



Cisco Unified Contact Center Enterprise Design Guide, Release 10.5(1)

First Published: 2014-06-18

Last Modified: 2016-02-12

Americas Headquarters

Cisco Systems, Inc.

170 West Tasman Drive

San Jose, CA 95134-1706

USA

<http://www.cisco.com>

Tel: 408 526-4000

800 553-NETS (6387)

Fax: 408 527-0883

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

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

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

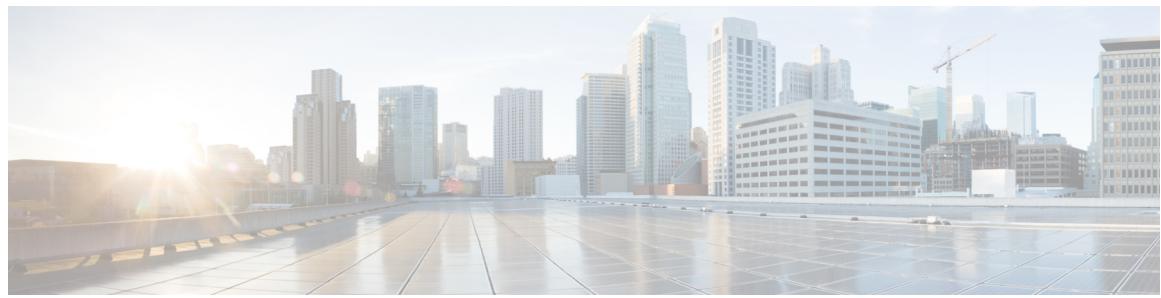
NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE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

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

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.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: <http://www.cisco.com/go/trademarks>. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

© 2003-2014 Cisco Systems, Inc. All rights reserved.



CONTENTS

P r e f a c e

Preface **xix**

Change History **xix**

About This Guide **xx**

Audience **xx**

Organization of This Guide **xx**

Related Documents **xxii**

Obtaining Documentation and Submitting a Service Request **xxii**

Field Alerts and Field Notices **xxii**

Documentation Feedback **xxii**

Conventions **xxii**

C H A P T E R 1

Architecture Overview **1**

Architecture Overview **1**

Solution Components **3**

 Cisco Unified Communications Manager **3**

 IPv6 **3**

 Cisco Voice Gateways **4**

 Agent Phones **4**

 Finesse and Multiline **5**

 Cisco Unified Customer Voice Portal **5**

 Cisco Unified IP IVR **6**

 Unified Intelligent Contact Manager **7**

 Time Synchronization **7**

 Cisco Unified Contact Center Enterprise **8**

 Unified CCE Software Components **8**

 Redundancy and Fault Tolerance **10**

 Peripheral Gateway and PIMs **11**

Unified CCE Agent Desktop Options	12
Administration and Data Server and Administration Client	13
Administration Server and Administration Client	13
Real-time Data Server	15
Historical Data Server And Detail Data Server	16
Unified CCE Reporting	16
Cisco Unified Intelligence Center	16
Unified Contact Center Management Portal	16
JTAPI Communications	17
Multichannel Subsystems: EIM and WIM	19
Cisco Outbound Option	19
Cisco Unified Mobile Agent	20
Serviceability	20
Diagnostic Tools	20
Network Management Tools	20
Combining IP Telephony and Unified CCE in the Same Unified Communications Manager Cluster	21
Combining IP Telephony and Unified CCE Extensions on the Same IP Phone	21
Agent Phones in Countries with Toll-Bypass Regulations	22
Queuing in a Unified CCE Environment	23
Transfers and Conferences in Unified CCE Environments	23
Dial Plans	24
Sample Unified CCE with Unified CVP Dial Plan	24

CHAPTER 2

Deployments 27

Unified CCE Base Model	27
Unified CCE Base Model Architecture	28
Unified CCE Base Model Components	29
Unified CCE Base Model Design Requirements	31
Unified CCE Base Model Variations	32
Enterprise Unified CCE Peripheral	33
Unified CCE Administration and Data Server	33
Agent Type Deployment Scenarios	37
Local Agent	37
Local Agent Architecture	38

Local Agent Components	38
Local Agent Benefits	39
Local Agent Design Requirements	39
Remote Offices	40
Remote Office with Agents	41
Remote Office with Agents Architecture	41
Remote Office with Agents Components	42
Remote Office with Agents Benefits	42
Remote Office with Agents Design Requirements	43
Remote Office with Agents and Voice Gateway	44
Remote Office with Agents and Voice Gateway Architecture	45
Remote Office with Agents and Voice Gateway Components	45
Remote Office with Agents and Voice Gateway Benefits	46
Remote Office with Agents and Voice Gateway Design Requirements	46
Home Agent with Broadband	48
Unified Mobile Agent	48
Unified Mobile Agent Architecture	49
Unified Mobile Agent Components	50
Unified Mobile Agent Benefits	50
Unified Mobile Agent Design Requirements	50
Centralized Data Center	54
Geographically Redundant Data Centers	55
Geographically Redundant Data Centers with Clustering over WAN	56
Geographically Redundant Data Centers with Distributed Unified Communications Manager Clusters	57

CHAPTER 3

Design Considerations for High Availability	59
High Availability Designs	59
High Availability and Virtualization	61
Data Network Design Considerations	61
Public and Private Network Connections	62
Unified Communications Manager Design Considerations	63
Unified Communications Manager Redundancy	64
Unified Communications Manager Load Balancing	65
Unified CVP Design Considerations	66

Unified IP IVR Design Considerations	67
High Availability Through Call Forwarding	68
High Availability Through Call Flow Routing Scripts	68
Cisco Web and E-Mail Interaction Manager Design Considerations	69
Unified CCE Integration	69
Load-Balancing Considerations	69
Failover Management	70
Cisco Outbound Option Design Considerations	70
SIP Dialer Design Considerations	71
Agent Peripheral Gateway Design Considerations	72
Agent PG Deployment for Unified Communications Manager Cluster	74
Central Controller Design Considerations	77
Common Processes That Support Failovers	78
Failure Detection Methods	78
Device Majority	78
PG Weight	78
Record Keeping During Failovers	79
Call Survivability	79
Unified CCE Failovers During Network Failures	79
Response to Private Network Failures	80
Response to Public Network Failures	81
Failures Between Unified Communications Managers	81
Failures Between Data Centers in Clustering over WAN	81
Failures to Agent Sites in Clustering over WAN	82
Response to Failures of Both Networks	82
Unified CCE Failovers During Single-Component Failures	83
Agent PG Fails	83
Subscriber Without CTI Manager Link to Agent PG Fails	84
CTI Manager with Agent PG Link Fails	86
Voice Response Unit PG Fails	88
Logger Fails	88
Administration and Data Server Fails	89
CTI Server Fails	90
Cisco Finesse Server Fails	92
Finesse Behavior When Other Components Fail	92

CTI OS Server Fails	93
Unified CCE Failovers During Multicomponent Failures	93
Agent PG and CTI Manager Fail	93
Unified Communications Manager Subscriber and CTI Manager Fail	94
Other Considerations for High Availability	96
Reporting Considerations	96

CHAPTER 4**Features** **97**

Precision Routing	97
Precision Routing Attributes	97
Precision Routing Limitations	98
Agent Greeting	98
Agent Greeting Phone Requirements (for Local Agents Only)	98
Agent Greeting Functional Limitations	99
Whisper Announcement with Agent Greeting	99
Whisper Announcement	99
Whisper Announcement Audio File	100
Whisper Announcement with Transfers and Conference Calls	100
Whisper Announcement Functional Limitations	100
Congestion Control	101
Deployment Types	102
Congestion Treatment Mode	105
Congestion Control Levels and Thresholds	105
Real-Time Capacity Monitoring	106
Congestion Control Configuration	107
Congestion Level Notification	107
Call Treatment for Outbound Option	107
Special Operating Condition	107

CHAPTER 5**Unified CCE Desktop Deployment Scenarios** **109**

Desktop Architecture	109
Peripheral Gateway and CTI Server	109
Cisco Finesse Server	109
CTI Object Server	111
Agent Desktops	112

Supervisor Desktops	113
Agent Mobility	113
Desktop Solutions	114
Cisco Finesse Desktop Solution	115
Finesse REST API	115
Finesse Agent Desktop	116
Finesse Supervisor Desktop	116
Finesse Administration Console	116
Finesse Multiserver Support	117
Load Balancing for Finesse	117
CTI OS Desktop Toolkit Solution	118
CTI OS Desktop Toolkit SDK and User Applications	119
Cisco Unified CRM Connector for Siebel Solution	122
Silent Monitoring	122
Unified Communications Manager Silent Monitoring	123
Cisco Finesse	124
CTI OS	124
CTI OS-Based Silent Monitoring	125
Clusters	128
Connection Profiles	128
Cisco Remote Silent Monitoring	128
RSM Platform Considerations	129
CTI OS Server	130
CTI Server	130
Voice Response Unit	130
Agent Phones	131
RSM Server Considerations	131
RSM Component Interaction	132
RSM Deployment Models	133
Single Site	133
Multisite WAN	135
Single Cluster with Multiple PG/CTI OS	139
Multiple Cluster with Multiple PG/CTI OS	140
Remote Silent Monitoring Bandwidth Requirements	140
Agent Phone Bandwidth Figures	141

RSM Codec Support	142
Failover Redundancy and Load Balancing	142
Host-Level Security	144
Transport or Session-Level Security	144
Support for Mobile Agent, IP Communicator, and Other Endpoints	144
Deployment Considerations	145
Citrix and Microsoft Terminal Services	145
Cisco Finesse	145
CTI OS Toolkit Desktop	145
NAT and Firewalls	146
Cisco Finesse and NAT	146
CTI Toolkit Desktop and NAT	146
Cisco Finesse and CTI OS Agents on the Same PG	146
IP Phone and IP Communicator Support	147
Cisco Jabber Support	147
Desktop Latency	148

CHAPTER 6

Cisco Outbound Option Description	149
Cisco Outbound Option Feature Description	149
Cisco Outbound Option Processes	150
Benefits of Cisco Outbound Option	150
Cisco Outbound Option Deployment Considerations	151
SIP Dialer Deployment Considerations	151
Outbound Dialing Modes	152
Call Flow for Agent-Based Campaign	152
Call Flow for Transfer to VRU Campaign	154
Cisco Outbound Option for Unified CCE	156
Enterprise Deployments	156
Single Gateway Deployment for SIP Dialer	156
Multiple Gateway Deployment for SIP Dialer	157
Clustering Over the WAN	158
Distributed Deployments	158
Distributed Deployment for Agent-Based Campaign	158
Distributed Deployment for Transfer to Unified CVP Campaign	160
Distributed Deployment for Transfer to IP IVR Campaign	161

Unified Contact Center Enterprise Deployments	163
Configure Cisco Outbound Option for Unified CCE	163
Calculate Number of Dialer Ports	163
Voice Gateway Considerations	163
Agent PG Considerations	164
Unified Communications Manager Considerations	164
Cisco Unified SIP Proxy Considerations	164
Unified CVP Considerations	164
Unified IP IVR Considerations	164
Unified Mobile Agent Considerations	165
SIP Dialer Throttling Considerations	165
Single Gateway Deployment	165
Multiple Gateway Deployment	166
SIP Dialer Recording	166
Call Transfer Timelines	167
High Availability Design for SIP Dialer	167
Campaign Manager and Import	167
SIP Dialer	168
CTI Server and Agent PG	168
Cisco Unified SIP Proxy Server	168
Cisco Outbound Option for Unified Mobile Agents	169
References	169

CHAPTER 7

Cisco Unified Mobile Agent	171
Cisco Unified Mobile Agent Architecture	171
Connection Modes	172
Call-by-Call Connection Mode	172
Nailed Connection Mode	173
Mobile Agent Connect Tone for Nailed Connection Mobile Agent	174
Supported Mobile Agent and Caller VoIP Endpoints	174
Agent Location and Call Admission Control Design	175
Dial Plan Design	176
Music on Hold Design	176
Codec Design	177
DTMF Considerations with Mobile Agent	177

Cisco Unified Border Element Considerations with Mobile Agent **178**

Cisco Unified Mobile Agent Interfaces **178**

Cisco Agent Desktop **178**

Cisco Agent Desktop Silent Monitoring and Recording **179**

CTI OS **180**

CTI OS Silent Monitoring **181**

Cisco Finesse **182**

Cisco Finesse Mobile Agent Silent Monitoring **183**

Customer Relationship Management Integrations **183**

Unified Mobile Agent with Cisco Outbound Option **183**

Cisco Unified Mobile Agent Fault Tolerance **184**

Unified Mobile Agent Sizing **184**

CHAPTER 8**Video Contact Center **185****

Video Contact Center Overview **185**

Video Contact Center Features **186**

Video Remote Expert **187**

Video Topologies **188**

Video Call Flows **190**

Video Remote Expert Support for Contact Center Features **190**

Video Infrastructure **190**

CHAPTER 9**Securing Unified CCE **193****

Introduction to Security **193**

Security Layers **194**

Platform Differences **195**

Security Design Elements **195**

Other Security Guides **196**

Network Firewalls **197**

TCP/IP Ports **197**

Network Firewall Topology **197**

Network Address Translation **198**

Active Directory Deployment **198**

AD Site Topology **199**

Organizational Units **199**

Application-Created OUs	199
AD Administrator-Created OUs	199
IPSec Deployment	201
Host-Based Firewall	202
Configuring Server Security	203
Unified Contact Center Security Wizard	203
Virus Protection	203
Antivirus Applications	203
Configuration Guidelines	203
Intrusion Prevention	204
Patch Management	204
Security Patches	204
Automated Patch Management	204
Endpoint Security	205
Agent Desktops	205
Unified IP Phone Device Authentication	206
Media Encryption (SRTP) Considerations	207
IP Phone Hardening	207

CHAPTER 10

Sizing Contact Center Resources	209
Sizing Contact Center Resources	209
Contact Center Basic Traffic Terminology	209
Contact Center Resources and the Call Timeline	212
Erlang Calculators as Design Tools	213
Erlang-C	213
Erlang-B	214

CHAPTER 11

Sizing Unified CCE Components and Servers	215
Sizing Considerations for Unified CCE	215
Core Unified CCE Components	215
Operating Conditions	216
Additional Sizing Factors	219
Peripheral Gateway and Server Options	223
Agent Greeting Sizing Considerations	224
Central Controller	224

Peripheral Gateway	225
Communications Manager	225
Mobile Agent	225
CVP and VXML Gateway	225
Whisper Announcement Sizing Considerations	225
Throttling During Precision Queue Changes	225
System Performance Monitoring	226
Summary	227

CHAPTER 12**Sizing Cisco Unified Communications Manager Servers** **229**

Sizing Unified Communications Manager Clusters for Unified CCE	229
Cluster Sizing Concepts	230
Cisco Unified Collaboration Sizing Tool	231
Cluster Guidelines and Considerations	231
Unified Communications Manager Redundancy	234
Load Balancing for Unified Communications Manager	235
Deployment of Agent PG in Unified Communications Manager Cluster	235
Sizing Considerations for Unified Mobile Agent	236

CHAPTER 13**Bandwidth Provisioning and QoS Considerations** **237**

Bandwidth Provisioning and QoS Considerations for Unified CCE	237
Unified CCE Network Architecture Overview	238
Network Segments	239
IP-Based Prioritization and QoS	241
UDP Heartbeat and TCP Keep-Alive	242
HSRP-Enabled Network	243
RSVP	243
Traffic Flow	244
Public Network Traffic Flow	244
Private Network Traffic Flow	244
Bandwidth and Latency Requirements	245
Quality of Service	245
Where to Mark Traffic	246
How to Mark Traffic	246
QoS Configuration	249

QoS Enablement in Unified CCE	249
QoS Configuration on Cisco IOS Devices	249
QoS Performance Monitoring	252
Bandwidth Provisioning	252
Bandwidth Requirements for Unified CCE Public and Private Networks	252
Public Network Bandwidth	252
Private Network Bandwidth	252
Bandwidth Requirements for Clustering over WAN	255
Bandwidth Requirements for Finesse Client to Finesse Server	257
Auto Configuration	258
Options for Gateway PG and Unified CCE	258
Outbound Option Bandwidth Provisioning and QoS Considerations	259
Distributed SIP Dialer Deployment	260
Agent-Based Campaign – No SIP Dialer Recording	260
Agent-Based Campaign – SIP Dialer Recording	261
Transfer-To-VRU Campaign – No SIP Dialer Recording	262
Transfer-To-VRU Campaign – SIP Dialer Recording	263
Bandwidth Requirements and QoS for Agent and Supervisor Desktops	264
Bandwidth Requirements for CTI OS Agent Desktop	264
CTI-OS Client/Server Traffic Flows and Bandwidth Requirements	265
Silent Monitoring Bandwidth Usage	265
CTI OS Server Bandwidth Calculator	266
Bandwidth reductions for CTI OS Server and CTI OS Agent Desktop	266
Bandwidth Requirements for an Administration and Data Server and Reporting	267
Bandwidth Requirements for Cisco EIM/WIM	267
Bandwidth and Latency Requirements for the User List Tool	267

CHAPTER 14

Cisco Unified Contact Center Management Portal	269
Unified Contact Center Management Portal	269
Unified CCMP Architecture	270
Portal Interfaces	270
Deployment Modes	271
Lab Deployment	271
Standard Deployments	271
Resilient Deployments	271

Parent/Child Deployment	272
Unified CCE Administration and Data Server	272
Roles	272
Administration Server (Configuration-Only Administration Server)	273
Systems That Exceed Published Limits	273
Software Compatibility	273
Reporting	273
Bandwidth Requirements	274
References	274

APPENDIX A**Acronym List** 275

APPENDIX B**System Requirements and Constraints** 287

Introduction to the Unified CCE Reference Designs	287
Unified CCE Reference Designs	288
Unsupported Configurations in the Unified CCE Reference Designs	289
Configuration Limits and Scalability Constraints for Unified CCE Reference Designs	289
Component and Feature Support for Unified CCE Reference Designs	296
Configuration Limits and Scalability Constraints for Non-Reference Designs	309
Administration and Data Server Limits by Deployment Type	314
Standard Operating Conditions	315
Data Store Configurations	318
Workstation Specifications	318
All-Event Client and Monitor-Mode Connection Limits	318
G.711 Audio Codecs Support	319
Codec Support in CVP	320
Codec Support in Unified CCE	320
Mixed Environments Not Supported	321
Solution Component and Feature Availability by Deployment Type	321
Congestion Control Limits by Deployment Type	322
Scalability Impacts of Components and Features	323
Notes on Unified ICM/Unified CCE Components	324
Administration & Data Server Deployment Capacities and Requirements	325
CTI OS Server	326
Silent Monitor Service for CTI OS	326

Cisco Unified Web and E-Mail Interaction Manager	326
Cisco Finesse Server	327
Unified Contact Center Management Portal	327
E.164 Dial Plan Design Considerations	328

APPENDIX C**Parent/Child 329**

Parent/Child Architecture	329
Parent/Child Components	330
Unified ICM (Parent) Data Center	330
Unified CCE Call Center (Child) Site	330
Unified CCE Gateway PGs at Data Center	331
Bandwidth for Unified CCE Gateway PG to Central Controller	332
Bandwidth for Unified CCE Gateway PG to System PG	332
Unified CCE System Peripheral	333
Parent/Child Limitations	334
Active Directory Deployments for Parent/Child	335
Unified CCMP for Parent/Child Deployments	336
Parent/Child Deployments Across Sites	337
Geographically Redundant Child Data Centers (Using Unified IP IVR)	339
Geographically Redundant Child Data Centers with CoW	340
Unified IP IVR-Based Child Data Centers with Distributed Unified Communications Manager	341
Unified CCE High Availability with Unified ICM	341
Parent/Child Call Flows	342
Typical Inbound PSTN Call Flow	342
Post-route Call Flow	343
Parent/Child Fault Tolerance	343
Unified CCE Child Loses Connection to Unified ICM	343
Unified CCE Gateway PG Cannot Connect to Unified ICM	344
Reporting and Configuration Impacts	345
Other High Availability Considerations	345
Traditional ACD Integration	345
Hybrid Deployment with Fixed PSTN Delivery	345
Hybrid Deployment with Unified CVP	345
ACD Integration and Parent/Child Deployments	346

Traditional VRU Integration	346
Integration Through PBX Transfer	346
Integration Through PSTN Transfer	347
Integration Through VRU Double Trunking	348
Unified Communications Manager Transfer and VRU Double Trunking	349

APPENDIX D**Cisco Agent Desktop** **351**

CAD Base Services	351
Cisco Agent Desktop Solution	352
CAD User Applications	353
CAD Application Features	353
Cisco Agent Desktop	354
Cisco Agent Desktop IP Phone Agent	355
Cisco Supervisor Desktop	356
Cisco Desktop Administrator	357
Cisco Desktop Monitoring Console	357
CAD Silent Monitoring and Recording	357
CAD-Based Monitoring	358
Desktop Monitoring	358
Server Monitoring	358
Mobile Agent Monitoring	358
Fault Tolerance for CAD-Based Monitoring and Recording	359
Load Balancing for CAD-Based Monitoring and Recording	359
Cisco Agent Desktop Presence Integration	359
Cisco Agent Desktop and NAT	361
Support for IP Phones and IP Communicator	363
Cisco Agent Desktop and Citrix	363
Support for Mix of CAD and CTI OS Agents on the Same PG	364
High Availability for Cisco Agent Desktop	364
Cisco Agent Desktop Failover Scenarios	364
CAD IP Phone Agent	365
Replacement of MSDE with SQL Database Server	366
Cisco Agent Desktop Component Sizing	366
Cisco Agent Desktop Operating Conditions	366
CTI OS for Cisco VXI	370

Cisco Agent Desktop Base Services	370
Cisco Agent Desktop VoIP Monitor Service	370
Cisco Agent Desktop Recording and Playback Service	370
Peripheral Gateway and Server Options for Cisco Agent Desktop	371
Bandwidth Requirements for Cisco Agent Desktop	372
Silent Monitoring Bandwidth Usage	372
Cisco Agent Desktop Applications Bandwidth Usage	375
Cisco Agent Desktop Service Placement	378
Miscellaneous Deployment Considerations	379



Preface

- [Change History, page xix](#)
- [About This Guide, page xx](#)
- [Audience, page xx](#)
- [Organization of This Guide, page xx](#)
- [Related Documents, page xxii](#)
- [Obtaining Documentation and Submitting a Service Request, page xxii](#)
- [Field Alerts and Field Notices, page xxii](#)
- [Documentation Feedback, page xxii](#)
- [Conventions, page xxii](#)

Change History

Changes	Date
<p>Clarified discussion of PG Failover mechanism in "Design Considerations for High Availability."</p> <p>Updated limits for precision queues. Added information on throttling during precision queue changes.</p> <p>Updated links to some of the bandwidth calculators.</p> <p>Removed design information about the deprecated SCCP Dialer.</p>	Initial release for 10.5
<p>Added information on the Unified CCE Reference Designs to the "System Requirements and Constraints" appendix.</p> <p>Added information on Finesse behavior during failovers to the "Design Considerations for High Availability" chapter.</p> <p>Added information about Finesse support for XenApp to the "Unified CCE Desktop Deployment Scenarios" chapter.</p>	July 16, 2014

Changes	Date
Added information about All-Event client and Monitor-Mode connection limits to the "System Requirements and Constraints" appendix.	August 26, 2014
Changed limits on active agents per Unified CM cluster to reflect recent improvements.	March 4, 2015
Added section "Load Balancing for Finesse" to the "Unified CCE Desktop Deployment Scenarios" chapter.	July 6, 2015
Updates to All-Event Client limits	September 25, 2015
Added notes on G.711 bandwidth for remote offices	January 5, 2016
Updated Desktop Latency section.	February 12, 2016
Corrected version to remove 11.0 information that was added by mistake.	November 30, 2016

About This Guide

This guide provides design considerations and guidelines for deploying Cisco Unified Contact Center Enterprise (Unified CCE). This guide assumes that you are familiar with basic contact center terms and concepts. Successful deployment of Unified CCE also requires familiarity with the information presented in the *Cisco Collaboration System Solution Reference Network Designs*.

Audience

This guide is primarily for contact center designers and system administrators.

Organization of This Guide

Section	Content
Architecture Overview	Introduces the components in a Unified CCE deployment.
Deployments	Describes the standard Unified CCE model deployments.
Design Considerations for High Availability	Discusses the Unified CCE features that provide for high availability in your contact center. This chapter also provides pointers on how to design your contact center to support the high availability features.
Features	Discusses the Precision Routing, Agent Greeting, Whisper Announcement, and Congestion Control features.

Section	Content
Unified CCE Desktop Deployment Scenarios	Discusses the Cisco Finesse and CTI OS desktops and how to design your contact center for them.
Cisco Outbound Option Descriptions	Describes the Cisco Outbound Option feature for calling campaigns and how to design your contact center to use the feature.
Cisco Unified Mobile Agent	Describes the Unified Mobile Agent feature which enables agents to work from remote locations.
Video Contact Center	Describes deployments that include video channels in a Unified CCE contact center.
Securing Unified CCE	Provides an introductory discussion of designing security into your contact center.
Sizing Contact Center Resources	Provides an general introduction to sizing resources for a contact center.
Sizing Unified CCE Components and Servers	Discusses sizing the Unified CCE components for your contact center. This chapter also discusses the impact of some optional features on component sizing.
Sizing Cisco Unified Communications Manager Servers	Discusses sizing the Unified Communications Manager components of your contact center.
Bandwidth Provisioning and QoS Considerations	Discusses bandwidth, latency, and quality of service design considerations for your contact center.
Cisco Unified Contact Center Management Portal	Describes the Unified Contact Center Management Portal.
Acronym List	Lists some common industry and Cisco-specific acronyms that you might encounter.
System Requirements and Constraints	Details Unified CCE system requirements and limits, the standard operating conditions that Unified CCE uses for tests, and other important specifications.
Parent/Child	Discusses design considerations for Parent/Child deployments.
Cisco Agent Desktop	Describes the Cisco Agent Desktop and how to design your contact center for it.

Related Documents

Subject	Link
Descriptions of Unified CCE features, including configuration and use.	<i>Cisco Unified Contact Center Enterprise Features Guide</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-feature-guides-list.html

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 delivers content directly to your desktop using a reader application. The RSS feeds are a free service.

Field Alerts and Field Notices

Cisco products may be modified or key processes may be determined to be important. These are announced through use of the Cisco Field Alerts and Cisco Field Notices. You can register to receive Field Alerts and Field Notices through the Product Alert Tool on Cisco.com. This tool enables you to create a profile to receive announcements by selecting all products of interest.

Log into www.cisco.com and then access the tool at <http://www.cisco.com/cisco/support/notifications.html>.

Documentation Feedback

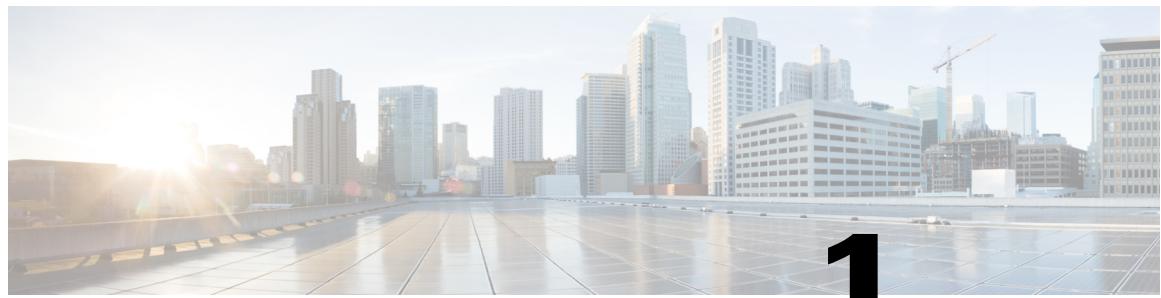
To provide comments about this document, send an email message to the following address:
contactcenterproducts_docfeedback@cisco.com

We appreciate your comments.

Conventions

This document uses the following conventions:

Convention	Description
boldface font	<p>Boldface font is used to indicate commands, such as user entries, keys, buttons, and folder and submenu names. For example:</p> <ul style="list-style-type: none"> • Choose Edit > Find. • Click Finish.
<i>italic</i> font	<p>Italic font is used to indicate the following:</p> <ul style="list-style-type: none"> • To introduce a new term. Example: A <i>skill group</i> is a collection of agents who share similar skills. • A syntax value that the user must replace. Example: IF (<i>condition, true-value, false-value</i>) • A book title. Example: See the <i>Cisco Unified Contact Center Enterprise Installation and Upgrade Guide</i> .
window font	<p>Window font, such as Courier, is used for the following:</p> <ul style="list-style-type: none"> • Text as it appears in code or that the window displays. Example: <code><html><title>Cisco Systems, Inc. </title></html></code>
< >	<p>Angle brackets are used to indicate the following:</p> <ul style="list-style-type: none"> • For arguments where the context does not allow italic, such as ASCII output. • A character string that the user enters but that does not appear on the window such as a password.



1

CHAPTER

Architecture Overview

- [Architecture Overview, page 1](#)
- [Solution Components, page 3](#)
- [Cisco Unified Contact Center Enterprise, page 8](#)
- [Combining IP Telephony and Unified CCE in the Same Unified Communications Manager Cluster, page 21](#)
- [Combining IP Telephony and Unified CCE Extensions on the Same IP Phone, page 21](#)
- [Agent Phones in Countries with Toll-Bypass Regulations, page 22](#)
- [Queuing in a Unified CCE Environment, page 23](#)
- [Transfers and Conferences in Unified CCE Environments, page 23](#)
- [Dial Plans, page 24](#)

Architecture Overview

Cisco Unified Contact Center Enterprise (Unified CCE) is a solution that delivers intelligent call routing, network-to-desktop Computer Telephony Integration (CTI), and multichannel contact management to contact center agents over an IP network. Unified CCE combines software IP automatic call distribution (ACD) functionality with Cisco Unified Communications to enable companies to deploy an advanced, distributed contact center infrastructure rapidly.

You must be familiar with the basic Unified Communications architecture and functionality as described in the *Cisco Unified Communications and Collaboration Solutions Design Guidance* at [http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-system/products-implementation-design-guides-list.html](http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-system-products-implementation-design-guides-list.html). Make sure that you become familiar with the concepts described in that manual for topics such as routes, labels, and dialed numbers.

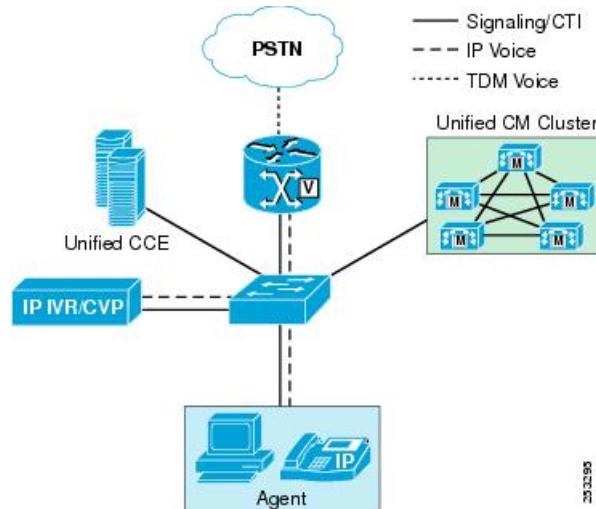
This design guide describes the deployment models and their implications including scalability, fault tolerance, and interaction between the solution components.

The Unified CCE product integrates with Cisco Unified Communications Manager, Cisco Unified IP IVR, Cisco Unified Customer Voice Portal, Cisco VoIP Gateways, and Cisco Unified IP Phones. Together these products provide contact center solutions to achieve intelligent call routing, multichannel ACD functionality, voice response unit (VRU) functionality, network call queuing, and consolidated enterprise-wide reporting.

Unified CCE can optionally integrate with Cisco Unified Intelligent Contact Manager to network with legacy ACD systems while providing a smooth migration path to a converged communications platform.

The Unified CCE solution is designed for implementation in both single and multisite contact centers. Unified CCE uses your existing Cisco IP network to lower administrative expenses and to include branch offices, home agents, and knowledge workers in your contact center. The following figure illustrates a typical Unified CCE setup.

Figure 1: Typical Unified CCE Solution Deployment



The Unified CCE solution consists primarily of four Cisco software products:

- Unified Communications infrastructure: Cisco Unified Communications Manager
- Queuing and self-service: Cisco Unified Customer Voice Portal (Unified CVP) or Cisco Unified IP Interactive Voice Response (Unified IP IVR)
- Contact center routing and agent management: Unified CCE. The major components are CallRouter, Logger, Peripheral Gateway, and the Administration & Data Server/Administration Client.
- Agent desktop software: Cisco Finesse, CTI Toolkit Desktop (CTI OS), or integrations with third-party customer relationship management (CRM) software through Cisco Unified CRM Connector.

The solution is built on the Cisco IP Telephony infrastructure, which includes:

- Cisco Unified IP Phones
- Cisco Voice Gateways
- Cisco LAN/WAN infrastructure

The following sections discuss each of the software products in more detail and describe the data communications between each of those products. For more information about a particular product, see the specific product documentation available online at cisco.com.

**Note**

Cisco Packaged CCE is a predesigned, bounded deployment model of Unified CCE. If your contact center requirements fit the boundaries of the Packaged CCE solution, you can enjoy the advantages of the simplified management interface, smaller hardware footprint, and reduced time to install. Packaged CCE customers also benefit from the comprehensive feature set of Unified CCE and Unified CVP. Packaged CCE supports up to 1000 agents, and uses Cisco Unified Intelligence Center for comprehensive reporting and Cisco Finesse desktop software. For more information, see <http://www.cisco.com/c/en/us/support/customer-collaboration/packaged-contact-center-enterprise/tsd-products-support-series-home.html>.

Solution Components

Cisco Unified Communications Manager

Cisco Unified Communications Manager is a software application that controls the Voice Gateways and IP phones, thereby providing the foundation for a VoIP solution. Unified Communications Manager runs on Cisco Unified Computing System (UCS) hardware or a specification-based equivalent. The software running on a VM is referred to as a Unified Communications Manager server. Multiple Unified Communications Manager servers can be grouped into a cluster to provide for scalability and fault tolerance. Unified Communications Manager communicates with the gateways using standard protocols such as Media Gateway Control Protocol (MGCP) and Session Initiation Protocol (SIP). Unified Communications Manager communicates with the IP phones using SIP or Skinny Call Control Protocol (SCCP). For details on Unified Communications Manager call processing capabilities and clustering options, see the latest version of the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Unified Communications Manager communicates with Unified CCE through the Java Telephony Application Programming Interface (JTAPI). In a fault-tolerant design, a Unified Communications Manager cluster is capable of supporting thousands of agents. However, the number of agents and the number of busy hour call attempts (BHCA) supported within a cluster varies and must be sized according to Cisco guidelines.

Typically, when designing a Unified CCE solution, you first define the deployment scenario, including the arrival point (or points) for voice traffic and the location (or locations) of the contact center agents. After defining the deployment scenario, you can determine the sizing of the individual components within the Unified CCE design, including how many Unified Communications Manager servers are needed within a cluster.

Cisco Service Advertisement Framework Call Control Discovery (SAF CCD) allows Unified Communications Manager to advertise directory number ranges that it owns and discover and dynamically create routes for other Unified Communications Manager clusters. SAF CCD replaces the need for gatekeepers and SIP proxies.

Related Topics

[Sizing Cisco Unified Communications Manager Servers, on page 229](#)

IPv6

Unified CCE supports interoperability with an IPv6-enabled Unified Communications Manager cluster. All of the Unified CCE components run with IPv4 enabled including Unified CVP, Unified IP IVR, the CTI OS

agent desktops, and agent phones. The Unified CCE Agent PG uses IPv4 to integrate with Unified Communications Manager CTI Manager.

Caller phones or voice gateways can run either IPv4 or IPv6. If the caller's environment is IPv6 only, you must use Media Terminating Point (MTP) resources for call treatment (Unified CVP VXML gateways and Unified IP IVR).

IPv6 is not supported for agent phones. The agent phone might support dual stack IPv4 or IPv6, but the agent sign-in is denied unless you register the phone as IPv4. You must configure Mobile Agent and Outbound Option endpoints (CTI ports and Dialer ports) as IPv4 devices.

Cisco Voice Gateways

When you select Voice Gateways for a Unified CCE deployment, it is important to select Voice Gateways that satisfy not only the number of required PSTN trunks but also the busy hour call completion rate on those trunks. Busy hour call completion rates per PSTN trunk are typically higher in a contact center than in a normal office environment. For Cisco Catalyst Communications Media Module (CMM) Voice Gateways used in pure contact center deployments, provision a maximum of four T1/E1 interfaces to ensure that the call processing capacity of the Voice Gateway is satisfactory.

Agent Phones

Unified CCE supports both monitoring of a single line for all agent devices (Single-Line Agent Mode) and monitoring multiple agent lines when Multi-Line Agent Mode is enabled for the Peripheral. Multi-Line Agent Mode provides the following capabilities:

- Monitoring and reporting of calls on all lines on the phone.
- Other than call initiation, all other call control on the non-ACD extensions is supported from multiline capable desktops. Calls initiated from the hard phone can be controlled after initial call setup.
- Requires a busy trigger of 1 (no call waiting), although calls can be forwarded to other extensions on the phone when busy.
- Requires a maximum of two call appearances.
- Supports a maximum of four lines per phone, one ACD line and up to three non-ACD lines.
- Shared lines are supported on non-ACD lines but not on ACD Lines.
- Shared lines are only supported on non-ACD lines. You can configure multiple devices with a shared non-ACD line, but can only log in an Agent into one of them. Shared lines are supported to enable agents with phones facilities at both, home and work, to be able to share a voicemail line.



