture

Product	Product Description	Recommended Software Version
Cisco Unified Communications Manager and IM and Presence Service	Call control, instant messaging, and presence services	11.0(1)
Cisco Unity Connection	Voicemail services	11.0(1)
Cisco Expressway-C and Expressway-E	Mobile and remote access and business-to-business communications	X8.6
Cisco Prime License Manager	Single management point for licensing	11.0(1)
Cisco Prime Collaboration Deployment	Installs Unified CM cluster with IM and Presence Service and Unity Connection cluster	11.0(1)
Cisco TelePresence Conductor	Video conferencing resource management	4.0
Cisco TelePresence Server	Audio and video conferencing resources	4.2
Cisco ISR and ASR	PSTN gateway, SRST, and external connectivity to the Internet	IOS 15.5(2)T for ISR IOS XE 3.15 for ASR
Cisco IP Phone 7811	General office use, single-line phone	10.3(1)
Cisco IP Phone 8800 Series	General office use	10.3(1)
Cisco Unified IP Conference Phone 8831	IP conference phone	10.3(1)
Cisco Jabber	Soft client with integrated voice, video, voicemail, and instant messaging and presence functionality for mobile devices and personal computers	Jabber 11.0
Cisco DX Series	Personal TelePresence endpoint for the desktop	10.2(4)
Cisco TelePresence MX Series	TelePresence multipurpose room endpoint	CE 8.0[1]
Cisco TelePresence SX Series	Integrator Series TelePresence endpoint	CE 8.0[1]

Table A-1 Products and Software Versions for Enterprise Collaboration 11.x Preferred Architecture (continued)

Product	Product Description	Recommended Software Version
Cisco TelePresence IX Series	Immersive TelePresence room system	IX 8.1
Cisco TelePresence Management Suite (TMS)	Scheduling, web conferencing integration, and other advanced video features	15.0

[1] MX Series and SX Series endpoints require the forthcoming release of CE 8.0 firmware. The release of this firmware load is expected by the end of 2015. This firmware load is required for the functionality described in the Bandwidth Management chapter, and specifically for the QoS markings set by the endpoint. If firmware load CE 8.0 is not used, then the QoS settings for the MX Series and SX Series endpoints will not function as described in the Bandwidth Management chapter.