CISCO

Implementing Cisco IP Telephony and Video, Part 1 (CIPTV1)

Foundation Learning Guide

(CCNP Collaboration Exam 300-070 CIPTV1)

Third Edition

Akhil Behl
Berni Gardiner
Josh Finke

ciscopress.com

FREE SAMPLE CHAPTER

SHARE WITH OTHERS

Implementing Cisco IP Telephony and Video, Part 1 (CIPTV1) Foundation Learning Guide

CCNP Collaboration Exam 300-070 CIPTV1, Third Edition

Akhil Behl, CCIE No. 19564

Berni Gardiner, CSI, CCNP Voice

Josh Finke, CCIE No. 25707

Cisco Press

800 East 96th Street

Indianapolis, Indiana 46240 USA

Implementing Cisco IP Telephony and Video, Part 1 (CIPTV1) Foundation Learning Guide

CCNP Collaboration Exam 300-070 CIPTV1, Third Edition

Akhil Behl, Berni Gardiner and Josh Finke

Published by:
Cisco Press
800 East 96th Street
Indianapolis, IN 46240 USA

Printed in the United States of America

First Printing September 2016

Library of Congress Cataloging-in-Publication Number: 2016946274

ISBN-13: 978-1-58714-451-6

ISBN-10: 1-587-14451-4

Warning and Disclaimer

This book is designed to provide information about Cisco Unified IP Telephony and Video administration and to provide test preparation for the CCNP Collaboration Exam 300-070 CIPTV1. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an "as is" basis. The authors, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc. cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Special Sales

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact intcls@pearson.com.

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through email at feedback@ciscopress.com. Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

Editor-in-Chief: Mark Taub

Product Line Manager: Brett Bartow

Alliances Manager, Cisco Press: Ron Fligge

Acquistions Editor: Michelle Newcomb

Managing Editor: Sandra Schroeder

Development Editor: Marianne Bartow, Eleanor Bru

Senior Project Editor: Tracey Croom

Copy Editor: Warren Hapke

Technical Editors: Akhil Behl, Berni Gardiner

Editorial Assistant: Vanessa Evans

Cover Designer: Chuti Prasertsith

Composition: codeMantra

Indexer: Erika Millen

Proofreader: Srimathy

About the Authors

Akhil Behl is a pre-sales manager with a leading service provider. His charter involves an overarching technology portfolio encompassing IoT, collaboration, security, infrastructure, service management, cloud, and data center. He has thirteen-plus years of experience working in leadership, advisory, business development, and consulting positions with various organizations; leading global accounts; while driving business innovation and excellence. Previously, he was in a leadership role with Cisco Systems.

Akhil has a Bachelor of Technology degree in electronics and telecommunications from MAIT College, IP University, Delhi, India, and a master's degree in business administration from Symbiosis Institute, Pune, India. Akhil holds dual CCIE in Collaboration and Security, PMP, ITIL, VCP, TOGAF, CEH, ISO/IEC 27002, and many other industry certifications.

He has published several research papers in national and international journals, including IEEE, and has been a speaker at prominent industry forums such as Interop, Enterprise Connect, Cloud Connect, Cloud Summit, Cisco Sec-Con, IT Expo, Computer Society of India, Singapore Computer Society, CommunicAsia, Total Security Conference, and Cisco Networkers.

Akhil is the author of the following Cisco Press books:

- *CCIE Collaboration Quick Reference*
- *Securing Cisco IP Telephony Networks*
- *Implementing Cisco IP Telephony and Video (Part 2)*

He is a technical editor for Cisco Press and other publications. Akhil can be reached at akbehl@technologist.com

Berni Gardiner is an independent telecommunications consultant and a long-time certified Cisco Instructor. Berni began her career in the software development arena in the 1980s and moved into the service provider arena in 1990, collaborating on building the first commercial ISP in her home province of Prince Edward Island, Canada. Building on the success of the provincial network, Berni was key in developing one of the first Canadian national ISP offerings.

Berni became a Certified Cisco Systems Instructor in 1998 and continues to combine contract instruction and course development with a career in telecommunications consulting. Her primary focus is in the collaboration product line and Quality-of-Service implementations. Berni holds a number of certifications including CCSI and CCNP Voice.

She has authored a number of white papers and blogs for Global Knowledge. She can be reached at bernigardiner@hotmail.com.

Josh Finke, CCIE No. 25707, is the engineering and services manager for Iron Bow Technologies, a Cisco Gold and Master Unified Communications Partner. Josh was previously a lead instructor and director of operations for Internetwork Expert, a leading CCIE training company. Josh has multiple certifications, including the Cisco Voice CCIE, CCNP, CCDP, CCNA, CCDA, and Cisco Meeting Place Specialist. Josh specializes in Cisco UC, routing & switching, and network design. Josh started working with Cisco networking technologies in 2000 and later became one of the youngest Voice CCIEs in the world. He lives with his wife in Seattle, Washington.

Dedications

I would like to dedicated this book first to my family, my wonderful and beautiful wife Kanika and my lovely sons Shivansh and Shaurya, for their love, patience, sacrifice, and support while writing this book. They have been very kind and supporting as always during my journey to write yet another book. Moreover, my loving wife Kanika has been pivotal while writing the book. She reviewed my work and suggested amendments and improvements.

To my parents, Vijay Behl and Ravi Behl, for their continuous love, encouragement, guidance, and wisdom. To my brothers, Nikhil Behl and Ankit Behl, who have always been there to support me in all my endeavors. To all my extended family and friends, thank you for the support and love during my journey.

And I would like to thank God for all his blessings in my life.

—Akhil

I would like to dedicate this book to Ralph for his patience and support during the late hours and weekend writing marathons. To my children and grandchildren, thank you for understanding the occasional hours and days when mom and grandmom became unavailable to join in with family activities. All of your support and encouragement carried me through this project.

—Berni

Acknowledgments

Akhil Behl:

I would like to thank the following amazing people and teams for helping me write this book.

A special Thank You to the Cisco Press editorial team: Brett Bartow—Executive Editor, for seeing the value and vision in the proposed title and providing me the opportunity to write this title; Michelle Newcomb—Acquisitions Editor; Marianne Bartow—Development Editor; Ellie Bru—Development Editor, and Vanessa Evans—Editorial Assistant, for their support and guidance throughout the writing of this book. It is my sincere hope to work again with them in the near future. And my gratitude and thanks to everyone else in the Cisco Press production team, for their support and commitment.

I would like to thank my mentors and my peers who have guided me and stood by me all these years. Thank you to all my managers and peers from Cisco who have been supportive of what I wanted to do and helped me achieve it.

And lastly but most importantly, to all those special people—my relatives and my friends; who stood by me during the highs and lows of life.

Berni Gardiner:

I would like to acknowledge and thank the Cisco Press editorial team: Brett Bartow for providing me the opportunity to join this project, Michelle Newcomb, Ellie Bru and Marianne Bartow for patiently keeping me on track and Vanessa Evans for taking care of the business end of things. Thank you to my co-authors for their comments and directions. Thank you to the unseen team members who work behind the scenes to put together the finished product. All of your help has been tremendously appreciated.

Contents at a Glance

Contents

Command Syntax Conventions

The conventions used to present command syntax in this book are the same conventions used in Cisco's Command Reference. The Command Reference describes these conventions as follows:

- **Boldface** indicates commands and keywords that are entered literally as shown. In actual configuration examples and output (not general command syntax), boldface indicates commands that are manually input by the user (such as a show command).
- *Italics* indicate arguments for which you supply actual values.
- Vertical bars (|) separate alternative, mutually exclusive elements.
- Square brackets [ ] indicate optional elements.
- Braces { } indicate a required choice.
- Braces within brackets [{ }] indicate a required choice within an optional element.

Note This book covers multiple operating systems, and a differentiation of icons and router names indicate the appropriate OS that is being referenced.

Reader Services

Register your copy at www.ciscopress.com/title/9781587144516 for convenient access to downloads, updates, and corrections as they become available. To start the registration process, go to www.ciscopress.com/register and log in or create an account*. Enter the product ISBN 9781587144516 and click Submit. When the process is complete, you will find any available bonus content under Registered Products.

*Be sure to check the box that you would like to hear from us to receive exclusive discounts on future editions of this product.

Introduction

Professional career certifications have been a critical part of the computing IT industry for many years and will continue to become more important. Many reasons exist for these certifications, but the most popularly cited reason is that of credibility and the knowledge to get the job done.

All other considerations held equal, a certified employee/consultant/job candidate is considered more valuable than one who is not. CIPTV1 sets stage with the above objective in mind and helps you learn and comprehend the topics for the CCNP Collaboration CIPTV1 exam. At the same time, it prepares you for real world configuration of Cisco's Audio and Video technology.

Goals and Methods

The most important goal of this book is to provide you with knowledge and skills in Cisco Collaboration solution, with focus on deploying the Cisco Unified Communications Manager (CUCM).

CUCM features, CUCM-based call routing, Cisco IOS Voice Gateways, Cisco Unified Border Element (CUBE), and Quality of Service (QoS). All of these are associated and relevant to building and maintaining a robust and scalable Cisco Collaboration solution. Subsequently, another obvious goal of this book is to help you with the Cisco IP Telephony and Video (CIPTV) Part 1 Exam, which is part of the Cisco Certified Network Professional Voice (CCNP) Collaboration certification. The methods used in this book are designed to be helpful in both your job and the CCNP Collaboration exam. This book provides questions at the end of each chapter to reinforce the chapter's concepts and content.