Note

"Non ACD Line impact configuration" in PG Explorer is not supported on a shared Non-ACD Line.

-
- Call Park is not supported on ACD and non-ACD lines.
 - Unified CCE may not be backward compatible with third-party CTI applications when Multi-Line Agent Mode is enabled. Multiline support must be validated with the third-party vendor.

**Note**

Use of the Join Across Line (JAL) and Direct Transfer Across Line (DTAL) phone features is deprecated. Do not use these features in any new deployment.

For a list of supported agent phones, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.

The Agent Greeting feature relies on the phone's Built-In Bridge (BIB) functionality to play back the greeting to both the caller and the agent.

Agent Greeting requires:

- The phones have the BIB feature.
- The phones must be running the latest firmware version delivered with Unified Communications Manager.
- The phones must be configured as BIB enabled in Unified Communications Manager.

**Note**

These requirements apply to local agents only.

For a list of agent phones and required firmware to support the Agent Greeting feature, see the *Compatibility Matrix for Unified CCE*.

Finesse and Multiline

Finesse supports the configuration of multiple lines on agent phones if Unified CCE is configured for multiline. You can configure one or more secondary lines on an agent phone. However, the Finesse server blocks any events that are sent by the CTI server as a result of operations on an agent's secondary line. The Finesse server does not publish these events to the Finesse clients. Information about calls that agents handle on their secondary lines does not appear on the Finesse desktop.

If your agents use 8900 Series or 9900 Series phones, you must enable Multi-Line on the Unified Communications Manager peripheral. Because this configuration option is a peripheral-wide option, if you enable Multi-Line for even one agent who uses an 8900 Series or 9900 Series phone, you must enable it for all agents.

You must configure all phones with the following settings:

- Set Maximum number of calls to 2.
- Set Busy trigger to 1.

Cisco Unified Customer Voice Portal

Unified Customer Voice Portal (Unified CVP) is a software application that runs on Cisco Unified Computing System (UCS) hardware or specification-based equivalents. CVP provides prompting, collecting, queuing, and call control services using standard web-based technologies. The CVP architecture is distributed, fault tolerant, and highly scalable. With the CVP system, voice is terminated on Cisco IOS gateways that interact with the Unified CVP application server using VoiceXML (speech) and SIP (call control).

The Unified CVP software is tightly integrated with the Cisco Unified CCE software for application control. CVP interacts with Unified CCE using the Voice Response Unit (VRU) Peripheral Gateway Interface. The Unified CCE scripting environment controls the execution of building-block functions such as play media, play data, menu, and collect information. The Unified CCE script can invoke external VoiceXML applications for execution by the CVP VoiceXML Server, an Eclipse and J2EE-based scripting and web server environment. VoiceXML Server is suited for sophisticated and high-volume VRU applications and it can interact with custom or third-party J2EE-based services. These applications can return results and control to the Unified CCE script when complete. Advanced load balancing across all CVP solution components can be achieved by Cisco Content Services Switch (CSS) and Cisco IOS Gatekeepers or Cisco Unified Presence SIP Proxy Servers.



Important Unified CCE deployments with multiple clusters do not support Unified CVP's Enhanced Location Call Admission Control feature.

Unified CVP can support multiple grammars for prerecorded announcements in several languages. CVP can optionally provide automatic speech recognition and text-to-speech capability. CVP can also access customer databases and applications through the Cisco Unified CCE software.

Unified CVP also provides a queuing platform for the Unified CCE solution. Voice and video calls can remain queued on CVP until they are routed to a contact center agent (or external system). The system can play back music or videos while the caller is on hold. When Unified CCE routes the call to an agent, the agent can send videos to a caller from the agent desktop application. For more information, see the latest version of the *Design Guide for Cisco Unified Customer Voice Portal* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-implementation-design-guides-list.html>.

Unified CVP also supports Agent Greeting recording and playback when integrated with Unified CCE. A preinstalled CVP VXML application is provided to allow agents to record and manage their greetings. Unified CCE instructs the CVP to play back the agent's specific greeting to the caller and agent when the agent answers the call.

Unified CVP also supports the Whisper Announcement feature to play a prerecorded announcement to the agent when the agent answers the call.

Cisco Unified IP IVR

The Unified IP IVR provides prompting, collecting, and queuing capability for the Unified CCE solution. Unified IP IVR does not provide call control as Unified CVP does because it is behind Unified Communications Manager and under the control of the Unified CCE software by way of the Service Control Interface (SCI). When an agent becomes available, the Unified CCE software instructs the Unified IP IVR to transfer the call to the selected agent phone. The Unified IP IVR then requests Unified Communications Manager to transfer the call to the selected agent phone.

Unified IP IVR is a software application that runs on Cisco Unified Computing System (UCS) hardware or a specification-based equivalent. You can deploy multiple Unified IP IVR servers with a single Unified Communications Manager cluster under control of Unified CCE.

Unified IP IVR has no physical telephony trunks or interfaces like a traditional VRU. The telephony trunks are terminated at the Voice Gateway. Unified Communications Manager provides the call processing and switching to set up a g.711 or G.729 Real-Time Transport Protocol (RTP) stream from the Voice Gateway to the Unified IP IVR. The Unified IP IVR communicates with Unified Communications Manager through the Java Telephony Application Programming Interface (JTAI), and the Unified IP IVR communicates with

Unified CCE through the Service Control Interface (SCI) with a VRU Peripheral Gateway or System Peripheral Gateway.

For deployments requiring complete fault tolerance, a minimum of two Unified IP IVRs are required.

Related Topics

[Sizing Contact Center Resources, on page 209](#)

[Design Considerations for High Availability, on page 59](#)

Unified Intelligent Contact Manager

Unified CCE may be deployed with Unified ICM to form a complete enterprise call management system. Unified ICM interfaces with ACDs from various vendors (including Cisco Unified CCE), VRUs (including Cisco Unified IP IVR and Unified CVP), and telephony network systems to efficiently and effectively treat customer contacts such as calls and e-mail regardless of where the resources are located in the enterprise.

Unified CCE may be deployed in a *hybrid* model with Unified ICM where the CallRouter, Logger, Administrative & Data Servers, and other components are shared between Unified ICM and Unified CCE.

Alternatively, Unified CCE may be deployed in a parent/child model where Unified ICM is the parent and Unified CCE is the child. This closely resembles the deployment model of Unified ICM with ACDs from other vendors. It is used for a highly scalable deployment because it provides CallRouters, data servers, and so forth for each product; although there are more components to manage and maintain. It is also used for a distributed model where isolation is needed between Unified ICM and Unified CCE, such as in an outsourced operation.

Related Topics

[Parent/Child , on page 329](#)

Time Synchronization

To ensure accurate operation and reporting, all the components in your contact center solution must use the same value for the time. You can synchronize the time across your solution using a Simple Network Time Protocol (SNTP) server. The following table outlines the needs of various component types in your solution.



Important Use the same NTP sources throughout your solution.

Type of component	Notes
Domain controllers	Domain controllers must all point to the same NTP servers.
ESXi hosts	All ESXi hosts must point to the same NTP servers as primary domain controllers.
Windows components in the contact center domain	Windows machines in the domain point to, and are automatically in synch with, the primary domain controller for NTP. They require no configuration for NTP.

Type of component	Notes
Windows components not in the contact center domain	Follow the Microsoft documentation to synchronize directly with the NTP server.
Non-Windows components	Components such as Unified Intelligence Center, Finesse, Social Miner, and Unified Communications must point to the same NTP servers as the domain controllers.
Cisco Integrated Service Routers	To provide accurate time for logging and debugging, use the same NTP source as the solution for the Cisco IOS Voice Gateways.

Cisco Unified Contact Center Enterprise

Cisco Unified Contact Center Enterprise (CCE) is the software application that provides the contact center features, including agent state management, agent selection, call routing and queue control, VRU control, CTI Desktop screen pops, and contact center reporting. Unified CCE runs in virtual environments on Cisco Unified Computing System (UCS) hardware or specification-based equivalents. Unified CCE relies on the Microsoft Windows Server 2008 R2 operating system software and the Microsoft SQL Server 2008 database management system. With the flexibility of virtual machines, you can scale and size Unified CCE software to meet various deployment requirements.

Related Topics

[Sizing Unified CCE Components and Servers, on page 215](#)

Unified CCE Software Components

This topic describes the main components of the Unified CCE product. Following sections describe some key concepts and terminology and go into more detail on some of the components.

The Cisco Unified CCE software is a collection of components that can run on multiple virtual machines (VMs). The number and type of components that can run on one server is primarily based on busy hour call attempts (BHCA) and the size of the VM. Other factors that impact the hardware sizing are the number of agents, the number of skill groups per agent, the number of Unified IP IVR ports, the number of VRU Script nodes in the routing script, Expanded Call Context (ECC) usage, and which statistics the agents need at their desktops.

The core Unified CCE software components are listed here and described in greater detail later in this chapter.

Table 1: Core Unified CCE Software Components

Unified CCE Software Components	Description
CallRouter	Makes all routing decisions on how to route a call or customer contact. Often just referred to as the “Router” in the context of Unified CCE components. The Router is a part of the Central Controller.
Logger	The database server that stores contact center configuration data and temporarily stores historical reporting data for distribution to the data servers. The Logger is a part of the Central Controller.
Cisco Finesse Server	The Finesse server connects Finesse desktops to the CTI server on the Agent PG.
CTI Object Server (CTI OS)	CTI interface for CTI OS Agent Desktops.
Peripheral Gateway (PG)	Interfaces to various <i>peripheral</i> devices, specifically to Unified Communications Manager, VRU (Unified IP IVR or Unified CVP), or Multichannel products (EIM and WIM for email and chat). The PG includes one or more Peripheral Interface Managers (PIMs) for the specific device interfaces.
Agent PG	PG that has a Unified Communications Manager PIM.
Unified Communications Manager Peripheral Interface Manager (PIM)	Part of a PG that interfaces to a Unified Communications Manager cluster by using the JTAPI protocol.
VRU PIM	Part of a PG that interfaces to the Unified IP IVR or Unified CVP through the Service Control Interface (SCI) protocol.
MR PIM	Part of a PG that interfaces to call center Multimedia products, specifically EIM and WIM for email and chat.
CTI Server	Part of the PG that interfaces to CTI OS and provides an open CTI interface, which is useful for some server-to-server communications and Finesse and third-party CTI applications
Administration & Data Server	Configuration interface and real-time and historical data storage (for example, for reporting). There are several different deployment models described later in this chapter.
Administration Client	Configuration interface. This component has a subset of the functionality of the Administration & Data Server. It is a <i>lighter weight</i> deployment because it does not require a local database and it is deployed to allow more places from which to configure the solution.

Unified CCE Software Components	Description
Cisco Unified Intelligence Center (Unified Intelligence Center)	Provides web browser-based real-time and historical reporting. Unified Intelligence Center also works with other Cisco Unified Communications products.

The combination of CallRouter and Logger is called the Central Controller. When the CallRouter and Logger modules run on the same VM, the server is referred to as a Rogger.

Redundancy and Fault Tolerance

You deploy the CallRouter and Logger in a paired redundant fashion. The two sides of the redundant deployment are referred to as Side A and Side B. For example, CallRouter A and CallRouter B are redundant instances of the CallRouter running on two different VMs. In normal operation, both sides are running. When one side is down, the configuration is running in stand-alone mode. These modes are occasionally referred to as duplex and simplex modes.



Note

Stand-alone (simplex) deployments of the CallRouter and Logger are not supported in production environments. You *must deploy* these components in redundant pairs.

The two sides are for redundancy, not load-balancing. Either side can run the full load of the solution. The A and B sides both execute the same set of messages and produce the same result. In this configuration, there logically appears to be only one CallRouter. The CallRouters run in synchronized execution across the two VMs, which mean both sides process every call. During a failure, the surviving Call Router picks up the call midstream and continue processing in real time and without user intervention.

The Peripheral Gateway (PG) components run in hot-standby mode, meaning that only one PG is active and controlling Unified Communications Manager or the appropriate peripheral. When the active side fails, the surviving side automatically takes over processing of the application. During a failure, the surviving side runs in simplex mode until the redundant side is restored and the configuration automatically returns to redundant operation.

The CTI OS component provides fault tolerance through a pair of servers that operate together and back up each other. There is no notion of an active and passive server, or of a primary and secondary server. Both servers are always active. Clients can connect to either server. During the failure of any one server, clients can automatically reconnect to the alternate server.

The Administration & Data Servers, which handle configuration and real-time data, are deployed in pairs for fault tolerance. You can deploy multiple pairs for scalability. The Administration & Data Servers for historical data follow an n+1 architecture for redundancy and scalability. Each Administration & Data Server has a Logger (Side A or B) as its preferred and primary data source.

Related Topics

[Administration and Data Server and Administration Client, on page 13](#)

Peripheral Gateway and PIMs

For each Unified Communications Manager cluster in the Unified CCE environment, there is a Unified Communications Manager PIM on an Agent Peripheral Gateway (PG). For scalability requirements, some deployments may require multiple PIMs for the same cluster. You must deploy each PIM on a different Agent PG. Each Agent PG with a Unified Communications Manager PIM can support a maximum of 2000 agents.

For each Agent PG, there is one CTI Server component and one or more CTI OS components to communicate with the desktops associated with the phones for that cluster.



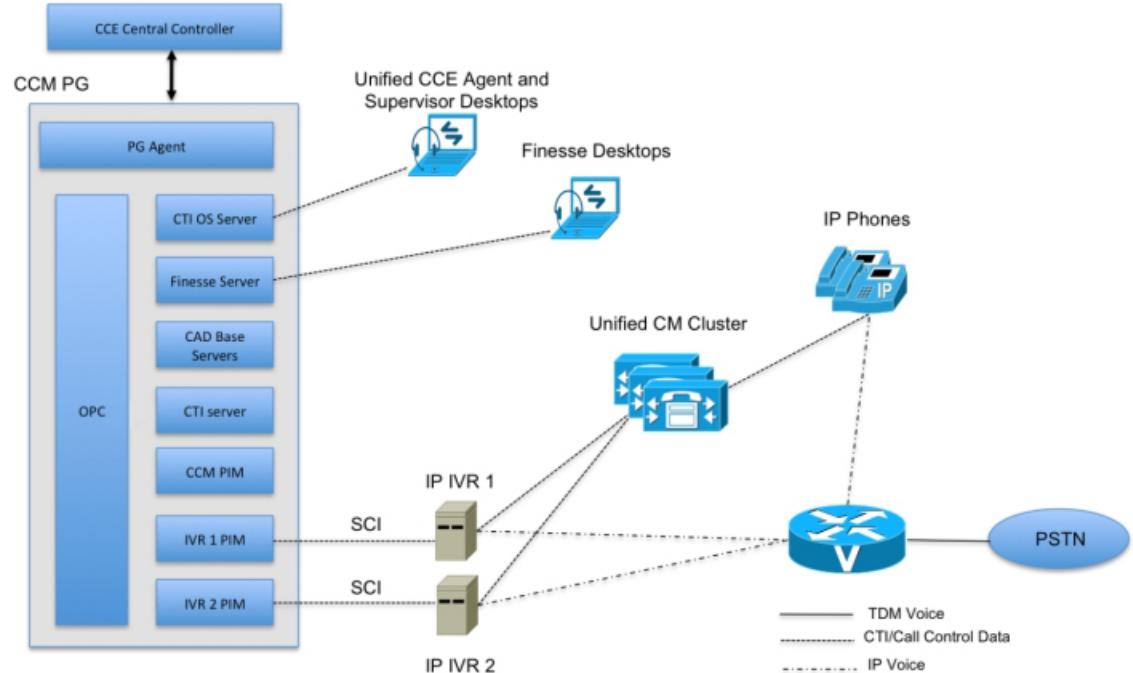
Note

The CTI OS components on Side A and Side B are simultaneously active to load-balance desktop communication.

For each Unified IP IVR or CVP Server, there is one VRU PIM. VRU PIMs may be part of the Agent PG. Often, the Unified Communications Manager PIM, the CTI Server, the CTI OS, and multiple VRU PIMs may run on the same VM.

Internal to the PG is a process called the PG Agent which communicates to the Central Controller. Another internal PG process is the Open Peripheral Controller (OPC), which enables the other processes to communicate with each other and is also involved in synchronizing PGs in redundant PG deployments. The following figure shows the communications among the various PG software processes.

Figure 2: Communications Among Peripheral Gateway Software Processes



In larger, multisite (multi-cluster) environments, multiple Agent PGs are usually deployed. When multiple clusters are deployed, Unified CCE tracks all the agents and calls centrally. Unified CCE is able to route the

calls to the most appropriate agent independent of the site or cluster that they are using, thus making them all appear to be part of one logical enterprise-wide contact center with one enterprise-wide queue.

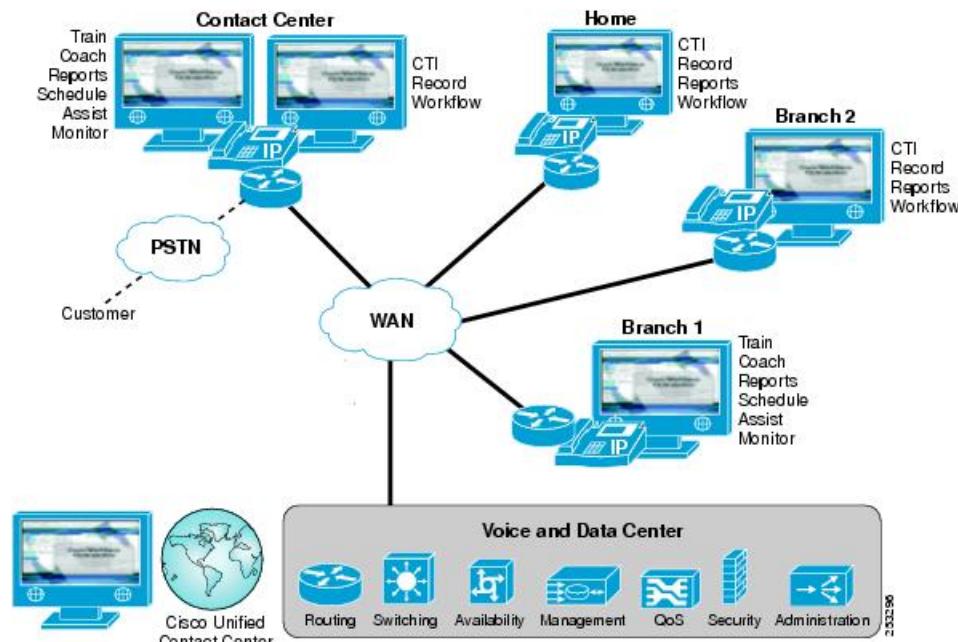
Unified CCE Agent Desktop Options

Cisco offers the following interfaces for Unified CCE agents:

- **Cisco Finesse Desktop**—Cisco Finesse is a web-based desktop solution that allows for the extension of the desktop through standardized web components. Cisco Finesse offers:
 - a browser-based solution
 - an extensible desktop interface using standard OpenSocial gadgets
 - server features available to applications using documented REST APIs
- **Cisco CTI OS Desktop Toolkit**—The CTI OS Desktop Toolkit provides a software toolkit for building custom desktops, desktop integrations into third-party applications, or server-to-server integrations to third-party applications.

These integrated solutions enable call control (Answer, Drop, Hold, Un-Hold, Blind or Warm Transfers, and Conferences), outbound calls, consultative calls, and delivery and manipulation of Call Context Data (CTI screen pop).

Figure 3: Variety of Agent Interfaces for Unified CCE



Agents who use a third-party CRM user interface that is connected through a CRM Connector can be supervised using a CTI OS Desktop Toolkit-based supervisor desktop.

Related Topics

[Unified CCE Desktop Deployment Scenarios, on page 109](#)

Administration and Data Server and Administration Client

Administration & Data Servers have several roles: Administration, Real-time data server, Historical Data Server, and Detail Data Server. A Unified CCE deployment must have Administration & Data Servers to fill these roles. The servers may be deployed in the following combinations to achieve the needed scalability with the minimum number of VMs:

- Administration Server and Real-Time Data Server (AW)
- Configuration-Only Administration Server
- Administration Server, Real-Time and Historical Data Server, and Detail Data Server (AW-HDS-DDS)
- Administration Server and Real-Time and Historical Data Server (AW-HDS)
- Historical Data Server and Detail Data Server (HDS-DDS)



Note See the [Deployments, on page 27](#) chapter for more details on deployment options and requirements.

-
- An Administration Client (formerly known as a *client AW*) serves the administration role but is deployed as a client to an Administration Server for scalability. The Administration Client may view and modify the configuration and receive real-time reporting data from the AW, but it does not store the data itself and does not have a database.

Install each Administration & Data Server on a separate VM for production systems to ensure no interruptions to the real-time call processing of the CallRouter and Logger processes. For lab or prototype systems, the Administration & Data Server can be installed on the same VM as the CallRouter and Logger.

The AW acts as the authentication server for Cisco Finesse. In a Finesse deployment, the AW is mandatory and must run in high-availability mode (both a primary and backup AW).

For information about data storage in virtualized deployments, see the *Virtualization for Unified CCE DocWiki* at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE.

Administration Server and Administration Client

The Administration Server, Configuration-Only Administration Server, and Administration Client provide a Configuration Manager tool that is used to configure Unified CCE. The configuration options include, for example, the ability to add agents, add skill groups, assign agents to skill groups, add dialed numbers, add call types, assign dialed numbers to call types, or assign call types to routing scripts.

The Administration Server and Administration Client also have the tool Script Editor, which is used to build routing scripts. Routing scripts specify how to route and queue a contact (that is, the script identifies which skill group or agent handles a particular contact).

Precision Routing is available in Unified CCE. To configure Precision Routing, you can use a web-based application, Unified CCE Administration. For more information about Precision Routing, see the *Cisco Unified Contact Center Enterprise Features Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-feature-guides-list.html>. For third-party configuration, you can use REST API. For more information about using API to configure Precision Routing, see the *Cisco Unified*

Contact Center Enterprise Developer Reference Guide at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-programming-reference-guides-list.html>.

The Administration Server and Configuration-Only Administration Server also support the following configuration tools:

- Internet Script Editor Server—HTTPS (default protocol) connection for Script Editor clients

For information about these and other configuration tools, see the *Administration Guide for Cisco Unified Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-maintenance-guides-list.html>.

The Administration Server is deployed as part of the Administration and Real-time Data Server, known as AW. AWs are deployed in pairs for fault tolerance. During normal operation, the *primary AW* communicates directly with the Central Controller for configuration data and the *secondary AW* connects to the primary AW for the data. If the primary AW fails, the secondary AW connects to the Central Controller. Both types of AW store the configuration and real-time data in the AW Database, or AWDB. Each AW can be deployed in the same location as, or remote from, the Central Controller. A secondary AW need not be co-located with the primary AW.

**Note**

The Unified Contact Center Management Portal (Unified CCMP) and the Unified CCE Administration web tool require a connection with the primary AW. If you connect with the secondary AW, you see errors when saving configuration changes.

Multiple Administration Clients can be deployed and connected to either primary or secondary AWs. An Administration Client must be geographically local to its AW.

**Note**

Administration Clients and Administration Workstations can support remote desktop access. But, only one user can access a client or workstation at a time. Unified CCE does not support simultaneous access by several users on the same client or workstation.

Configuration Only Administration Servers are the same as AWs, but without the real-time data. As such, Administration Clients cannot connect to them and they cannot display real-time data in Script Editor.

Figure 4: Communication Between Central Controller and Administration & Data Server

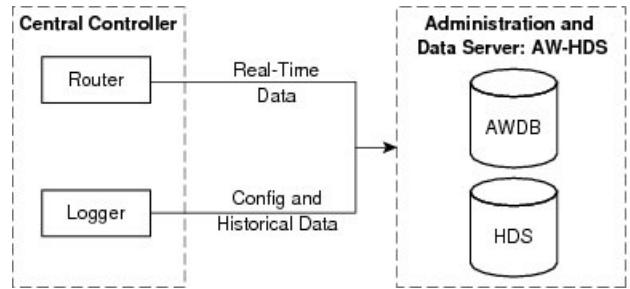
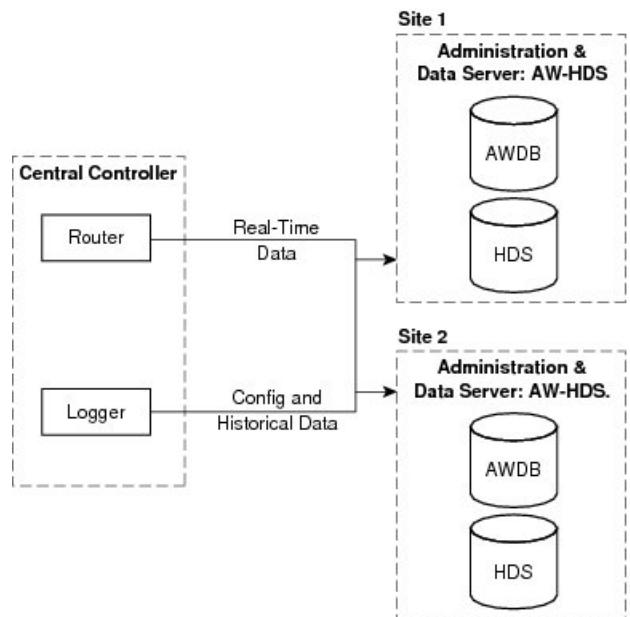


Figure 5: Communication Between Central Controller and multiple Administration & Data Servers



AWs, Configuration-Only Administration Servers, and Administration Clients may operate only as a single instance on a given VM. In a hosted environment, multiple instances may be installed and configured and the Select Administration Instance tool may be used to switch between the instances.

Real-time Data Server

The Real-Time Data Server portion of the AW uses the AW database to store real-time data along with the configuration data. Real-time reports combine these two types of data to present a near-current transient snapshot of the system.

Historical Data Server And Detail Data Server

The Historical Data Server (HDS) and Detail Data Server (DDS) are used for longer-term historical data storage. The HDS stores historical data summarized in 15 or 30 minute intervals and is used for reporting. DDS stores detailed information about each call or call segment and is used for call tracing. Data may be extracted from either of these sources for warehousing and custom reporting.

Typically these Data Servers are deployed with a primary AW as a single server serving all three roles (AW-HDS-DDS). In very large deployments, it might be desirable to separate them for scalability.

Unified CCE Reporting

The Unified CCE Reporting solution provides an interface to access data describing the historical and real-time states of the system.

The reporting solution consists of the following components:

- Cisco Unified Intelligent Center—Reporting user interfaces
- Configuration and Reporting Data—Contained on one or more Administration & Data Servers



Note

Reporting concepts and data descriptions are described in the *Reporting Concepts for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>. (This description is independent of the reporting user interface being used.)

Cisco Unified Intelligence Center

Cisco Unified Intelligence Center (Unified Intelligence Center) is an advanced reporting product used for Unified CCE and other products. This platform is a web-based application offering many Web 2.0 features, high scalability, performance, and advanced features such as the ability to integrate data from other Cisco Unified Communications products or third-party data sources. Unified Intelligence Center incorporates a security model that defines different access and capabilities for specific users.

Unified Intelligence Center Standard is included with Unified CCE. Unified Intelligence Center Premium is an optional product with additional features. See the *Cisco Unified Intelligence Center User Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-intelligence-center/products-user-guide-list.html>. Install Unified Intelligence Center on a separate VM; it cannot be co-resident with other Unified CCE components.

Unified Contact Center Management Portal

The Unified Contact Center Management Portal provides a simple-to-use web-based user interface to streamline the day-to-day provisioning and configuration operations performed by a contact center manager, team lead, or administrator. The management portal provides the following key benefits:

- Simple-to-use web user interface for performing basic tasks such as moving, adding, or modifying phones, agents, skill groups, and teams and other common contact center administrative functions for an IP contact center
- Unified Configuration; that is, tenant provisioning of both the applicable IP contact center elements and the Cisco Unified Communications Manager components through a single task-based web interface
- Partitioned System supporting multiple business units with complete autonomy
- Hierarchical Administration supporting multiple business-level users where each user is defined with specific roles and responsibilities
- Audit Trail Reports that detail configuration changes and usage by all users of the management portal

Related Topics

[Cisco Unified Contact Center Management Portal, on page 269](#)

JTAPI Communications

For JTAPI communications to occur between Unified Communications Manager and external applications such as Unified CCE and Unified IP IVR, configure a JTAPI user ID and password with Unified Communications Manager. When the Unified Communications Manager PIM or Unified IP IVR starts, they use the JTAPI user ID and password to sign in to the cluster. The Unified Communications Manager PIM or Unified IP IVR application sign-in process establishes JTAPI communications between the cluster and the application. Each Unified IP IVR server requires a separate JTAPI user ID. A Unified CCE deployment with one cluster and two Unified IP IVRs requires three JTAPI user IDs: one for Unified CCE and two for the two Unified IP IVRs. Use one PG user for each PG pair.

The Unified Communications software includes a module called the CTI Manager. The CTI Manager communicates through JTAPI to applications such as Unified CCE and Unified IP IVR. Every subscriber within a cluster can execute an instance of the CTI Manager process. But, the Unified Communications Manager PIM on the Agent PG communicates with only one CTI Manager (and thus one node) in the cluster. The CTI Manager process passes CTI messages to and from other nodes within the cluster.

For example, a deployment uses subscriber 1 to connect to a Voice Gateway (VG) and uses subscriber 2 to communicate with Unified CCE through the CTI Manager. When a call for the contact center arrives at the VG, subscriber 1 sends an intra-cluster message to subscriber 2. Subscriber 2 sends a route request to Unified CCE to determine how to route the call.

Each Unified IP IVR also communicates with only one CTI Manager within the cluster. The PIM and the two Unified IP IVRs from the example can communicate with different CTI Managers or they can all communicate with the same CTI Manager. However, they each use a different user ID. The user ID is how the CTI Manager tracks the different applications.

When the PIM is redundant, only one side is active and in communication with the cluster. The PIM on Agent PG A communicates with the CTI Manager on another subscriber. Unified IP IVR is not deployed in redundant pairs. But, if its primary CTI Manager is out of service, Unified IP IVR can fail over to another CTI Manager within the cluster.

The JTAPI communications between the cluster and Unified CCE include three distinct types of messaging:

- Routing control
Routing control messages provide a way for the cluster to request routing instructions from Unified CCE.

- Device and call monitoring

Device monitoring messages provide a way for the cluster to notify Unified CCE about state changes of a device (phone) or a call.

- Device and call control

Device control messages provide a way for the cluster to receive instructions from Unified CCE on how to control a device (phone) or a call.

A typical Unified CCE call includes all three types of JTAPI communications within a few seconds. When a new call arrives, Unified Communications Manager requests routing instructions from Unified CCE. For example, when a subscriber receives the routing response from Unified CCE, the subscriber sends the call to an agent phone. The subscriber notifies Unified CCE that the phone is ringing. That notification enables the answer button on the agent desktop. When the agent clicks the answer button, Unified CCE instructs the subscriber to make the phone go off-hook and answer the call. In order for the routing control communication to occur, subscriber needs a CTI Route Point. You associate a CTI Route Point with a specific JTAPI user ID. Through this association, the subscriber knows which application provides routing control for that CTI Route Point. Dialed Numbers (DNs) are then associated with the CTI Route Point. Then, the subscriber can generate a route request to Unified CCE when a new call to that DN arrives.



Note

You cannot use the DN for a CTI Route Point on a different CTI Route Point in another partition. Ensure that DNs are unique across all CTI Route Points on all partitions.

To monitor and control the phones, associate the phones with a JTAPI user ID in Unified Communications Manager. When you use Extension Mobility or Extension Mobility Cross Cluster, you can associate an Extension Mobility device profile instead. In a Unified CCE environment, you associate the IP phones or the corresponding Extension Mobility device profiles with Unified CCE JTAPI user IDs. When an agent desktop signs in, the PIM requests a subscriber to allow the PIM to begin monitoring and controlling that phone. Until the agent signs in, the subscriber does not allow Unified CCE to monitor or control that phone. If the device or the corresponding Extension Mobility device profile is not associated with a Unified CCE JTAPI user ID, then the agent sign-in request fails.

Using Extension Mobility Cross Cluster (EMCC), when a Unified CCE PIM phone registers to the local cluster after Extension Mobility sign in, the phone looks like an agent situated across a WAN. The Unified CCE peripheral manages the agent devices based on the Extension Mobility profile rather than on a phone device in the Application User on the cluster. For more information, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-implementation-design-guides-list.html>.

You can associate Extension Mobility devices using two methods; either by device or by user profile. Associate the Extension Mobility profile to the CCE Application User on Unified Communications Manager.

Configuring the EM Profile, instead of the device, provides more flexibility in which phones agents can use in the call center. Configuring the phone device limits which devices the agents can use. The option that you use in a contact center depends on the customer business case.

Because Unified IP IVR also communicates with the cluster using the JTAPI protocol, the same three types of communications also occur with Unified IP IVR. Unlike Unified CCE, the Unified IP IVR provides both the application itself and the devices being monitored and controlled.

The devices that Unified CCE monitors and controls are the physical phones. The Unified IP IVR does not have physical ports like a traditional VRU. Unified IP IVR ports are logical ports called CTI Ports. For each CTI Port on Unified IP IVR, there needs to be a CTI Port device defined in Unified Communications Manager.

Unlike a traditional PBX or telephony switch, Unified Communications Manager does not select the Unified IP IVR port to which it sends the call. When a call is made to a DN that is associated through a CTI Route Point with a Unified IP IVR JTAPI user, the subscriber asks the Unified IP IVR which CTI Port handles the call. If Unified IP IVR has an available CTI Port, Unified IP IVR responds to the routing control request with the device identifier of the CTI Port to handle that call.

SIP sends Dual Tone Multi-Frequency (DTMF) digits, however Unified IP IVR and Unified Communications Manager only support out-of-band DTMF digits. JTAPI messages from the cluster notify Unified IP IVR of caller-entered DTMF digits. The cluster uses an MTP resource to convert in-band signaling to out-of-band signaling. CTI ports only support out-of-band DTMF digits. If your deployment includes SIP phones or gateways, provision sufficient MTP resources to support the conversion. The Mobile Agent feature also requires extra MTP resources for this conversion.

The following scenarios are examples of Unified IP IVR device and call control. When an available CTI Port is allocated to the call, a Unified IP IVR workflow starts within Unified IP IVR. When the workflow executes the accept step, a JTAPI message is sent to the subscriber to answer the call for that CTI Port. When the Unified IP IVR workflow wants the call transferred or released, the workflow again instructs the subscriber on what to do with that call.

When a caller releases the call while interacting with the Unified IP IVR, the VG detects the caller release. The VG notifies the subscriber with the Media Gateway Control Protocol (MGCP), which then notifies the Unified IP IVR with JTAPI. When the VG detects DTMF tones, the VG notifies the subscriber through H.245 or MGCP, which then notifies the Unified IP IVR through JTAPI.

In order for the CTI Port device control and monitoring to occur, associate the CTI Port devices on Unified Communications Manager with the appropriate Unified IP IVR JTAPI user ID. If you have two 150-port Unified IP IVRs, you have 300 CTI ports. Associate half of the CTI ports with JTAPI user Unified IP IVR 1, and associate the other half of the CTI ports with JTAPI user Unified IP IVR 2.

While you can configure Unified Communications Manager to route calls to Unified IP IVRs on its own, Unified CCE routes calls to the Unified IP IVRs in a Unified CCE environment (even if you have only one Unified IP IVR and all calls require an initial VRU treatment). Doing so ensures proper Unified CCE reporting. For deployments with multiple Unified IP IVRs, this routing practice also allows Unified CCE to load-balance calls across the multiple Unified IP IVRs.

Related Topics

[Design Considerations for High Availability, on page 59](#)

Multichannel Subsystems: EIM and WIM

Unified CCE has the capability to provide a multichannel contact center with E-mail Interaction Manager (EIM) and Web Interaction Manager (WIM).

For design information about these products, see the *Cisco Unified Web and E-Mail Interaction Manager Solution Reference Network Design Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-email-interaction-manager/products-implementation-design-guides-list.html>.

Cisco Outbound Option

Agents can handle both inbound and outbound contacts, which helps in optimizing contact center resources. The Cisco Outbound Option enables the multifunctional contact center to take advantage of Cisco Unified

CCE enterprise management. Contact center managers in need of outbound campaign solutions can take advantage of the enterprise view that Cisco Unified CCE maintains over agent resources.

Related Topics

[Cisco Outbound Option for Unified CCE, on page 156](#)

Cisco Unified Mobile Agent

Cisco Unified CCE provides the capability for an agent to use any PSTN phone and a quality high-speed data connection between the agent desktop and the CTI OS server.

Related Topics

[Cisco Unified Mobile Agent, on page 171](#)

Serviceability

Diagnostic Tools

Unified CCE has a built-in web-based (REST-like) interface for diagnostics called the Diagnostic Framework, which is resident on every Unified CCE server. The Analysis Manager functionality integrated with the Unified Communications Manager Real-Time Monitoring Tool (RTMT) is provided as the client-side tool to collect diagnostic information from this diagnostic framework. In addition to the Analysis Manager, a command line interface—Unified System CLI tool—is available that allows a client to access the diagnostic framework on any Unified Communications server. (A user need not use Remote Desktop first to gain access to use the CLI.)

Using the Analysis Manager, the administrator connects to one or more Unified Communications devices to set trace levels, collect trace and log files, and gather platform and application configuration data as well as version and license information. The Analysis Manager is the one tool that allows administrators to collect diagnostic information from all Cisco Unified Communications applications and devices.

The Analysis Manager offers local user and domain security for authentication and secure HTTP to protect data exchanged by it and the diagnostic framework.

For more information about the Unified CCE Diagnostic Framework (that runs on every Unified CCE server), see the *Serviceability Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

Network Management Tools

Unified CCE is managed using the Simple Network Management Protocol (SNMP). Unified CCE devices have a built-in SNMP agent infrastructure that supports SNMP v1, v2c, and v3 and it exposes instrumentation defined by the CISCO-CONTACT-CENTER-APPS-MIB. This MIB provides configuration, discovery, and health instrumentation that can be monitored by standard SNMP management stations. Unified CCE provides a rich set of SNMP notifications that alerts administrators of any faults in the system. Unified CCE also provides a standard syslog event feed (conforming to RFC 3164) for those administrators who want to take advantage of a more verbose set of events.

For managing a Unified Communications deployment, customers are encouraged to use the Cisco Unified Operations Manager (Unified Operations Manager) product. Unified Operations Manager is a member of the Cisco Unified Communications family of products and provides a comprehensive and efficient solution for network management, provisioning, and monitoring of Cisco Unified Communications deployments.

Unified Operations Manager monitors and evaluates the current status of both the IP communications infrastructure and the underlying transport infrastructure in your network. Unified Operations Manager uses open interfaces such as SNMP and HTTP to remotely poll data from different devices in the IP communications deployment. In addition to the infrastructure, Unified Operations Manager offers capabilities specific to Unified Communications applications including Unified CCE. For more information, see the Unified Operations Manager documentation at <http://www.cisco.com/c/en/us/support/cloud-systems-management/prime-unified-operations-manager/tsd-products-support-series-home.html>.

For more information about configuring the Unified CCE SNMP agent infrastructure and the syslog feed, see the *SNMP Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

For details on the health monitoring and management capabilities of Unified CCE, review the *Serviceability Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

Combining IP Telephony and Unified CCE in the Same Unified Communications Manager Cluster

It is possible for a Unified Communications Manager cluster to support Cisco Unified IP phones with both normal IP Telephony (office) extensions and Unified CCE (call center) extensions. When running dual-use Unified Communications Manager clusters with both IP Telephony and Unified CCE extensions, it is important to ensure all elements of the solution are compatible, as documented in the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.

It is also important to note that many contact center environments have very stringent maintenance windows. Unified CCE agents process far more calls than typical administrator phone users in a Unified Communications Manager cluster, so their device weight (or the amount of processing power required per agent) is higher than a typical business phone user. For example, an administrator-only cluster might be able to support 20,000 phones, but a Unified CCE cluster might support only a fraction of those as agents because of the higher call volume and messaging that Unified Communications Manager is required to maintain to support those agents.

Because of these software and environmental limitations, it might sometimes be advantageous to separate the Unified Communications Manager clusters for IP Telephony extensions from the Unified Communications Manager clusters for Unified CCE extensions. It is important to consider the environment where Unified CCE is being deployed to determine whether a separate Unified Communications Manager cluster is advantageous.

Combining IP Telephony and Unified CCE Extensions on the Same IP Phone

Unified CCE supports only one agent ACD line on the IP phone, which does not have voice-mail or call forwarding defined so that Unified CCE can manage and control all calls sent to the agent on this line. Typically, the agent extension is not used as the agent's DID or personal line. A separate line can be assigned to the agent's phone for that purpose and can be configured with voice-mail and other calling features.

The position of the line on the phone determines which line is answered or used if the agent just picks up the handset. In a typical call center, the ACD line is the first line on the phone to make it easier for the agent to answer inbound ACD calls and ensure that calls the agent makes using the phone are tracked by the system as external calls. The agent's state changes based on the activity on this line. If the agent picks up the phone to place a call, the agent is put into not-ready mode and the Unified CCE does not route a call to that agent.

In some cases, the agents are knowledge workers or they do not take as many ACD calls as they do normal extension calls. The call center manager does not track their phone activity that is not ACD related, and it might be inconvenient for those users to always get the ACD line first when they want to pick up a DID call instead. In this case, the order of the lines might best be reversed, placing the ACD line on the last (or bottom) line appearance on the phone and placing the DID or normal extension on the first line on the phone. This arrangement allows users to pick up the phone and answer the first line as well as use this line for all calls they place. To answer an ACD call, they have to select that line on the phone or use the agent desktop to answer that line appearance directly. Also, be aware that they have to manage their agent state and go into not-ready mode manually when they want to place a call on their normal extension so that Unified CCE does not attempt to route a call to them while they are on the other line.

It is possible to have a deployment where the agent extension is the same as the agent's DID or personal line. When call waiting is configured on the agent phone, agent-to-agent calls can interrupt a customer call. To prevent this from happening, agent-to-agent routing can be used and the agent-to-agent routing script can be set up to queue or reject the call if the agent is busy. This is a good option if there is a need to see all agent activity and to avoid all interruptions for the agent. The configuration involves using CTI Route Points in Unified Communications Manager instead of the agent DID to send the calls to Unified CCE for agent-to-agent routing. For ease of configuration and to reduce the number of CTI route points, the Unified Communications Manager wildcard feature can be used, although Unified CCE requires distinct routing DNs (one for each agent).



Note

You cannot use the DN for a CTI Route Point on a different CTI Route Point in another partition. Ensure that DNs are unique across all CTI Route Points on all partitions.

Agent Phones in Countries with Toll-Bypass Regulations

Telecommunications regulations in some countries, such as India, require the voice infrastructure to be partitioned logically into two systems:

- One for Closed User Group (CUG) or Voice over IP (VoIP) to enable communications across the boundaries within an organization
- Another one to access the local PSTN

To comply with such regulations, agents had one line to access customer calls and a different phone for VoIP access to teammates or experts located outside the contact center.

The Logical Partitioning feature in Cisco Unified Communications Manager provides the same capability to control calls and features based on specific allowed or forbidden configurations. A common telephony system in a contact center provides access to both the PSTN and VoIP networks. Therefore, the system requires configurations to provide controlled access and to avoid toll bypass.

You can enable and configure the Logical Partitioning feature to prevent toll-bypass calls. With the feature, agents in a Unified CCE system can use the same phone for receiving customer calls and for making or receiving VoIP calls with other employees. Although this feature eliminates the need for agents to have a

second phone, contact center managers can choose to have a dedicated line or phone for customer calls and can allocate a different line or phone for other calls.

Queuing in a Unified CCE Environment

Call queuing can occur in three distinct scenarios in a contact center:

- New call waiting for handling by initial agent
- Transferred call waiting for handling by a second (or subsequent) agent
- Rerouted call because of ring-no-answer, waiting for handling by an initial or subsequent agent

When planning your Unified CCE deployment, it is important to consider how to handle queuing and requeuing.

Call queuing in a Unified CCE deployment requires use of an VRU platform that supports the SCI interface to Unified CCE. Unified IP IVR is one such platform. Cisco offers another VRU platform, Unified CVP, which you can use as a queuing point for Unified CCE deployments. For more information, see [Deployments, on page 27](#).

In a Unified CCE environment, a VRU provides voice announcements and queuing treatment while waiting for an agent. Unified CCE provides the control over the type of queuing treatment for a call through the SCI interface. The Run VRU Script node in a Unified CCE routing script is the component that causes Unified CCE to instruct the VRU to play a particular queuing treatment.

While the VRU plays the queuing treatment to the caller, Unified CCE waits for an available agent with the skill defined in the routing script. When an agent with the defined skill becomes available, Unified CCE reserves that agent. Unified CCE instructs the VRU to transfer the voice path to that agent phone.

Transfers and Conferences in Unified CCE Environments

Transfers and conferences are commonly used in contact centers. They require special attention to ensure the proper system resources are available and configured correctly. Some points to consider when designing your contact center's handling of transfers and conferences:

- When Call Recording is enabled in the DN configuration for an agent phone, the codec is not renegotiated when establishing a conference. As a result, if two phones are connected using g.722 and a conference call is initiated, the codec is not renegotiated to g.711 and a hardware conference bridge or transcoder is required.
- To maintain call context during single-step transfers, route to agents or configured routing DNs on the same peripheral.
- To execute a network transfer from Unified CVP, configure a routing DN on the same peripheral.

Related Topics

[Deployments, on page 27](#)

Dial Plans

Whatever your deployment model for the Unified CCE, it is always helpful to have a dial plan before you begin. The Dial Plan you use is associated with telephone networks and dialing patterns. Many books about dial plans are available. Although written for a different Cisco product, "[Implementing Cisco Unified Communications: Introducing Dial Plans](#)" provides an introduction to the concepts and shows the necessity of having a dial plan.

The following section provides a simple, sample dial/configuration plan that you can use as a model.

Sample Unified CCE with Unified CVP Dial Plan

The following table provides a sample dial plan for a Unified CCE system that uses the Unified CVP as its VRU application for queuing.


Note

The dots in some of the values shown represent data that you enter when you configure the gateway. For example, in the value "55551291.." the dots represent any number from 00 through 99.

Dial Peer	Extension	Destination-Pattern or Incoming called number	Type or Service	Session Target
Voice Gateway				
51291		55551291..	VoIP	Unified CVP
512919		5555129199T	bootstrap	
Unified CVP				
Dialed Number	Extension	Destination	Type	Target
5129199>	5555129199	5555129199 <correlation ID>	VRU label	Voice Gateway
51291>	55551291[00 - 99]	55551291[00 - 99]	CVP Route Point	Unified CVP
512>	55512....	555512....	Device	Unified Communications Manager
Unified CM (Route Pattern)				
Dialed Number	Extension	Destination	Type	Target

5129101!	5555129101	5555129101 <correlation ID>	Unified Communications Manager Label	Unified CVP
	1[000 - 9999]	5555121[000 - 999]	Agent Extensions	Unified Communications Manager
	2[000 - 9999]	2[000 - 999]	Route Points	Unified Communications Manager



Deployments

- [Unified CCE Base Model , page 27](#)
- [Enterprise Unified CCE Peripheral, page 33](#)
- [Unified CCE Administration and Data Server, page 33](#)
- [Agent Type Deployment Scenarios, page 37](#)
- [Centralized Data Center, page 54](#)
- [Geographically Redundant Data Centers, page 55](#)

Unified CCE Base Model

Cisco Unified Contact Center Enterprise (Unified CCE) is a solution that delivers intelligent call routing, network-to-desktop Computer Telephony Integration (CTI), and multichannel contact management over an IP network to contact center agents. Unified CCE adds software to create an IP automatic call distribution (ACD) onto a Cisco Unified Communications framework. This unified solution allows companies to rapidly deploy an advanced, distributed contact center infrastructure.

You can configure Unified CCE to sort customer contacts. Unified CCE monitors resource availability and delivers each contact to the most appropriate resource in the enterprise. The system profiles each customer contact using related data such as dialed number and calling line ID, caller-entered digits, data submitted on a web form, and information obtained from a customer database lookup. Simultaneously, the system monitors the resources available in the contact center to meet customer needs, including agent skills and availability, voice-response-unit (VRU) status, and queue lengths.

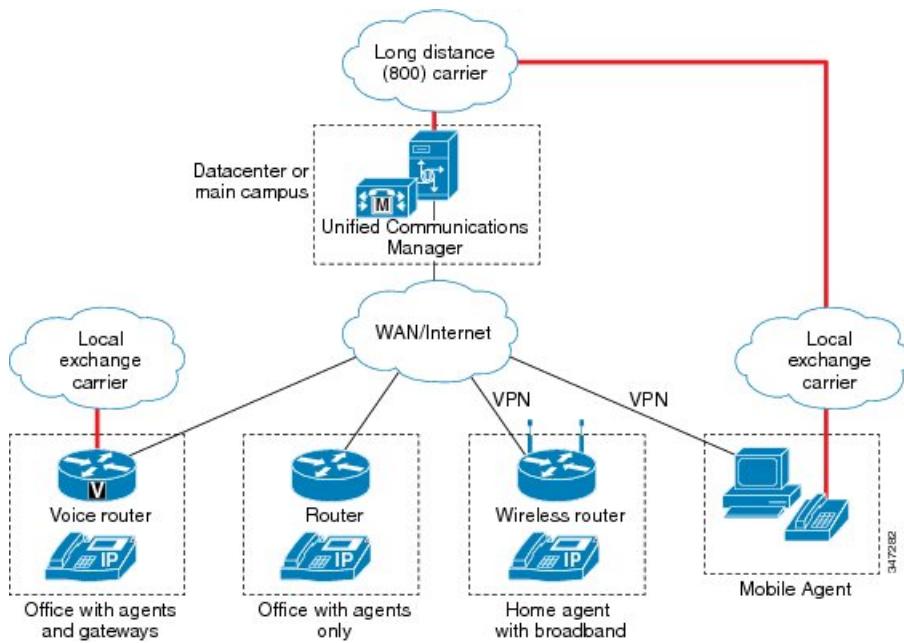
Unified CCE allows you to smoothly integrate inbound and outbound voice applications with Internet applications such as real-time chat, web collaboration, and email. This integration enables a single agent to support multiple interactions simultaneously regardless of which communications channel the customer chooses.

The Unified CCE base model includes a common set of features that apply across supported Unified CCE models.

Unified CCE Base Model Architecture

The following figure shows the logical view of the Unified CCE base model. Agents that are local to the data center are not shown.

Figure 6: Unified CCE Base Model—Logical View



Unified CCE Base Model Components

The Unified CCE base model includes multiple components. The data center hosts the main solution components: routing services, call control, voice response unit (VRU), and reporting.

Figure 7: Data Center Components—Logical View

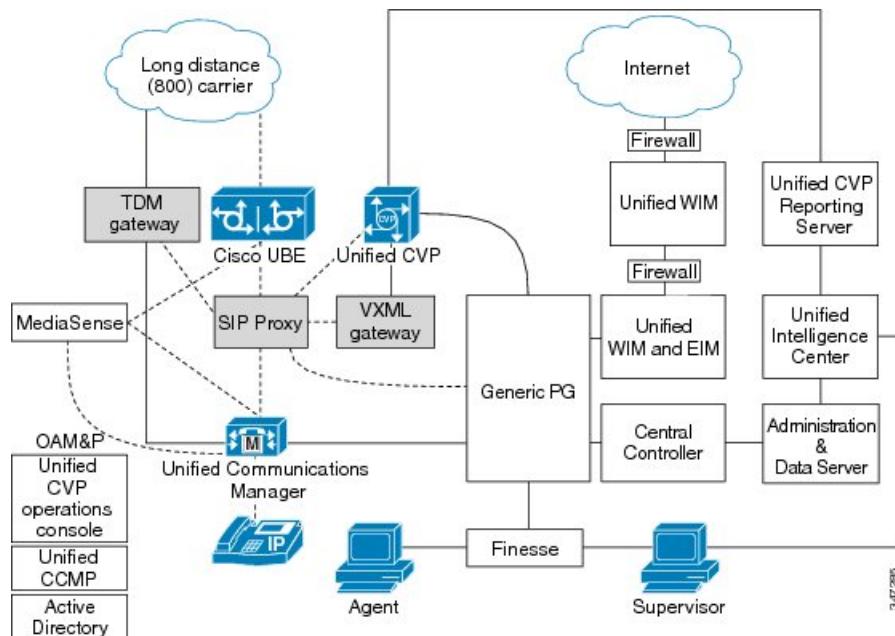


Table 2: Unified CCE Base Model Components

Component	Description
Unified CCE Call Router	Makes all routing decisions on how to route a call or customer contact. Often referred to as the <i>Router</i> in the context of Unified CCE components. The Router is a part of the Central Controller.
Unified CCE Logger	The database server that stores contact center configuration data and temporarily stores historical reporting data for distribution to the data servers. The Logger is a part of the Central Controller.
Administration & Data Server	Configuration interface and real-time and historical data storage (for example, for reporting).

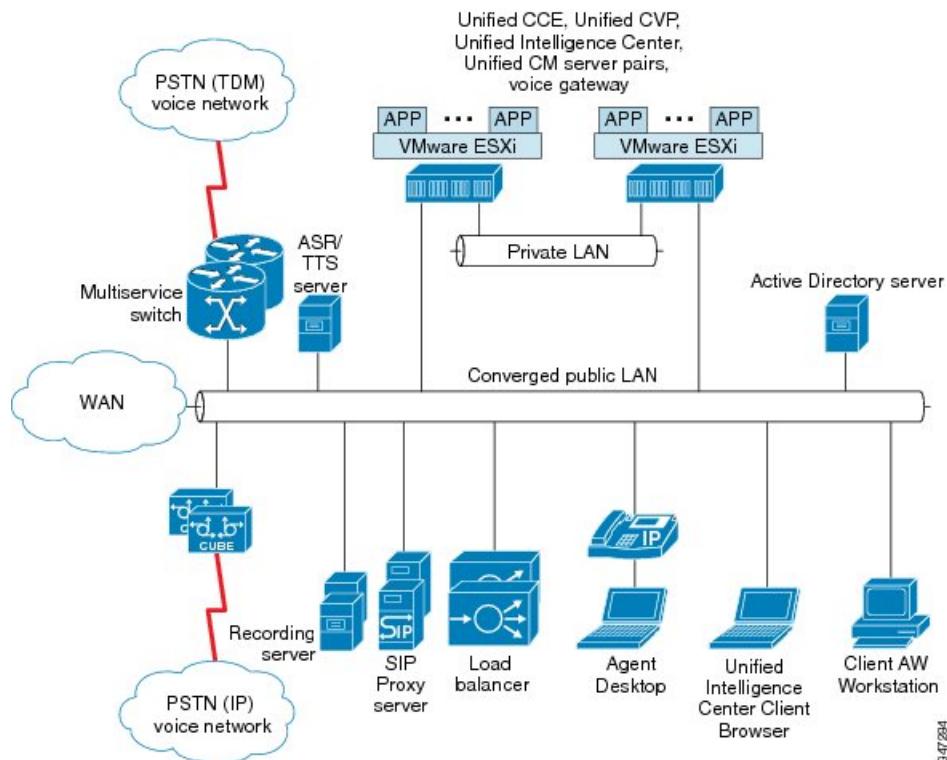
Component	Description
Peripheral Gateway (PG)	The PG is an interface to various peripheral devices, specifically to Unified Communications Manager, VRU (Unified Customer Voice Portal or Unified IP IVR), or Multichannel products (Unified Web Interaction Manager and Unified E-Mail Interaction Manager) for email and chat.
Desktop Server	Colocated with Unified CCE PG. Desktop choices: Finesse; CTI OS; Cisco Agent Desktop.
Unified Intelligence Center	Provides web browser-based real-time and historical reporting. Unified Intelligence Center also works with other Cisco Unified Communications products.
Unified Communications Manager	Unified Communications Manager is the call-processing component of the Cisco Unified Communications System. It extends enterprise telephony features and capabilities to IP phones, media processing devices, VoIP gateways, mobile devices, and multimedia applications.
Cisco Voice Gateways	A voice gateway provides the connection between the PSTN and the Unified Communications Manager framework.

Table 3: Unified CCE Optional Components

Optional component	Description
Cisco Unified Contact Center Management Portal (Unified CCMP)	Unified CCMP is a browser-based management application designed for contact center system administrators, business users, and supervisors. Unified CCMP is a dense multitenant provisioning platform that overlays the Unified CCE, Unified Communications Manager, and Unified CVP equipment.
Automatic Speech Recognition and Text-to-Speech (ASR/TTS)	ASR allows callers to speak words or phrases to choose menu options. TTS converts plain text (UNICODE) into speech.

Optional component	Description
Multichannel (Unified WIM and EIM)	<p>Cisco Unified E-Mail Interaction Manager (Unified EIM) allows organizations to intelligently route and process inbound emails, webform inquiries, faxes, and letters.</p> <p>Cisco Unified Web Interaction Manager (Unified WIM) provides agents with a comprehensive set of tools for serving customers in real time. It allows call center agents to provide immediate personalized service to customers through text chat messaging and page-push abilities. Agents can also use Unified WIM to assist customers using web chat.</p>

Figure 8: Data Center Components—Physical View



Unified CCE Base Model Design Requirements

Unified CCE is designed for contact centers that serve business-critical functions. This requirement means that the Unified CCE base model builds in high availability.

The Unified CCE base model design require the following:

- Unified CCE deployments must be redundant

- Unified Communications Manager subscriber nodes 1:1 redundancy
- Unified CVP N + 1 redundancy, or 1:1 for geographic redundancy
- Voice gateways N + 1 for geographic redundancy

The following information is not provided in the Unified CCE base model:

<ul style="list-style-type: none"> • Data infrastructure for the LAN • Type of voice gateways • Number of voice gateways and trunks 	See <i>Cisco Campus Design Guides</i> at http://www.cisco.com/en/US/netsol/ns815/networking_solutions_program_home.html and <i>Cisco Collaboration System Solution Reference Network Designs</i> at http://www.cisco.com/go/uucsnd
<ul style="list-style-type: none"> • Number of Unified Communications Manager servers and number and type of OVA • Number of Unified CVP servers 	See the <i>Virtualization for Unified CCE DocWiki</i> at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE to determine the number and type of required OVAs.
Amount of digital signal processor (DSP) resources required to place calls on hold, complete consultative transfers, and conferences.	See <i>Cisco Collaboration System Solution Reference Network Designs</i> to size these resources.



Note

Use 2-vCPU OVAs for Unified Communications Manager servers. Do not use 1-vCPU OVAs for Unified Communications Manager in an Unified CCE deployment.

Related Topics

[Design Considerations for High Availability, on page 59](#)

Unified CCE Base Model Variations

Unified CCE has two base model variations. One variation connects the voice gateways to the line side of a PBX instead of the public switched telephone network (PSTN). Another variation connects to multiple PSTNs and a PBX from the same single site. For example, a deployment can have trunks from a local PSTN, a toll-free PSTN, and a traditional PBX/ACD.

These variations do not specify the following:

- Type of signaling (for example, ISDN, multifrequency signaling, R1) to use between the PSTN and the voice gateway.
- Signaling (SIP or MGCP) to use between the voice gateway and Unified Communications Manager.

Related Topics

[ACD Integration and Parent/Child Deployments, on page 346](#)

[Traditional VRU Integration, on page 346](#)

Enterprise Unified CCE Peripheral

In Enterprise Unified CCE peripheral deployments, the Unified CCE software treats the VRU and Unified Communications Manager as separate peripherals. This means that you must route once for each peripheral. When a call touches the peripheral, Termination Call Detail records are created for each peripheral. When moving calls between peripherals, you must use translation routes.

You can deploy the Unified Communications Manager PG and VRU PG independently, or you can deploy the Unified Communications Manager and VRU in a Generic PG with separate PIMs for Unified Communications Manager and VRU.

These deployments provide a large degree of configuration flexibility. For example, this deployment uses either a Unified CVP or a Unified IP IVR attached to the VRU peripheral. You can load balance between multiple VRUs.

Unified CCE Administration and Data Server

Several Administration & Data Server deployments with different roles are available. The deployments have varying functionality and can handle varying amounts of reporting data.

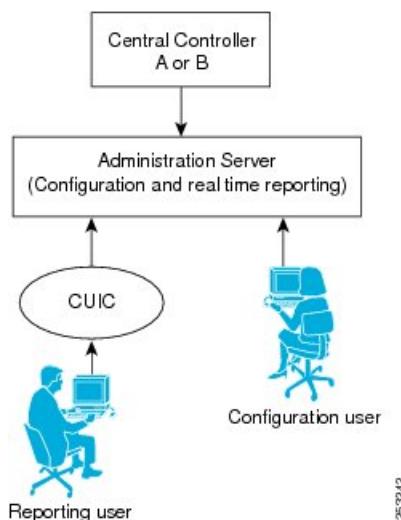
Roles

The Administration & Data Servers are classified into the following roles based on the system configuration and the call load that it can handle:

Administration Server (Configuration and Real-Time Reporting)

This role is similar to the former Distributor AW model which provides the capability for configuration changes as well as real-time reporting. The real-time reporting is supported using Cisco Unified Intelligent Center (Reporting client). No historical reporting is supported.

Figure 9: Configuration and Real-Time Reporting

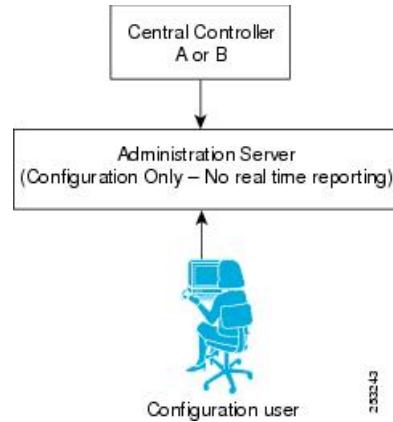


Administration Server (Configuration-Only AW)

In this Administration & Data Server deployment role, the HDS is not enabled and real-time reporting is turned OFF. This distributor deployment provides the capability for configuration changes only. No real-time and historical reporting is supported.

This deployment role allows CCMP to configure a specific Unified CCE Customer Instance. The load is low enough on such a lightweight Administration & Data Server that a single server is sufficient if deployed using VMWare.

Figure 10: Configuration-Only AW



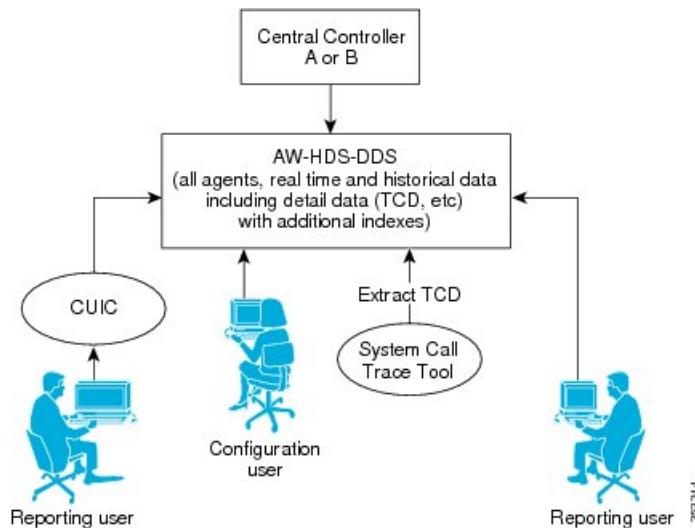
Administration Server, Historical Data Server, and Detail Data Server (AW-HDS-DDS)

This Administration & Data Server deployment role is similar to the existing Distributor AW with HDS model which provides the capability for configuration changes as well as both real-time and historical reporting. The real-time and historical reporting is supported using Cisco Unified Intelligence Center (Unified Intelligence Center Reporting client). Call detail and call variable data are supported for custom reporting data extraction to feed historical data.



Note Unified Intelligence Center is not part of the out-of-the-box solution.

Figure 11: Administration Server, Historical Data Server, and Detail Data Server (AW-HDS-DDS)



Administration Server And Historical Data Server (AW-HDS)

This Administration & Data Server deployment role provides the capability for configuration changes as well as both real-time and historical reporting. Real-time and historical reporting are supported using Cisco Unified Intelligence Center Reporting user.

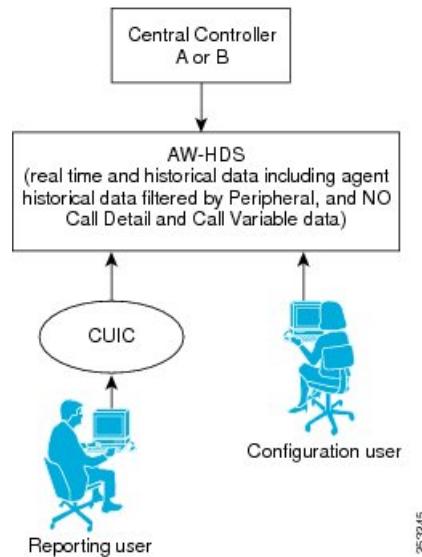


Note Cisco Unified Intelligence Center is not part of the out-of-the-box solution.

In addition, the following features are disabled and not supported:

- Call Detail, Call Variable, and Agent State Trace data
- Custom reporting data extraction

Figure 12: Administration Server And Historical Data Server (AW-HDS)



Historical Data Server and Detail Data Server (HDS-DDS)

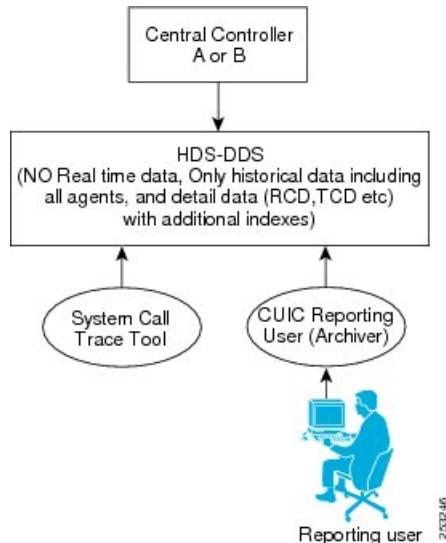
The HDS-DDS deployment model is used specifically for data extraction and for custom reports for call detail (TCD and RCD) only.

In addition, the following features are disabled and not supported:

- Real-time data reporting
- Ability to make configuration changes

This deployment role is limited to one per Logger side.

Figure 13: Historical Data Server And Detail Data Server (HDS-DDS)



Agent Type Deployment Scenarios

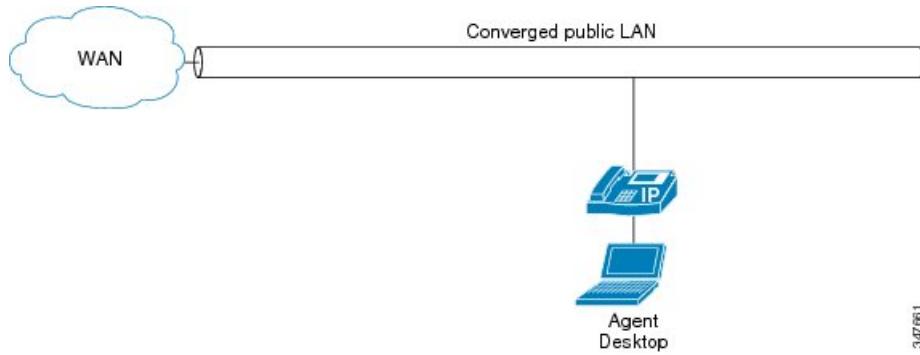
Local Agent

A data center can contain all of the Unified CCE base model components. In the local agent deployment scenario, the agents, supervisors, and administrators are local to the data center.

Local Agent Architecture

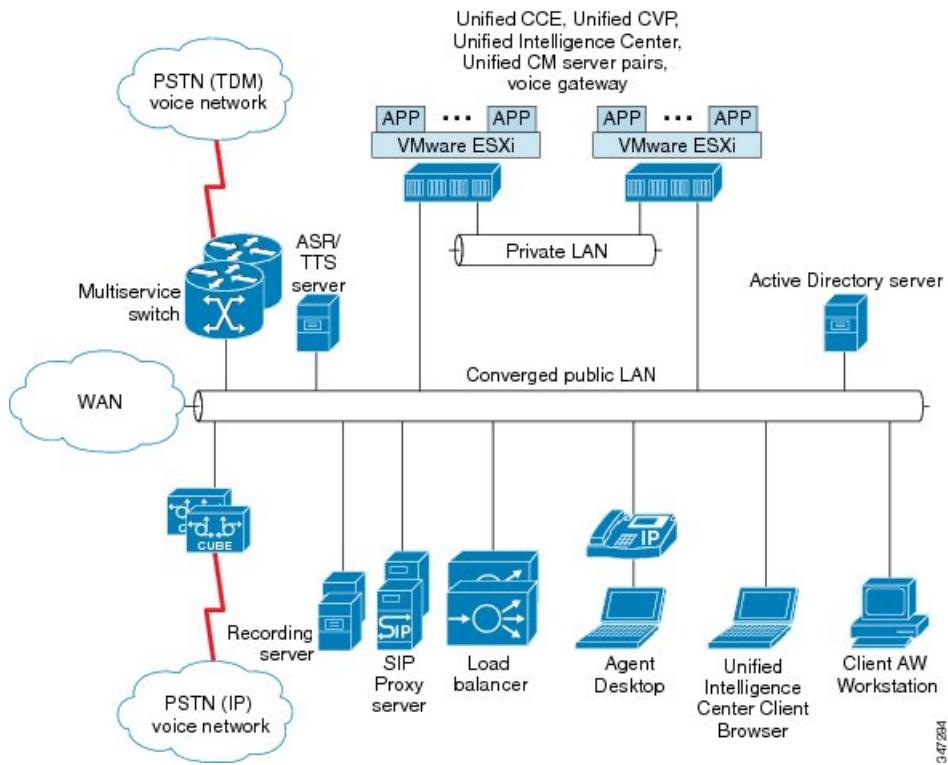
The following figure shows the logical view of a local agent.

Figure 14: Local Agent—Logical View



The following figure shows the physical view of a local agent.

Figure 15: Local Agent—Physical View



Local Agent Components

The local agent deployment scenario includes the following components:

- Unified CCE base model components
- Unified Intelligence Center browser clients for local access to reporting
- Administration tools: Unified CCE configuration tools, Internet Script Editor, or the local Administrative Workstation
- Optional third-party recording server for VoIP capture of agent or customer calls
- Agent phones for Silent Monitoring support:
 - Unified Communications Manager-based
 - CTI OS-based

Local Agent Benefits

The local agent deployment scenario provides the following benefits:

- Does not require location-based call admission control
- Simple codec setup

Local Agent Design Requirements

The following table describes the design requirements for a local agent.

Table 4: Local Agent Design Requirements

	Requirement	Notes
Infrastructure	Location-based call admission control is not required	Local agents use LAN bandwidth, which is typically sufficient for all Unified CCE traffic.
Desktop	Finesse CTI OS Cisco Agent Desktop Customer Relationship Management	
Codec	Transcoding is not required.	If all agents are local to the data center (no required WAN connectivity), you do not need to use G.729 or any other compressed RTP stream.
Recording	Unified Communications Manager-based BIB Cisco MediaSense provides recording using Cisco Unified Border Element media forking	MediaSense provides audio-only call recording.

	Requirement	Notes
Silent Monitoring	Unified Communications Manager-based BIB CTI OS-based	

The following table describes the media resources for a local agent.

Table 5: Local Agent Media Resources

Resource	Method	Notes
Music on Hold	Unicast Unified Communications Manager	
Conference bridges	IP phone with BIB Hardware-based, located at voice gateways	
Media Termination Points	Not supported	
Transcoders	Hardware-based, located at voice gateways	Required for SIP trunks with alaw.

Remote Offices

Remote agent support provides Computer Telephony Integration (CTI), contact distribution, and reporting capabilities to remote agents in branch offices or at home, through either a broadband network connection or their home telephone line. Unified CCE provides identical user interfaces and feature functions to agents regardless of agent location.

The Unified Mobile Agent feature gives the contact center the flexibility to adapt to a fast-moving mobile workforce. Agents can choose their destination phone number during login time and change the number as often as they want. Agents can be on any phone device on any third-party switch infrastructure.

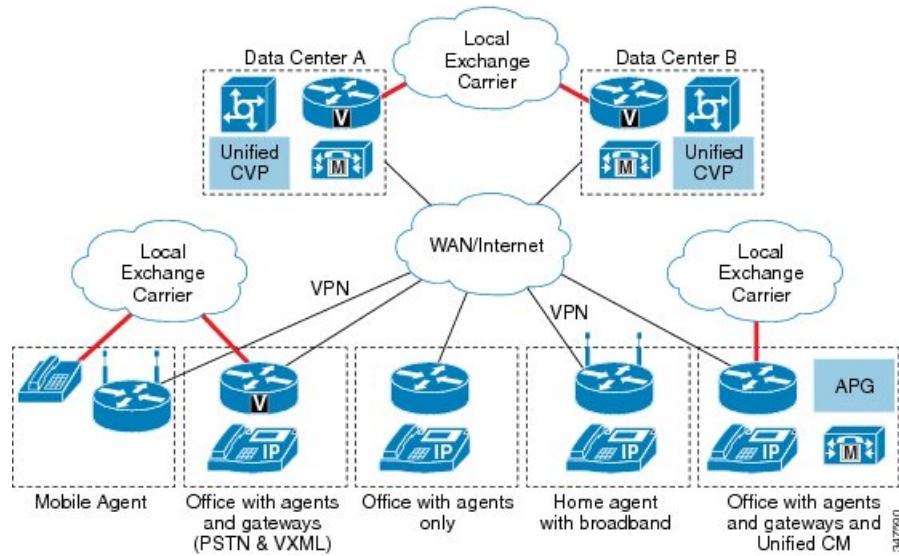
Unified CCE remote office features help companies to better use existing and on-demand resources and fully extend CTI functions across the extended enterprise.

There are several types of agent locations:

- Office with agents only
- Office with agents and voice gateway
- Office with agents, gateways, and Unified Communications Manager
- Home agent with broadband
- Unified Mobile Agent

The following figure shows the agent remote office types.

Figure 16: Agent Remote Offices



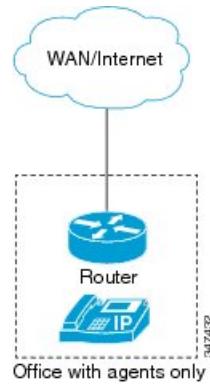
Remote Office with Agents

A remote office with agents is located either at the central office or at a branch office.