The organization of this book helps you discover the exam topics that you need to review in more depth, fully understand and remember those details, and test the knowledge you have retained on those topics. This book does not try to help you pass by memorization, but truly learn and understand the topics by going in-depths of the very concepts and architecture of Cisco Collaboration. The Cisco IP Telephony Part 1 Exam is one of the foundation topics in the CCNP Collaboration Certification. The knowledge contained in this book is vitally important for you to consider yourself a truly skilled Cisco Collaboration engineer or professional. The book helps you pass the Implementing Cisco IP Telephony and Video Part 1 exam by using the following methods:

Helps you discover which test topics you have not mastered

Provides explanations and information to fill in your knowledge gaps

Connects to real-world case studies and scenarios which are useful beyond the exam in the real life implementation tasks

Who Should Read This Book?

This book is written to be both a general CUCM book as a foundation for Cisco Collaboration and a certification preparation book. It provides you with the knowledge required to pass the CCNP Voice Cisco IP Telephony and Video Exam for in CCNP Collaboration Exams Series CIPT Part 1.

Why should you want to pass the CCNP Voice Cisco IP Telephony exam? The first CIPT test is one of the milestones toward getting the CCNP Voice certification. The CCNP Collaboration could mean a raise, promotion, new job, challenge, success, or recognition. But ultimately you determine what it means to you. Certifications demonstrate that you are serious about continuing the learning process and professional development. Today's technology is evolving at a rapid rate. It is impossible to stay at the same level while the technology around you is constantly advancing. Engineers must continually retrain themselves, or will find themselves with out-of-date commodity-based skill sets. In a fast growing technology like Collaboration; where new solutions are presented and created every day, it is most vital to keep to the pace of change.

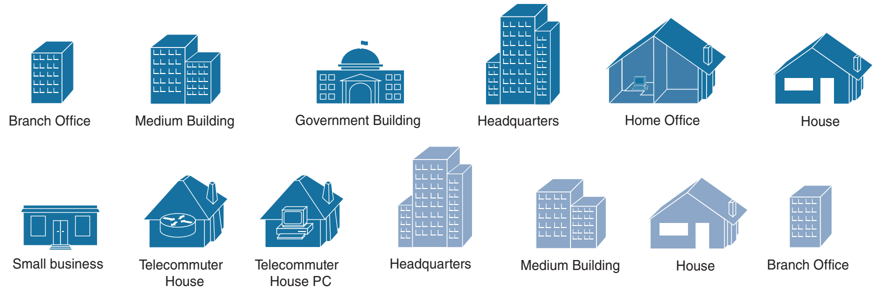
How This Book Is Organized

- **Chapter 1, "Understanding Cisco Unified Communications Manager Architecture,"** sets the stage for this book by introducing the very central focus of the Cisco Collaboration solution—CUCM. This chapter covers the nuts and bolts of CUCM architecture and gives an overview of CUCM deployment models.
- **Chapter 2, "Cisco Unified Communications Manager Deployment Models,"** gives an insight to the CUCM deployment models; which help you understand where and why you should position a certain deployment model in a Cisco Collaboration solution as well as the merits and limitations of each model. This helps you comprehend the content not just for the exam but also for real life customer consulting and architecture definition of a Cisco Collaboration solution.
- **Chapter 3, "Cisco Unified Communications Manager Services and Initial Configuration Settings,"** gives an overview of the various initial settings that must be done to bring a CUCM server/cluster online and make it useable for a Cisco Collaboration solution. Some settings are very critical from a design and deployment perspective while others from a functional perspective and all of these are covered in detail.
- **Chapter 4, "Deploying Endpoints and Users in Cisco Unified Communications Manager,"** gives an insight to deploying users and multitude of endpoints in the gambit of Cisco Collaboration solution to support small to medium to large enterprise deployments.
- **Chapter 5, "Deploying IP Phone Services in Cisco Unified Communications Manager,"** helps lay a solid foundation of IP Phone services; which in any successful deployment is necessary for offering state-of-art-services to the end users.
- **Chapter 6, "An Overview of Dial Plan Design and Implementation in Cisco Unified Communications Manager,"** describes the various dial plan elements and gives an overview of the dial plan pertinent to CUCM. This chapter discusses a dial plan from an internal dial plan to a globalized + E.164-based dial plan and lays the foundation for call routing.
- **Chapter 7, "Implementing Cisco Unified Communications Manager Call Routing and Digit Manipulation,"** gives an insight to call routing elements such as route patterns, route groups as well as cover the basis of digit manipulation both from an internal and external call perspective. Call routing and digit manipulation are some of the most basic yet complex constructs in a dial plan which are covered at length in this chapter.

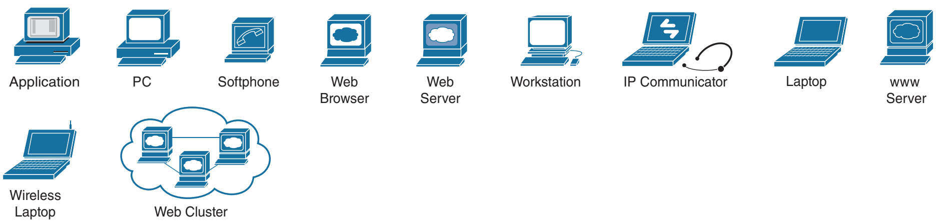
- **Chapter 8 "Implementing Calling Privileges in Cisco Unified Communications Manager,"** gives an insight to deployment locks and keys (partitions and Calling Search Spaces) which form the basis of allowing and disallowing internal or external calling access for the users.
- **Chapter 9, "Implementing Call Coverage in Cisco Unified Communications Manager,"** explains the concepts and implementation of various call coverage mechanisms at play in CUCM based audio and video solutions.
- **Chapter 10, "Implementing Media Resources in Cisco Unified Communications Manager,"** discusses the concept and implementation of various media resources ranging from audio media call resources to video call media resources. These media resources enable what would otherwise be a very daunting task of mixing audio/video streams or playing around with a range of codecs, and so on.
- **Chapter 11, "Cisco Video Conferencing,"** describes the deployment various video conferencing options and tools (platforms) available in Cisco Collaboration solution. The chapter lays the foundation for Cisco TelePresence Conductor, Cisco TelePresence Server, and discusses other platforms that enable rich media conferencing experience.
- **Chapter 12, "Quality of Service in Cisco Collaboration Solution,"** expands on the basics of Quality of Service (QoS) and defines the QoS tools, mechanisms, and ways in which audio or video calls can be handled in much better way as opposed to non-preferential treatment.
- **Chapter 13, "Implementing Cisco IOS Voice Gateways and Cisco Unified Border Element,"** discusses the very basis of how a Cisco Collaboration solution connects with the outside world such as PSTN and IT Service Provider. This chapter details the various voice and video protocols at play in a Cisco Collaboration solution and the role of Cisco Voice Gateways and Cisco Unified Border Element (CUBE). Moreover, the chapter discusses the features by which intuitive user and administrative experience are offered by these platforms.
- **Appendix A, "Answers to the Review Questions,"** allows you to check the validity of your answers at the end of each chapter as you review the questions.