Remote Office with Agents Architecture

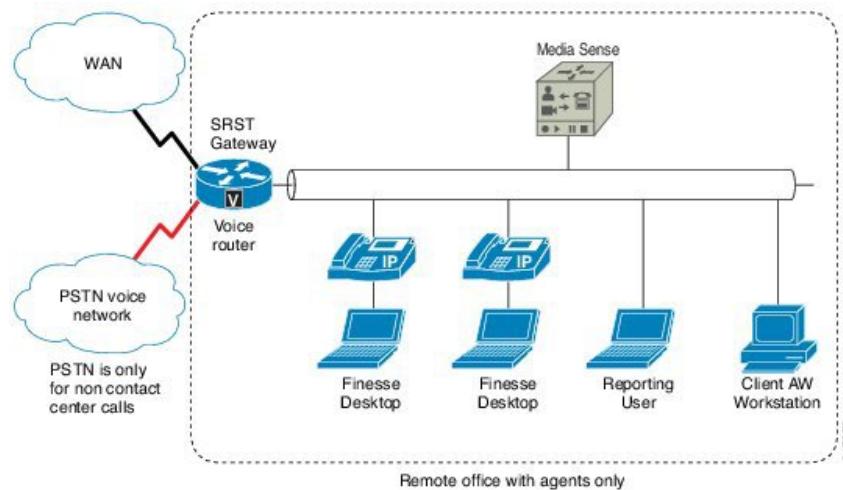
The following figure shows the logical view of a remote office with agents.

Figure 17: Remote Office with Agents—Logical View



The following figure shows the physical view of a remote office with agents.

Figure 18: Remote Office with Agents—Physical View



Remote Office with Agents Components

A remote office with agents includes the following components:

- Unified Intelligence Center browser clients for local access to reporting
- Administration tools: Unified CCE configuration tools, Internet Script Editor, or the local Administrative Workstation
- Optional third-party recording server for VoIP capture of agent or customer calls
- Agent phones with BIB for Unified Communications Manager-based Silent Monitoring support

Remote Office with Agents Benefits

A remote office with agents provides the following benefits:

- Requires only a small data switch and router, IP phones, and agent desktops at remote sites for a few agents.
- Requires only limited system and network management skills at remote sites.
- Small remote sites and offices do not require PSTN trunks, except for local POTS lines for emergency services (911) in the event of a WAN link loss.
- PSTN trunks for incoming traffic connect to data centers for efficiency.
- Unified CCE queue points (Unified CVP or Unified IP IVR) are aggregated for efficiency.
- Does not use VoIP WAN bandwidth while calls queue. Calls extend over the WAN only when an agent is available for the caller.

Remote Office with Agents Design Requirements

The following table describes the design requirements for a remote office with agents.

Table 6: Remote Office with Agents Design Requirements

	Requirement	Notes
Infrastructure	Location-based call admission control	Unified Communications Manager location-based call admission control failure results in a disconnected routed call. You must provision adequate bandwidth to the remote sites and design a Quality of Service WAN.
	Bandwidth	<p>You must plan bandwidth capacity for the following traffic:</p> <ul style="list-style-type: none"> • RTP (caller to agent) • Unified Communications Manager signaling to IP phones • Client desktop to PG (CTI data) • ISE client to ISE server • Administration Client • Unified Intelligence Center client to Unified Intelligence Center server • Silent Monitoring RTP • Recording RTP (if there is no recording server in the remote office) • Music on Hold traffic for calls that are on hold when you use Unified Communications Manager Unicast Music on Hold <p>Note Adequate bandwidth and QoS provisioning are critical for client desktop to PG links.</p>
	Customer contact numbers	Customers might need to dial a long-distance number rather than a local PSTN number to reach the central office. You can offer customers a toll-free number, but the contact center incurs toll-free charges.

	Requirement	Notes
Desktop	Finesse CTI OS desktop Cisco Agent Desktop Customer Relationship Management	
Codec	G.711 or G.729a	G.711 requires significantly more bandwidth.
Recording	BIB	Cisco MediaSense provides audio-only call recording. Audio forking requires Unified Border Element.
Silent Monitoring	Unified Communications Manager-based BIB	

The following table describes the media resources for a remote office with agents.

Table 7: Remote Office with Agents Media Resources

Resource	Method	Notes
Music on Hold	Unicast using Unified Communications Manager	
Conference bridges	Hardware-based, located at voice gateways	Conference bridges use local Unified Survivable Remote Site Telephony (SRST).
Media Termination Points	Hardware-based, located at voice gateways	For Unified Mobile Agents, MTPs are required only at the data center.
Transcoders	Hardware-based, located at voice gateways	Transcoders use local Unified SRST.

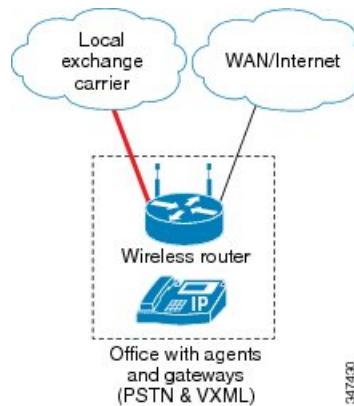
Remote Office with Agents and Voice Gateway

The remote office with agents and voice gateway model is appropriate for a company with many small sites that each require local PSTN trunks for incoming calls. This model provides local PSTN connectivity for local calling and access to local emergency services.

Remote Office with Agents and Voice Gateway Architecture

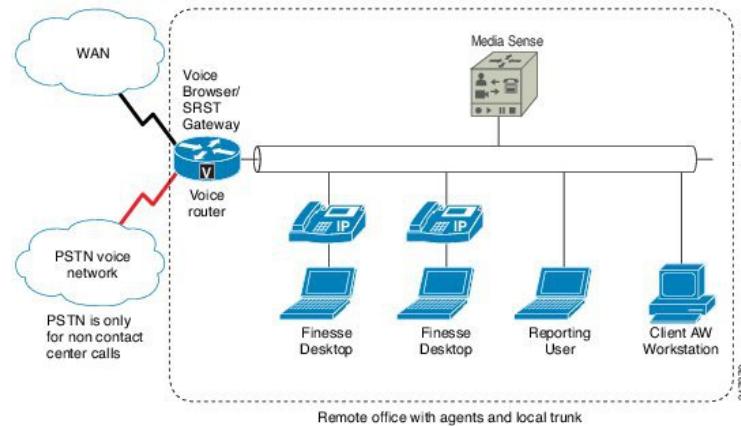
The following figure shows the logical view of a remote office with agents and voice gateway.

Figure 19: Remote Office with Agents and Voice Gateway—Logical View



The following figure shows the physical view of a remote office with agents and voice gateway.

Figure 20: Remote Office with Agents and Voice Gateway—Physical View



Remote Office with Agents and Voice Gateway Components

A remote office with agents and voice gateway includes the following components:

- Integrated Services Router (ISR) voice gateway for ingress voice customer calls under Unified CVP control with local PSTN. Unified SRST backup requires trunks.
- Unified Intelligence Center browser clients for local access to reporting.

- Administration tools: Unified CCMP browser clients, Internet Script Editor, or the local Administrative Workstation.
- Optional third-party recording server for VoIP capture of agent or customer calls.
- Agent phones with BIB for Unified Communications Manager-based Silent Monitoring support.

Remote Office with Agents and Voice Gateway Benefits

A remote office with agents and voice gateway provides the following benefits:

- Requires only limited systems management skills for remote sites because most servers, equipment, and system configurations are managed from a centralized location.
- Does not require WAN RTP traffic for calls that arrive at the remote site and are handled by agents at the remote site.
- Unified CVP uses the VoiceXML browser in Cisco IOS on the voice gateway to provide call treatment and queueing at the remote site. This call treatment and queueing eliminates the need to move the call over the VoIP WAN to a central queue and treatment point.

Remote Office with Agents and Voice Gateway Design Requirements

The following table describes the design requirements for a remote office with agents and voice gateway.

Table 8: Remote Office with Agents and Voice Gateway Design Requirements

	Requirement	Notes
Infrastructure	Location-based call admission control	Unified Communications Manager location-based call admission control failure results in a disconnected routed call. You must provision adequate bandwidth to the remote sites and design a QoS WAN.

	Requirement	Notes
	Bandwidth	<p>You must plan bandwidth capacity for the following traffic:</p> <ul style="list-style-type: none"> • RTP for calls transferred to other remote offices, or if calls are not restricted to the remote office where the calls arrive. • Unified Communications Manager signaling to IP phones • Client desktop to PG (CTI data) • Unified Intelligence Center client to Unified Intelligence Center server • Silent Monitoring RTP • Recording RTP (if a recording server is not located in the remote office) • VXML gateway (VXML documents and VXML file retrieval) • Music on Hold for calls that are on hold when you use Unified Communications Manager Unicast Music on Hold • ISE client to server • Administration client to the Administration Server and Real-Time Data Server
Desktop	Finesse CTI OS desktop Cisco Agent Desktop Customer Relationship Management	
Codec	G.711 or G.729a	G.711 requires significantly more bandwidth.

	Requirement	Notes
Recording	BIB	MediaSense provides audio-only call recording. Audio forking requires Unified Border Element.
Silent Monitoring	Unified Communications Manager-based BIB	

The following table describes the media resources for a remote office with agents and voice gateway.

Table 9: Remote Office with Agents and Voice Gateway Media Resources

Resources	Method	Notes
Music on Hold	Unicast using Unified Communications Manager	
Conference bridges	Hardware-based, located at voice gateways	Conference bridges use local Unified SRST.
Media Termination Points	Hardware-based, located at voice gateways	For Unified Mobile Agents, MTPs are required only at the data center.
Transcoders	Hardware-based, located at voice gateways	Transcoders use local Unified SRST.

Home Agent with Broadband

For connectivity requirements, see Unified Communications Manager bandwidth and latency in the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

VPN router capabilities must include QoS for desktops. Applications such as Unified Intelligence Center, desktop, and additional call flows such as recording, require bandwidth calculation.

Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

[Sizing Unified CCE Components and Servers, on page 215](#)

[Sizing Cisco Unified Communications Manager Servers, on page 229](#)

Unified Mobile Agent

There are two types of remote agents:

- Home agent with broadband
- Unified Mobile Agent

Unified Mobile Agent supports call center agents using phones that Unified CCE does not directly control. A mobile agent can be physically located either outside or inside the contact center.

- Outside the contact center: The agent uses an analog phone in the home or a cell phone.
- Within the contact center: The agent uses an IP phone connection that Unified CCE or Unified Communications Manager does not control.

In addition, a mobile agent can be available through different phone numbers at different times; the agent enters the phone number at login time. The agent can access Unified Mobile Agent using any phone number, as long as the agent can dial the number using the Unified Communications Manager dial plan.

System administrators configure the Unified Mobile Agent to use a nailed (permanent) or call-by-call connection. 10.0.

Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

[Cisco Unified Mobile Agent, on page 171](#)

[Unified Mobile Agent, on page 48](#)

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

Unified Mobile Agent Architecture

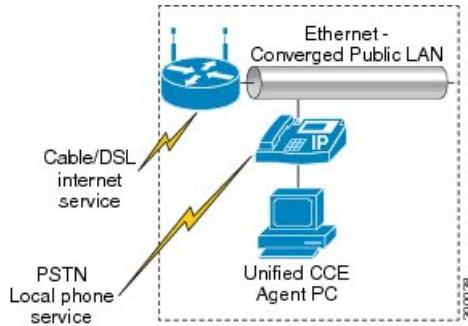
The following figure shows the logical view of Unified Mobile Agent.

Figure 21: Unified Mobile Agent—Logical View



The following figure shows the physical view of Unified Mobile Agent.

Figure 22: Unified Mobile Agent—Physical View



Unified Mobile Agent Components

The Unified Mobile Agent deployment scenario includes the following components:

- Cisco Virtual Office 871 cable/DSL router for secure VPN data connectivity to the data centers (no voice)
- Agent uses local phone with traditional local phone service to accept inbound calls
- Agent CTI desktops connect to Cisco Virtual Office 871 cable/DSL router
- Administration tools: Unified configuration tools, Internet Script Editor, or the local Administrative Workstation

Unified Mobile Agent Benefits

The Unified Mobile Agent deployment scenario provides the following benefits:

- Unified Mobile Agent can send calls to any PSTN or cell phone—and thereby extend the reach of a centralized IP contact center.
- Contact centers can hire skilled employees where they live and integrate remote workers into geographically dispersed teams with access to equivalent corporate applications.
- Contact centers can reduce startup costs by bringing temporary agents online during seasonal high call volume. Agents can choose their destination phone number during sign-up time and change the number as often as they want, giving the contact center the flexibility to adapt to a fast-moving mobile workforce.
- Contact center agents with central-site-equivalent access to applications and services in geographically dispersed locations (such as home offices) creates a built-in backup plan to keep business processes functioning in unforeseen circumstances.

Unified Mobile Agent Design Requirements

The following table describes the design requirements for Unified Mobile Agent.

Table 10: Unified Mobile Agent Design Requirements

	Requirement	Notes
Configuration	Dial plan	If you want mobile agents to use a dedicated gateway, configure the dial plan so that all calls from the CTI ports go through a specific gateway at the data center regardless of which phone number is called.
		Define the local CTI port Directory number (DN), which is the routing label when the agent is selected.
		To keep the mobile agent logged in, set the values for both the Maximum Call Duration timer and Maximum Call Hold timer to 0. To configure these timers, use the Unified Communications Manager Administration web page for service parameters using Unified Communications Service.
		The Cisco Unified Mobile Agent connect tone provides an audible indication when a call is delivered to the nailed connection mobile agent. The connection tone is two beeps, which the nailed connection mobile agent hears when answering a call. This feature is turned off by default. Use the PG registry key PlayMACConnectTone to enable the Cisco Unified Mobile Agent connect tone.
	SIP trunk (CUBE)	CUBE dynamically changes the media port during the call. If you use the Mobile Agent feature, the SIP trunk that connects to the agent endpoint requires MTP resources.

	Requirement	Notes
Codec	G.711 or G.729	Ingress and egress voice gateways can be G.711 or G.729 but not a mix of both. All CTI ports for a PG must advertise the same codec type. All mobile agents should use the same codec, but local agents on the supervisor's team can use a mix of codecs.
Infrastructure	DNS	You must have a DNS entry for the mobile agent desktop. If you do not have a DNS entry for the mobile agent desktop, the agent cannot connect to a CTI server.
	Firewall	In a deployment with a firewall, if an agent in a nailed connection mode is idle longer than the firewall idle timeout value, the firewall can block the media stream when the firewall idle timeout expires. To prevent the firewall from blocking the media stream, increase the firewall idle timeout value.
	Bandwidth	Minimum supported bandwidth speed: <ul style="list-style-type: none"> • 256 kbps upload • 1.0 Mbps download Use bandwidth calculators to ensure that you provide sufficient bandwidth. QoS is enabled only at the remote agent router edge. Currently, service providers do not provide QoS.

	Requirement	Notes
	Latency	The mobile agent round-trip delay to the Unified CCE data center must not exceed 150 ms. The mobile agent jitter delay must not exceed 60 ms.
	Voice gateways	Use egress gateways for mobile agents.
	Call control	Use RONA when a mobile agent is logged in and ready, but is unavailable to pick up a call. Only blind transfer and conference are supported if a mobile agent on one PG calls a mobile agent on another PG, and both PGs are connected to the same Unified Communications Manager cluster.
	Phones	Disable agent phone call features such as call waiting, call forwarding, and voicemail.
	Agent workstation	Set up the mobile agent workstation to use DHCP.
	Security	Enable security features on the remote agent router.
Desktop	Finesse CTI OS desktop Cisco Agent Desktop	Finesse does not support Switched Port Analyzer (SPAN) port silent monitoring.
Recording	SPAN port	Recording server in the data center. MediaSense is not supported.
Silent Monitoring	SPAN port	For CTI OS desktop, use CTI OS silent monitor server in the data center. For CAD, use CAD SPAN port monitoring.

The following table describes Unified Mobile Agent media resources.

Table 11: Unified Mobile Agent Media Resources

Resource	Method	Notes
Music on Hold	Unified Communications Manager unicast	If the Music on Hold server is not set up to stream using a G.729 codec, then you must set up a transcoder to enable outside callers to receive Music on Hold.
Conference bridges	Voice gateways in the data center	Agent greeting requires a conference bridge.
Media Termination Points	Voice gateways in the data center	<p>Assign two MTPs for each Unified Mobile Agent:</p> <ul style="list-style-type: none"> • MTP for remote CTI port • MTP for local CTI port <p>CTI ports do not support in-band Dual-Tone Multifrequency (DTMF) RFC 2833. The MTPs perform the conversion.</p> <p>Do not place MTPs at the egress gateway.</p> <p>If you use SIP trunks, you must configure Media Termination Points (MTPs).</p> <p>Enabling the use of an MTP on a trunk affects all calls that traverse that trunk, even non contact center calls. Ensure that the number of available MTPs can support the number of calls traversing the trunk.</p>
Transcoders	Voice gateways in the data center	All mobile agents must have the same codec: G.711 or G.729.

Centralized Data Center

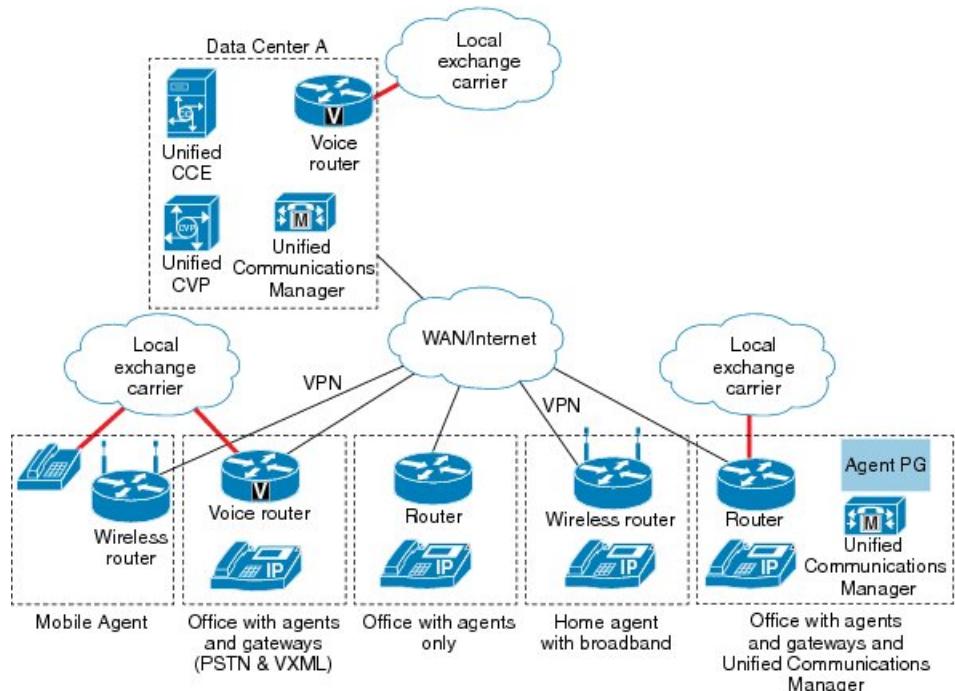
A centralized data center can contain all of the Unified CCE base model components. In a centralized data center, the agents, supervisors, and administrators are local to the data center. A centralized data center can also include multiple agent locations.

In addition to local agents, a centralized data center includes the following agent locations:

- Unified Mobile Agent

- Remote office with PSTN and VXML gateways (distributed voice gateways)
- Remote office with only agents and phones
- Remote office with at home agents using a VPN
- Remote office with agents and gateways and Unified Communications Manager

Figure 23: Centralized Data Center



Geographically Redundant Data Centers

Globalization, security, and disaster recovery considerations are driving business to diversify locations across multiple regions. In addition, organizations want to distribute workloads between computers, share network resources effectively, and increase the availability of critical applications. Geographically redundant data centers split critical applications across two data centers. Enterprises deploy geographically redundant data centers to minimize planned or unplanned downtime and share data across regions.

Geographically redundant data centers have a minimum of two load balancers, one in each data center. You can use two load balancers for each data center for local redundancy.

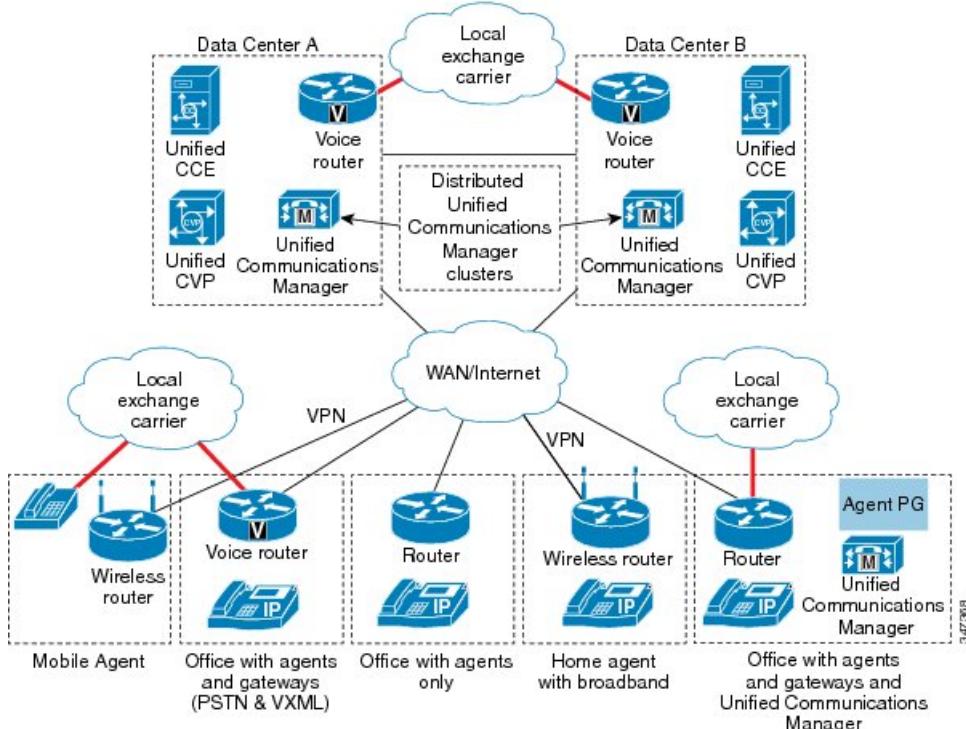
Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

Geographically Redundant Data Centers with Clustering over WAN

The following figure shows geographically redundant data centers with clustering over the WAN.

Figure 24: Geographically Redundant Data Centers with Clustering over WAN



Geographically redundant data centers provide clustering over the WAN, distributed Unified Communications Manager clusters, and 1:1 redundancy for Unified CVP, SIP proxy, voice gateways, and Cisco Unified Intelligence Center for example.

Latency requirements across the high-availability (HA) WAN must meet the current Cisco Unified Communications requirements for clustering over the WAN. Unified Communications Manager Release 6.1 or later allows a maximum latency of 40 ms one way (80-ms round-trip).

If you use a single fault tolerant network that carries all your traffic, for example, Multiprotocol Label Switching (MPLS) or SONET, keep the public and private traffic on separate routes within the network and respect standard latency and bandwidth.

If you use Unified IP IVR in geographically redundant data centers, see the Parent/Child appendix.

Related Topics

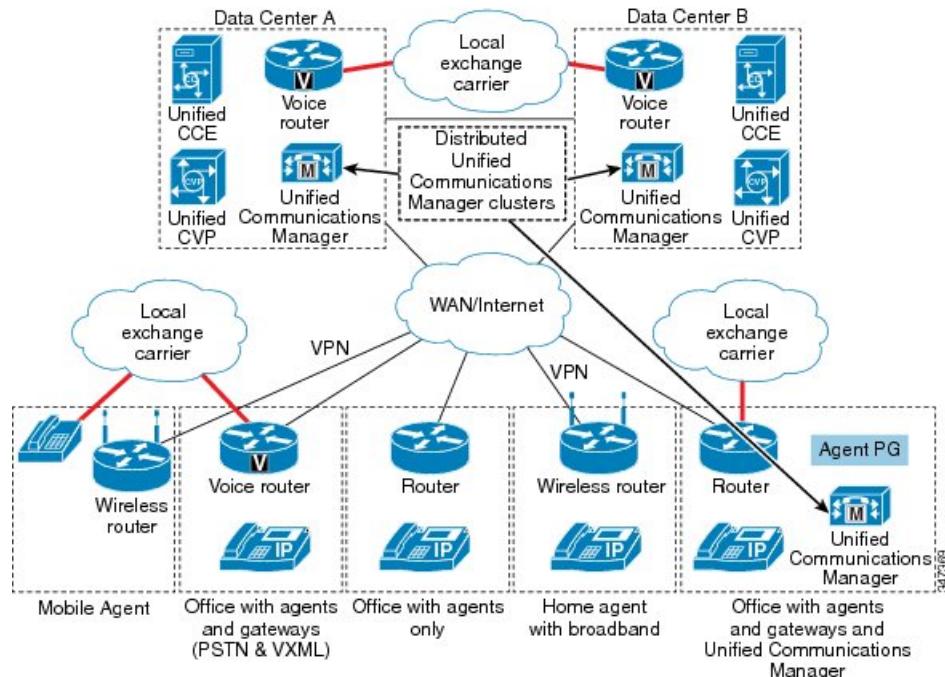
[Geographically Redundant Child Data Centers \(Using Unified IP IVR\), on page 339](#)

Geographically Redundant Data Centers with Distributed Unified Communications Manager Clusters

If you have a remote office with agents, gateways, and Unified Communications Manager clusters, the Unified Communications clusters at the data centers are typically independent. In this distributed call processing model, each data center has its own Unified Communications cluster, with its own agents and PG pairs.

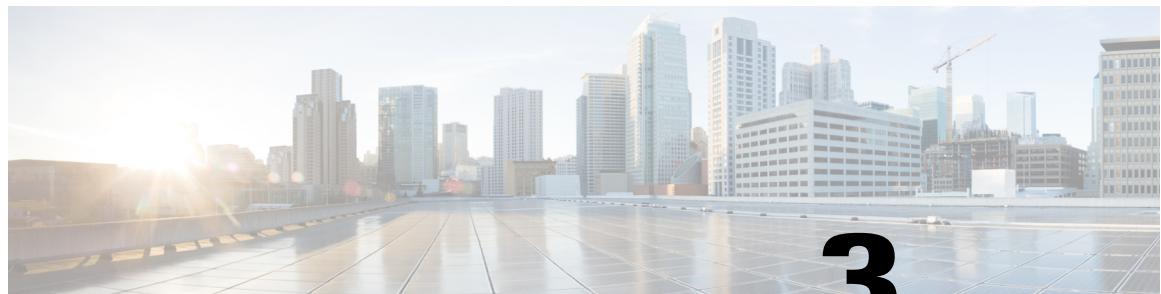
The following figure shows three Unified Communications Manager clusters. The remote office has a WAN connection back to the data centers. Each Unified Communications Manager cluster is independent, with its own agents and PG pairs. Each data center uses subscribers that are local to the data center because JTAPI is not supported over the WAN. For example, data center A cannot use the subscribers in data center B. The Unified CCE central controller, Unified Intelligence Center, load balancer, SIP proxy server, and Unified CVP are located in the data centers. TDM and VXML voice gateways are located at the remote office with local PSTN trunks.

Figure 25: Geographically Redundant Data Centers with Distributed Unified Communications Manager Clusters



Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)



3

CHAPTER

Design Considerations for High Availability

- [High Availability Designs, page 59](#)
- [High Availability and Virtualization, page 61](#)
- [Data Network Design Considerations, page 61](#)
- [Unified Communications Manager Design Considerations, page 63](#)
- [Unified CVP Design Considerations, page 66](#)
- [Unified IP IVR Design Considerations, page 67](#)
- [Cisco Web and E-Mail Interaction Manager Design Considerations, page 69](#)
- [Cisco Outbound Option Design Considerations, page 70](#)
- [Agent Peripheral Gateway Design Considerations, page 72](#)
- [Central Controller Design Considerations, page 77](#)
- [Common Processes That Support Failovers, page 78](#)
- [Unified CCE Failovers During Network Failures, page 79](#)
- [Unified CCE Failovers During Single-Component Failures, page 83](#)
- [Unified CCE Failovers During Multicomponent Failures, page 93](#)
- [Other Considerations for High Availability, page 96](#)

High Availability Designs

Cisco Unified Contact Center Enterprise (Unified CCE) products incorporate high availability features in all standard deployments. Every production deployment must include redundancy for the core Unified CCE components. The redundant components are designed to fail over automatically and recover without manual intervention. Your design can include more than that basic high availability capability. A successful Unified CCE deployment requires a team with experience in data and voice internetworking, system administration, and Unified CCE application design and configuration.

Each change to promote high availability comes at a cost. That cost can include more hardware, more software components, and more network bandwidth. Balance that cost against what you gain from the change. How

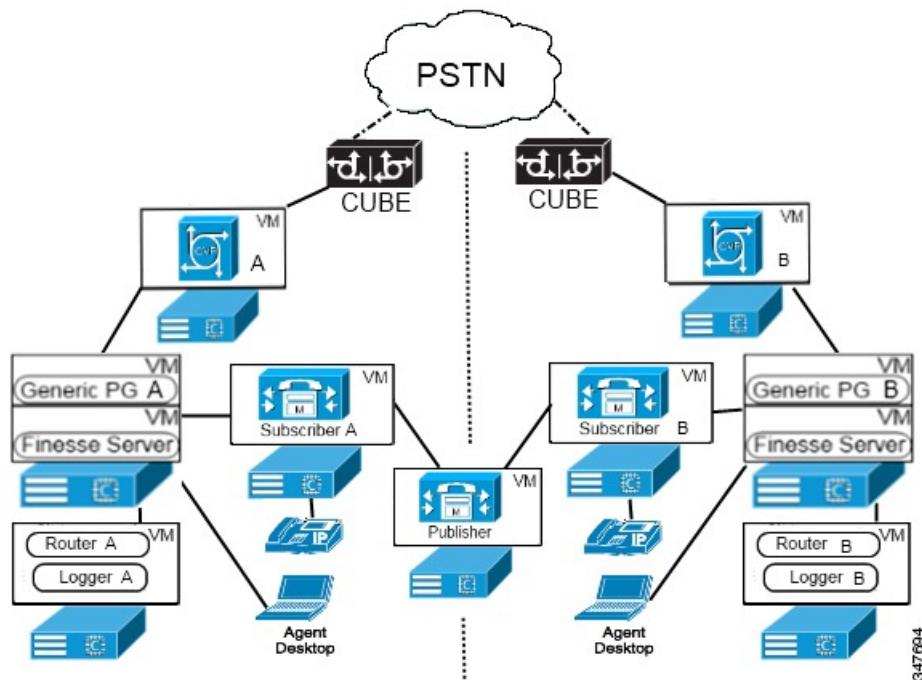
critical is preventing disconnects during a failover scenario? Is it acceptable for customers to spend a few extra minutes on hold while part of the system recovers? Would the customer accept losing context for some calls during a failure? Can you invest in greater fault tolerance during the initial design to position the contact center for future scalability?

**Note**

This guide focuses on design of the contact center itself. Your contact center operates in a framework of other systems. This guide cannot provide complete information about every system that supports your contact center. The guide concentrates on the Cisco Unified CCE products. When the guide must discuss another system, it does not offer a comprehensive view. For more information about the complete Cisco Unified Communications product suite, see the Cisco solutions design documents at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

The following figure shows a fault-tolerant Unified CCE single-site deployment.

Figure 26: Unified CCE Component Redundancy



This design shows how each Unified CCE component is duplicated for redundancy. All Unified CCE deployments use redundant Unified Communications Manager, Unified CCE, and Cisco Unified Customer Voice Portal (Unified CVP) or Unified IP IVR components. Because of the redundancy, a Unified CCE deployment can lose half of its core systems and be still operational. In that state, a Unified CCE deployment can handle calls by rerouting them through Unified CVP to either a VRU session or an agent that is connected to the still-operational components. Where possible, deploy Unified CCE so that no devices, call processing, or CTI Manager services are running on the Cisco Unified Communications Manager publisher.

To enable automatic failover and recovery, pairs of redundant components interconnect over private network paths. The components use TCP keepalive messages at 100-ms intervals for failure detection. The Unified Communications Manager uses a cluster design for failover and recovery. Each cluster contains a Unified Communications Manager publisher and multiple Unified Communications Manager subscribers. Agent

phones and computers register with a primary target but automatically reregister with a backup target if the primary fails.

High Availability and Virtualization

In a virtualized deployment, place components carefully to maintain high availability. The mechanisms that support high availability are the same. But, you must distribute components to minimize multiple failovers from a single failure. When you deploy on Direct Attached Storage (DAS) only systems, consider the following points:

- Failure of a VM brings down all the components that are installed on the VM.
- Failure of a physical server brings down all the VMs that are installed on that VMware vSphere Host.

Deployments on systems with shared storage can use some of the VMware High Availability features for greater resiliency. For specific information about supported VMware features, see the *Unified Communications in a Virtualized Environment* at http://docwiki.cisco.com/wiki/Unified_Communications_in_a_Virtualized_Environment.

To minimize the impact of hardware failures, follow these guidelines:

- Avoid placing a primary VM and a backup VM on the same physical server, chassis, or site.
- Avoid placing all the active components in a failover group on the same physical server, chassis, or site.
- Avoid placing all VMs with the same role on the same physical server, chassis, or site.

Also consider which components can be coresident and which components must be coresident on the same VMs. For more information about placement of components in virtual environments, see the *Virtualization for Unified CCE DocWiki* at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE.

Data Network Design Considerations

Highly available contact center designs start with the network infrastructure for data, multimedia, and voice traffic. A single point of failure in your network infrastructure devalues any other high availability features that you design into the contact center. Begin from the PSTN and ensure that incoming calls have multiple paths for reaching Unified CVP for initial treatment and queuing.

Ideally, design with at least two SIP trunks each connecting to a separate Cisco Unified Border Element (Cisco UBE). If any Cisco UBE or SIP trunk fails, the PSTN can route all traffic through the remaining SIP trunks. The PSTN route either by configuring all the SIP trunks as a large trunk group or by configuring rerouting or overflow routing to the other SIP trunks. You can also connect a redundant Cisco UBE to each SIP trunk to preserve capacity when a Cisco UBE fails and the SIP trunk is still functional.

In some areas, the PSTN does not provide multiple SIP trunks to a single site. In that case, you can connect the SIP trunk to a Cisco Unified SIP Proxy. Then, you could connect multiple Cisco UBEs to the Unified SIP Proxy to provide some redundancy.

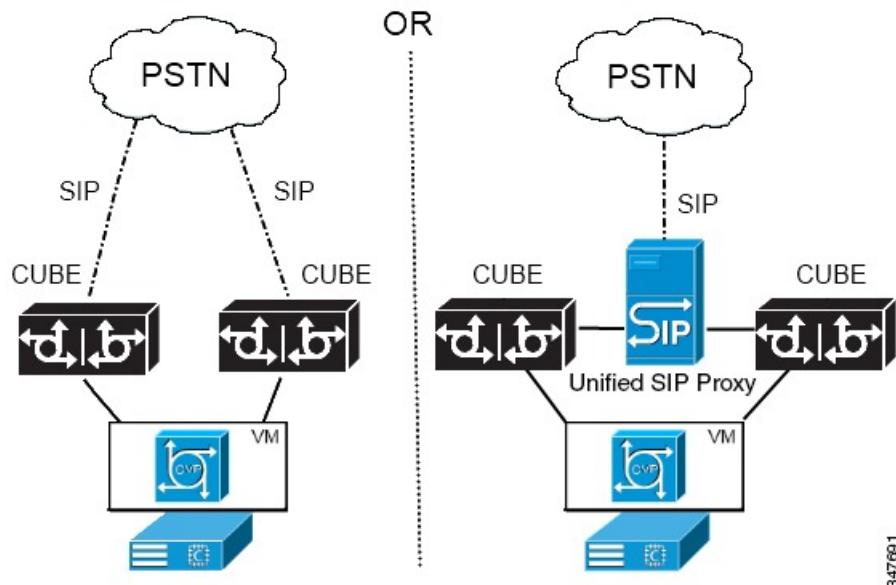
The Cisco UBE passes calls to Unified CVP for initial treatment and queuing. Register each Cisco UBE with a separate Unified CVP for load balancing. For further fault tolerance, you can register each Cisco UBE with a different Unified CVP as a backup. If a Cisco UBE cannot connect with a Unified CVP, you can also use TCL scripts to provide some call processing. A TCL script might reroute the calls to another site or dialed number or play a locally stored .wav file to the caller and end the call.

**Note**

In systems that use Cisco Unified IP IVR instead of Unified CVP, the call flows are different. But your design must still support redundant paths from the call ingress point to the queuing and treatment process.

For more information about Cisco UBE, Unified CVP, and voice networks in general, see the *Cisco Collaboration System Solution Reference Network Designs* at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

Figure 27: High Availability Ingress Points



347691

Voice gateways using the Cisco Unified Survivable Remote Site Telephony (SRST) option for Unified Communications Manager follow a similar failover process. If the gateway is cut off from its controlling subscriber, the gateway fails over into SRST mode. The failover drops all voice calls and resets the gateway into SRST mode. Phones rehome to the local SRST gateway for local call control.