Icons Used in This Book

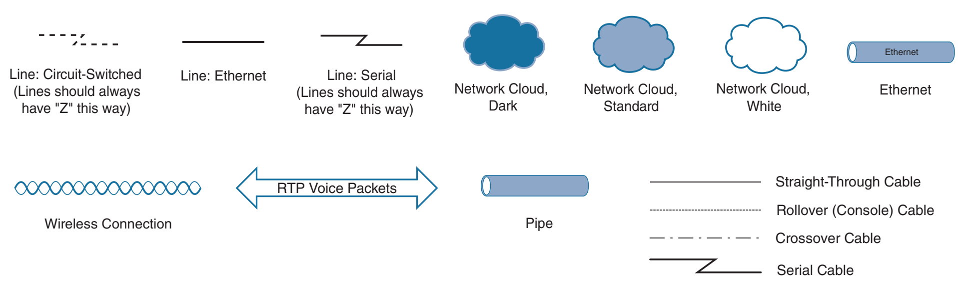
Buildings



Computers and Hardware

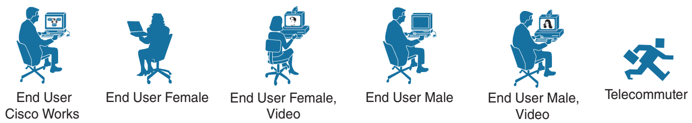


Connections

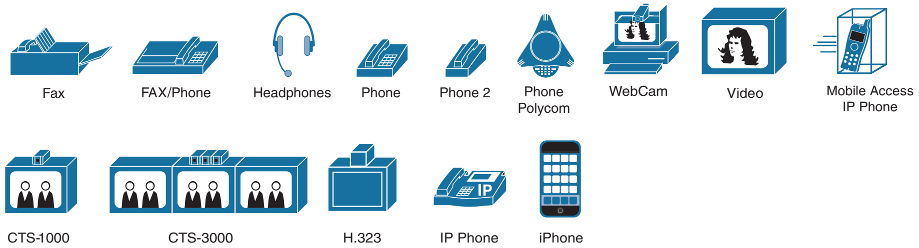


Firewalls

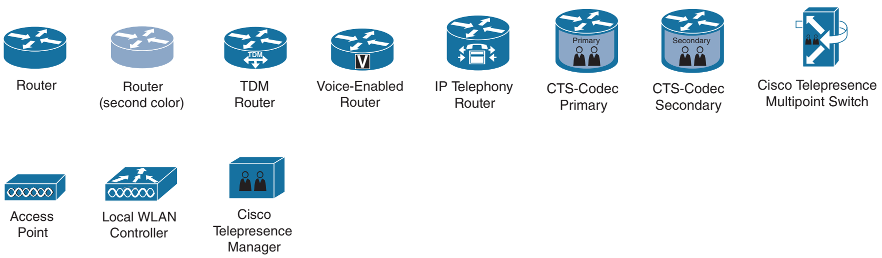
People



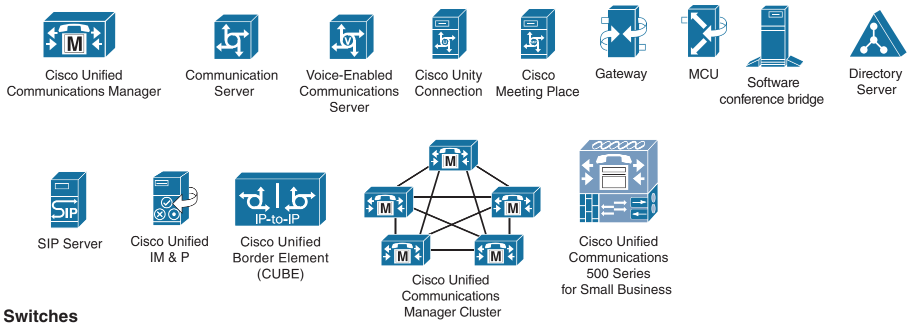
Phones, Multimedia, and Communications



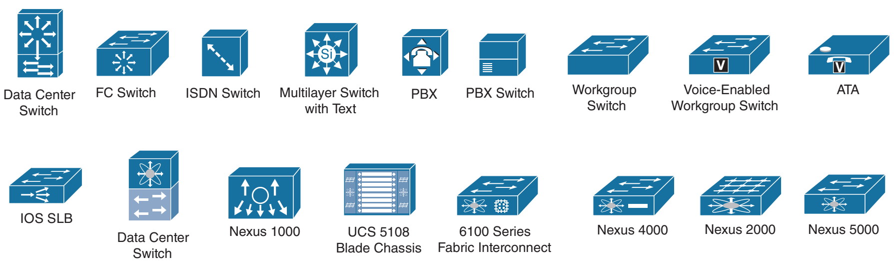
Routers



Servers



Switches



Chapter 2

Cisco Unified Communications Manager Deployment Models

This chapter introduces the Cisco Unified Communications Manager (CUCM) deployment models and architectures that ensure redundancy and provide high availability for call processing and other services. The different redundancy models explored in this chapter can be applied to the different deployment models to provide fault tolerance for CUCM and its services.

Chapter Objectives

Upon completing this chapter, you will understand the CUCM deployment and redundancy options and be able to meet the following objectives:

- Identify the supported CUCM deployment options.
- Describe the characteristics of a CUCM single-site deployment, and identify the reasons for choosing this deployment option.
- Describe the characteristics of a CUCM multisite deployment with centralized call processing, and identify the reasons for choosing this deployment option.
- Describe the characteristics of a CUCM multisite deployment with distributed call processing, and identify the reasons for choosing this deployment option.
- Describe the characteristics of a CUCM multisite deployment with clustering over the WAN, and identify the reasons for choosing this deployment option.
- Describe the Cisco Collaboration Edge solution for teleworkers and remote workers
- Explain how call-processing redundancy is provided in a CUCM cluster, and identify the requirements for different redundancy scenarios.

Cisco Collaboration Network Overview

In a typical Cisco collaboration network, there can be multiple possibilities from campus to remote sites. Figure 2-1 gives an overview of a typical large enterprise Cisco collaboration campus network where the Cisco collaboration services are available in the campus (headquarters) network.

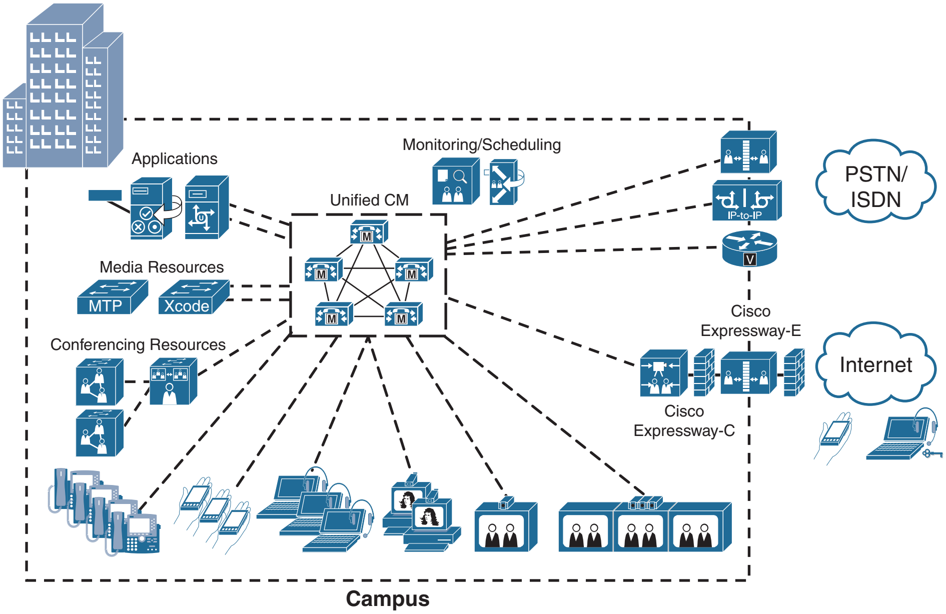


Figure 2-1 *Cisco Collaboration Solution Campus Deployment in a Large Enterprise*

Figure 2-2 shows the campus and a branch (or remote) site; with a subset of campus collaboration services available at the branch/remote site.

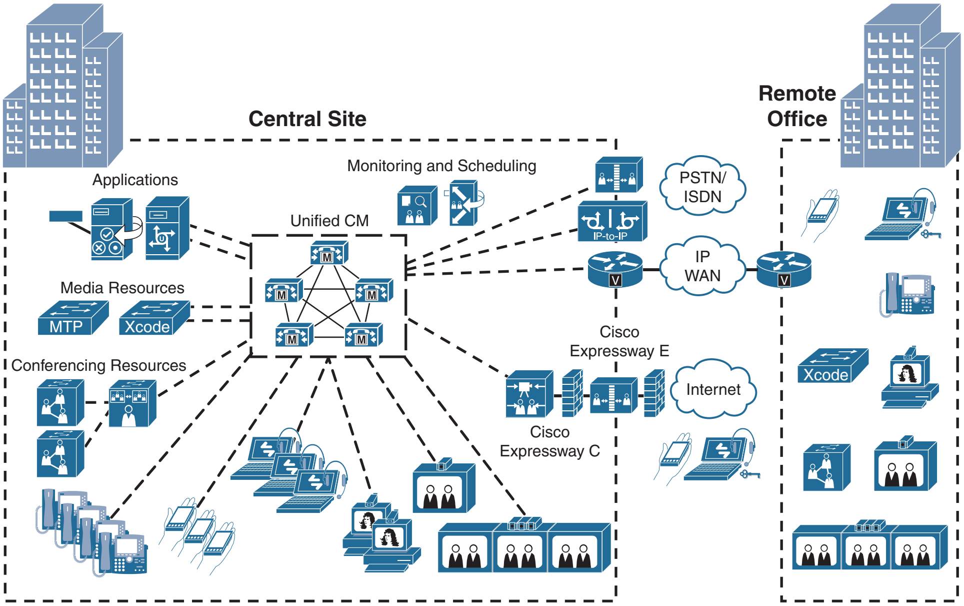


Figure 2-2 *Cisco Collaboration Solution Deployment at Campus and Branch in a Large Enterprise*

As discussed previously, the collaboration network and the associated collaboration services vary from one organization to another. Some of the factors considered are:

- Number of branch or remote sites
- Call control configuration (centralized/distributed)
- Services available for branch or remote sites
- Teleworking options

The following sections cover CUCM deployment models to support various organization/network/service requirements.

CUCM: Single-Site/Campus Deployment

As illustrated in Figure 2-3, the single-site model for CUCM consists of a CUCM cluster located at a single site or campus with no telephony services provided over a WAN.

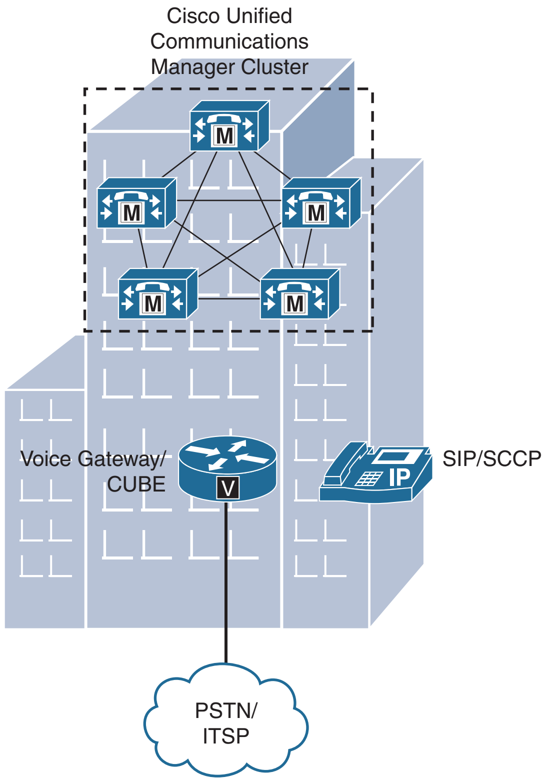


Figure 2-3 *Single-Site Deployment*

All CUCM servers, applications, and digital signal processor (DSP) resources are located in the same physical location or at multiple physical buildings with local-area networks (LAN) or metropolitan-area network (MAN)–based connectivity. LANs are normally defined as having connectivity speeds of 1000 Mbps (1 Gbps) and above, while MANs are typically in the multi-megabit range. In this model, calls beyond the LAN or MAN use the public switched telephone network (PSTN). Besides the voice gateway, Cisco Unified Border Element (CUBE) can also be used to connect all PSTN traffic via IT Service Provider (ITSP) cloud.

Note ITSP-based PSTN connectivity leverages Session Initiation Protocol (SIP), which is the most popular and prevalent endpoint and media gateway protocol. SIP is described in detail later in this book.

Each cluster supports a maximum of 40,000 IP phones. If there is a need to deploy more than 40,000 IP phones in a single-site configuration, multiple clusters can be implemented inside a LAN or within a MAN and connected through intercluster trunks. Gateway trunks that connect directly to the PSTN manage external calls. If an IP WAN exists between sites, it is used to carry data traffic only; no telephony services are provided over the WAN.

> **Note** Cisco Business Unit (BU)-supported configurations are available for mega-cluster implementations that can support up to 80,000 devices with 21 servers in a single cluster. Such configurations are subject to review by Cisco Account Team and Cisco BU.

Design Guidelines for Single Site/Campus Model

To accommodate future scalability, Cisco recommends that best practices specific to the distributed and centralized call-processing models be used in a single-site deployment.

Current calling patterns within the enterprise must be understood. How and where are users making calls? If calling patterns indicate that most calls are intrasite, using the single-site model will simplify dial plans and avoid having to provision additional dedicated bandwidth for voice across the IP WAN.

Because Voice over Internet Protocol (VoIP) calls are within the LAN or campus network, it is assumed that bandwidth is not a concern. Using G.722 or G.711 codecs for all endpoints will eliminate the need for DSP resources for transcoding, and those resources can be allocated to other functions, such as conferencing and Media Transfer Protocols (MTPs).

All off-net calls will be diverted to the PSTN (via voice gateway or CUBE) or sent to the legacy private branch exchange (PBX) for call routing if the PSTN resources are being shared during migratory deployments.

To ensure successful operations, a network infrastructure designed for high-availability, fault-tolerant connectivity options should be utilized. In addition, reliable Power over Ethernet (PoE), quality of service (QoS) mechanisms, and monitoring services are recommended. When designing a single campus deployment, do not oversubscribe CUCM to scale larger installations. A single-site deployment does not always equate to a single cluster. If the site has more than 40,000 IP phones, install multiple clusters and configure ICTs between the clusters (or provision mega-cluster).

Benefits of Centralized Call Processing Model

A single infrastructure for a converged network solution provides significant cost benefits and enables CUCM to take advantage of the many IP-based applications in the enterprise.

Single-site deployment allows each site to be completely self-contained. Calls between sites will be routed over the PSTN. Extra provisioning of WAN bandwidth is not needed. Dial plans are also easier to provision. There are no service issues in the event of an IP WAN failure or insufficient bandwidth, and there is no loss of call-processing service or functionality.

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

- Ease of deployment
- A common infrastructure for a converged solution
- Simplified dial plan
- No transcoding resources are required, due to the use of a single codec

Multisite Deployment with Centralized Call Processing

The multisite deployment with centralized call-processing model consists of a centralized CUCM cluster that provides services for many sites and uses the IP WAN to transport IP telephony traffic between the sites.

The IP WAN also carries call-control signaling between the CUCM cluster at the central site and the IP phones at the remote sites.

Figure 2-4 illustrates a typical centralized call-processing deployment, with a CUCM cluster at the central site or data center and a QoS-enabled IP WAN to connect all the sites. The remote sites rely on the centralized CUCM cluster to manage their call processing. Applications such as voice mail and interactive voice response systems are typically centralized as well to reduce the overall costs of administration and maintenance.

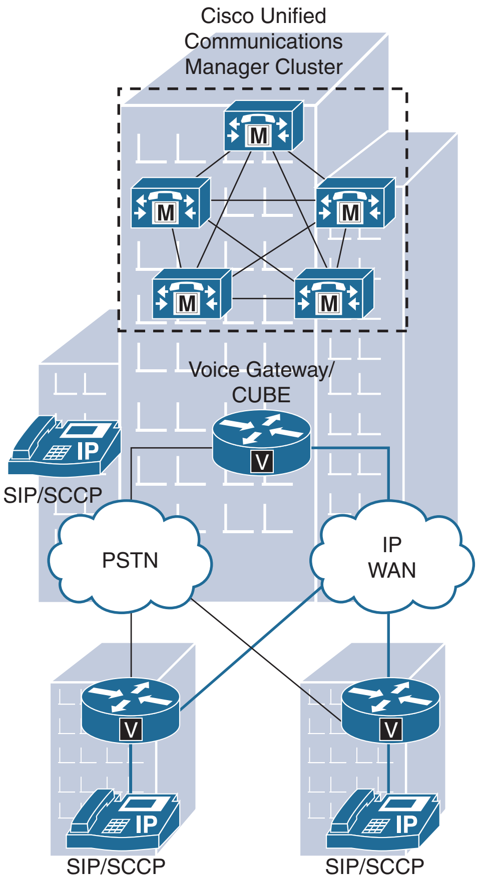


Figure 2-4 *Centralized Multisite Deployment*

The Cisco Unified Survivable Remote Site Telephony (SRST) and E-SRST features that are available in Cisco IOS gateways provide call-processing services to remote IP phones during a WAN outage. When the IP WAN is down, the IP phones at the remote branch office can register to the local Cisco Unified SRST router. The Cisco Unified SRST router can process calls between registered IP phones and send calls to other sites through the PSTN. Figure 2-5 gives an overview of remote site SRST/E-SRST deployment with centralized call processing. The same arrangement however, will work if there are different CUCM clusters (distributed call processing or clustering over WAN) with one or more remote sites.

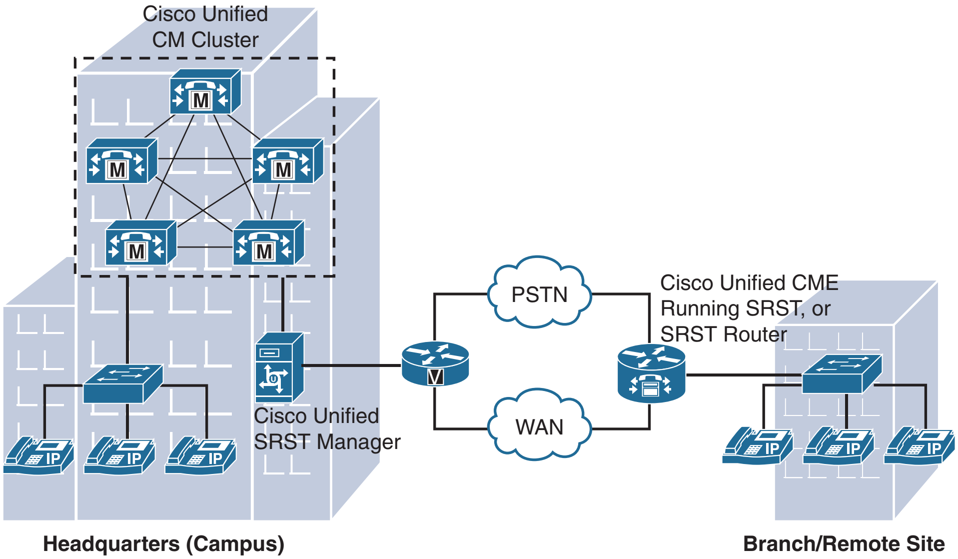


Figure 2-5 *Cisco Unified SRST/E-SRST Deployment with Centralized Call Processing*

> **Note** Topics of SRST, E-SRST, CAC, and AAR are discussed in detail in *Implementing Cisco IP Telephony and Video, Part 2* (CIPTv2).

To avoid oversubscribing the WAN links with voice traffic, causing deterioration of the quality of established calls, Call Admission Control (CAC) is used to limit the number of calls between the sites.

Centralized call-processing models can take advantage of automated alternate routing (AAR) features. AAR allows CUCM to dynamically reroute a call over the PSTN if the call is denied because of CAC.