While running in SRST mode, Unified CCE operates as if the agents have no CTI connection from their desktops. The Unified CCE routing application detects the agents as not ready and sends no calls to these agents. When the gateway and subscriber reestablish their connection, the subscriber takes control of the gateway and phones again, allowing the agents to reconnect.

Public and Private Network Connections

Unified CCE components use a public network and a private network to communicate. These networks must be separate physical networks. For high availability, include redundant connections in your public network. Ideally, each connection uses a different carrier.

If QoS and bandwidth are configured correctly, your design can merge a public or private WAN link with other corporate traffic. If you use a link that merges non-contact-center traffic, keep the public and private traffic on different networks. However, never split private network traffic onto low-priority and high-priority data paths. The same link must carry all private network traffic for a given component. Sending low-priority and high-priority traffic on different links disables the component failover behavior. Similarly, all low- and

high-priority traffic from each peripheral gateway to the low- and high-priority addresses of the call router must take the same path.

During a public network failure, you can temporarily fail over the public Unified Communications Manager traffic to the private network. Size the private network to accommodate the extra traffic. When the public traffic fails over to the private network, restore the public network as quickly as possible to return to normal operation. If the private network also fails, Unified CCE instability and data loss can occur, including the corruption of one Logger database.

A SONET fiber ring is a highly resilient network with built-in redundancy. You can send the public and private traffic over the same SONET ring under normal operations or following a network failover. A separate link for the private traffic is not required in this case. Also, two routers are required on each side of the WAN for redundancy. Under normal operations, use one router for the Unified CCE public traffic and use the other router for the Unified CCE private traffic. The other rules described in this section also apply.

Figure 28: Network Architecture with SONET Ring



Unified Communications Manager Design Considerations

After you design the data network, design the Cisco Unified Communications infrastructure. Before you can deploy any telephony applications, you need the Unified Communications Manager cluster and CTI Manager in place to dial and receive calls.

For details on the architecture of all these services, see the *Cisco Collaboration System Solution Reference Network Designs* at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

High availability design for a cluster requires that you understand how the Unified Communications Manager, CTI Manager, and CallManager services interact. Unified Communications Manager uses the CTI Manager service to handle all its CTI resources. CTI Manager acts as an application broker that abstracts the physical binding of applications to a particular Unified Communications Manager server. The CallManager service registers and monitors all the Cisco Unified Communications devices.

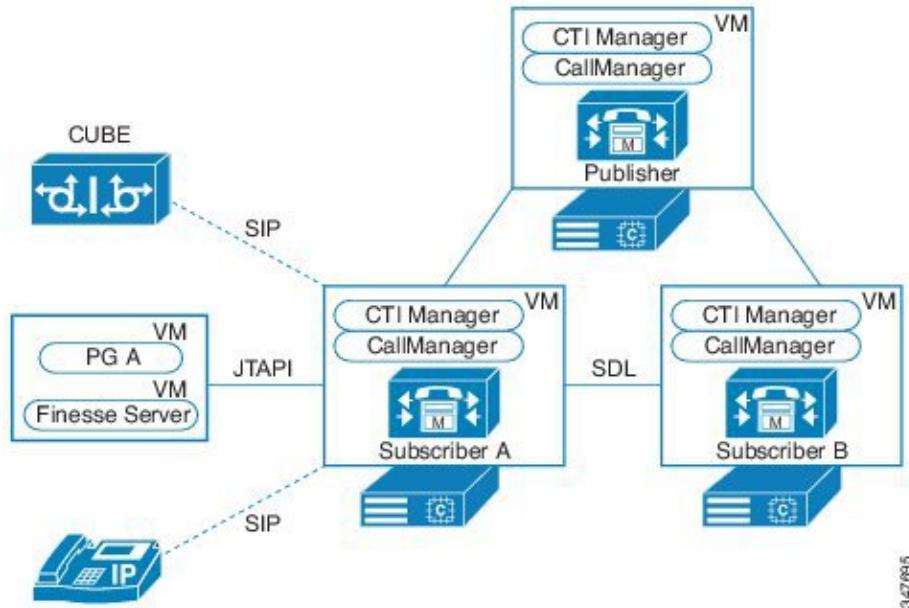
The CTI Manager accepts messages from the Agent PG, a CTI application, and sends them to the appropriate resource in the cluster. The CTI Manager acts like a JTAPI messaging router using the Cisco JTAPI link to communicate with Agent PGs. The JTAPI client library in Cisco Unified Communications Manager connects to the CTI Manager instead of connecting directly to the CallManager service.

CallManager service acts as a switch for all the Cisco Unified Communications resources and devices in the system. The CallManagers on each Unified Communications Manager server link themselves across the public network with the Signal Distribution Layer (SDL). This link keeps the cluster in sync. Each CTI Manager connects with the Unified Communications Manager and CallManager services on its server. CTI Managers do not connect directly with other CTI Managers in the cluster.

Agent PGs use a CTI-enabled user account in Unified Communications Manager, typically called the JTAPI user or PG user. The Agent PGs sign in to the CTI Manager to connect to the devices that are associated to that user. If the local CallManager services the appropriate device, the CTI Manager handles the request for that device. If the device is not resident on its local subscriber, then the CallManager service forwards the request to the appropriate subscriber through the private link to the other CallManager services.

The following figure shows the connections in a cluster.

Figure 29: Connections in Unified Communications Manager Cluster



For high availability, distribute device registrations across all the subscribers in the cluster. If you concentrate the registrations on a single subscriber, the traffic puts a high load on that subscriber. The memory objects that the Agent PGs use to monitor registered devices also add to the device weights on the subscribers.

If the PG that is connected to that subscriber fails, the redundant PG that takes over sends all the requests to another subscriber. Then, the local CallManager service must route the CTI Manager messaging for those requests across the cluster to the original subscriber. The additional messaging in this failover condition creates greater load on the cluster.

Unified Communications Manager Redundancy

Some Unified Communications Manager deployments use a 2:1 redundancy scheme. Each pair of primary subscribers shares a single backup subscriber. But, because of the higher phone usage in contact centers and to simplify upgrade processes, Unified CCE uses a 1:1 redundancy scheme for subscribers. Each primary subscriber requires its own backup subscriber.

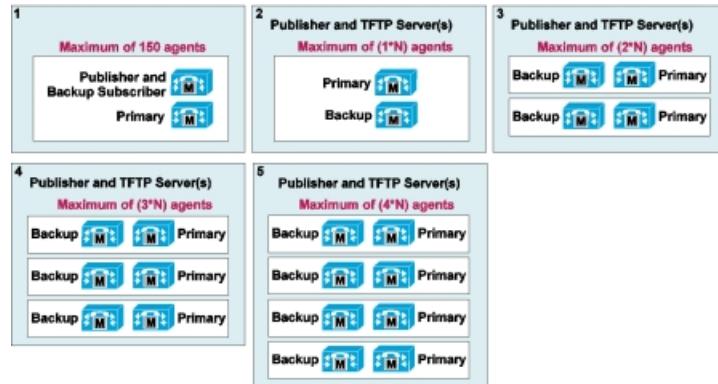
For details on other cluster deployment and redundancy options, see the latest version of the *Cisco Collaboration System Solution Reference Network Designs* at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

The following figure shows different size clusters. Option 1 supports up to 150 Unified CCE agents. Options 2 to 5 illustrate increasingly larger clusters. In this figure, the value of N depends on the components that your contact center uses:

- For Unified CCE deployments with Unified Communications Manager and Unified CVP, N is equal to 2000.

- For deployments with Unified IP IVR, N is equal to 500.

Figure 30: Redundancy Configuration Options



Unified Communications Manager Load Balancing

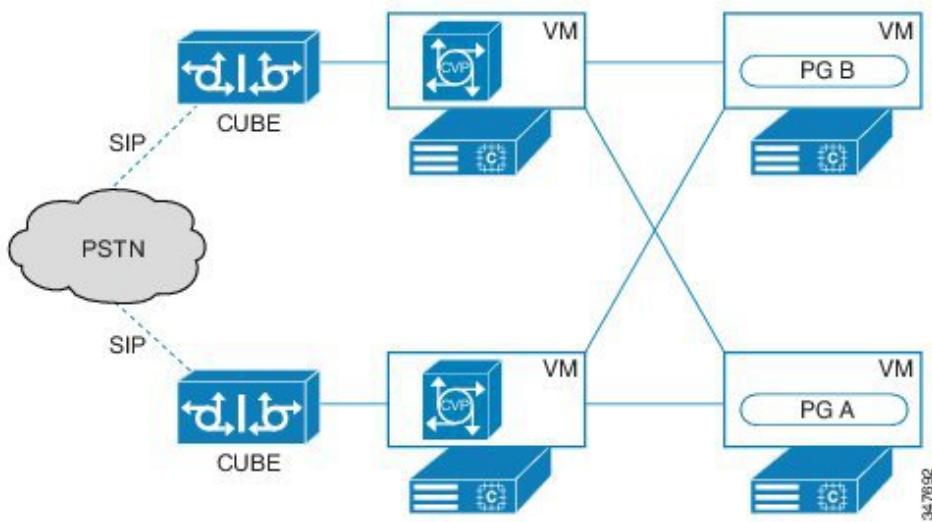
The 1:1 redundancy scheme for Unified Communications Manager subscribers lets you balance the devices over the primary and backup subscriber pairs. Normally, a backup subscriber has no devices registered unless its primary subscriber is unavailable.

You can enable load balancing through Unified Communications Manager redundancy groups and device pool settings. You can move up to half of the device load from the primary to the secondary subscriber. In this way, you can reduce by half the impact of any server becoming unavailable. To minimize the effect of any outage, distribute all devices and call volumes equally across all active subscribers.

Unified CVP Design Considerations

The Contact Center Enterprise Reference Designs use Unified CVP for call treatment and queuing. Unified CVP uses SIP for call control, rather than relying on Unified Communications Manager for JTAPI call control.

Figure 31: Unified CVP High Availability Deployment



Unified CVP can use the following system components:

- Cisco Unified Border Element (CUBE) supports the transition to SIP trunking. CUBE provides interworking, demarcation, and security services between the PSTN and your contact center.
- Cisco Voice Gateway (VG) terminates TDM PSTN trunks to transform them into IP-based calls on an IP network. CVP uses specific Cisco IOS Voice Gateways that support SIP to enable more flexible call control. VGs controlled by Unified CVP can also use the Cisco IOS built-in Voice Extensible Markup Language (VoiceXML) Browser to provide caller treatment and call queuing. CVP can also leverage the gateway's Media Resource Control Protocol (MRCP) interface to add automatic speech recognition (ASR) and text-to-speech (TTS) functions.
- Unified CVP Server provides call control signaling when calls are switched between the ingress gateway and another endpoint gateway or a Unified CCE agent. The CVP Server also provides the interface to the Unified CCE VRU Peripheral Gateway (PG). The CVP Server translates specific Unified CCE VRU commands into VoiceXML code for rendering on the Unified CVP VG. The CVP Server can communicate with the gateways using SIP as part of the solution.
- Unified CVP Media Server acts as a web server that provides predefined audio files to the voice browsers as part of their VoiceXML processing. You can cluster media servers using the Cisco Content Services Switch (CSS) products. With clustering, you can pool multiple media servers behind a single URL for access by all the voice browsers.
- Unified CVP Server hosts the Unified CVP VoiceXML runtime environment. The VoiceXML service creation environment uses an Eclipse toolkit browser in the Unified CVP Call Studio Application. The runtime environment executes the dynamic VoiceXML applications and processes Java and Web Services calls for external systems and database access.

- Cisco Unified SIP Proxy servers in a Unified CVP deployment select voice browsers and associate them with specific dialed numbers. When a call comes into the network, the VG queries the Unified SIP Proxy to determine where to send the call based on the dialed number.

**Important**

Unified CCE deployments do not support Unified CVP's Enhanced Location Call Admission Control feature.

These methods can increase the high availability of Unified CVP:

- To provide automatic call balancing across the Unified CVP Servers, add redundant Unified CVP Servers under control of the Unified CCE PGs.
- To handle conditions where the gateway cannot contact the Unified CVP Server, add survivability TCL scripts to the gateway. For example, you can redirect calls to another Unified CVP Server on another Unified CVP-controlled gateway.
- To load balance audio file requests across multiple Unified CVP Media Servers and VoiceXML URL access across multiple servers, add a Cisco Content Server.

For more information about these options, review the Unified CVP product documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/tsd-products-support-series-home.html>.

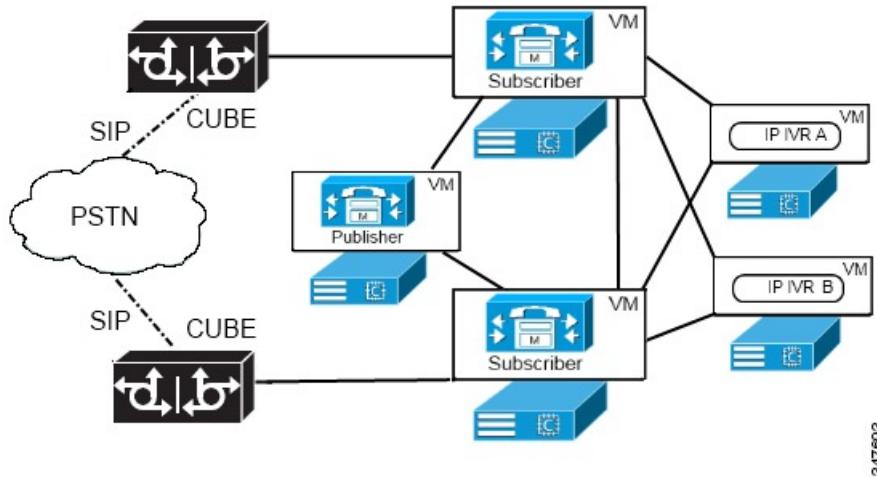
Unified IP IVR Design Considerations

Cisco Unified IP IVR can establish JTAPI connections with two CTI Managers on different subscribers in the Unified Communications Manager cluster. This feature enables Unified IP IVR redundancy at the CTI Manager level. You can gain more redundancy by deploying multiple Unified IP IVR servers. Multiple Unified IP IVR servers enable call routing scripts to load balance calls between the available IP IVR resources.

The following figure shows two Unified IP IVR servers set up redundantly with a cluster. Set up the Unified IP IVR group with each server connected to the CTI Manager on a different Unified Communications Manager

subscriber. Then, add a second CTI Manager as a backup to each Unified IP IVR server. If the primary CTI Manager fails, the Unified IP IVR server fails over to the backup CTI Manager.

Figure 32: IP IVR High Availability Deployment



High Availability Through Call Forwarding

You can use the following call forwarding features in Unified Communications Manager to manage Unified IP IVR port usage:

- Forward Busy—Forwards calls to another port or route point when Unified Communications Manager detects that the port is busy.
- Forward No Answer—Forwards calls to another port or route point when Unified Communications Manager detects that a port did not pick up a call within the timeout period.
- Forward on Failure—Forwards calls to another port or route point when Unified Communications Manager detects a port failure caused by an application error.

When using the call forwarding features, do not establish a path back to the CTI port that initiated the call forwarding. Such paths create loops when all the Unified IP IVR servers are unavailable.

High Availability Through Call Flow Routing Scripts

You can use Unified CCE call flow routing scripts to support high availability. Check the Unified IP IVR Peripheral Status with a call flow routing script before sending calls to Unified IP IVR. This check prevents calls from queuing to an inactive Unified IP IVR. For example, create a Translation Route to the Voice Response Unit (VRU) to select the Unified IP IVR with the most idle ports. This method distributes the calls evenly on a call-by-call basis. You can modify this method to load balance ports across multiple Unified IP IVRs. This method can address all the Unified IP IVRs on the cluster in the same Translation Route or Send to VRU node.

**Note**

If the Unified IP IVR server fails, all calls at the Unified IP IVR are dropped. Minimize the impact of such failures by distributing calls across multiple Unified IP IVR servers. In Unified IP IVR, a default script prevents loss of calls if the Unified IP IVR loses the link to the VRU PG.

Cisco Web and E-Mail Interaction Manager Design Considerations

The Cisco Web and E-Mail Interaction Manager provides web and email interaction management through a common set of web servers and pages for agents and administrators. It integrates with the Unified CCE platform to provide universal queuing of contacts to agents from different media channels.

For more architectural and design information, see the *Cisco Unified Web and E-Mail Interaction Manager Solution Reference Network Design Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-email-interaction-manager/products-implementation-design-guides-list.html>.

Unified CCE Integration

To integrate with Unified CCE, Unified WIM and EIM adds a Services server running the External Agent Assignment Service (EAAS) and the Listener Service. The EAAS interacts with the Media Routing Peripheral Gateway (MR PG) through the Media Routing interface. The Listener Service interacts with the Agent PG through the Agent Resource Management (ARM) interface.

The Unified WIM and EIM application server connects with the Unified CCE Administration and Data Server to import relevant configuration data. The Unified WIM and EIM application server also maps the configuration to Unified WIM and EIM objects in the Cisco Interaction Manager database.

For certain deployments of Unified CCE, the MR PG of Unified CCE can reside on the Services server.

Load-Balancing Considerations

You can load balance the web service component of a Unified WIM and EIM deployment to serve many agents who simultaneously access the application. You can set up the web (or web and application) servers behind the load balancer with a virtual IP address. When an agent accesses the Unified WIM and EIM with the virtual IP address, the load balancer sends a request to one of the servers behind the address. The load balancer then sends a response back to the agent. In this way, from a security perspective, the load balancer also serves as a reverse proxy server.

The load balancer must support sticky sessions with cookie-based persistence. Before allowing agents access after every scheduled maintenance task, verify that all Web and application servers are available to share the load. If you allow access without all servers being available, the sticky connection feature can cause an overload on the first Web and application server. Using other parameters, you can define a load-balancing algorithm to meet the following objectives:

- Equal load balancing
- Isolation of the primary Web and application server

- Sending fewer requests to a low-powered Web and application server

The load balancer monitors the health of all Web and application servers in the cluster. During a failure, the load balancer removes the given server from the available pool of servers.

Failover Management

Unified WIM and EIM supports clustered deployments. These deployments ensure high availability and performance through transparent replication, load balancing, and failover. To handle failure conditions within integrated deployments of Unified WIM and EIM and Unified CCE, use the following methods:

- Implement multiple Web and Application servers. If the primary server goes down, the load balancer can mitigate the failure by routing requests to alternate servers. The load balancer detects application server failure and redirects requests to another application server. The alternate application server creates new user sessions and the agents have to sign in again to the Unified Web and E-Mail Interaction Manager.
- Dynamically add or remove servers from the online cluster to accommodate external changes in demand or internal changes in infrastructure.
- Use redundant Unified CCE components, such as MR PIMs and Agent PIMs, to enable Unified WIM and EIM service failovers.

The single points of failure in Unified WIM and EIM include the following:

- A JMS server failure
- A Services server failure
- A Database server failure

Cisco Outbound Option Design Considerations

The Cisco Outbound Option enables your contact center to place outgoing calls for a calling campaign. The major components of the Cisco Outbound Option are the following:

- Outbound Option Campaign Manager manages the dialing lists and rules associated with the calls. This component always resides on the Side A Logger platform. You cannot install a redundant copy on the Side B Logger.
- Outbound Option Dialer performs the dialing tasks on behalf of the Campaign Manager. The Dialer emulates a set of IP phones for Unified Communications Manager to make the outbound calls. The Dialer detects the called party and manages the interaction tasks with the Cisco Finesse or CTI OS server to transfer the call to an agent. The Dialer also communicates with the Media Routing Peripheral Gateway (MR PG). Each Dialer has its own peripheral interface manager (PIM) on the MR PG.
- The Outbound Option Import component does not run as a redundant pair. Install this component with the Side A Logger.
- The MR PG accepts route requests from noninbound voice systems such as the Unified Outbound Option or the Multichannel products.

For more information about the Cisco Outbound Option, see the *Outbound Option Guide for Unified Contact Center Enterprise and Hosted* at http://www.cisco.com/en/US/products/sw/custcosw/ps524/products_installation_and_configuration_guides_list.html.

You can deploy the Dialers in either of these methods:

- Coresident on a VM with the MR PG and the Agent PG for the Unified Communications Manager.
- Coresident on a VM with just the MR PG.

The system can support multiple Dialers across the enterprise, but the central Campaign Manager controls all the Dialers. Redundant pairs of SIP Dialers operate in a warm-standby mode similar to the PG fault-tolerance model.

To improve high availability in the Cisco Outbound Option:

- Deploy the MR PGs in redundant pairs.
- Deploy a redundant pair of SIP Dialers for each redundant Agent PG pair. Use the redundant SIP Dialer in the Campaign Manager to enable automatic fault recovery during a failure.
- Deploy redundant Voice Gateways for outbound dialing. The redundant gateways ensure that the Dialers have enough trunks available to place calls if a gateway fails. In some instances where outbound calling is the primary application, you can dedicate these gateways to outbound calling only.

**Note**

The SCCP Dialer is deprecated as of Unified CCE Release 10.0(1).

Do not use the SCCP Dialer in any new deployments. The SCCP Dialer will be removed in a coming release.

SIP Dialer Design Considerations

The Outbound Option with SIP Dialer provides high availability through fault-tolerant design in the SIP Dialer, the Agent PG, and Unified SIP Proxy. Many components in the Outbound Option with SIP Dialer run as redundant pairs.

The Campaign Manager supports a single active Dialer per peripheral. Configure one SIP Dialer, but use the same Dialer Name when installing both parts of the redundant Dialer pair. The peripheral setup program allows you to input the Dialer name in the setup page for each SIP Dialer. Only register two SIP Dialers under each name. The Campaign Manager rejects the registration of any additional Dialers that use the same name.

The Campaign Manager activates one SIP Dialer in the Ready state from its registered SIP Dialer pool. If the activated SIP Dialer changes state from Ready to Not Ready or loses its connection, the Campaign Manager activates the standby SIP Dialer. The Campaign Manager returns all outstanding records to Pending status after a timeout period.

If the active SIP loses connection to the CTI Server, Agent PG, or SIP server, the SIP Dialer fails over. The SIP server can be a Voice Gateway (VG) or Unified SIP Proxy. The SIP Dialer uses a heartbeat mechanism to verify its connection to the VG or Unified SIP Proxy. Configure each SIP Dialer to connect to a different VG or Unified SIP Proxy.

During the failover, the SIP Dialer sends all active and pending customer records to the Campaign Manager. If the Campaign Manager is not available, the Dialer closes them internally.

The Unified SIP Proxy server provides weighted load balancing and redundancy in a multiple-gateway deployment by configuring each gateway as part of the Server group configuration. If a gateway is overloaded or loses its WAN link to the PSTN network, Unified SIP Proxy can resend an outbound call to the next available gateway.

Unified SIP Proxy supports the Hot Swappable Router Protocol (HSRP). This protocol provides network redundancy by allowing two Unified SIP Proxy servers to test each other for connectivity continuously.

Because the Campaign Manager and SIP Dialer already include warm-standby functionality, configuring HSRP for Unified SIP Proxy adds undesirable complexity for Outbound Option. Do not use the HSRP configuration for the Unified SIP Proxy servers that are dedicated for Outbound Option.

Agent Peripheral Gateway Design Considerations

The Agent PG communicates with the Unified Communications Manager cluster through the CTI Manager. An Agent PG can control agent phones and CTI route points anywhere in the cluster. The Agent PG registers with the CTI Manager on a Unified Communications Manager subscriber in the cluster. The CTI Manager accepts all JTAPI requests from the PG for the cluster. When the PG requests a phone or route point on another subscriber, the CTI Manager forwards the request to the other subscriber using the CallManager SDL links.

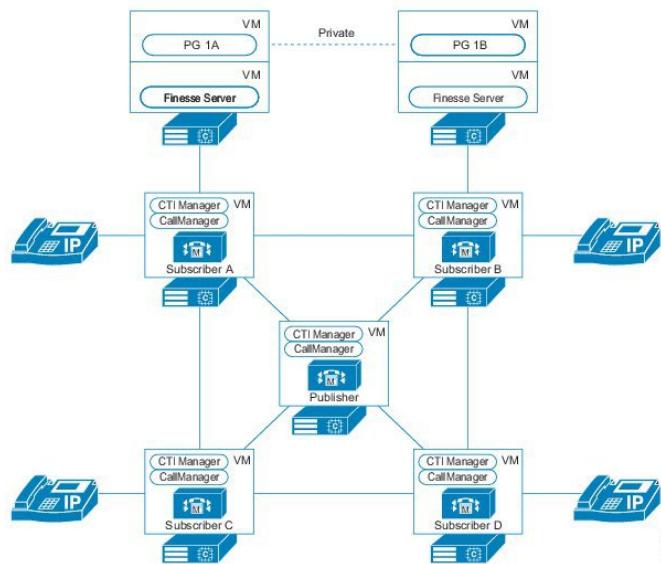
**Note**

This section uses *Agent PG* to describe any PG that includes the Unified Communications Manager PIM. The Agent PG can be a Generic PG or a Unified Communications Manager PG. For example, the following figure uses Generic PGs to connect to the subscribers. Those Generic PGs are acting as Agent PGs.

A fault-tolerant design deploys Unified CCE Agent PGs in a redundant configuration, because a PG only connects to the cluster through a single CTI Manager. If that CTI Manager fails, the PG cannot communicate with the cluster. A redundant PG provides a second pathway through a different CTI Manager on a different subscriber in the cluster.

The minimum design for a high-availability cluster is one publisher and two subscribers. If the primary subscriber fails, the devices rehome to the secondary subscriber and not to the publisher for the cluster.

Figure 33: High Availability Design for Unified Communications Manager Cluster



The redundant PG servers keep in synchronization through a private network that is isolated from the public network. If the two servers run on different physical machines at the same site, you can create the private network by connecting an Ethernet Cross-Over Cable between their private-network NICs. If the two PG servers are geographically distributed, use a separate WAN connection for the private network. To avoid a single point of failure in the network, do not use the same circuits or network gear as for the public network.

Within the Agent PG, the JTAPI Gateway and Agent PG PIM manage the connectivity to the cluster. The JTAPI Gateway handles the JTAPI socket connection protocol and messaging between the PIM and the CTI Manager. The PIM manages the interface between Unified CCE, the JTAPI Gateway, and the cluster. It requests specific objects to monitor and handle route requests from the cluster. The PG starts the JTAPI Gateway and PIM automatically as node-managed processes. The PG monitors the processes and automatically restarts them if they fail.

During PG installation, download the JTAPI Gateway from the cluster to ensure compatibility. Whenever you upgrade either the PG or Unified Communications Manager, remove and reinstall the JTAPI Gateway.

The JTAPI services from both redundant Agent PGs sign in to the CTI Manager after initialization. Agent PG Side A signs in to the primary CTI Manager; Agent PG Side B signs in to the secondary CTI Manager. Only one PG in each pair actively registers and monitors phones and CTI route points. The redundant PG runs in hot-standby mode. The redundant PG signs into the secondary CTI Manager only to initialize the interface and make it available for a failover. This arrangement significantly decreases the time for the failover.

During system start, the PG that first connects to the Unified CCE Call Router server and requests configuration information is the active PG. The Call Router ensures that the PG side that has the best connection becomes active. The nominal designations of “Side A” and “Side B” do not affect which PG becomes active. During a PG failover caused by a private link failure, a weighting mechanism chooses which PG is active to minimize the impact on the contact center.

The PIM startup process registers all CTI route points first, which is done at a rate of five route points per second. For systems with large numbers of CTI route points, it can take several minutes before the system allows any agents to sign in. You can reduce this time by distributing the devices over multiple PIM interfaces to the cluster.

If calls arrive at the CTI Route Points before the PIM is operational, the calls fail unless you set up the route points with a recovery number. Place the recovery number in their Call Forward on Unregistered or Call Forward on Failure setting. For example, you can set the recovery numbers to the Cisco Unity voicemail system for the Auto Attendant.

Active PG Shutdowns

Avoid shutting down an active peripheral gateway service in your production environment. This causes a service interruption of a minute or more while the other side connects and activates. The length depends on the size of the configuration and the type of peripheral. For example, the VRU peripheral can take less time. The other side for the VRU might take 30 seconds or less to reactivate.

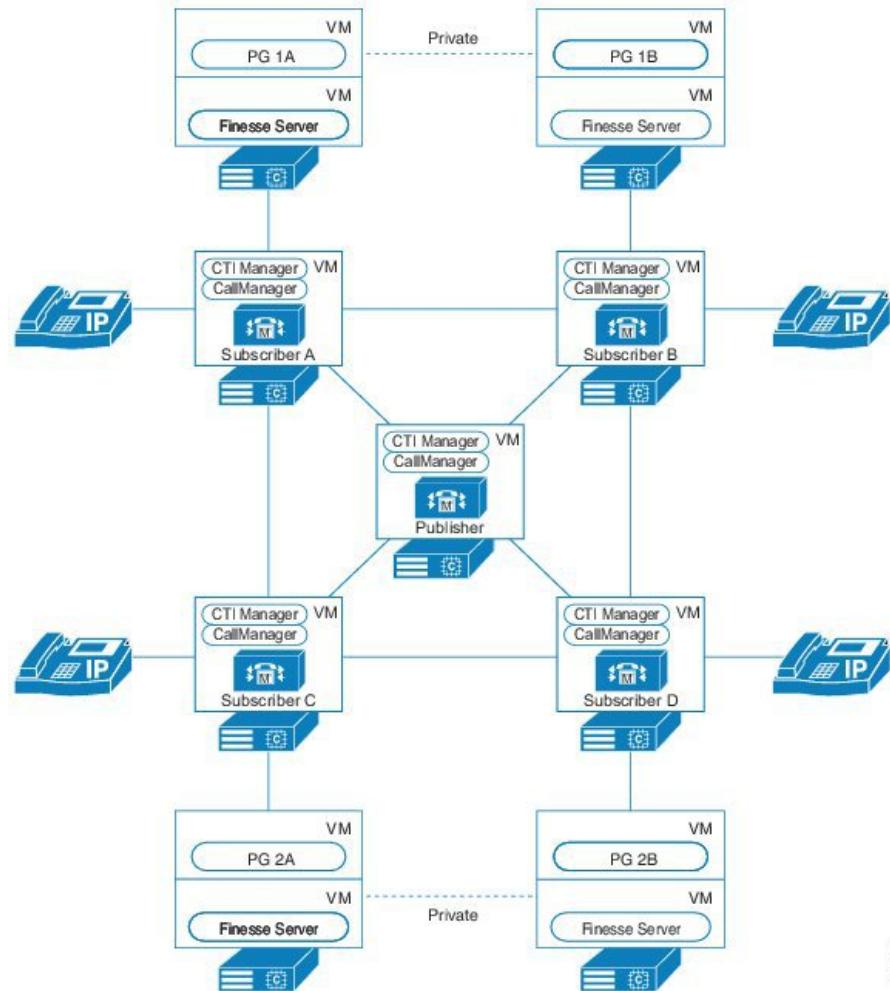
Agent PG Deployment for Unified Communications Manager Cluster

You can deploy Agent PGs in a Unified Communications Manager cluster in either of the following ways:

- Deploy one Agent PG pair for every two Unified Communications Manager subscribers. Each subscriber includes a CTI Manager. Each Agent PG connects to the CTI Manager running on a different subscriber.

The following diagram shows a deployment with two Agent PG pairs that are connected to a cluster with four subscribers.

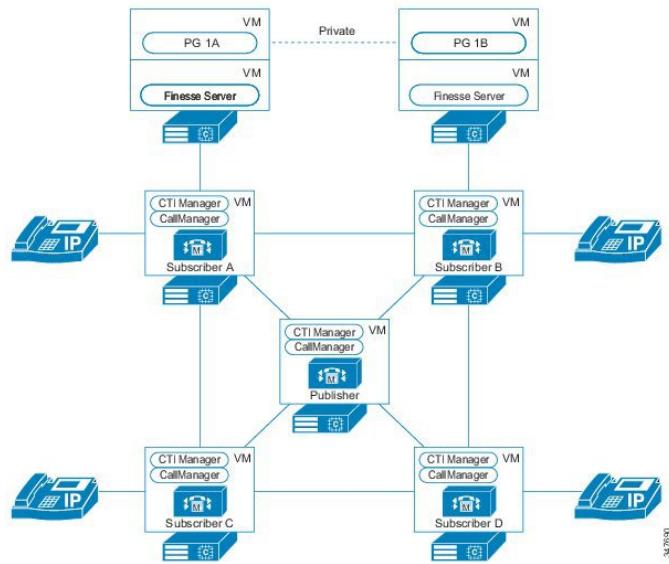
Figure 34: 2 Agent PG Pairs for Unified Communications Manager Cluster



- Deploy a single Agent PG pair for the entire cluster. In this deployment, only the CTI Managers on two subscribers have a direct link to the Agent PG pair. You can spread agent phone registrations among all the subscribers, not just the subscribers that directly connect to the Agent PG pair. The other subscribers

send and receive Agent PG messages through a connected subscriber. The following diagram shows a deployment with a single Agent PG pair that is connected to a cluster with four subscribers.

Figure 35: Single Agent PG Pair for Entire Unified Communications Manager Cluster

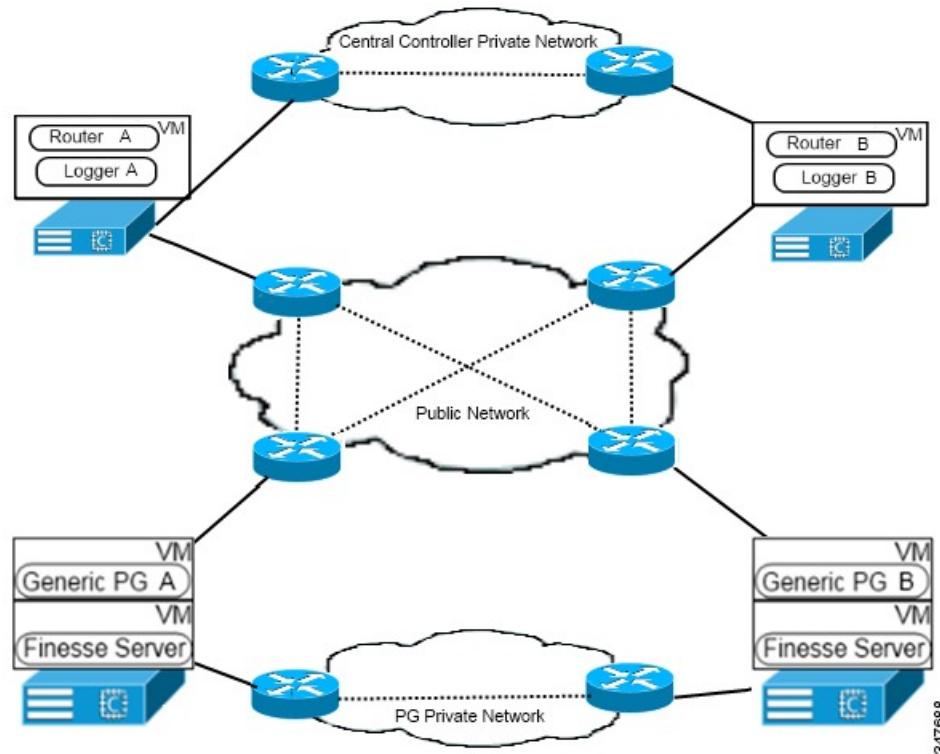


This model reduces the PG server count for the contact center, but you must adhere to the PG sizing constraints. You can also create teams that span across many subscribers because all calls pass through a single Agent PG pair. These teams enable supervisors to monitor agent phones registered across the cluster. However, this deployment can impose slightly higher resource utilization on the cluster.

Central Controller Design Considerations

You can geographically distribute redundant Unified CCE servers or locate them at the same physical site. In a production deployment, the Call Router, Logger, and Database Server must connect over a private network that is isolated from the public network.

Figure 36: High Availability Design for Central Controller



For simplicity, this figure represents the Central Controller as a single server. Most designs have a set of servers sized to support the Unified CCE agent count and call volume. The Central Controller includes the following redundant servers:

- Call Router provides call routing instructions based on real-time conditions. Redundant pairs of Call Routers maintain synchronized records in memory.
- Logger and Database Server is the repository for all configuration and scripting information as well as recent historical data that the system collects.

The redundant Logger pairs connect with the redundant Call Router pairs: Side A to Side A, and Side B to Side B. Each Call Router reads and writes data only to its connected Logger. Because the Call Routers run in lock step through synchronized messages, the data produced and written to both Loggers is identical.

**Note**

Some designs install the Call Router and Logger on the same VM. That combination is sometimes called a Rogger.

When the servers are located at the same site, configure them with a second virtual NIC for the private network connection and isolate the private connections. When the servers are geographically separated, use a separate WAN connection for the private network. To avoid a single point of failure in the network, do not use the same circuits or network gear as for the public network.

Common Processes That Support Failovers

Failover scenarios can arise either from software component failures or from network failures. The following sections describe common processes that support component failover behavior.

Failure Detection Methods

Unified CCE uses the Message Delivery Subsystem (MDS) to send synchronization messages. The private network uses TCP keepalive messages that are generated at 100-ms intervals. If no TCP keepalive messages arrive for 500 ms, the system decides that either a network or component failure occurred.

The public network uses the UDP heartbeat mechanism between PGs and the Central Controller. Redundant components generate UDP heartbeats at 100-ms intervals. Routers and PGs generate UDP heartbeats at 400-ms intervals. In both cases, the system decides a failure occurred after missing five UDP heartbeats.