Design Guidelines for Multisite WAN Model with Centralized Call Processing

Consider the following best practice guidelines when implementing a multisite WAN model with centralized call processing:

- Use a maximum of 2000 locations per CUCM cluster.
- Use a maximum of 2100 H.323 devices (gateways, multipoint control units, trunks, and clients) or 1100 MGCP gateways per CUCM cluster.
- Minimize delay between CUCM and remote locations to reduce voice cut-through delays.
- Use enhanced locations CAC mechanism in CUCM to provide CAC into and out of remote branches. Locations can support a maximum of 40,000 IP phones per cluster when CUCM runs on the largest supported cluster. Another option is to use Resource Reservation Protocol (RSVP)-based CAC between locations.
- Choose appropriate platform for SRST support. There is no limit to the number of IP phones at each individual remote branch. However, the capability that the Cisco Unified SRST feature provides in the branch router limits remote branches to a maximum of 1500 Cisco IP phones on a Cisco 3945E Integrated Services Router during a WAN outage or failover to SRST. Other platforms have different (lower) limits.
- Use high-bandwidth audio (for example, G.711 or G.722) between devices in the same site (intrasite), but low-bandwidth audio (for example, G.729) between devices in different sites (intersite).
- Use high-bandwidth video (for example, 1.5 Mbps with 4CIF or 720p, to 2 Mbps with 1080p) between devices in the same site, but low-bandwidth video (for example, 384 kbps with 448p or CIF) between devices at different sites.
- Use a minimum of 1.5 Mbps or greater WAN link speed. Video is not recommended on WAN connections that operate at speeds lower than 1.5 Mbps.

If a distributed call-processing model is more suitable for the business needs of a customer, the choices include installing a CUCM cluster at the remote branch or running CUCM Express on the branch router.

Benefits of Multisite Deployment with Centralized Call Processing Model

A multisite deployment with centralized call processing saves PSTN costs for intersite calls by using the IP WAN instead of the PSTN. The IP WAN can also be used to bypass toll charges by routing calls through remote site gateways that are closer to the PSTN number that is dialed. This practice is known as Tail End Hop Off (TEHO). TEHO is not permitted in some countries, and local regulations should be verified before implementing TEHO.

This deployment model maximizes the utilization of available bandwidth by allowing voice traffic to share the IP WAN with other types of traffic. Deploying QoS and CAC ensures voice quality. AAR reroutes calls over the PSTN if CAC denies the calls because of oversubscription.

Cisco Extension Mobility can be used within the CUCM cluster, allowing roaming users to use their directory numbers at remote phones as if they were at their home phones.

When the multisite WAN with centralized call-processing deployment model is used, CUCM administration is centralized, and therefore simpler, compared with a multisite WAN with distributed call-processing model where multiple clusters must be separately administered.

Multisite Deployment with Distributed Call Processing

The model for a multisite WAN deployment with distributed call processing consists of multiple independent sites, each with its own CUCM cluster.

An IP WAN carries voice traffic between the distributed clusters. CUCM Session Management Edition (SME) cluster or SIP proxy servers can be used to provide intercluster call routing and dial plan aggregation in multisite distributed call-processing deployments. Cisco CUCM Session Management Edition (SME) is the recommended trunk and dial plan aggregation platform in multisite distributed call processing deployments. SME is essentially a CUCM cluster with trunk interfaces only and no IP endpoints. It enables aggregation of multiple unified communications systems, referred to as leaf systems.

Cisco CUCM SME may also be used to connect to the PSTN and third-party unified communications systems such as PBXs and centralized unified communications applications. Figure 2-6 illustrates a distributed multisite deployment.

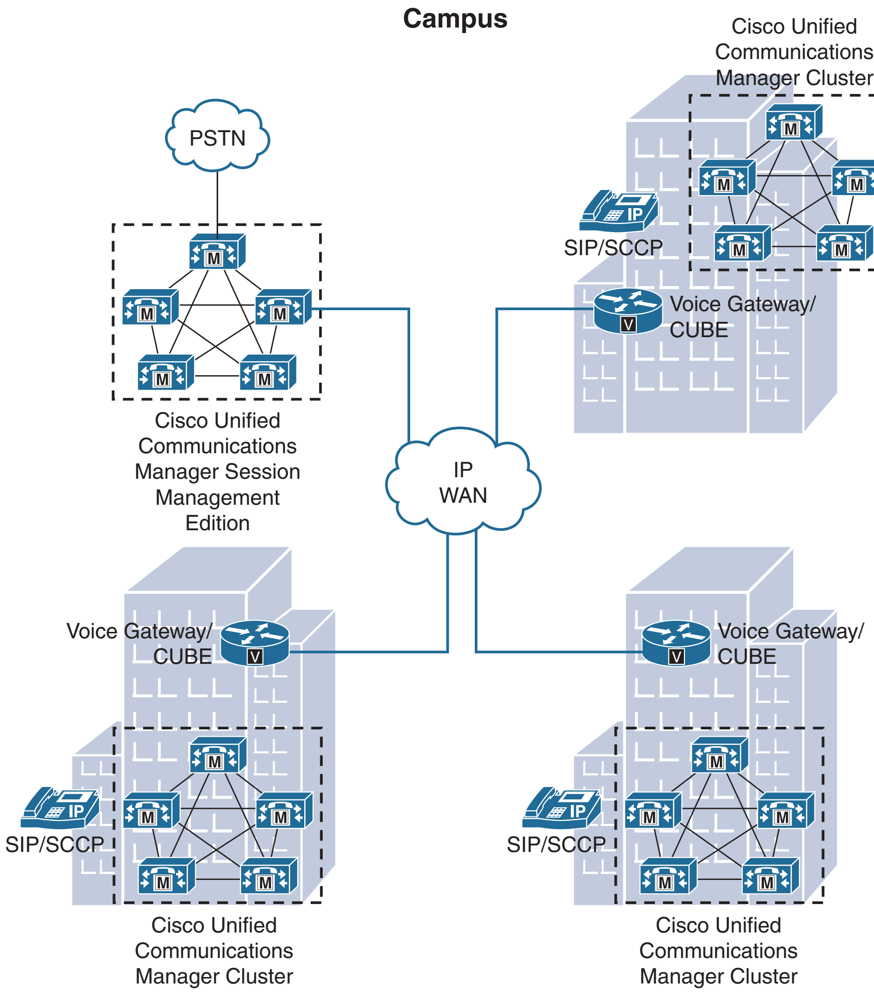


Figure 2-6 *Distributed Multisite Deployment*

Design Guidelines for Multisite Deployment with Distributed Call Processing Model

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

- A centralized platform for trunk and dial plan aggregation is commonly deployed. This platform is typically a Cisco Unified Communications Session Management Edition (SME) cluster, although an SIP proxy server (for example, Cisco Unified SIP Proxy (CUSP)) could also be used to provide intercluster call routing and dial plan aggregation in multisite distributed call-processing deployments.

- Centralized services such as centralized PSTN access, centralized voice mail, and centralized conferencing are available. These services can be deployed centrally, thus benefiting from centralized management and economies of scale. Services that need to track end-user status (for example, Cisco IM and Presence) must connect to the CUCM cluster for the users that they serve.
- The use of high-bandwidth audio (for example, G.711 or G.722) between devices within the same site, but low-bandwidth audio (for example, G.729) between devices in different sites.
- The use of high-bandwidth video (for example, 1.5 Mbps with 4CIF or 720p, to 2 Mbps with 1080p) between devices in the same site, but low-bandwidth video (for example, 384 kbps with 448p or CIF) between devices at different sites.
- The use of se a minimum of 1.5 Mbps or greater WAN link speed. Video is not recommended on WAN connections that operate at speeds lower than 1.5 Mbps.
- Call admission control is achieved through Enhanced Locations CAC or RSVP.

Benefits of Multisite Deployment with Distributed Call Processing Model

The multisite deployment with distributed call-processing model is a superset of both the single-site and multisite WAN with centralized call processing models.

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

- PSTN call cost savings are possible when the IP WAN is used for calls between sites.
- In this model, you can use the IP WAN to bypass toll charges by routing calls through remote site gateways, closer to the PSTN number that is dialed—that is, TEHO.
- Maximum utilization of available bandwidth is possible by allowing voice traffic to share the IP WAN with other types of traffic.

Clustering over the IP WAN

Cisco supports CUCM clustered over an IP WAN. Figure 2-7 shows the publisher and two subscribers at one location while another pair of subscribers from the same cluster resides at a different location. The QoS-enabled IP WAN connects the two sites. Note the requirement of a round trip time less than 80 ms between the sites. This requirement is in support of database replication occurring between the publisher and all the subscribers in the cluster.

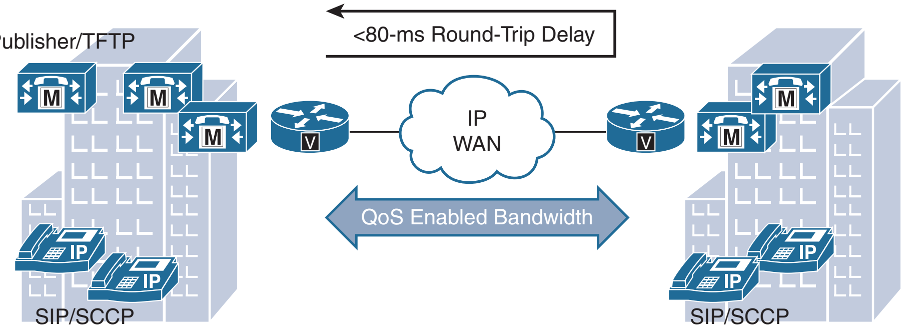


Figure 2-7 *Clustering over the WAN*

Some of the characteristics of this model include:

- Applications and CUCM servers of the same cluster can be distributed over the IP WAN.
- The IP WAN carries intracluster server communication and signaling.
- Limited number of sites:
  - Two to four sites for local failover (two CUCM servers per site)
  - Up to eight sites for remote failover across the IP WAN (one CUCM server per site).

The cluster design is useful for customers who require more functionality than the limited feature set that is offered by Cisco Unified SRST. This network design also allows remote offices to support more IP phones than SRST if the connection to the primary CUCM is lost.

Design Guidelines for Clustering over WAN Deployment Model

Although the distributed single-cluster call-processing model offers some significant advantages, it must adhere to these strict design guidelines:

- Two CUCM servers in a cluster must have a maximum round-trip delay of 80 ms between them. Because of this strict guideline, this design can be used only between closely connected, high-speed locations.
- A minimum of 1.544 Mbps (T1) of bandwidth is required for Intra-Cluster Communication Signaling (ICCS) between each site and every other site that is clustered over the WAN. This bandwidth supports up to 10,000 busy hour call attempts (BHCAs) within the cluster. The BHCA represents the number of call attempts that are made during the busiest hour of the day.
- In addition to the bandwidth required for ICCS traffic, a minimum of 1.544 Mbps (T1) of bandwidth is required for database and other inter-server traffic between the publisher and every subscriber node within the cluster.

- Up to eight small sites are supported using the remote failover deployment model. Remote failover allows you to deploy one server per location. (A maximum of eight call-processing servers are supported in a cluster.) If CUCM fails, IP phones register to another server over the WAN. Therefore, Cisco Unified SRST is not required in this deployment model (although it is supported). The remote failover design may require significant additional bandwidth, depending on the number of telephones at each location.

Benefits of Clustering over WAN Deployment Model

Clustering over the IP WAN provides a combination of the benefits of the two multisite deployment models to satisfy specific site requirements.

Although there are stringent requirements, clustering over the IP WAN offers these advantages:

- Single point of administration for users for all sites within the cluster
- Feature transparency
- Shared line appearances
- Cisco Extension Mobility within the cluster
- A unified dial plan

The clustering over IP WAN design is useful for customers who want to combine these advantages with the benefits that are provided by a local call-processing agent at each site (intrasite signaling is kept local, independent of WAN failures) and require more functionality at the remote sites than is provided by Cisco Unified SRST. This network design also allows remote offices to support more Cisco IP phones than SRST (1500 IP phones using Cisco 3945E Integrated Services Routers) in the event of WAN failure.

These features make clustering across the IP WAN ideal as a disaster-recovery plan for business continuance sites or as a single solution for up to eight small or medium sites.

Collaboration Edge Deployment Model

With increasing focus on teleworking and remote workers, enterprise collaboration resources are required to be extended beyond traditional collaboration borders. This border between an enterprise Unified Communications network and the outside world is referred to as the Collaboration Edge. Collaboration Edge services offer access to enterprise network resources from the outside world via multiple mechanisms. The users can be teleworkers working from home, mobile workers with LTE or Wi-Fi Internet access, or users using collaboration applications such as Jabber to make and receive calls to and from the PSTN or enterprise network. Figure 2-8 gives an overview of a Collaboration Edge solution.

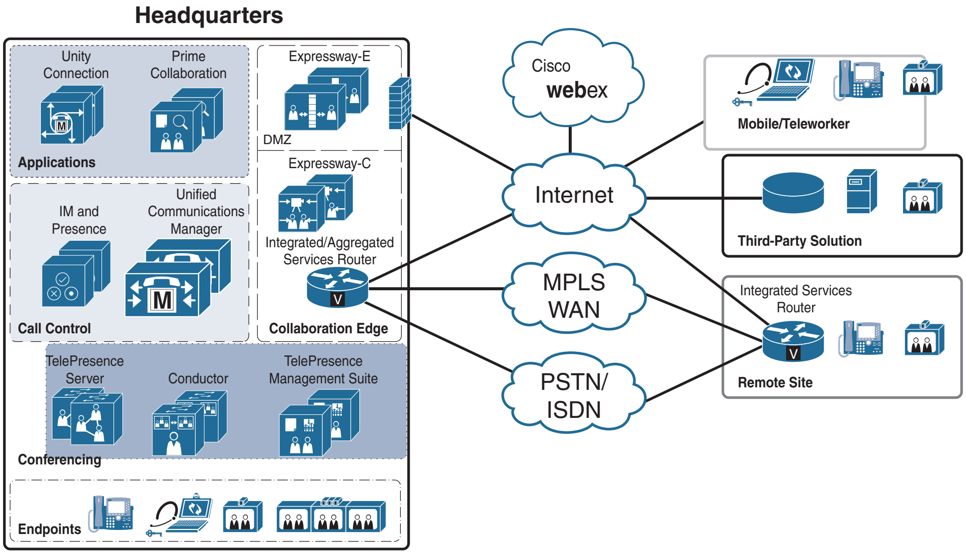


Figure 2-8 *Cisco Collaboration Edge Solution Overview*

The Collaboration Edge solution depends on the requirements of an organization and the technology an organization wishes to leverage. For example, the remote collaboration client access can be categorized into four main categories:

- **VPN-based access:** With endpoints capable of supporting traditional IPsec client or AnyConnect client.
- **VPN-less access:** With clients that traverse the firewall without any VPN client, for example Cisco Expressway solution.
- **Business-to-business communications:** Leveraging CUBE for B2B audio and video calls/conferencing.
- **IP PSTN access:** Leveraging ITSP SIP trunks instead of traditional PSTN trunks. CUBE yet again plays an important and integral part in connecting the enterprise network to ITSP.

Note Cisco Collaboration Edge solution using Cisco Expressway is addressed in *Implementing Cisco IP Telephony and Video Part 2*. VPN based access is out of scope of this text. For more information on VPN-based access refer to *Securing Cisco IP Telephony Networks*. B2B and IP PSTN access is covered in Chapter 13, "Implementing Cisco IOS Voice Gateways and Cisco Unified Border Element."

The next section addresses CUCM call processing redundancy.

CUCM Call-Processing Redundancy

A cluster is a set of networked servers that can be configured to provide specific services per server. Some cluster servers can be configured to provide CUCM services while other servers can provide Computer Telephony Integration (CTI), Trivial File Transfer Protocol (TFTP), and other media services such as conferencing or music on hold (MOH) These services can be provided by the subscribers and the publisher and can be shared by all servers.

Clustering provides several benefits. It allows the network to scale to up to 40,000 endpoints, provides redundancy in case of network or server failures, and provides a central point of administration. CUCM also supports clusters for load sharing. Database redundancy is provided by sharing a common database, whereas call-processing redundancy is provided by CUCM groups.

A cluster consists of one publisher and a total maximum of 20 servers (nodes) running various services, including TFTP, media resources, conferencing, and call processing. You can have a maximum of eight nodes for call processing (running the Cisco CallManager service).

For a quick recap, a CUCM cluster has a CUCM publisher server that is responsible for replicating the database to the other subscriber nodes in the cluster. The publisher stores the call detail records, and is typically used to make most of configuration change, except starting with CUCM 8.0 where database modifications for user facing call processing features are made on the subscriber servers. The subscriber servers replicate the publisher's database to maintain configuration consistency across the members of the cluster and facilitate spatial redundancy of the database.

To process calls correctly, CUCM needs to retrieve configuration settings for all devices. These settings are stored in a database using an IBM Informix Dynamic Server (IDS). The database is the repository for information such as service parameters, features, device configurations, and the dial plan.

The database replicates nearly all information in a star topology (one publisher, many subscribers). However, CUCM nodes also use a second communication method to replicate run-time data in a mesh topology as shown in Figure 2-9 (every node updates every other node). This type of communication is used for dynamic information that changes more frequently than database changes. The primary use of this replication is to communicate newly registered phones, gateways, and DSP resources, so that optimum routing of calls between members of the cluster and the associated gateways occurs.

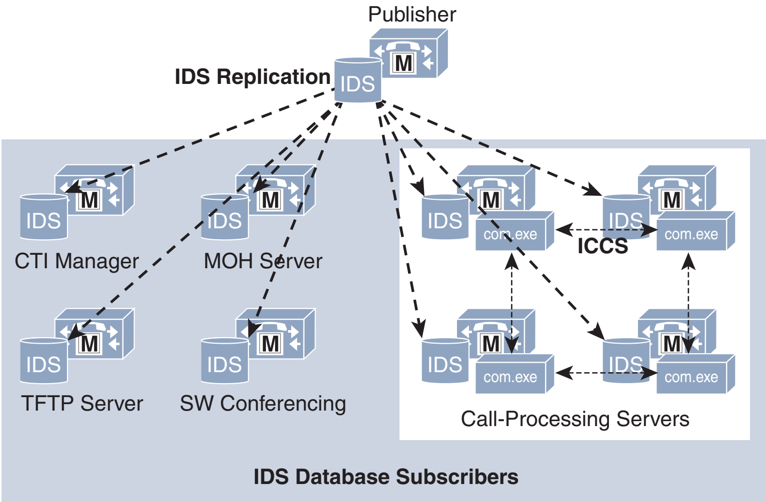


Figure 2-9 *Cisco Unified Communications Manager Database Replication Overview*

Database replication is fully meshed between all servers within a cluster. Static configuration data, because it is created through moves, adds, and changes, is always stored on the publisher and replicated one way from the publisher to each subscriber in the cluster. However, user-facing feature data, for example, Cisco Extension Mobility features, is writeable on a subscriber and are replicated from an updated subscriber to all other servers. All nonuser-facing feature data can be written only to the publisher database and is replicated from the publisher to all subscribers.