Device Majority

Device majority determines whether a Call Router enters a disabled state. The Call Router checks for device majority when it loses its connection with its redundant Call Router. Each Call Router determines device majority for itself. None, one, or both Call Routers can have device majority simultaneously.

To have device majority, a Call Router must meet one of these conditions:

- The Call Router is the Side A router and it can communicate with *at least half* of its total enabled PGs.
- The Call Router is the Side B router and it can communicate with *more than half* of its total enabled PGs.

PG Weight

During a failover for a private link failure, a weighted value determines which PG becomes the enabled PG. The number and type of active components on each side determines the weighted value of the PG. The weight assigned to each component reflects the recovery time of that component and the disruption to the contact center when the component is down. Agent PIMs have higher weights than VRU PIMs and the CTI Server. The component weights are not configurable.

Record Keeping During Failovers

The call data that gets records during a failover depends on which component fails. Depending on the failure condition, some call data is lost. The router can lose access to in-progress calls because of the failure. The in-progress calls are still active, but the Call Router responds as if the calls have dropped. In most cases, the Agent PG creates a Termination Call Detail (TCD) record in the Unified CCE database at this point.

Calls that are already connected to an agent can continue during a failover. The Agent PG creates another TCD record for such calls when they end.

Call Survivability

Call survivability during failovers varies depending on your deployment and which components fail:

- In Unified CVP deployments, the routing dialog in the Central Controller stops and calls under Unified CVP control get treatment from the survivability TCL script in their ingress Voice Gateways. If the survivability scripts redirect the calls to another active Unified CCE component, the call appears as a "new call" to the system with no relationship to the original call for reporting or tracking purposes.
- In Unified IP IVR deployments, survival scripting on the VRU can do some call treatment as the call ends, but does not send the call back to another active Unified CCE component.
- During Agent PG failures, calls survive. An agent with a hard phone can manage the call if the deployment allows. If the agent signs back in while the call is still active, the agent gets reconnected to the in-progress call.

Unified CCE Failovers During Network Failures

Network failures simultaneously affect any components that send traffic across the affected network. Unified CCE components use both private and public network links to communicate.

The traffic on the private network performs these functions:

- State transfer during component startup
- Synchronization of redundant pairs of Call Routers
- Synchronization of redundant pairs of PGs

The public network carries the rest of the traffic between the contact center components: voice data, call context data, and reporting data. The public network includes all the public network links between the Unified CCE components.



Note

In virtualized contact centers, network failures can arise from failures in the virtual environment, like a virtual NIC, or from failures of physical resources.

Response to Private Network Failures

When a private network fails, the contact center quickly enters a state where, depending on system topology, one or both sides transition into isolated-enabled operation. The isolated operation continues until the Call Routers detect the restoration of the private network. The redundant pairs of Routers and PGs then resynchronize and resume normal operation.

Assume that Side A is the pair-enabled Call Router and Side B is the pair-disabled Call Router. When the private network fails, the Side A Call Router behaves as follows:

- If the Side A Call Router has device majority, the Call Router transitions to the isolated-enabled state and continues handling traffic.
- If the Side A Call Router does not have device majority, the Call Router transitions to the isolated-disabled state and stops processing traffic.

When the private network fails, the Side B Call Router behaves as follows:

- If the Side B Call Router does not have device majority, the Call Router transitions to the isolated-disabled state and does not process traffic.
- If the Side B Call Router does have device majority, the Call Router enters a test state. The Router instructs its enabled PGs to contact the Side A Call Router over the public network. Then, the Side B Call Router responds as follows:
 - If no PG can contact the Side A Call Router to determine its state, the Side B Call Router transitions to the isolated-enabled state and begins handling traffic. This case can result in both Side A and Side B running in isolated-enabled state.
 - If any PG contacts the Side A Call Router and finds the Call Router in the isolated-disabled state, the Side B Call Router transitions to the isolated-enabled state and begins handling traffic.
 - If any PG contacts the Side A Call Router and finds the Call Router in the isolated-enabled state, the Side B Call Router transitions to the isolated-disabled state and does not process traffic.

During Call Router failover processing, any Route Requests that are sent to the Call Router are queued until the surviving Call Router is in isolated-enabled state. A Call Router failure does not affect any in-progress calls that have already been routed to a VRU or an agent.

The Logger shuts down when its Call Router goes idle. Each Logger communicates only with its own Call Router. If the private network connection is restored within 12 hours, the isolated-enabled Call Router's Logger uses its data to resynchronize the other Logger. If the private network connection remains down for more than 12 hours, manually resynchronize the Loggers using the process described in the *Administration Guide for Cisco Unified Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-maintenance-guides-list.html>.

In each redundant pair of PGs, there is also an enabled PG and a disabled PG. At system start, the first PG to connect becomes the enabled PG. However, after a private network failure, the PG with the greatest weight in the redundant pair becomes the enabled PG. The other PG becomes the disabled PG.

Response to Public Network Failures

Highly available networks generally include redundant channels for the public network. When one channel fails, another channel takes over seamlessly. The contact center detects a public network failure only when all channels fail between two components.

**Note**

In contact centers without redundant public networks, the contact center detects a failure when the single channel fails.

How the contact center responds to a public network failure depends on number and function of the sites and how the sites are linked. The following sections look at some of the more common or significant scenarios.

Failures Between Unified Communications Managers

The scenario that can cause the most problems involves the subscribers losing their public link. Because the functioning private network keeps the Call Routers and Agent PGs in synch, the Call Routers can still detect all agent devices. In this situation, a Call Router can choose an agent device that is registered on the subscriber on the other side of the public network failure. The local CVP cannot pass the connection information to the agent device on the other side of the public network failure. The call fails, but the routing client marks the call as routed to the agent device on the remote subscriber.

Failures Between Data Centers in Clustering over WAN

In a clustering over the WAN deployment, you need a highly available, highly resilient WAN with low latency and sufficient bandwidth. The public network is a critical part of Unified CCE fault tolerance. A highly available WAN is fully redundant with no single points of failure, usually across separate carriers. During a partial failure of the WAN, the redundant link needs the capability to handle the full load for the data centers within the QoS parameters. As an alternative to redundant WANs, you can employ a SONET fiber ring. For more information about designing highly available, highly resilient WANs, see the [Overall WAN Architecture page](#) in the Cisco Design Zone.

If the public network fails between the data center locations, the system responds in this manner:

- 1 The Unified Communications Manager subscribers detect the failure. The subscribers continue to function locally with no impact to local call processing and call control. However, any calls that were set up over the public network fail.
- 2 The Call Routers and Agent PGs detect the failure. The Agent PGs automatically realign their data communication stream to their local Call Router. The local Call Router then passes data to the Call Router on the other side over the private network to continue call processing. The altered data path does not cause a failover of the Agent PGs or the Call Router.

The impact of the public network failure on agents depends on where their phones and desktops registered:

- The most common case is that the agent desktop and agent phone are both registered to the Agent PG and a subscriber on the same side (Side A for this example). When the public link between the data centers fails, the agent can continue handling calls normally.
- In some cases, the agent desktop (Side A for this example) and the agent phone (Side B for this example) can end up registered on different sides. In those cases, the CTI Manager directs phone events over the

Response to Failures of Both Networks

public network to the Agent PG on the opposite side. When the public network between the data centers fails, the phone does not rehome to Side A of the cluster. The phone remains operational on Side B. The Agent PG on Side A cannot detect this phone. Because the Unified CCE can no longer direct calls to the agent phone, Unified CCE automatically signs out the agent.

- Normally, the redundant desktop server pair load balances agent desktop connections. So, half of the desktops register on a desktop server that connects to the active CTI Server Peripheral Gateway (CG) across the public network. When the public network fails, the desktop server loses connection with the remote CG. The desktop server disconnects the active agent desktops to force them to rehome to the redundant desktop server at the remote site. The agent desktop automatically uses the redundant desktop server. The agent desktop remains disabled until it connects to the redundant desktop server.

Failures to Agent Sites in Clustering over WAN

The Unified CCE model for clustering over the WAN assumes that the Unified CCE agents are remotely located at multiple sites. Each agent site requires access to both data centers through the public network for redundancy. In a complete network failure, these connections also provide basic SRST functionality, so that the agent site can still make emergency (911) calls.

If the Unified CCE agent site loses the public network connection to one of the data centers, the system responds in this manner:

- 1 Any IP phones that are homed to the Unified Communications Manager subscribers at the disconnected data center automatically rehome to subscribers at the other data center. To use the rehoming behavior, configure a redundancy group.
- 2 Agent desktops that are connected to the desktop server at that disconnected data center automatically realign to the redundant server at the other data center. (Agent desktops are disabled during the realignment process.)

If the Unified CCE agent site loses the public network connection to both of the data centers, the system responds in this manner:

- 1 The local Voice Gateway (VG) detects the failure of the communications path to the cluster. The VG then goes into SRST mode to provide local dial-tone functionality.
- 2 With Unified CVP, the VGs detect the loss of connection to the Unified CVP Server. Then, the VGs execute their local survivability TCL script to reroute the inbound calls.
- 3 If an active call came in to the disconnected agent site on a local PSTN connection, the call remains active. But, the Agent PG loses access to the call and creates a TCD record.
- 4 The Finesse server (or CTI OS server) detects the loss of connectivity to the agent desktop and automatically signs the agent out of the system. While the IP phones are in SRST mode, they cannot function as Unified CCE agents.

Response to Failures of Both Networks

Individually, parts of the public and private networks can fail with limited impact to the Unified CCE agents and calls. However, if both of these networks fail at the same time, the system retains only limited functionality. This failure is considered catastrophic. You can avoid such failures by careful WAN design with built-in backup and resiliency.

A simultaneous failure of both networks within a site shuts down the site.

If both the public and private networks simultaneously fail between two sites, the system responds in this manner:

- 1 Both Call Routers check for device majority. Each router enters isolated-enabled mode if the router has device majority or isolated-disabled mode if the router does not have device majority.
- 2 The PGs automatically realign their data communications, if necessary, to their local Call Router. A PG that cannot connect to an active Call Router becomes inactive.
- 3 The Unified Communications Manager subscribers detect the failure and continue to function locally with no impact to local call processing and call control.
- 4 Any in-progress calls that are sending active voice path media over the public WAN link fail with the link. When the call fails, the PG creates a TCD record for the call.
- 5 In a clustering over the WAN deployment, the Unified Communications Manager subscribers on each side operate with access only to local components.
- 6 The Unified CCE call routing scripts automatically route around the off-line devices using peripheral-on-line status checks.
- 7 Agents with both their phones and desktops registered with local subscribers are not affected. All other agents lose some or all functionality while their phones and desktops rehome. Those agents might also find themselves signed out, depending on the exact system configuration.
- 8 Unified CCE does not route new calls that come into the disabled side. But, you can redirect or handle those calls with the standard Unified Communication Manager redirect on failure for their CTI route points or with the Unified CVP survivability TCL script in the ingress Voice Gateways.

Unified CCE Failovers During Single-Component Failures

Unified CCE components run in redundant pairs. If a component fails, its counterpart takes over processing. These failovers happen automatically and often have little or no impact on active agents. The following sections discuss how Unified CCE responds to component failures.



Note

The following sections use "Agent PG" to describe the Peripheral Gateway that has a Unified Communications Manager PIM installed. Different Unified CCE deployments can install the Unified Communications Manager PIM in various combinations. For example, you can install the PIM on a standalone PG or together with other components on a Generic PG.

Agent PG Fails

This scenario shows recovery from a PG Side A failure.

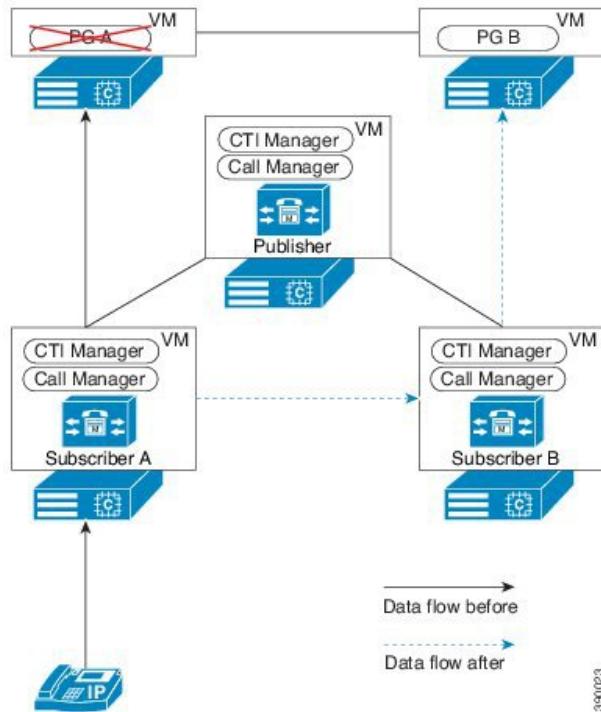
The following conditions apply to this scenario:

- Unified Communications Manager subscriber A has the primary CTI Manager.
- For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.

Subscriber Without CTI Manager Link to Agent PG Fails

The following figure shows a failure on PG Side A and a failover to PG Side B. All CTI Manager and Unified Communications Manager services continue running normally.

Figure 37: Agent PG Side A Fails



Failure recovery occurs as follows:

- 1 PG Side B detects the failure of PG Side A.
- 2 PG Side B registers all dialed numbers and phones. Call processing continues through PG Side B.
- 3 Phones and gateways stay registered and operational with subscriber A; they do not fail over.
- 4 The in-progress calls remain active on agent phones, but the agents cannot use phone services, like transfers, until the agents sign back in.
- 5 During the failover to PG Side B, the states of unoccupied agents and their desktops can change depending on their configuration. Options for three-party calls can be affected. In some cases, agents have to sign back in or manually change their state after the failover completes.
- 6 After recovery from the failure, PG Side B remains active and uses the CTI Manager on subscriber B. The PG does not fail back to Side A, and call processing continues on the PG Side B.

Subscriber Without CTI Manager Link to Agent PG Fails

In a Unified Communications Manager cluster supporting 2000 agents, you have one Unified Communications Manager publisher and four Unified Communications Manager subscribers. Each subscriber can support 500 agents. Each Agent PG can support only one CTI Manager connection. While each subscriber has a CTI

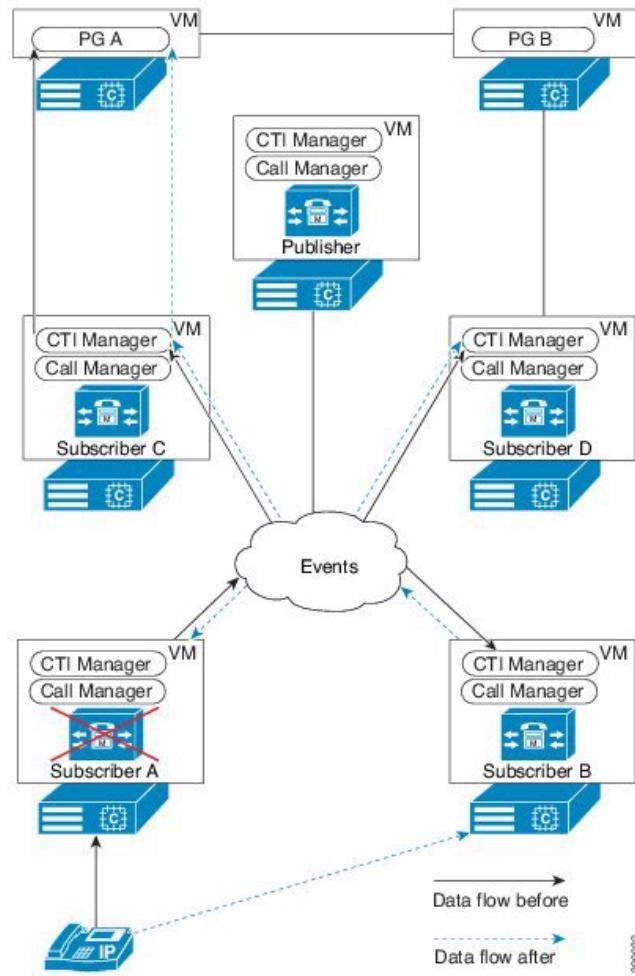
Manager, only two subscribers can connect to the Agent PGs. You would have to add another pair of Agent PGs to enable all the subscribers in this cluster to connect directly to an Agent PG.

The following figure shows a failure on subscriber A, which does not have a direct connection to an Agent PG.

The following conditions apply to this scenario:

- For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.
- Subscribers C and D connect to the Agent PGs and each run a local instance of CTI Manager to provide JTAPI services for the PGs.

Figure 38: Unified Communications Manager Without Link to Agent PG Fails



Failure recovery occurs as follows:

- 1 If subscriber A fails, its registered phones and gateways rehome to the backup subscriber B.

CTI Manager with Agent PG Link Fails

- 2 Agent PG Side A remains active and connected to the CTI Manager on subscriber C. The PG does not fail over, because the JTAPI-to-CTI Manager connection has not failed. But, the PG detects the phone and device registrations automatically switching from subscriber A to subscriber B.
- 3 Call processing continues for any devices that are not registered to subscriber A.
- 4 While the agent phones are not registered, the Agent PG disables the agent desktops. This response prevents the agents from using the system without a subscriber connection. The Agent PG signs the agents out during this transition to avoid routing calls to them.
- 5 Call processing resumes for the phones after they reregister with their backup subscriber.
- 6 In-progress calls continue on phones that were registered to subscriber A, but the agents cannot use phone services, like transfers, until the agents sign back in.
- 7 When the in-progress call ends, that phone reregisters with the backup subscriber. The Agent PG signs the agents out during this transition to avoid routing calls to them.
- 8 When subscriber A recovers, phones and gateways rehome to it. You can set up the rehoming on subscribers to return groups of phones and devices gracefully over time. Otherwise, you can require manual intervention during a maintenance window to redistribute the phones to minimize the impact to the call center. During this rehoming process, the CTI Manager notifies the Agent PG of the registrations switching from subscriber B back to the original subscriber A.
- 9 Call processing continues normally after the phones and devices return to their original subscriber.

CTI Manager with Agent PG Link Fails

In a Unified Communications Manager cluster supporting 2000 agents, you have one Unified Communications Manager publisher and four Unified Communications Manager subscribers. Each subscriber can support 500 agents. Each Agent PG can support only one CTI Manager connection. While each subscriber has a CTI Manager, only two subscribers can connect to the Agent PGs. You would have to add another pair of Agent PGs to enable all the subscribers in this cluster to connect directly to an Agent PG.

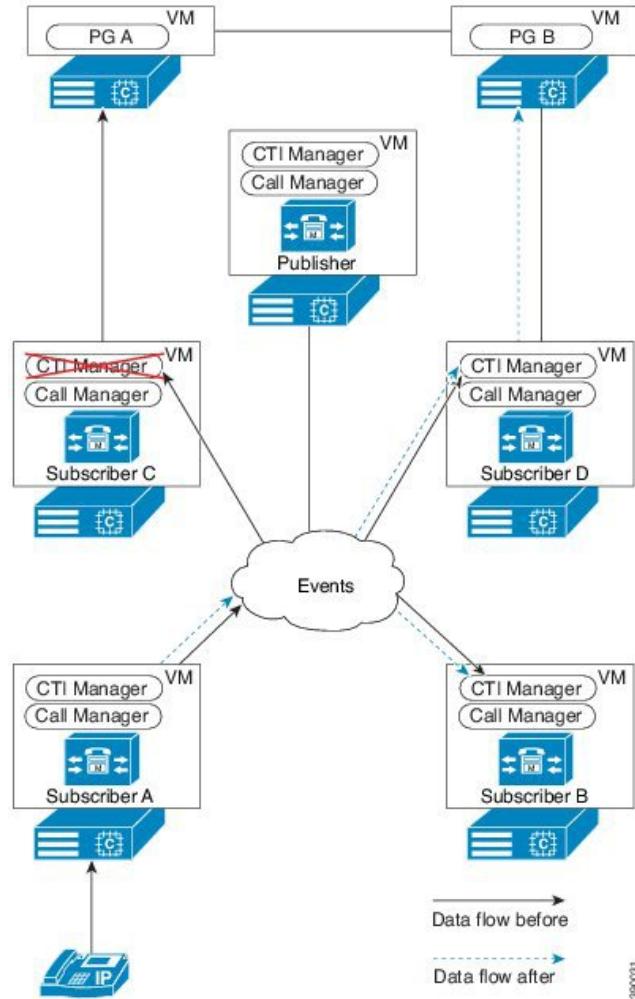
The following figure shows the failure of a CTI Manager with a connection to the Agent PG. Only subscribers C and D are configured to connect to the Agent PGs.

The following conditions apply to this scenario:

- For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.

- The CTI Managers on subscribers C and D provide JTAPI services for the Agent PGs.

Figure 39: CTI Manager with Agent PG Connection Fails



Failure recovery occurs as follows:

- When the CTI Manager on subscriber C fails, the Agent PG Side A detects the failure and induces a failover to PG Side B.
- Agent PG Side B registers all dialed numbers and phones with the CTI Manager on subscriber D and call processing continues.
- In-progress calls stay active, but the agents cannot use phone services, like transfers, until the agents sign back in.
- When the CTI Manager on subscriber C recovers, Agent PG Side B continues to be active and uses the CTI Manager on subscriber D. The Agent PG does not fail back in this model.

Unlike Unified CVP, Unified IP IVR depends on the CTI Manager for call control. In Unified IP IVR deployments, failure of a CTI Manager with an Agent PG connection causes the Unified IP IVR JTAPI

Voice Response Unit PG Fails

subsystem to shut down. This shutdown causes the Unified IP IVR server to drop all voice calls that the server is processing.

Then, the JTAPI subsystem restarts and connects to the CTI Manager on the backup subscriber. The Unified IP IVR reregisters all the CTI ports that are associated with the Unified IP IVR JTAPI user. After all the Unified Communications Manager devices are successfully registered, the server resumes its Voice Response Unit (VRU) functions and handles new calls.

Voice Response Unit PG Fails

When a Voice Response Unit (VRU) PG fails, calls in progress or queued in Unified CVP do not drop. The Survivability TCL script in the Voice Gateway redirects the calls to a secondary Unified CVP or a number in the SIP dial plan, if available.

In deployments using Unified IP IVR, all calls that are currently queued or in treatment drop unless the deployment includes one of the following:

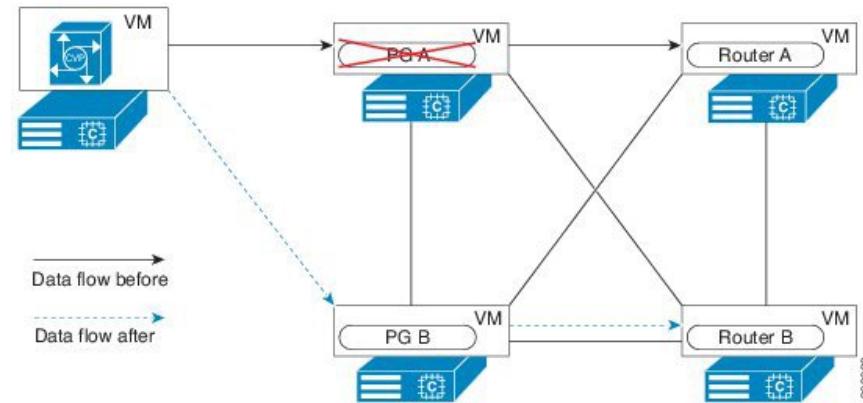
- A default script application defined for Unified IP IVR
- Recovery numbers defined for the CTI Ports in Unified Communications Manager.

After failover, the redundant VRU PG connects to the Unified CVP or Unified IP IVR and begins processing new calls. On recovery of the failed VRU PG side, the currently running VRU PG continues to operate as the active VRU PG. Redundant VRU PGs enable Unified CVP or Unified IP IVR to function as an active queue point or to provide call treatment.

**Note**

Unless Unified IP IVR deployments have a redundant VRU PG, a VRU PG failure still blocks the use of a functional Unified IP IVR server.

Figure 40: VRU PG Fails



Logger Fails

The Unified CCE Logger and Database Server maintains the system database for the configuration (agent IDs, skill groups, call types) and scripting (call flow scripts). The server also maintains the recent historical

data from call processing. The Loggers receive data from their local Call Router. Because the Call Routers are synchronized, the Logger data is also synchronized.

A Logger failure has no immediate impact on call processing. The redundant Logger receives a complete set of call data from its local Call Router. If the failed Logger is restored within 12 hours, the Logger automatically requests all the transactions for when it was off-line from the backup Logger. The Loggers maintain a recovery key that tracks the date and time of each entry recorded in the database. The redundant Logger uses these keys to identify the missing data.

If the Logger was off line for more than 12 hours, the system does not automatically resynchronize the databases. In this case, the system administrator can manually resynchronize the Loggers using the Unified ICMDA application. The manual process allows you to choose a convenient time to transfer the data across the private network.

The Logger replication process sends data from the Logger database to the HDS database on the Administration and Data Servers. The replication process also automatically replicates each new row that is written to the Logger database when the synchronization takes place.

In deployments that use Cisco Outbound Option, the Campaign Manager is loaded only on the primary Logger. If that platform is out of service, any outbound calling stops while the Logger is down.

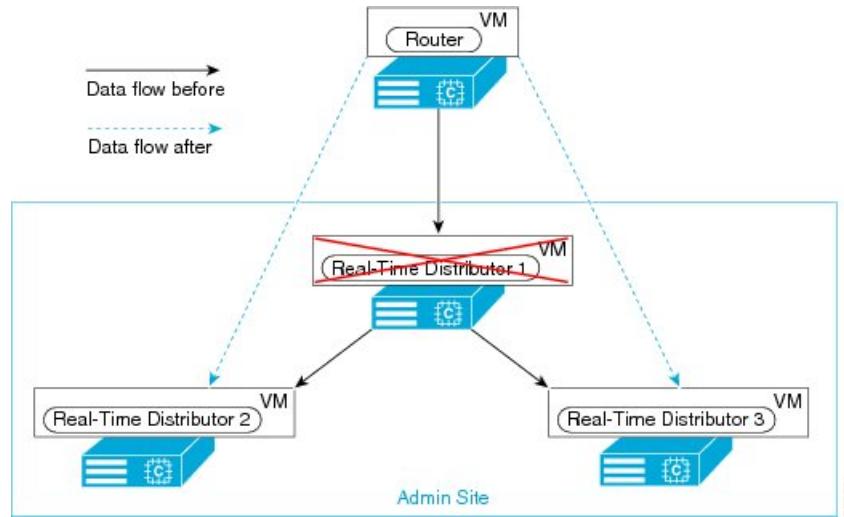
Administration and Data Server Fails

The Administration and Data Server provides the user interface to the system for making configuration and scripting changes. The server can also host the web-based reporting tool and the Internet Script Editor. Unlike other Unified CCE components, the Administration and Data Server does not operate in redundant pairs. If you want to provide redundancy for the functions on this server, you can include more Administration and Data Servers in your design. But, there is no automatic failover behavior.

The Administration and Data Server receives a real-time feed of data from across Unified CCE from the Call Router through a Real-Time Distributor. If you have several Administration and Data Servers at the same site, you can configure the Real-Time Distributors into a single Admin Site. The Admin Site has a primary distributor and one or more secondary distributors. The primary distributor registers with the Call Router and receives the real-time feed across the network from the router. The secondary distributors use the primary distributor as their source for the real-time feed. This arrangement reduces the number of real-time feeds that the router supports and saves bandwidth.

If the primary real-time distributor fails, the secondary real-time distributors register with the router for the real-time feed as shown in the following figure. Administration clients that cannot register with the primary or secondary Administration and Data Server cannot perform any tasks until the distributors are restored.

Figure 41: Primary Real-Time Distributor Fails



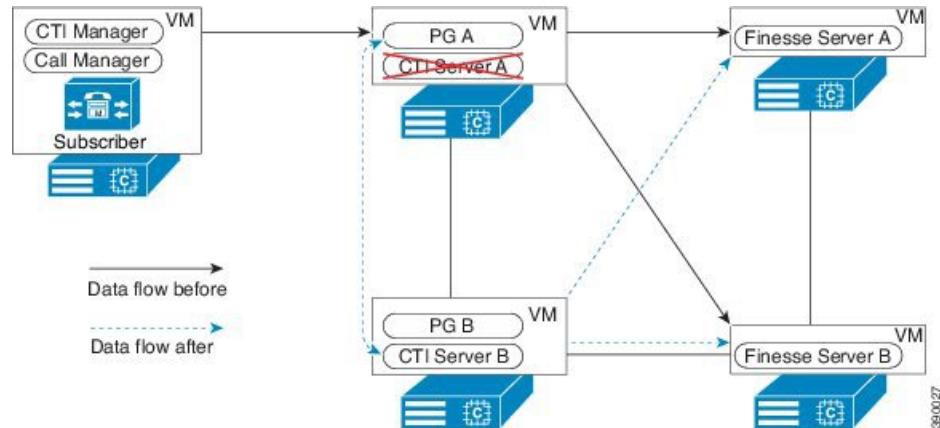
In some deployments, the Administration and Data Server also hosts an interface for the Cisco Unified Contact Center Management Portal (Unified CCMP). In those deployments, when the Administration and Data Server is down, any configuration changes that are made to the Unified CCE or Unified CCMP systems are not passed over the interface.

CTI Server Fails

The CTI Server monitors the Agent PG traffic for specific CTI messages (such as call ringing or off-hook events). The CTI Server makes those messages available to CTI clients such as the Cisco Finesse server or CTI OS server. The CTI Server also processes third-party call control messages (such as make call or answer call) from the CTI clients. The CTI Server sends those messages through the Agent PG to Unified Communications Manager for processing.

You deploy the CTI Server in redundant pairs. Each half of the redundant pair is coresident on a VM with one half of a redundant Agent PG pair. On failure of the active CTI Server, the redundant CTI Server becomes active and begins processing call events.

Figure 42: CTI Server Fails



Both the Finesse server and CTI OS server are clients of the CTI Server. The desktop server, rather than the CTI Server, maintains agent state during a failover. Both Finesse and CTI OS partially disable agent desktops when the CTI Server fails. In some cases, an agent must sign in again after the failover completes.



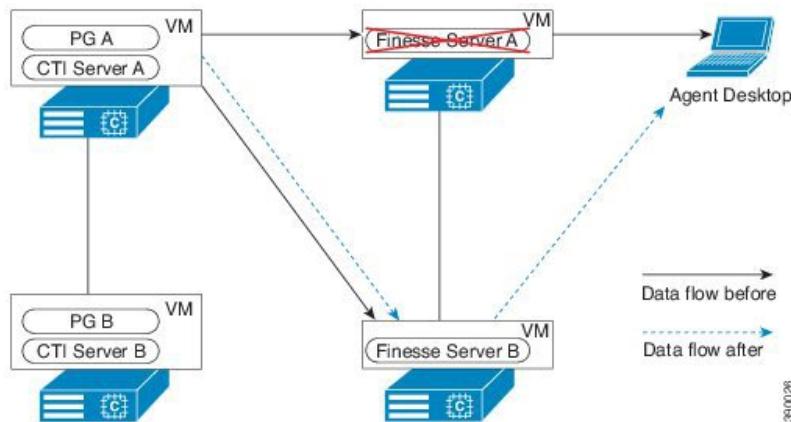
Note

If no clients are connected to the active CTI Server, a mechanism is in place to force a failover after a preset period. This failover occurs to isolate any spurious reasons that prevent the CTI clients from connecting to the active CTI Server.

Cisco Finesse Server Fails

You deploy the Cisco Finesse server in redundant pairs in dedicated virtual machines. Both Finesse servers run in active mode all the time.

Figure 43: Finesse Server Fails



When a Cisco Finesse server fails, failure recovery occurs as follows:

- 1 Agent desktops that are signed in to the server detect a loss of connection and fail over to the redundant server.
- 2 Agents are automatically signed in on the new server after the failover.
- 3 Third-party applications that use the Cisco Finesse REST API must perform the failover within their application logic to move to the redundant server.
- 4 If Cisco Tomcat failed with the Cisco Finesse server, the system attempts to restart Cisco Tomcat before restarting the Cisco Finesse server.
- 5 When the failed server restarts, new agent desktop sessions can sign in on that server. Agent desktops that are signed in on the redundant server remain on that server.

Finesse Behavior When Other Components Fail

The following sections describe Finesse behavior when other Unified CCE components fail.

Agent PG Fails

Both Finesse servers connect to the active Agent PG. If the active Agent PG fails, Finesse tries to connect to the alternate Agent PG. Finesse may go out of service briefly while attempting to connect to the alternate Agent PG. Agents may see a red banner on the Finesse desktop when it loses connection, followed by a green banner when it reconnects.

CTI Server Fails

Both Finesse servers connect to the active CTI Server. If the active CTI Server fails, Finesse tries to connect to the alternate CTI Server. Finesse may go out of service briefly while attempting to connect to the alternate CTI Server. Agents may see a red banner on the Finesse desktop when it loses connection, followed by a green banner when it reconnects.

Administration & Data Server Fails

Finesse uses the Administration & Data Server to authenticate agents. The Finesse administrator configures the settings for the primary Administration & Data Server (and optionally, the backup Administration & Data Server) in the Finesse administration user interface. If the primary Administration & Data Server fails and a backup Administration & Data Server is not configured, Finesse agents cannot sign in to the desktop. Agents who are signed in when the failover occurs can no longer perform operations on the desktop.

If the backup Administration & Data Server is configured, Finesse tries to connect to the backup server. After Finesse connects to the backup Administration & Data Server, agents can sign in and perform operations on the desktop.

CTI OS Server Fails

You deploy the CTI OS server in redundant pairs. Each half of the redundant pair is coresident on a VM with one half of a redundant Agent PG pair. Unlike the PG processes that run in hot-standby mode, both of the CTI OS server processes run in active mode all the time.

When a CTI OS server fails, failure recovery occurs as follows:

- 1 Agent desktops that are signed in to the server detect a loss of connection and fail over to the redundant server.
- 2 When the failed server restarts, new agent desktop sessions can sign in on that server. Agent desktops that are signed in on the redundant server remain on that server.

Unified CCE Failovers During Multicomponent Failures

When more than one component fails, Unified CCE might not fail over as seamlessly as during a single-component failure. The following sections discuss how Unified CCE responds to multicomponent failures.

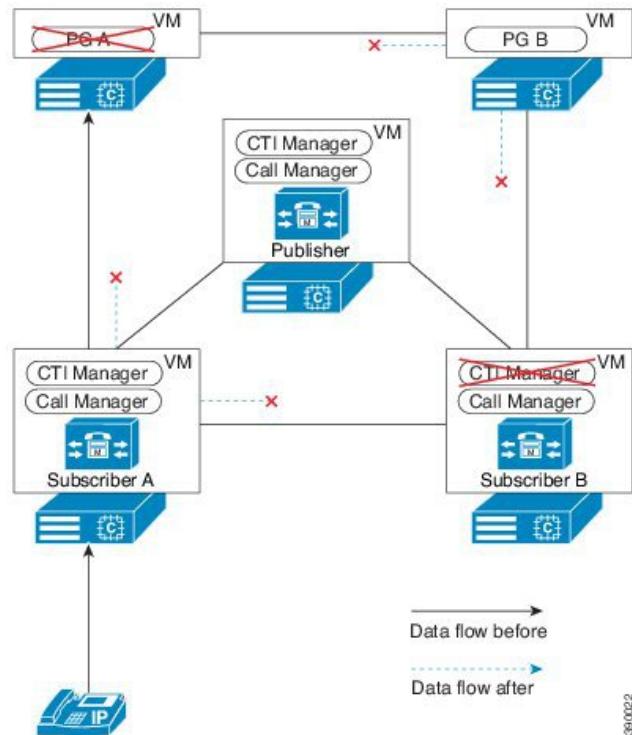
Agent PG and CTI Manager Fail

A CTI Manager connects only with its local subscriber and a single Agent PG. There is no direct communication with the other CTI Manager in the cluster. The CTI Managers are kept in synch by data from the other components.

Unified Communications Manager Subscriber and CTI Manager Fail

If the Agent PG on one side and the CTI Manager on the other side both fail, Unified CCE cannot communicate with the cluster. This scenario prevents the system from connecting to the agents on this cluster. The cluster remains disconnected until the Agent PG or the backup CTI Manager come back online.

Figure 44: Agent PG Cannot Connect to Backup CTI Manager



Unified Communications Manager Subscriber and CTI Manager Fail

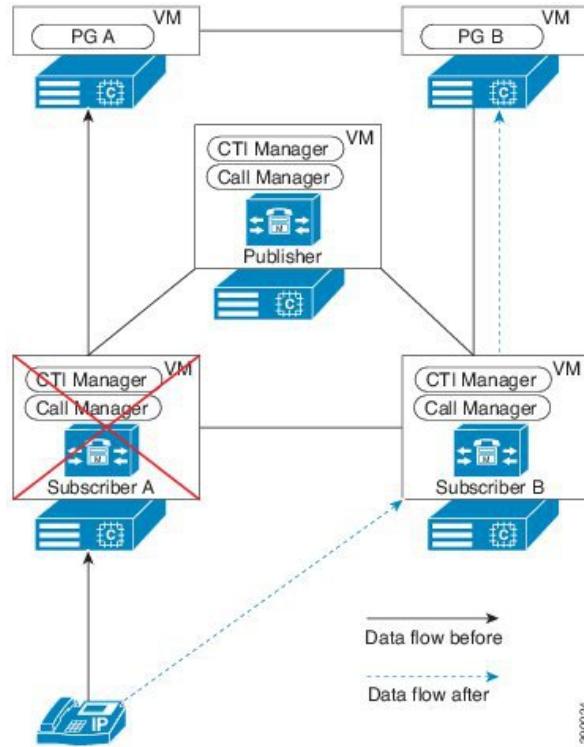
The scenario shows recovery from a complete failure of the Unified Communications Manager subscriber A server.