User-facing features are typically characterized by the fact that a user can enable or disable the feature directly on their phone by pressing one or more buttons, as opposed to changing a feature through a web-based GUI.

As illustrated in Figure 2-10, user-facing features that are listed below do not rely on the availability of the publisher. The dynamic user-facing feature data can be written to the subscribers to which the device is registered. The data is then replicated to all other servers within the cluster. By allowing the data to be written to the subscriber, the user-facing features can continue to function in the event of a publisher failure.

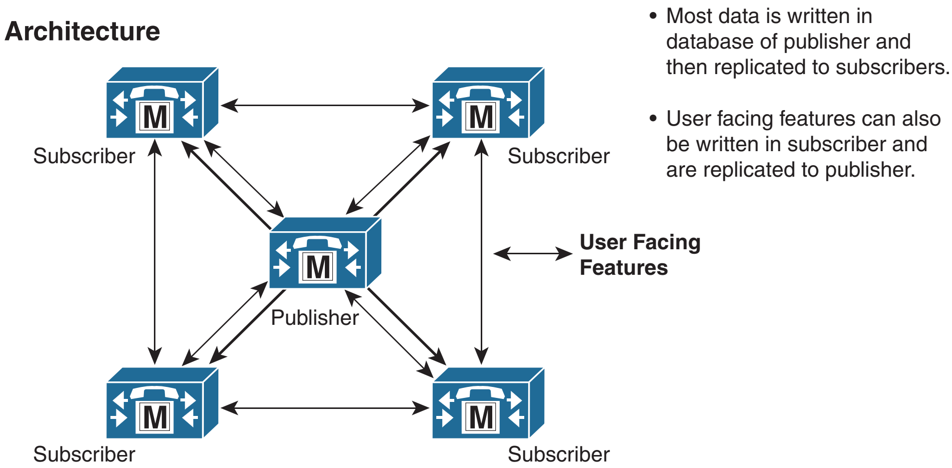


Figure 2-10 *User-Facing Feature Processing*

User-facing features are any features that can be enabled or disabled by pressing buttons on the phone and include the following:

- Call Forward All (CFA)
- Message Waiting Indicator (MWI)
- Privacy Enable/Disable
- Do Not Disturb (DND) Enable/Disable
- Cisco Extension Mobility Login
- Hunt-Group Logout
- Device Mobility
- CTI CAPF status for end users and application users

Therefore, most data (all nonuser-facing feature data) is still replicated in hub-and-spoke style (publisher to subscribers), while user-facing feature data is replicated bidirectionally between all servers.

Cisco Unified Communications Manager Groups: 1:1 Design

A 1:1 CUCM redundancy deployment design, as illustrated in Figure 2-11, guarantees that Cisco IP phone registrations never overwhelm the backup servers, even if multiple primary servers fail concurrently. This design provides high availability and simplifies the configuration. However, the 1:1 redundancy design has an increased server count compared with other redundancy designs and may not be cost-effective.

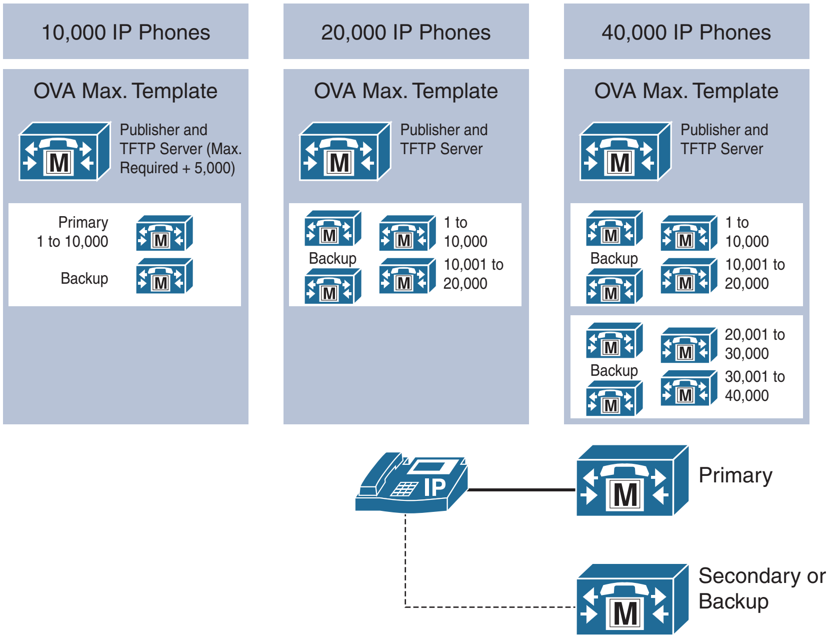


Figure 2-11 *1:1 Redundancy Design*

The other services (dedicated database publisher, dedicated TFTP server, or MOH servers) and media-streaming applications (conference bridge or MTP) may also be enabled on a separate server that registers with the cluster.

Each cluster must also provide the TFTP service, which is responsible for delivering IP phone configuration files to telephones, along with streamed media files, such as MOH and ring files. Therefore, the server that is running the TFTP service can experience a considerable network and processor load.

Depending on the number of devices that a server supports, you can run the TFTP service on a dedicated server, on the database publisher server, or on any other server in the cluster.

In Figure 2-11, an Open Virtualization Archive (OVA) template with the maximum number of users functions as the dedicated database publisher and TFTP server. In addition, there are two call-processing servers supporting a maximum of 10,000 Cisco IP phones. One of these two servers is the primary server; the other server is a dedicated backup server. The function of the database publisher and the TFTP server can be provided by the primary or secondary call-processing server in a smaller IP telephony deployment (fewer than 1000 IP phones). In this case, only two servers are needed in total.

When you increase the number of IP phones, you must increase the number of CUCM servers to support the IP phones. Some network engineers may consider the 1:1 redundancy design excessive because a well-designed network is unlikely to lose more than one primary server at a time. With the low possibility of server loss and the increased server cost, many network engineers choose a 2:1 redundancy design that is explained in the following section.

Cisco Unified Communications Manager Groups: 2:1 Design

Figure 2-12 shows a basic 2:1 redundancy design. While the 2:1 redundancy design offers some redundancy, there is the risk of overwhelming the backup server if multiple primary servers fail. In addition, upgrading the CUCM servers can cause a temporary loss of some services, such as TFTP or DHCP, because a reboot of the CUCM servers is needed after the upgrade is complete.

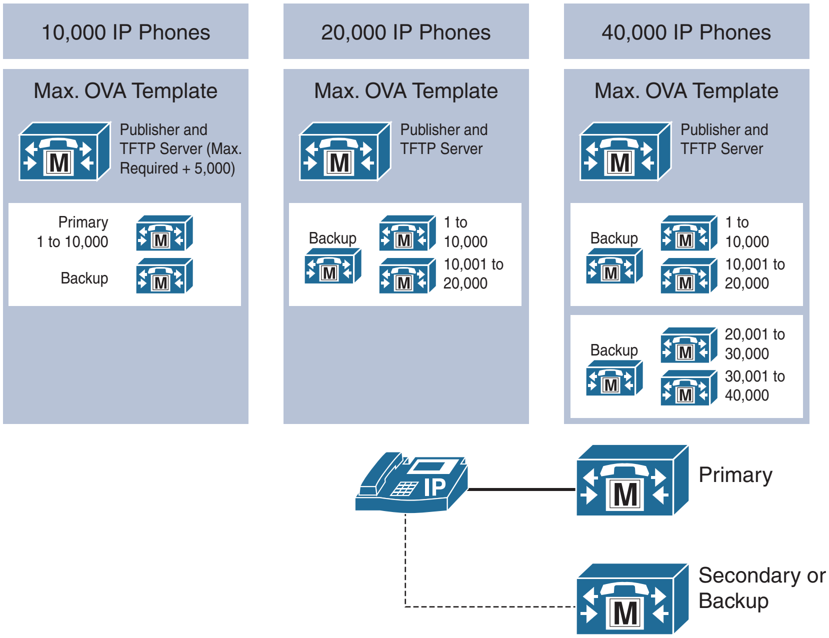


Figure 2-12 *2:1 Redundancy Design*

Network engineers use this 2:1 redundancy model in most IP telephony deployments because of the reduced server costs. If a virtual machine with the largest OVA template is used (shown in Figure 2-11), the server is equipped with redundant, hot-swappable power supplies and hard drives, and it is properly connected and configured, it is unlikely that multiple primary servers will fail at the same time, which makes the 2:1 redundancy model a viable option for most businesses.

As shown in the first scenario in Figure 2-12, when no more than 10,000 IP phones are used, there are no savings in the 2:1 redundancy design compared with the 1:1 redundancy design, simply because there is only a single primary server.

In the scenario with up to 20,000 IP phones, there are two primary servers (each serving 10,000 IP phones) and one secondary server. As long as only one primary server fails, the backup server can provide complete support. If both primary servers failed, the backup server would be able to serve only half of the IP phones.

The third scenario shows a deployment with 40,000 IP phones. Four primary servers are required to facilitate this number of IP phones. For each pair of primary servers, there is one backup server. As long as no more than two servers fail, the backup servers can provide complete support, and all IP phones will operate normally.

Cisco Voice Gateways and Cisco Unified Border Element

Because connectivity to the outside world is of utmost importance in Cisco Collaboration solution, this chapter wouldn't be complete without an overview and a brief discussion of Cisco IOS Voice Gateways and Cisco Unified Border Element (CUBE).

It is important to understand that both traditional voice gateways and CUBE have specific functions (with some degree of overlapping depending on deployment or design). Simply put, a voice gateway terminates time division multiplexing (TDM) signaling and transmits it by way of IP into the network or vice-versa. This allows calls to/from the PSTN network over traditional PSTN trunks, for example, ISDN T1, E1, and BRI trunks. A CUBE on the other hand terminates IP-to-IP calls, with the most common application being a SIP PSTN connection broker for enterprise network with ITSP. CUBE can do protocol interworking, address hiding, and multiple other functions described in the next section.

Note Cisco IOS voice gateways and CUBE and their functionalities, deployment options and protocols are described in detail in Chapter 13, "Implementing Cisco IOS Voice Gateways and Cisco Unified Border Element."

Cisco Voice Gateways

An access digital trunk gateway connects Cisco Unified Communications Manager to the PSTN or to a PBX via digital trunks such as Primary Rate Interface (PRI), Basic Rate Interface (BRI), or E1 R2 channel associated signaling (CAS). Digital E1 PRI trunks may also be used to connect to certain legacy voice mail systems.

Figure 2-13 gives an overview of an IOS voice gateway connecting the enterprise IP network to traditional PSTN network.

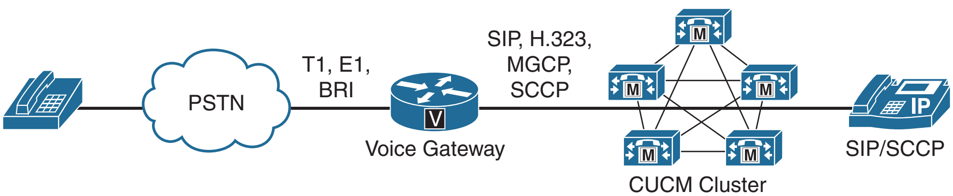


Figure 2-13 *Cisco IOS Voice Gateway Overview*

Gateways in a Collaboration network must meet the following core feature requirements:

- **Dual Tone Multifrequency (DTMF) relay capabilities:** DTMF relay capability, specifically out-of-band DTMF, separates DTMF digits from the voice stream and sends them as signaling indications through the gateway protocol (H.323, SCCP, MGCP, or SIP) signaling channel instead of as part of the voice stream or bearer traffic. Out-of-band DTMF is required when a low bit-rate codec is used for voice compression because the potential exists for DTMF signal loss or distortion.
- **Supplementary services support:** Supplementary services are typically basic telephony functions such as hold, transfer, and conferencing.
- **CUCM redundancy support:** CUCM clusters offer CUCM service and application redundancy. The gateways must support the ability to "re-home" to a secondary Cisco Unified Communications Manager in the event that a primary Cisco Unified Communications Manager fails. Redundancy differs from call survivability in the event of a Cisco Unified Communications Manager or network failure.
- **Fax/modem support:** Fax over IP enables interoperability of traditional analog fax machines with IP telephony networks. The fax image is converted from an analog signal and is carried as digital data over the packet network.

From a protocol perspective, CUCM supports the following gateway protocols:

- H.323
- Session Initiation Protocol (SIP)
- Media Gateway Control Protocol (MGCP)
- Skinny Client Control Protocol (SCCP)

Cisco Unified Border Element (CUBE)

Cisco Unified Border Element (CUBE) facilitates simple and cost-effective connectivity between enterprise unified communications with the PSTN world by leveraging Session Initiation Protocol (SIP) trunks to the IT Service Provider (ITSP), also known as the SIP Service Provider. A CUBE is primarily an IP-to-IP gateway that helps connect two or more similar or dissimilar networks, while offering a host of features that a regular voice gateway cannot offer. For example, a CUBE router can connect an H.323 network to SIP network or vice-versa, or a SIP network to a SIP provider. The following are some of the features that CUBE offers:

- Security demarcation, firewalling, DOS protection, and VPN services
- Signaling, protocol, and media interworking (H.323–SIP, SIP–H.323, SIP-SIP)
- Transcoding
- DTMF relay

- Media and signaling control and monitoring
- QoS and bandwidth management
- Co-existence/co-operation with TDM trunking
- Business-to-Business (B2B) audio and video communications

Figure 2-14 gives an overview of CUBE playing a role in B2B communications and connecting Enterprises 1 and 2 to PSTN via ITSP.

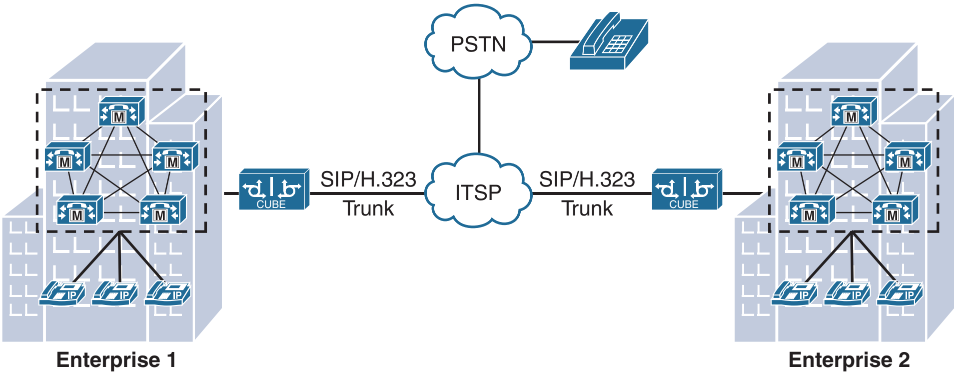


Figure 2-14 *CUBE in B2B Communications*

Chapter Summary

The following list summarizes the key points that were discussed in this chapter:

- Supported CUCM deployment models are Single-Site (Campus), Multisite with Centralized Call Processing, Multisite with Distributed Call Processing, and Clustering over the IP WAN.
- In the Single-Site deployment model, the CUCM, applications, and DSP resources are at the same physical location; all offsite calls are handled by the PSTN.
- The Multisite with Centralized Call Processing model has a single CUCM cluster. Applications and DSP resources can be centralized or distributed. The IP WAN carries call-control signaling traffic, even for calls within a remote site.
- The Multisite with Distributed Call Processing model has multiple independent sites, each with a CUCM cluster; the IP WAN carries traffic only for intersite calls.
- Clustering over the WAN provides centralized administration, a unified dial plan, feature extension to all offices, and support for more remote phones during failover, but it places strict delay and bandwidth requirements on the WAN.
- Clustering provide redundancy. A 1:1 redundancy design offers the highest availability but requires the most resources and is not as cost-effective as 2:1 redundancy.

Reference

For additional information, refer to the following:

- http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/cucm/srnd/collab10/collab10/models.html

Review Questions

Use the questions here to review what you learned in this chapter. The correct answers are found in Appendix A, "Answers to the Review Questions."

1. What is the maximum number of phones supported per CUCM cluster?
   a. 10,000
   b. 7500
   c. 30,000
   d. 40,000

2. How is call admission control handled in the Centralized Call Processing model?
   a. QoS
   b. H.323 gateway
   c. H.323 gatekeeper
   d. CUCM locations
   e. CUCM regions

3. What technology is used in the Centralized Call Processing model to reroute a call to a remote destination if there is not enough bandwidth to accommodate the call?
   a. Automated alternate routing
   b. Call admission control
   c. Quality of service
   d. Intercluster trunks

4. What technology is used to bypass toll charges by routing calls through remote-site gateways, closer to the PSTN number dialed?
   a. Automated alternate routing
   b. Tail-end hop-off
   c. Extension mobility
   d. Call admission control

5. Which call-processing model requires the use of SRST to provide backup for IP phones?
   a. Single-Site model
   b. Centralized multisite model
   c. Distributed multisite model
   d. Clustering over the WAN model

6. Gatekeepers are used within which call-processing model?
   a. Single-Site model
   b. Centralized model
   c. Distributed model
   d. Clustering over the WAN model

7. What is the maximum round-trip time requirement between CUCM servers in the Clustering over the WAN model?
   a. 20 ms
   b. 150 ms
   c. 80 ms
   d. 300 ms

8. What is the minimum amount of bandwidth that must be dedicated to database replication in the Clustering over the WAN model?
   a. 900 kbps
   b. 1.544 Mbps
   c. 80 kbps
   d. 2.048 Mbps

9. What platform is recommended to be used as a trunk and dial plan aggregation element?
   a. Cisco Unified SRST
   b. CallManager Express
   c. CUCM Session Management Edition
   d. Cisco Prime Collaboration

10. True or false? Clustering over the WAN allows for up to 20 sites, each with its own subscriber to provide local call control capabilities.
    a. True
    b. False

Index

Symbols

A

B

C

D

E

F

G

H

I

J-K

L

M

N

O

P

Q

S

T

U

V

W-X-Y-Z