The following conditions apply to this scenario:

- Subscriber A has the primary CTI Manager.

- For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.

Figure 45: Unified Communications Manager and CTI Manager Fail



Failure recovery occurs as follows:

- When subscriber A fails, all inactive registered phones and gateways reregister to subscriber B.
- The in-progress calls remain active, but the agents cannot use phone services, like transfers.
- Agent PG Side A detects a failure and induces a failover to Agent PG Side B.
- Agent PG Side B becomes active and registers all dialed numbers and phones. Call processing continues.
- As each in-progress call ends, that agent phone and desktop reregister with the backup subscriber. The exact state of the agent desktop varies depending on the configuration and desktop.
- When subscriber A recovers, all idle phones and gateways reregister to it. Active devices wait until they are idle before reregistering to the primary subscriber.
- Agent PG Side B remains active using the CTI Manager on subscriber B.
- After recovery from the failure, the Agent PG does not fail back to Side A of the redundant pair. All CTI messaging is handled using the CTI Manager on subscriber B which communicates with subscriber A to obtain phone state and call information.

Other Considerations for High Availability

A Unified CCE failover can affect other parts of the solution. Some failure scenarios can result in the loss of data that other products use.

Reporting Considerations

The Unified CCE reporting feature uses real-time, 5 minute, and reporting-interval (15 or 30 minute) data to build its reporting database. At the end of each 5 minute and reporting interval, each Peripheral Gateway gathers its local data and sends it to the Call Routers. The Call Routers process the data and send the data to their local Logger for historical data storage. The Logger replicates the historical data to the HDS/DDS database.

The PGs provide buffering (in memory and on disk) of the 5-minute data and reporting-interval data. The PGs use this buffered data to handle slow network response and automatic retransmission of data after network services are restored. If both PGs in a redundant pair fail, you can lose the 5-minute data and reporting-interval data that was not sent to the Central Controller.

When agents sign out, all their reporting statistics stop. When the agents next sign in, the real-time statistics for the agents start from zero. Depending on the agent desktop and what an agent is doing during a failure, some failovers can cause the contact center to sign out agents. For more information, see the *Reporting Concepts for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-user-guide-list.html>.



CHAPTER 4

Features

- Precision Routing, page 97
- Agent Greeting, page 98
- Whisper Announcement, page 99
- Congestion Control, page 101

Precision Routing

Precision Routing is an enhanced routing strategy that you can use as an alternative to, or in conjunction with, skill group routing. Skill group routing uses predefined groups. All of the members in the group can handle a particular type of call. A supervisor manually assigns an agent to the groups for which the agent can handle incoming calls.

Precision Routing creates attribute definitions, assigns attribute values to each agent, and uses routing scripts to dynamically find the agents with the necessary attributes to handle a call. Some example attributes are a proficiency level with a language, qualification in selling a product, or the location of the agent. By accurately exposing each agent's skills through this system, you can route the calls to each agent in ways that bring more value to your business.

For more information about Precision Routing, see the *Cisco Unified Contact Center Enterprise Features Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-feature-guides-list.html>.

Precision Routing Attributes

Attributes identify a call routing requirement, such as language, location, or agent expertise. Each precision queue can have up to 10 unique attributes, and these attributes can be used in multiple terms. You can create two types of attributes: Boolean or proficiency. Use Boolean attributes to identify an agent attribute value as true or false. For example, you can create a Boston attribute specifying that the agent is located in Boston. Use proficiency attributes to establish a level of expertise in a range from 1 to 10, with 10 being the highest level of expertise. For example, a native speaker likely has a higher value for their language proficiency than a non-native speaker.

When you create a precision queue, you identify which attributes are parts of that queue and then implement the queue in scripts. When you assign new attributes to an agent, the attribute values automatically associate the agent with any precision queue with matching criteria.

Precision Routing Limitations

Precision Routing is currently available for voice agents only. Precision Routing does not support non-voice media.

Precision Routing is available only for Agent PGs on CCE.

Cisco Outbound Option does not support Precision Routing. However, agents who participate in an outbound campaign or non-voice activities (through the use of Skill Groups) can also handle inbound calls from a precision queue.

Precision Routing does not support Unified CVP's Courtesy Callback feature. The complexity of Precision Queues makes calculating accurate Estimated Wait Times difficult. Courtesy Callback depends on Estimated Wait Times. Instead of Courtesy Callback, use nonpreemptive callbacks through the Agent Request interface.

Agent Greeting

The Agent Greeting feature lets an agent record a message that plays automatically to callers when they connect to the agent. The greeting message can welcome the caller, identify the agent, and include other useful contextual information. With Agent Greeting, each caller can receive a clear, well-paced, language-appropriate, and enthusiastic introduction. This feature saves the agent from having to repeat the same introductory phrase for each call. The feature also gives the agent a moment to review the desktop screen pop-ups while the greeting plays.

Recording a greeting is much the same as recording a message for voice mail. Depending on how you set up the call center, agents can record different greetings that play for different types of callers (for example, an English greeting for English speakers or an Italian greeting for Italian speakers).

By default, greeting play is enabled when agents log in to their agent desktop. Agents can turn greeting play off and on as necessary.

Agent Greeting Phone Requirements (for Local Agents Only)

Agent Greeting is available to agents and supervisors who use IP Phones with Built-In Bridge (BIB). These agents are typically located within a contact center. Phones used with Agent Greeting must meet these requirements:

- The phones must have the BIB feature.



Note

If you disable BIB, the system attempts to use a conference bridge for agent greeting call flow and raises a warning event.

- The phones must be running firmware version CM 8.5(1) or greater. (Usually, phone firmware upgrades automatically when you upgrade your Unified Communications Manager installation.)

- See the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE for the list of supported Cisco Unified Call Center phone models.

Agent Greeting Functional Limitations

Agent Greeting is subject to these limitations.

- Agent Greeting does not support outbound calls made by an agent. The announcement plays for inbound calls only.
- Only one Agent Greeting file plays per call.
- Supervisors cannot listen to agent recorded greetings.
- Agent Greetings do not play when the router selects the agent through a label node.
- The default CTI OS Toolkit Agent desktop includes the Agent Greeting buttons. If you do not set up Agent Greeting, the Agent Greeting buttons do not execute any functionality. If you use the default desktop but do not plan to use Agent Greeting, you should remove the buttons.
- Agent Greeting supports Silent Monitoring (CTI OS and Unified CM-based) with this exception: For Unified-CM based Silent Monitoring, supervisors cannot hear the greetings themselves. If a supervisor clicks the Silent Monitor button in their CTI OS desktop while a greeting is playing, a message displays stating that a greeting is playing and to try again shortly.

Whisper Announcement with Agent Greeting

You can use Agent Greeting with the Whisper Announcement feature. Here are some things to consider when using them together:

- On the call, the Whisper Announcement always plays first.
- To shorten your call-handling time, use shorter Whisper Announcements and Agent Greetings than if you were using either feature by itself. A long Whisper Announcement followed by a long Agent Greeting equals a long wait before an agent actively handles a call.
- If you use a Whisper Announcement, your agents probably handle different types of calls: for example, “English-Gold Member-Activate Card,” “English-Gold Member-Report Lost Card,” “English-Platinum Member-Account Inquiry.” Therefore, you may want to ensure that greetings your agents record are generic enough to cover the range of call types.

Whisper Announcement

Whisper Announcement plays a brief, prerecorded message to an agent just before the agent connects with each caller. The announcement plays only to the agent; the caller hears ringing (based on existing ring tone patterns) while the announcement plays.

The content of the announcement can contain information about the caller that helps prepare the agent to handle the call. The information can include caller language preference, choices the caller made from a menu (Sales, Service), customer status (Platinum, Gold, Regular), and so on.

After you enable Whisper Announcement, you specify which announcements to play in the call routing scripts. The script determines which announcement to play based on various inputs, such as the dialed number, a customer ID look up in your customer database, or selections the caller made from a VRU menu.

For more information about Whisper Announcement, see *Cisco Unified Contact Center Enterprise Features Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-feature-guides-list.html>.

Whisper Announcement Audio File

You store and serve your Whisper Announcement audio files from the Cisco Unified Contact Center Enterprise (Unified CCE) media server. This feature supports only the wave (.wav) file type. The maximum play time for a Whisper Announcement is subject to a timeout. Playback terminates at the timeout regardless of the actual length of the audio file. The default timeout is 15 seconds. In practice, you may want your messages to be much shorter than that, 5 seconds or less, to shorten your call-handling time.

Whisper Announcement with Transfers and Conference Calls

When an agent transfers or initiates a conference call to another agent, the second agent hears an announcement if the second agent's number supports Whisper Announcement. In the case of consultative transfers or conferences, while the whisper plays, the caller hears whatever normally plays during hold. The first agent hears ringing. In the case of blind transfers, the caller hears ringing while the whisper announcement plays.

Whisper Announcement Functional Limitations

Whisper Announcement is subject to these limitations:

- Announcements do not play for outbound calls made by an agent. The announcement plays for inbound calls only.
- For Whisper Announcement to work with agent-to-agent calls, use the SendToVRU or TranslationRouteToVRU node before you transfer the call to the agent. Transfer the call to Unified CVP before you transfer the call to another agent. Then, Unified CVP can control the call and play the announcement, regardless of which node transfers the call to Unified CVP.
- Announcements do not play when the router selects the agent through a label node.
- CVP Refer Transfers do not support Whisper Announcement.
- Whisper Announcement supports Silent Monitoring (CTI OS and Unified CM-based) with this exception: For Unified Communications Manager-based Silent Monitoring, supervisors cannot hear the announcements themselves. The supervisor desktop dims the Silent Monitor button while an announcement plays.
- Only one announcement can play for each call. While an announcement plays, you cannot put the call on hold, transfer, or conference; release the call; or request supervisor assistance. These features become available again after the announcement completes.
- The codec settings for Whisper Announcement recording and the agent's phone must match. For example, if Whisper Announcement is recorded in G.711 ALAW, the phone must also be at G.711 ALAW. If Whisper Announcement is recorded in G.729, the phone must support or connect using G.729.

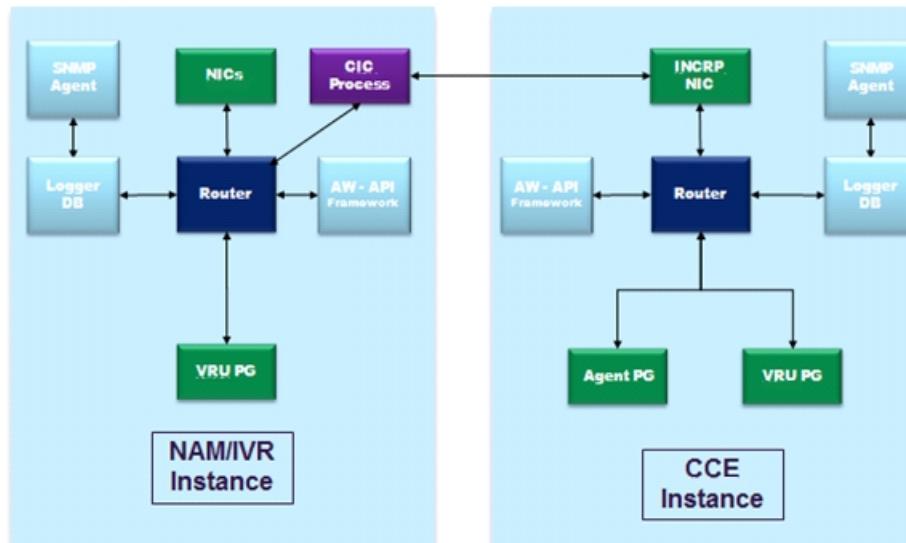
Congestion Control

The Congestion Control feature provides protection to the Central Control Router from overload conditions, due to high call rates. The main objective of congestion control is to keep the system running close to its rated capacity, when faced with extreme overload. The goal is to give satisfactory service to a smaller percentage of calls (your capacity) rather than a highly degraded service to all the calls, during an overloaded condition. This is achieved by restricting capacity on the system by rejection calls by the Routing Clients at the call entry point. Throttling the capacities ensures the service of those calls routed is successful, meaning no lates or timeouts.

The measured CPS at router is the trigger for identifying congestion in the system. For a given deployment, the supported capacity is set when the deployment type is selected. The router measures the new incoming call requests from all the routing clients and computes moving weighted average over a sample duration. If the average CPS exceeds beyond the thresholds for each level, the congestion levels are changed along with the reduction percentage. The congestion control algorithm utilizes three congestion levels and rejects/treats the incoming calls as per the reduction percentage for that level. The change in the congestion level is notified to the routing clients. The routing clients start rejecting/treating calls based on reduction percentage.

For every instance of ICM/CCE deployment, you select the type of deployment. As part of the deployment type selection, the CPS capacity is automatically set. In a multiple instance deployment like Network Application Manager/Customer ICM (NAM/CICM) as shown in the following diagram, the congestion control is done based on the call rate measured at each instance. The calls rejected are treated based on the congestion level in that instance. For example, if a call arrives at NAM instance through routing clients, if NAM router is congested, then the routing clients apply the congestion logic and reject/treat the calls. Similarly if the CICM is congested, and the NAM does a CICM lookup for routing any calls, such calls are subjected to congestion control at the INCRP routing client at the CICM instance.

Figure 46: NAM/CICM deployment



Deployment Types

After upgrading or installing the system, configure the system to a valid deployment type. The following table lists the supported deployment types with guidelines for selecting a valid deployment type. You can find the requirements referred to in the table on the *Virtualization for Unified CCE DocWiki* at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE.

Table 12: Deployment Types

Deployment Type Code	Deployment Name	Guidelines for Selection
0	Not Specified	This is a system default deployment type. User cannot select this option; is the default setting after fresh install/upgrade.
1	NAM	Select this deployment type for NAM instance in a Contact Director deployment. The system should be distributed deployment with Router and Logger installed on different VMs, which meets the specified requirements. No agents are allowed in this deployment type. If agents are configured and logged in, the capacity is adjusted to maximum capacity of an Enterprise Contact Center (Unified CCE 12000 Agents Router/ Logger).
2	IVR-ICM	Select this deployment type for ICM instance which is dedicated to self service call flows using Unified CVP or third party VRU systems. The system should be distributed deployment with Router and Logger installed on different VMs which meets the specified requirements. No agents are allowed in this deployment type. If agents are configured and logged in, the capacity is adjusted to maximum capacity of an Enterprise Contact Center (Unified CCE 12000 Agents Router/ Logger).

Deployment Type Code	Deployment Name	Guidelines for Selection
3	NAM Rogger	Select this deployment type for NAM instance in a Contact Director deployment. The Router and Logger co-located on a single VM meet the specified requirements. No agents are allowed in this deployment type. If agents are configured and logged in, the capacity is adjusted to maximum capacity of an Enterprise Contact Center (Unified CCE 12000 Agents Router/ Logger).
4	ICM Router/Logger	Select this deployment for type ICM Enterprise system where both Legacy TDM ACD PGs and CCE PGs are deployed. The system should be distributed deployment with Router and Logger installed on different VMs which meet the specified requirements.
5	Unified CCE 8000 Agents Router/Logger	Select this deployment for type CCE Enterprise system where only CCE PGs are deployed. The system should be distributed deployment with Router and Logger installed on different VMs which meet the specified requirements for 8000 CCE agents.
6	Unified CCE 12000 Agents Router/Logger	Select this deployment type for CCE Enterprise system where only CCE PGs are deployed. The system should be distributed deployment with Router and Logger installed on different VMs which meet the specified requirements for 12000 CCE agents.
7	Packaged CCE : CCE-PAC-M1	Select this deployment type when Packaged CCE: CCE-PAC-M1 is being deployed.

Deployment Type Code	Deployment Name	Guidelines for Selection
8	ICM Rogger	Select this deployment type for ICM Enterprise system where both Legacy TDM ACD PGs and CCE PGs are deployed. The Router and Logger are co-located on a single VM which meets the specified requirements.
9	Unified CCE 4000 Agents Rogger	Select this deployment type for CCE Enterprise system where only CCE PGs are deployed. The Router and Logger are co-located on a single VM which meets the specified requirements.
10	Packaged CCE : CCE-PAC-M1 Lab Only	Select this deployment type when Packaged CCE: CCE-PAC-M1 Lab is being deployed.
11	HCS-CC 1000 Agents	Select this deployment type when HCS-CC 1000 Agents is being deployed.
12	HCS-CC 500 Agents	Select this deployment type when HCS-CC 500 Agents is being deployed.
13	Unified CCE 450 Agents Progger	For all lab deployments, select this type although the Router, Logger, and PG are not on the same VM. Note This deployment type is not supported for production systems.
14	HCS-CC 4000 Agents	Select this deployment for Unified CCE system where only Unified CCE PGs are deployed. This deployment is for distributed systems with the Router and Logger on other servers that meet the requirements for 4000 Unified CCE agents.
15	HCS-CC 12000 Agents	Select this deployment type when HCS-CC 12000 Agents is being deployed.

**Note**

It is very important to set the proper deployment type for your solution during the configuration. If you select the wrong deployment type, your solution is either unprotected from overload or it rejects and treats calls based on incorrect capacity settings.

Congestion Treatment Mode

There are five options available to handle the calls that are to be rejected or treated due to congestion in the system. Contact center administrators can choose any of the following options to handle the calls:

- **Treat Call with Dialed Number Default Label**—The rejected calls are treated with the default label of the dialed number on which the new call has arrived.
- **Treat call with Routing Client Default Label**—The rejected calls are treated with the default label of the routing client on which the new call has arrived.
- **Treat call with System Default Label**—The rejected calls are treated with the system default label set in Congestion Control settings.
- **Terminate call with a Dialog Fail or RouteEnd**—Terminates the new call dialog with a dialog failure.
- **Treat call with a Release Message to the Routing Client**—Terminates the new call dialog with a release message.

The treatment options are set either at routing client or at global level system congestion settings. If the treatment mode is not selected at the routing client, the system congestion settings are applied for treating the calls.

**Note**

If you select the treatment option to return a label back to treat the call with an announcement, then the announcement system should be external to CCE instance. In any case the call that is treated should not be re-entered into system for further processing.

Congestion Control Levels and Thresholds

The Congestion Control algorithm works in three levels. Each level has onset and abatement values. When the average CPS exceeds one level's onset value, the system moves to a higher congestion level. For example, if the system is at level 0 and the CPS exceeds the Level 2 onset capacity, the system moves directly to Level 2. The congestion level reduces when the average CPS falls below the current level's abatement value. Congestion levels can rise several levels at once. However, the congestion level reduces only one level at a time.

Table 13: Congestion Levels

Congestion Levels	Threshold (Percent of Capacity)	Description
Level1Onset	110%	If the average CPS exceeds this value, the congestion level moves to Level 1.
Level1Abate	90%	If the average CPS goes below this value, the congestion level moves back to Level 0 (Normal operating Level).
L1Reduction	10%	The percentage of incoming calls that are rejected at Level 1 congestion.
Level2Onset	130%	If the average CPS exceeds this value, the congestion level moves to Level 2.
Level2Abate	100%	If the average CPS goes below this value, then the congestion level moves back to Level 1.
Level2Redution	30%	The percentage of incoming calls that are rejected in Level 2 congestion.
Level3Onset	150%	If the average CPS exceeds this value, the congestion level moves to Level 3.
Level3Abatement	100%	If the average CPS goes below this value, the congestion level moves back to Level 2.
Level3Reduction	Variable reduction from 100% to 30%	The percentage of incoming calls that are rejected in Level 3 congestion. Depending on the incoming call rate, the reduction percentage varies from 30% to 100% when the congestion level enters Level 3.

**Note**

You cannot configure the onset, abatement, and reduction settings. These values are defined as a percentage of the standard CPS capacity for the system.

Real-Time Capacity Monitoring

The “System Capacity Real-time” real-time report provides congestion level information to the user. This report is based on table System_Capacity_Real_Time in the *Database Schema Handbook for Cisco Unified Contact Center Enterprise* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-technical-reference-list.html>. The report provides the following views:

- Congestion Information View
- Rejection Percentage View
- Key Performance Indicators View
- System Capacity View

For a detailed description of these reports refer to the *Cisco Unified Contact Center Enterprise Reporting User Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-user-guide-list.html>.

Congestion Control Configuration

Congestion control settings can be configured using the Congestion Settings Gadget and Routing Client Explorer tool. The system level congestion control can be set using the congestion control settings. And the routing client level treatment options can be selected using the Routing Client Explorer tool.

Upon selecting the deployment type, the system starts computing the various metrics related to the congestion control and system capacity, and generate the real-time report. However, the system does not reject/treat the calls until you enable the **Congestion Enabled** option on the **Settings** tab of the **Unified CCE Administration** tool.

Congestion Level Notification

To notify the CCE administrators, the congestion level changes are logged to Windows application events. Additionally they are sent as SNMP TRAPs and are targeted to Network Management system as configured in Unified CCE SNMP Management settings.

Call Treatment for Outbound Option

Outbound Option is a special case for call treatment with Congestion Control. When Media Routing Peripheral Gateway (MRPG) is integrated for Outbound Option, the PG's routing client should be configured to always send the dialog failure. The rejected reservation calls are retried by the dialer after a specified interval of time.

Special Operating Condition

This feature is supported by ICM/CCE 9.0(1) and later releases, and requires both Central Controller and PGs to effectively control the overload. The pre 9.0(1) PG routing clients are not capable of rejecting the calls. When a large deployment is undergoing an upgrade to 9.0(1) release, there is a period of time when the PGs are still operated with pre 9.0 releases. In such cases, the PGs which are in 9.0(1) starts rejecting the calls but pre 9.0 PGs will not reject the call. Hence there may be a possibility where 9.0 PGs rejecting more call until all the PGs are upgraded.

The following options should be considered while upgrading the system:

- Upgrade all PG before enabling the congestion control - Enable congestion to reject/treat calls only after all the PGs in the system is upgraded to 9.0. This will ensure uniform rejection of calls across all the routing clients in system.

- Upgrade selected PGs: If there are options in the enterprise where few PGs can reject more calls, upgrade those PGs first and then the remaining PG.



CHAPTER 5

Unified CCE Desktop Deployment Scenarios

- [Desktop Architecture, page 109](#)
- [Desktop Solutions, page 114](#)
- [Silent Monitoring, page 122](#)
- [Deployment Considerations, page 145](#)

Desktop Architecture

Desktop applications typically run on agent or supervisor desktops, Administration & Data servers, or administration clients. Services that support the desktop applications can run on the Unified CCE Peripheral Gateway (PG) server or on their own server. For each PG, there is one set of active desktop services.

Peripheral Gateway and CTI Server

In the Unified CCE solution, you deploy the Peripheral Gateway (PG) in a redundant configuration, with a Side A and a Side B. You install the CTI server on the PG. Agent CTI services connect to one side or the other, depending on which side is active.

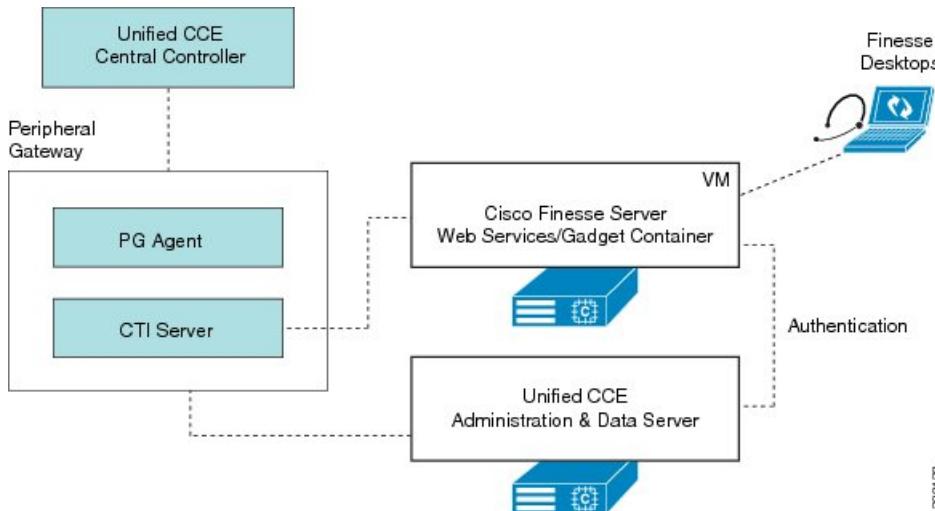
Desktop applications forward call control and agent requests to the Agent PG. The Agent PG processes agent state requests and updates the Central Controller for consideration in routing decisions. The Agent PG forwards call control requests to the Unified Communications Manager, which monitors and controls the phone end points. The Agent PG and desktop services keep the agent desktop application synchronized with the agent's IP phone state.

Cisco Finesse Server

You deploy the Cisco Finesse server on a dedicated VMware virtual machine (VM) that runs on the Cisco Voice Operating System (VOS) platform. The Finesse server is a required component for the Cisco Finesse desktop solution. The Cisco Finesse software is fault-tolerant and deploys on redundant VMs. Both Finesse servers are simultaneously active.

The Finesse server connects to the CTI server on the Agent PG. Authentication with Unified CCE is provided over a connection to the Administration & Data Server.

Figure 47: Cisco Finesse Desktop Architecture



Finesse requires that you deploy the Administration & Data Server with a backup Administration & Data Server. If the primary Administration & Data Server goes down, Finesse connects to the backup server for authentication so that agents can still sign in.

The Finesse server exposes supported client operations through a Representational State Transfer (REST) API. The REST API shields the developer from many of the details surrounding the CTI server wire protocol.

Configure the database on the Administration & Data Server to use Windows authentication. This user must be a domain user to access the Finesse database.

Finesse clients connect to the Finesse server over a web browser that points to the fully qualified domain name (FQDN) of the Finesse server.

You deploy the Finesse server in an active/active deployment, where both Finesse servers connect to the active CTI server on the Agent PG. The standard Cisco VOS replication mechanism provides redundancy.

Finesse Server Services

You can access the following Finesse services using the CLI:

- Cisco Finesse Notification service: This service is used for messaging and events. The Finesse desktop uses this service to view call events, agent state changes, and statistics.
- Cisco Tomcat service: This service contains all deployed Finesse applications. These applications include the following:
 - Finesse desktop application: This application provides the user interface for agents and supervisors.
 - Finesse REST API application: Finesse provides a REST API that enables client applications to access the supported server features. The REST API can use HTTP or HTTPS to transport application data. The REST API also provides a programming interface that third-party applications can use to interact with Finesse. See the Cisco Finesse documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/finesse/tsd-products-support-series-home.html> for more information on the REST API.

- Finesse administration application: This application provides the administrative operations for Finesse.
- Finesse Diagnostic Portal application: This application provides performance-related information for Finesse.

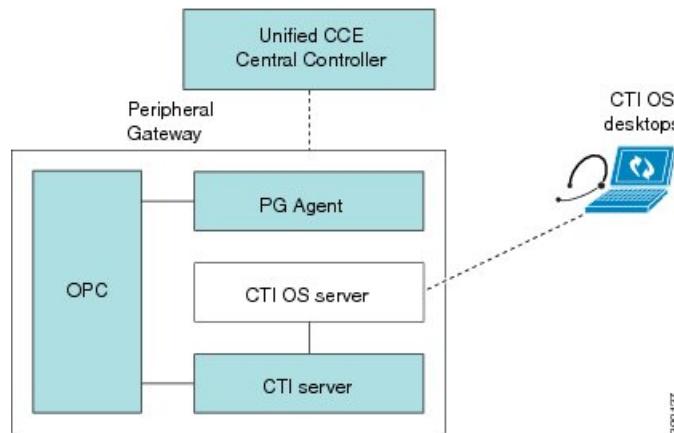
CTI Object Server

The CTI Object server (CTI OS) is a high-performance, scalable, fault-tolerant, server-based solution for deploying CTI applications. CTI OS is a required component for the CTI Toolkit Desktop. The CTI OS server runs as a redundant pair, one server on each VM that hosts an Agent PG.

Desktop applications pass communications, such as agent state change requests and call control, to the CTI OS server. CTI OS is a single point of integration for CTI Toolkit Desktops and third-party applications, such as Customer Relationship Management (CRM) systems, data mining, and workflow solutions.

The CTI Object server connects to the CTI server over TCP/IP and forwards call control and agent requests to the CTI server.

Figure 48: CTI OS Desktop Architecture



The CTI OS server also manages CTI Toolkit desktop configuration and behavior information, simplifying customization, updates, and maintenance, and supporting remote management.

CTI Object Server Services

- Desktop security: Supports secure socket connections between the CTI Object server on the PG and the agent, supervisor, or administrator desktop PC. Any CTI application built using the CTI OS Desktop Toolkit (CTI Toolkit) C++/COM CIL SDK can use the desktop security feature.



Note Desktop Security is not currently available in the .NET and Java CILs.

- Quality of Service (QoS): Supports packet prioritization with the network for desktop call control messages.

**Note**

QoS is not currently available in the .NET and Java CILs.

- Failover recovery: Supports automatic agent sign-in upon failover.
- Chat: Supports message passing and the text chat feature between agents and supervisors.
- Silent Monitoring: Supports VoIP monitoring of active calls. The CTI Object server communicates with the Silent Monitor Service (SMS) to start or stop the VoIP packet stream forwarding.

You deploy the CTI Object server in redundant pairs, one on Agent PG A and one on Agent PG B. Both CTI OS servers are active simultaneously. The CTI Toolkit desktop applications randomly connect to one of the two servers. If the connection to the original server fails, the desktops automatically fail over to the alternate server.

**Note**

The CTI OS server interfaces to any desktop application built using the CTI Toolkit SDK.

Related Topics

[Sizing Unified CCE Components and Servers, on page 215](#)

Agent Desktops

A Unified CCE deployment requires an agent desktop application. The agent uses this desktop for agent state control and call control. In addition to these required features, the desktop can provide other useful features.

Cisco offers the following primary types of Unified CCE agent desktop applications:

- Cisco Finesse: A browser-based agent desktop solution that provides a gadget-based architecture for extending base agent functionality.
- CTI Toolkit Desktop: An agent desktop application built with the CTI Toolkit. The desktop supports full customization and integration with other applications, customer databases, and Customer Relationship Management (CRM) applications.
- Cisco Unified CRM Connector for Siebel: A CTI driver for the Siebel Communication Server.

Cisco partners offer the following types of agent desktop applications:

- Partner agent desktops: Custom agent desktop applications are available through Cisco Technology Partners. These applications are based on the CTI Toolkit and are not discussed individually in this document. The Finesse REST API also enables partner desktop integration.
- Prepackaged CRM integrations: CRM integrations are available through Cisco Unified CRM Technology Partners. These integrations are based on the CTI Toolkit and are not discussed individually in this document.

Related Topics

[Cisco Agent Desktop, on page 351](#)

Supervisor Desktops

In addition to the agent desktop application, Cisco offers a supervisor desktop application. The contact center supervisor uses this application to monitor the agent state for members of their team. The supervisor desktop also allows the silent monitoring of agents during calls.

Cisco offers the following types of Unified CCE supervisor desktop applications:

- Cisco Finesse: A fully browser-based supervisor application that extends the base Finesse agent desktop with supervisor capabilities.
- CTI Toolkit supervisor desktop: A supervisor desktop application built with the CTI Toolkit. The desktop supports customization and integration with other applications, customer databases, and CRM applications.
- Supervisor desktop applications offered through Cisco partners.
- Prepackaged CRM integrations: CRM integrations are available through Cisco Unified CRM Technology Partners. These integrations are based on the CTI Toolkit and are not discussed individually in this document.



Note CAD also provides a supervisor desktop. For more information, see the CAD documentation.

Related Topics

[Cisco Agent Desktop, on page 351](#)

Agent Mobility

The Unified CCE deployment does not statically associate the agent desktop with any specific agent or IP phone extension. You configure agents and phone extensions within Unified CCE and associate them with a specific Unified Communications Manager cluster.

When agents sign in to their desktop, a dialog prompts for an agent ID or username, password, and the phone extension to use for that session. At that time, the agent ID, phone extension, and agent desktop IP address are dynamically associated. The association is released when the agent signs out.

This mechanism allows an agent to work (or hot-desk) at any workstation. The mechanism also allows agents to take their laptops to any appropriately configured Cisco Unified IP Phone and sign in from that device.

Agents can also sign in to other phones using the Cisco Extension Mobility feature. For more information about this feature, see the Extension Mobility section of the *Features and Services Guide for Cisco Unified Communications Manager* and *Features and Services Guide for Cisco Unified Communications Manager* at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-maintenance-guides-list.html>.

Desktop Solutions

Your choice of a desktop solution depends on the requirements of your contact center. The following table provides an abbreviated list of the functionality available in the various desktop applications. The table provides a starting point to help you determine the desktop that best meets your specific solution requirements.

You can find more information for each of the Cisco desktops in the following sections and in their respective product specifications at Cisco.com.

Desktop functionality	Cisco Finesse	CTI Toolkit	Cisco Unified CRM Connector for Siebel
Browser-based desktops	Yes		
Custom development	Yes (using standard web components such as HTML, JavaScript)	Yes (using C++, .NET, and Java)	
Desktop security	Yes Note Finesse supports HTTPS only for deployments of 1000 agents or less.	Yes	
Workflow automation	Yes		
Mobile (remote) agents	Yes	Yes	Yes
Silent monitoring	Yes	Yes	
Monitor mode applications		Yes	
Outbound calls	Yes	Yes	Yes
Microsoft Terminal Services support	No	Yes	
Citrix presentation server support		Yes	
Agent mobility	Yes	Yes	
Agent Greeting	Yes	Yes	

Cisco Finesse Desktop Solution

Cisco Finesse is a supervisor and agent desktop for use with Cisco Unified Contact Center solutions. Cisco Finesse provides the following:

- An agent and supervisor desktop that integrate traditional contact center functions into a thin-client desktop
- A browser-based desktop implemented through a Web 2.0 interface
- A single, customizable interface that gives contact center agents access to multiple assets and information sources
- Open Web 2.0 APIs that simplify the development and integration of applications

You install the Cisco Finesse server on a VM. Clients then use a web browser to point to the Finesse server. No Finesse software is installed on the clients, which speeds and simplifies installation and upgrade.

The following table lists the supported browsers and operating systems for Finesse clients (administration console, agent desktop, and supervisor desktop).

Operating System	Browser and Supported Versions
Windows 7	Internet Explorer 9.0 Internet Explorer 11.0 Note IE 11 requires Windows 7 SP1. Firefox (version 24 or later)
Windows 8.1	Internet Explorer 11.0 Firefox (version 24 or later)
Mac OS X	Firefox (version 24 or later)

The Finesse desktop application consists of the client and server components. The client is composed of standard web programming elements (HTML, JavaScript) that are distributed as gadgets using the OpenSocial 1.0 specification. You can configure the agent desktop to use Cisco and third-party gadgets through a layout management mechanism.

Cisco Finesse is part of a class of applications called Enterprise Mashups. An Enterprise Mashup is a web-centric method of combining applications on the client side. The gadget-based architecture of Finesse enables client-side mashup and easier integration. Version compatibility dependencies are reduced because gadget upgrades are handled independently.

You can customize the agent and supervisor desktops through the Finesse administration console. Administrators can define the tab names that appear on the desktops and configure which gadgets appear on each tab.

Finesse REST API

Finesse provides a REST API that allows client applications to access the supported server features. The REST API can use HTTP or secure HTTP (HTTPS) as the transport with XML payloads.

Finesse also provides a JavaScript library and sample gadget code that can help expedite third-party integration. You can find developer documentation for the REST API, the JavaScript library, and sample gadgets on the Cisco Developer Network at <https://developer.cisco.com/site/finesse/>.

Finesse Agent Desktop

The Finesse out-of-the-box agent desktop provides the following features:

- Basic call control (answer, hold, retrieve, end, and make call)
- Advanced call control (consultation, transfer after consult, conference after consult)
- Single-step transfer (agents can transfer a call without first initiating a consultation call)
- Queue statistics gadget (to view information about the queues to which the agent is assigned)
- Not Ready and Sign Out reason codes
- Phonebooks
- Workflows
- Mobile agent support
- Progressive, Predictive, and Preview Outbound



Note

Finesse does not support Direct Preview Outbound.

Finesse Supervisor Desktop

The Finesse supervisor features extend the agent desktop with more gadgets that are accessible to supervisors. These features include the following:

- Team performance gadget to view agent status
- Queue statistics gadget to view queue (skill group) statistics for the supervisor's queues
- Unified Communications Manager Silent Monitoring
- Barge-in
- Intercept
- Change agent state (a supervisor can sign out an agent, force an agent into Not Ready state, or force an agent into Ready state)

Finesse Administration Console

Cisco Finesse includes an administrative application that allows administrators to configure the following:

- Connections to the CTI server and the Administration & Data server database
- Cluster settings for VOS replication

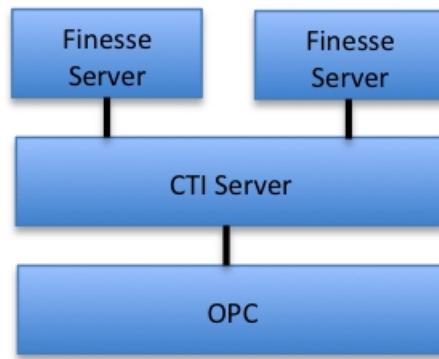
- Not ready and sign out reason codes
- Wrap-up reasons
- Phonebooks
- Workflows and workflow actions
- Call variable and ECC variable layouts
- Desktop layout
- Team resources

Reason codes, wrap-up reasons, phonebooks, workflows, and desktop layouts can be global (apply to all agents) or assigned to specific teams.

Finesse Multiserver Support

A single Agent PG supports two instances of Finesse. Each Finesse server can support the maximum of 2,000 users supported by the CTI server. This capacity enables one Finesse server to handle the full load if the other server fails. The total number of users between the two Finesse servers cannot exceed 2,000. Each Finesse server requires a single CTI connection, as shown in the following figure:

Figure 49: Multiple Finesse Servers



Finesse can be deployed according to coresidency policies outlined in the *Unified Communications Virtualization DocWiki* at http://docwiki.cisco.com/wiki/Unified_Communications_Virtualization.

Load Balancing for Finesse

After agents sign in to the Finesse desktop, the Finesse desktop client manages failover. For example, if a Finesse server goes out of service, the Finesse client automatically redirects and signs the agent in to the other Finesse server. The client can manage various network and server failure use cases. Given this client-side logic, the use of a load balancer after sign-in is not required nor supported.

However, the following are two scenarios in which you can use a load balancer with Finesse.

**Note**

These scenarios only apply to the Finesse desktop and not to Finesse IP Phone Agents.

When Agents Navigate to the Finesse Sign-In Page

If an agent attempts to navigate to a Finesse server that is down or not reachable, they cannot access the sign-in page. The agent receives an error and must manually sign in to the other Finesse server. To avoid this manual step, customers can use a load balancer using URL redirect mode to direct the agent to a Finesse server that is operational. One option is to use the Finesse SystemInfo REST API, which provides the status of the Finesse server. For details about this API, see the *Cisco Finesse Web Services Developer Guide*.

When you configure a load balancer to determine the status of the Finesse servers, the call flow is as follows:

- 1 When agents sign in to Finesse, they point their browsers to the load balancer.
- 2 The load balancer redirects the agent browser to an appropriate Finesse server.
- 3 The agent signs in to the Finesse server directly. At this stage, the load balancer is no longer part of the call flow.

When Customers Use the Finesse API Directly

If a customer uses the Finesse REST API directly, the Finesse client-side failover logic is not in the call flow. In this case, customers can opt to use a load balancer to manage high availability. This load balancer is considered part of a custom application which, like all custom applications, Cisco does not support. The customer or partner must provide the required support for the load balancer.

Before you configure the load balancer, remember that there are two connections between Finesse clients and the Finesse server:

- A REST channel for request and response
- An XMPP channel that the server uses to send notifications to the client

Both channels for a given client must connect to the same Finesse server.

You cannot connect the load balancer to the REST connection for one Finesse server and to the XMPP channel connection for the other Finesse server. This setup provides unpredictable results and is not supported.

CTI OS Desktop Toolkit Solution

The CTI OS Desktop Toolkit (CTI Toolkit) provides an out-of-the-box agent and supervisor desktop and contact center monitoring applications. You can use these applications as-is or customize them to meet the particular needs of your contact center.

The CTI Toolkit also provides an SDK for custom development of desktop applications. The CTI Toolkit supports C++, Java, and .NET development CILs and provides sample applications for customization.

The CTI toolkit offers advanced tools for integrating desktop applications with a database, CRM applications, and other contact center applications.

The CTI Toolkit offers the following contact center features:

- Collaboration: Supervisors can chat directly with agents. Agents can chat with supervisors or other team members (if enabled). Interactive collaboration allows for better communication within the contact

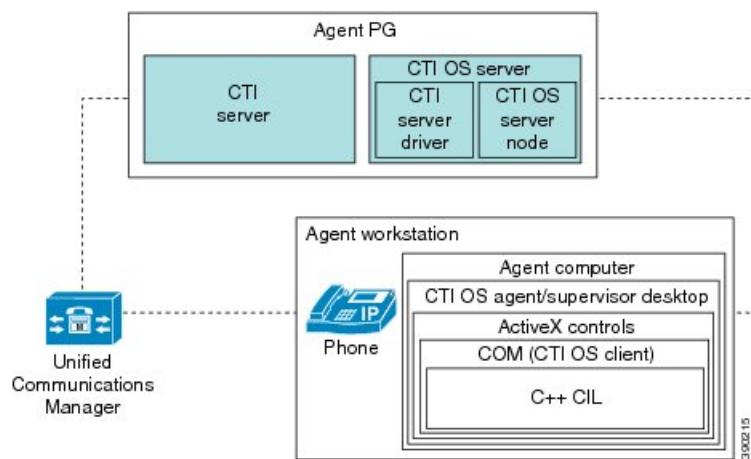
center, which can increase productivity and improve customer responsiveness. Collaboration can also help supervisors coach and train agents.

- Secure desktop connection: The CTI Toolkit provides desktop security between the agent and supervisor desktops and the CTI OS server.
- Silent monitoring: Supervisors can start silent monitoring sessions with agents who are on their team.

CTI OS Desktop Toolkit SDK and User Applications

The following figure illustrates the architecture of the CTI OS Desktop Toolkit.

Figure 50: CTI OS Toolkit Desktop Architecture



The CTI Toolkit provides the following user tools and applications.

CTI Toolkit Agent Desktop

The CTI Toolkit agent desktop is a Microsoft Windows Visual Basic application built on the COM CIL and ActiveX controls. The desktop runs on an agent PC and works with either a hardware IP phone or the Cisco IP Communicator. The CTI Toolkit agent desktop interfaces with the CTI OS server for call control and agent state events.

The CTI Toolkit agent desktop allows the agent to turn on or turn off Agent Greeting playback (Enable/Disable) on routed incoming calls to the desktop. The CTI Toolkit agent desktop also allows the agent to record new Agent Greetings.

The CTI Toolkit agent desktop supports desktop monitoring, which captures the voice stream on the agent IP phone. This feature allows the CTI Toolkit agent desktop to provide silent monitoring and call recording.

CTI Toolkit Supervisor Desktop

The CTI Toolkit supervisor desktop is a Microsoft Windows Visual Basic application built on the COM CIL and ActiveX controls. The CTI Toolkit supervisor desktop runs on a supervisor PC and interfaces with the CTI OS server for agent state change events and real-time statistics updates.

Supervisors can use the CTI Toolkit supervisor desktop to manage a team of agents. Supervisors can view real-time information about the agents on their team and interact with these agents. A supervisor can select

an agent to change that agent's state, view information specific to that agent, or chat with the agent. The supervisor can also silently monitor an agent's call and barge in or intercept that call. Agents can send requests for emergency assistance to the supervisor through the supervisor desktop.

In Unified CCE, supervisors can also be configured to act as agents. In this scenario, the standard set of agent phone controls is available on the supervisor desktop.

CTI Toolkit Outbound Desktop

The CTI Toolkit outbound desktop is a Microsoft Windows Visual Basic application built on the COM CIL and ActiveX controls. The outbound desktop runs on an agent PC and works with either a hardware IP phone or the Cisco IP Communicator. The CTI Toolkit outbound desktop interfaces with the CTI OS server for call control and agent state change events. In addition to the standard CTI Toolkit agent desktop controls, the outbound desktop provides a set of controls to manage outbound call campaigns. Unified CCE automatically manages outbound calls and the agent uses the additional controls to accept outbound calls.

CTI Toolkit Combo Desktop

The CTI Toolkit combo desktop is a Microsoft Windows .NET application that runs on the agent PC and works with either a hardware IP phone or the Cisco IP Communicator. The CTI Toolkit combo desktop interfaces with the CTI OS server for call control and agent state change events.

The combo desktop combines the functionality of the CTI Toolkit agent, supervisor, and outbound desktops into a single .NET application. The combo desktop source code is provided as a starting point for custom desktop development using the Microsoft .NET framework.

The CTI Toolkit combo desktop does not support Agent Greeting Enable/Disable or recording new Agent Greetings.

CTI Toolkit All-Agents Monitor

The CTI OS Desktop Toolkit provides a ready-to-run all-agents monitor application. With this application, an administrator can monitor agent login and state activity within the contact center. The CTI Toolkit All-Agents Monitor is a Windows application based on the C++ CIL.


Note

You can use the CTI Toolkit All-Agents Monitor application only if the number of skill groups per agent is less than 20.

CTI Toolkit All-Calls Monitor

The CTI OS Desktop Toolkit provides a ready-to-run all-calls monitor application. With this application, an administrator can monitor call activity within the contact center. The CTI Toolkit All-Calls Monitor is a Windows application based on the C++ CIL.


Note

You can use the CTI Toolkit All-Calls Monitor application only if the number of skill groups per agent is less than 20.

C++ CIL API

The C++ CIL API is a Windows SDK for developing C++ CTI applications. The CIL provides a set of header files and static libraries for building C++ CTI applications using Microsoft Visual Studio .NET. The C++ CIL supports a secure desktop connection between the agent PC and the CTI Object Server on the PG.

The CTI Toolkit C++ CIL supports Agent Greeting Enable/Disable and Agent Greeting Recording.

Java CIL API

The Java CIL API is a cross-platform library for developing Java CTI applications. This CIL does not provide APIs for Agent Greeting Enable/Disable or Agent Greeting Recording.

.NET CIL API

The .NET CIL API is a Windows SDK for developing custom .NET framework CTI applications. The CIL provides native .NET class libraries and a .NET combo desktop. The .NET combo desktop is a sample application build using the .NET CIL.

This CIL does not provide APIs for Agent Greeting Enable/Disable or Agent Greeting Recording.

COM CIL API

The COM CIL API is a set of COM Dynamic Link Libraries (COM DLL) for building Visual Basic 6.0 CTI applications. The CTI Toolkit agent and supervisor desktops are sample applications built with Visual Basic 6.0 using the COM CIL.

The CTI Toolkit COM CIL supports Agent Greeting Enable/Disable and Agent Greeting Recording.

CTI OS Runtime Callable Wrappers

The CTI OS runtime callable wrappers provide a set of .NET assemblies that allow you to use the COM CIL and ActiveX controls in native .NET applications.

ActiveX Controls

The CTI Toolkit includes a set of ActiveX controls to enable rapid application development. The ActiveX controls are UI components that enable easy drag-and-drop creation of custom CTI applications in a variety of container applications. Container applications include Microsoft Visual Basic .NET, Microsoft Internet Explorer, Microsoft Visual C++ 8.0, Borland Delphi, and other applications that support the OC96 ActiveX standard.

The CTI Toolkit includes the following ActiveX controls:

- Agent Greeting Enable/Disable Control
- Agent State Control
- Chat Control
- Emergency Assist Control
- Alternate Control
- Answer Control
- Bad Line Control
- Call Appearance Control

- Conference Control
- Hold Control
- Make Call Control (allows Agent Greeting recording)
- Reconnect Control
- Status Bar Control
- Record Control
- Transfer Control
- Agent Statistics Control
- Skill Group Statistics Control
- Agent Select Control
- Supervisor Control
- Silent Monitor Control

For more information about the CTI OS Desktop Toolkit, see the *CTI OS Developer Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/computer-telephony-integration-option/products-programming-reference-guides-list.html>.

Cisco Unified CRM Connector for Siebel Solution

The Cisco Unified CRM Connector for Siebel is a component that you install to enable integration of Unified CCE with the Siebel CRM environment. In this solution, the Siebel agent desktop provides the agent state and call control interface. The connector is built on top of the CTI OS Desktop Toolkit C++ CIL. The Siebel desktop uses the connector to communicate with the CTI OS server.

The Cisco Unified CRM Connector for Siebel does not support the following:

- Agent Greeting Enable/Disable
- Recording new agent greetings

For more information about the Cisco Unified CRM Connector for Siebel, see product documentation at http://www.cisco.com/en/US/products/ps9117/tsd_products_support_series_home.html. For more information about the Siebel eBusiness solution, see the [Siebel website](#).

Silent Monitoring

Silent monitoring allows supervisors to monitor the conversations of agents within their team. Supervisors cannot participate actively in the conversations and agents and callers are not aware that they are being monitored. Cisco Finesse and CTI Object Server (CTI OS) provide solutions support for silent monitoring.

Cisco Finesse supports Unified Communications Silent Monitoring only. You configure silent monitoring on Unified Communications Manager. No additional configuration is required on the Finesse server.

CTI OS supports two types of silent monitoring: Unified Communications Manager Silent Monitoring and CTI OS-based Silent Monitoring. You can configure the CTI OS server to use CTI OS-based Silent Monitoring

or Unified Communications Manager Silent Monitoring, or to disable silent monitoring. When supervisor desktops connect to the CTI OS server, the desktops download the configuration. The supervisor desktop invokes the configured type of silent monitoring when the Start Silent Monitor button is clicked. The CTI OS server uses the initial message from the supervisor desktop to drive either CTI OS- based Silent Monitoring or Unified Communications Manager Silent Monitoring.

For information about how to configure silent monitoring for CTI OS, see the *CTI OS System Manager Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>. For information about how to implement CTI OS based or Unified Communications Manager Silent Monitoring, see the *CTI OS Developer Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/computer-telephony-integration-option/products-programming-reference-guides-list.html>.

**Note**

Even for regions that are configured to use G.711, Unified CCE uses G.722 as the default codec for devices that support G.722. However G.722 is not supported with Silent Monitoring and Call Recording based on CTI OS or Unified Communications Manager. To disable this default, in Unified Communications Manager Administration, go to Enterprise Parameters and set Advertise G.722 Codec to disabled.

**Note**

If voice streams are encrypted, silent monitoring does not work correctly. Although the voice streams can still be captured, the silent monitoring service cannot decode them correctly.

Unified Communications Manager Silent Monitoring

Unified Communications Manager accomplishes silent monitoring with a call between the supervisor (monitoring) device and the agent (monitored) device. The agent phone mixes and sends the agent's conversation to the supervisor phone, where it is played out to the supervisor.

Unified CCE supports the Silent Monitoring functionality available in Unified Communications Manager. Unified Communications Manager Silent Monitoring supports only one silent monitoring session and one recording session for the same agent phone.

**Note**

Unified Communications Manager Silent Monitoring does not support mobile agents.

Unified Communications Manager Silent Monitoring can monitor any Unified CCE agent desktop, including Siebel, if the following conditions exist:

- The monitored agents uses a compatible Cisco Unified IP phone or Cisco IP Communicator. For more details, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.
- The contact center uses a compatible version of Cisco Unified Communications Manager. For more information, see the *Compatibility Matrix for Unified CCE*.

Supervisors can use any Cisco IP Phone, including Cisco IP Communicator, to silently monitor agents.

Unified Communications Manager Silent Monitoring works the same as other call control functionality provided by Unified Communications Manager (such as conference and transfer). When the silent monitoring session begins, the desktop sends a message through Unified CCE, through Unified Communications Manager, and out to the phones where silent monitoring is executed.

Messaging through Unified CCE and Unified Communications Manager impacts Unified CCE performance.


Note

We only support agent phones in Enterprise. We do not support agent phones behind a phone proxy or over the edge behind VCS Expressway.

Related Topics

[Sizing Unified CCE Components and Servers, on page 215](#)

Cisco Finesse

Cisco Finesse provides silent monitoring functionality through Unified Communications Manager Silent Monitoring. Finesse works with Unified Communications Manager Silent Monitoring as follows:

- 1 The supervisor application sends a REST request to the Finesse server to initiate silent monitoring.
- 2 The Finesse server sends the AgentSuperviseCall() message to Unified CCE to start the silent monitoring session.
- 3 Unified CCE sends the CallStartMonitor() message to Unified Communications Manager.
- 4 Unified Communications Manager instructs the supervisor phone to call the Built-In Bridge (BIB) on the agent phone.
- 5 The supervisor phone places the call to the BIB on the agent phone.
- 6 The agent phone forwards a mix of the agent voice stream and customer voice stream.
- 7 Unified Communications Manager sends call events for the silently monitored call to Unified CCE.
- 8 Unified CCE sends update events to the Finesse server.
- 9 The Finesse server sends XMPP updates to the Finesse supervisor application.

Finesse does not support silent monitoring of mobile agents. Supervisors cannot silent monitor mobile agents and mobile supervisors cannot perform silent monitoring.

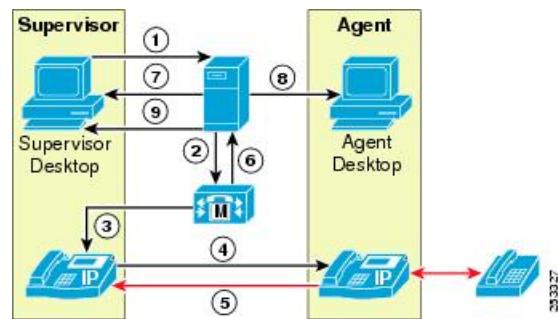
CTI OS

When the CTI OS server uses Unified Communications Silent Monitoring, CTI OS provides silent monitoring as follows:

- 1 The supervisor initiates Silent Monitoring by sending the Agent.SuperviseCall() message to Unified CCE.
- 2 Unified CCE sends the Call.startMonitor() message to Unified Communications Manager.
- 3 Unified Communications Manager instructs the supervisor phone to call the Built-In Bridge (BIB) in the agent phone.
- 4 The supervisor phone places the call to the BIB in the agent phone.
- 5 The agent phone forwards a mix of the agent voice stream and customer voice stream.

- 6 Call events for the silently-monitored call are sent from Unified Communications Manager to Unified CCE.
- 7 CTI OS sends a SilentMonitorStarted event to the supervisor desktop.
- 8 CTI OS sends a SilentMonitorStarted event to the agent desktop.
- 9 CTI OS sends call events for the silently-monitored call to the supervisor desktop.

Figure 51: Unified CM Silent Monitoring for CTI OS



CTI OS-Based Silent Monitoring

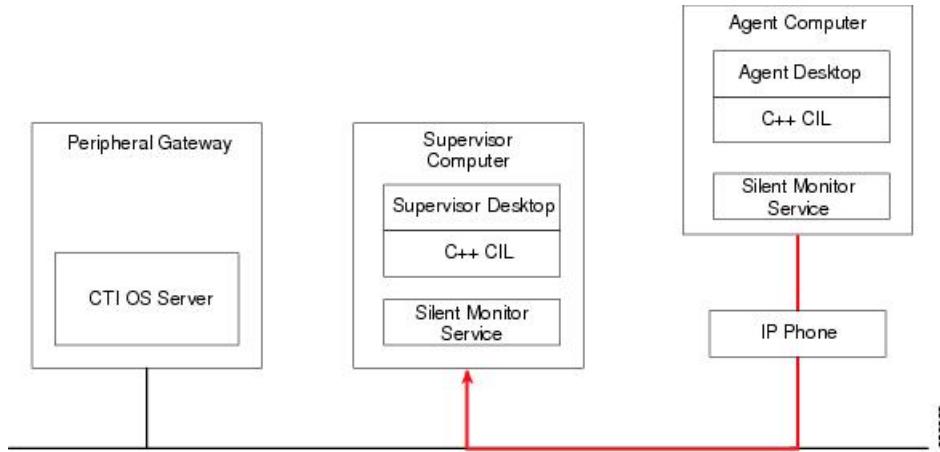
CTI OS-based Silent Monitoring uses one or more VoIP monitoring services located on the agent desktop (desktop-based monitoring) or on a separate VoIP monitor server (server-based monitoring). CTI OS uses desktop-based monitoring to support traditional Unified CCE agents and server-based monitoring to support mobile agents.

To allow for silent monitoring in a mobile agent or Citrix deployment, the silent monitoring functionality is not part of the CIL. This functionality resides in a separate silent monitoring service. You can then deploy the service where it can access the agent's voice stream and the supervisor's sound card.

The following figures show where to deploy the silent monitoring service for each deployment model. The red line in each diagram illustrates the path of the monitored voice stream.

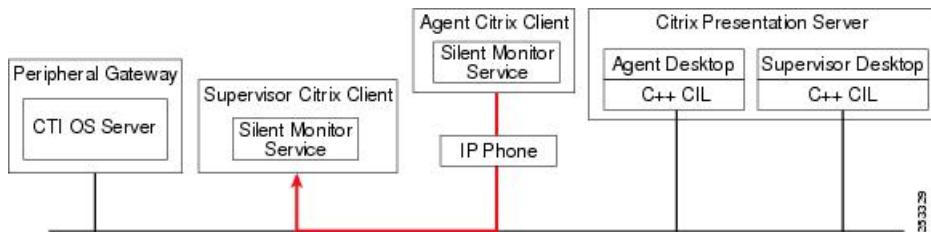
The following figure illustrates a deployment where the agent uses an IP phone. In this deployment, you configure silent monitoring the same whether the agent is a local agent or a mobile agent. The silent monitoring service runs alongside the CIL to provide silent monitoring functionality.

Figure 52: CTI OS-Based Silent Monitoring for Cisco Unified CCE when a Mobile or Local Agent Uses an IP Phone



The following figure illustrates a Citrix deployment where the agent uses an IP phone. The agent in this deployment can be local or mobile. The silent monitoring service is deployed on Citrix clients, where it can access the agent's voice stream and the supervisor's sound card. The CIL makes a connection to the silent monitoring service and sends instructions over a TCP connection to start and stop the silent monitoring sessions.

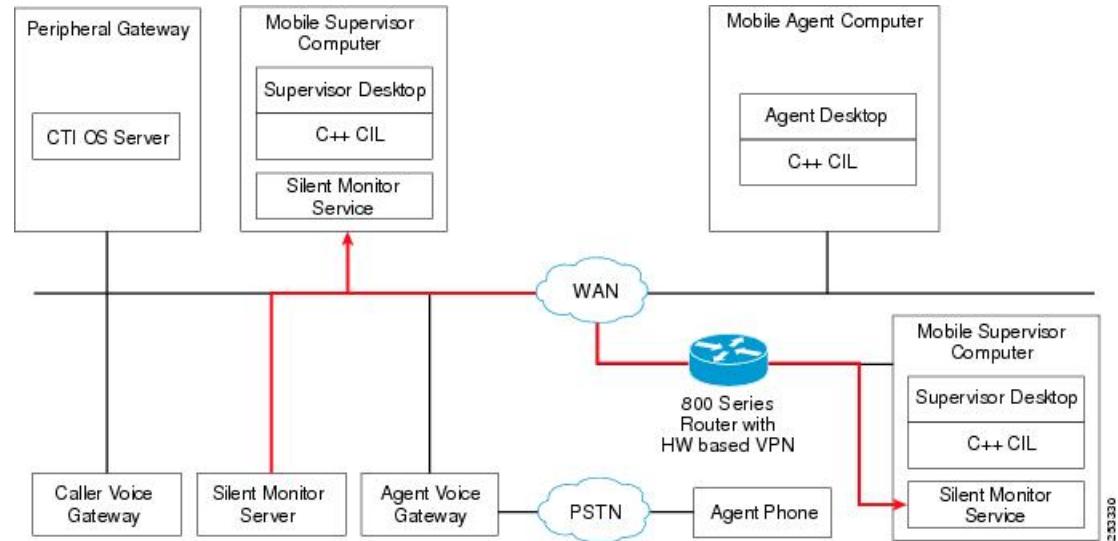
Figure 53: CTI OS-Based Silent Monitoring for Cisco Unified CCE with Citrix when a Mobile or Local Agent Uses an IP Phone



The following figure illustrates the deployment model for mobile agents using PSTN phones. In this model, one silent monitoring service is deployed on a switch's SPAN port to gain access to voice traffic passing through the agent gateway. Agents use the service attached to the SPAN port to forward their voice streams to the supervisor silent monitoring services. Local supervisors are deployed the same as Unified CCE supervisors. Remote supervisors are deployed the same as Unified CCE supervisors with one exception. A

Cisco 800 Series Router with a hardware-based VPN is required for the supervisor to receive agent voice streams.

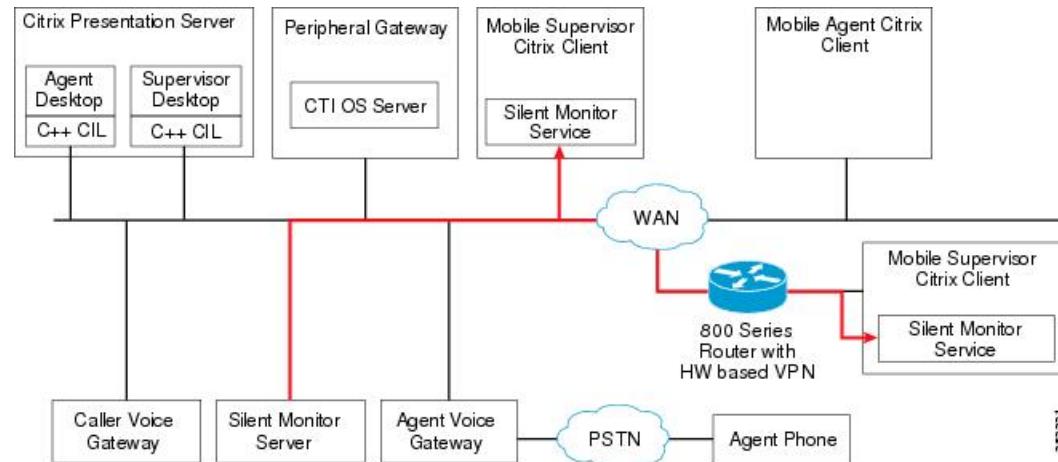
Figure 54: Silent Monitoring for a Mobile Agent Using a PSTN Phone



The following figure illustrates silent monitoring deployment for a mobile agent using a PSTN phone in a Citrix or Microsoft Terminal Services environment. One silent monitoring service is deployed on a switch's SPAN port to gain access to voice traffic that passes through the agent gateway. Agents use the service attached to the SPAN port to forward their voice streams to the supervisor silent monitoring services. Mobile agents need to run only their Citrix clients. Agent desktops running on the Citrix server connect to the silent monitoring server.

Local supervisors are deployed the same as Citrix Unified CCE supervisors. Remote supervisors are deployed the same as Citrix Unified CCE supervisors with one exception. A Cisco 800 Series Router with a hardware-based VPN is required for the supervisor to receive agent voice streams.

Figure 55: Silent Monitoring for a Mobile Agent Using a PSTN Phone with Citrix or Microsoft Terminal Services



In mobile agent deployments where the agents use PSTN phones, calls where the voice traffic does not leave the agent gateway cannot be silently monitored. These calls include agent-to-agent calls and agent consultations with other agents. Because the mobile agent solution requires separate gateways for callers and agents to ensure that voice traffic is put on the network, only calls between agents and customers can be reliably monitored.

Clusters

If mobile agent log in can be handled by one of two gateways, you can cluster two silent monitoring servers together. The clustered servers provide silent monitoring functionality regardless of the gateway that handles the call. A cluster-based (SPAN) deployment supports a maximum of two silent monitoring servers.

When a silent monitoring server receives a request to silently monitor an agent from the agent desktop, it forwards the request to its peer. Both silent monitor servers then attempt to detect the stream. When a server detects the agent voice stream, it forwards the stream to the supervisor's silent monitoring service.

For more information about deployment and configuration of the silent monitoring service, see the *CTI OS System Manager Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

Connection Profiles

In mobile agent deployments, agent desktops use a CTI OS connection profile to decide where and how to connect to the silent monitoring server. When an agent signs in, the agent desktop uses the following algorithm to determine the location of the silent monitoring service:

- 1 If a silent monitoring service is present in the connection profile, attempt to connect to it.
- 2 If no silent monitoring service is present, determine if the desktop is running under Citrix.
- 3 If the desktop is running under Citrix, connect to the silent monitoring service running at the IP address of the Citrix client.
- 4 If the desktop is not running under Citrix, connect to the silent monitoring service running at `localhost`.

Supervisor desktops use the following algorithm to find the silent monitoring service:

- 1 If the desktop is running under Citrix, connect to the silent monitoring service running at the IP address of the Citrix client.
- 2 If the desktop is not running under Citrix, connect to the silent monitoring service running at `localhost`.

If the `IPCCSilentMonitorEnabled` key is set to 0, the desktops do not attempt to connect to a silent monitoring service.

Cisco Remote Silent Monitoring

Cisco Remote Silent Monitoring (RSM) allows for real-time monitoring of agents as a dial-in service. You can use RSM with the CTI-OS Toolkit desktop solution.

The RSM solution consists of three components:

- VLEngine

- PhoneSim
- Callflow scripts for Unified CVP and Unified IP IVR

For more information about these components, see the *Cisco Remote Silent Monitoring Installation and Administration* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

RSM Platform Considerations

The RSM solution is highly integrated into the Cisco Unified CCE environment. RSM requires resources from other components of the Unified CCE platform to function. You must understand how RSM interacts with the rest of the environment to properly integrate RSM and to plan, provision, and manage capacity.

RSM interacts mainly with the Unified Communications Manager cluster.

The RSM server ties in to each Unified Communications Manager cluster in the environment that it is configured to use as follows:

- Simulated Phones

The RSM PhoneSim component requires that you create a Cisco Unified IP Phone 7941 device entry on the Unified Communications Manager cluster for each of the simulated phones (simphones) that it is configured to manage. For instance, a RSM system that is configured to handle up to 100 dialed-in supervisors monitoring agents on a particular Unified Communications Manager cluster must have at least 100 simphones. To the Unified Communications Manager cluster, these simphones appear as Cisco Unified IP Phone 7941 SIP phones. However, instead of being a physical phone device, the simphones are homed to PhoneSim and controlled by PhoneSim.

Compared to the usage profile of a physical phone device, the simphone usually puts a lighter load on the Unified Communications Manager cluster. The simphone exhibits only a small set of behaviors, consisting of:

- Registering with the Unified Communications Manager cluster when PhoneSim starts.
- Making a monitoring call to an agent's phone when a dialed-in supervisor requests to monitor that agent. The agent's phone then forks off a copy of the agent's conversation to the simphone.

- JTAPI

When RSM is integrated into the environment, a JTAPI user is created and associated with each agent phone device that can be monitored, as well as with each simphone device on the cluster.

When supervisor monitors an agent, the RSM server makes a JTAPI monitor request call to the Unified Communications Manager cluster that manages that agent's phone. Also, while RSM is in use, a JTAPI CallObserver is attached to each simphone device. The JTAPI CallObserver is also attached to an agent phone device, but only while the JTAPI monitor request is being issued to that device.

JTAPI connections may optionally be encrypted. However, encryption induces a slight performance penalty on the server itself when higher agent loads are utilized. For more information about enabling JTAPI connection security, see the *Cisco Remote Silent Monitoring Installation and Administration* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

CTI OS Server

RSM makes a persistent monitor-mode connection to each CTI OS server that it is configured to use. Certain platform events (such as call start, call end, and agent on hold) are streamed in real-time through this connection.

RSM makes an additional, short-lived agent-mode connection to each CTI OS server when a supervisor dials in and authenticates. This connection performs a corresponding login to CTI OS to validate the supervisor's credentials. RSM does not make this connection if the built-in authentication mechanisms of the RSM call flow (for example, the checkCredentials API call) are not used. If the login is successful, the RSM server requests the supervisor's team membership. After the request is returned, a logout is called and the connection terminates.



Note

The total supervisor count in Unified CCE must be spread across CTI OS desktop users and RSM. For example, in a 2,000 agent configuration, up to 200 agents can be supervisors. In this example, the total supervisor count between CTI OS and RSM must not exceed 200.

Optionally, CTI OS connections can be encrypted through the use of IP Sec configurations. However, this optional encryption induces a significant performance penalty on the server itself when higher agent loads are used. For more information about enabling CTI OS connection security, see the *Cisco Remote Silent Monitoring Installation and Administration* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

CTI Server

RSM also supports Unified CCE integration through the CTI protocol. In RSM configuration manager, you can pick either CTI or CTI OS for Unified CCE integration. When you choose between CTI and CTI OS integrations, consider these points:

- Finesse installations without any CTI OS servers require a CTI integration.
- To support 2000 Active Agents and 12000 Configured agents for each PG pair, use CTI integration for RSM. With CT IOS (using Java CIL), RSM can support 2000 Active agents and 8000 configured agents.
- In most deployments, the CTI OS server supports a maximum of five monitor-mode connections. The CTI server supports a maximum of seven All-Event clients. The combined maximum of monitor-mode connections and All-Event clients on an Agent PG is nine. On VMs created from the large 4-core OVAs, you can increase those limits if the All-Event Clients use Event Minimization in their CTI Server protocol integration.

Voice Response Unit

The RSM platform does not directly media-terminate inbound calls. Instead, supervisors dial in to a Unified CVP or Unified IP IVR-based Voice Response Unit (VRU) system. The VRU system runs call flow script logic that interacts with services hosted on the RSM server over HTTP. Therefore, if a given RSM installation supports up to 40 supervisors, a VRU must be present (as well as the necessary PRI/network resources) that offers the same level of support.

RSM calls often place higher loads on the VRU processor and memory than more traditional VRU-type calls. More traditional VRU call flows play shorter, and often cached or non-streamed prompts. These prompts are separated by periods of gathering caller input and silence. With RSM, the predominant caller activity is

monitoring agent calls. To the VRU, monitoring a call looks like the playback of a long-streaming audio prompt, which requires a relatively higher level of VRU processor involvement.

For Unified CVP deployments, supported VXML gateway models are listed in the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-technical-reference-list.html>.

When you provision a VRU for use by RSM, count each RSM call as 1.3 non-RSM calls for processor and memory usage. For example, for a VRU that can normally handle 40 concurrent calls, plan for it to handle only 30 RSM calls ($40/1.3=30$).

Also note that RSM makes extensive use of VXML Voice Browser functionality under both Unified CVP and Unified IP IVR.

RSM supports RTSP prompt streaming and does not require a dedicated VXML Gateway for Unified CVP installations. You do not need to configure the “ivr prompt streamed http option” in the VXML Gateway, which conflicts with Unified CVP IOS requirements. RSM scalability on Unified CVP supports 80 concurrent sessions on any Unified CVP-supported VXML Voice Gateway model and IOS version.

Agent Phones

RSM requires that agent phones have Built-In Bridge (BIB). The BIB allows the phone to fork off a copy of the current conversation stream to the RSM server. For an up-to-date list of phone models that support BIB feature, refer to the *Unified CM Silent Monitoring Recording Supported Device Matrix* at <https://developer.cisco.com/web/sip/wiki/-/wiki/Main/Unified+CM+Silent+Monitoring+Recording+Supported+Device+Matrix>

Cisco Unified Contact Manager provides a maximum of one active monitoring session per agent. The agent phone can handle only one active monitoring session and one active recording session at any given time. If a third-party recorder is recording an agent's conversation, the supervisor can still monitor the agent using the supervisor's desktop or RSM. However, if both an RSM-based supervisor and a supervisor desktop-based supervisor try to monitor an agent during the same time period, the request fails with the last one to try because it exceeds the monitoring limit.

RSM sets up only one monitoring session through Unified Communications Manager for a single monitored agent. If two or more RSM users make a request to monitor the agent's call at the same time, RSM forks the stream to cover all RSM users. More than two RSM -based supervisors can monitor the same agent. However, if multiple RSM servers exist in the environment that monitor the same agent, each server makes a separate monitoring call to that agent.

If the monitoring limit is reached for a specific agent and a dialed-in supervisor then attempts to monitor the same agent, the supervisor's request is denied. The supervisor receives an audio prompt from the system that states the agent cannot be monitored.

RSM Server Considerations

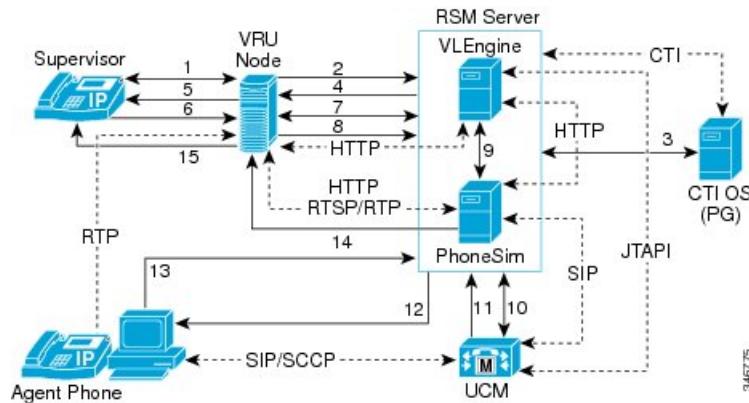
An RSM server can support 12000 active agents, 72000 configured agents, and a maximum of 80 concurrent monitoring sessions. RSM can support monitoring these agents distributed across multiple PGs and supports up to a six PG clusters configuration on each server. To support greater numbers, you can deploy additional RSM servers. In all supported RSM configurations, the VLEngine and PhoneSim components are installed on the same VM.

For more information about remote silent monitoring, see the *Cisco Remote Silent Monitoring Installation and Administration* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

RSM Component Interaction

The following figure illustrates the interactions that occur when a supervisor dials into an RSM-enabled platform and monitors an agent.

Figure 56: Remote Silent Monitor-Enabled Call Flow



The RSM call flow steps are as follows:

- 1 A supervisor calls in and the call is media-terminated on the VRU (Unified CVP or Unified IP IVR). The VRU runs the RSM callflow script to handle the call. The call begins with a request for the user to authenticate. The user then enters credentials.
- 2 The VRU makes a login request to RSM over HTTP.
- 3 The VLEngine component in RSM interacts with the CTI or CTI OS server to validate the authentication credentials.
- 4 The VLEngine replies to the VRU node over HTTP with the authentication result.
- 5 If the supervisor is authenticated, the script in the VRU plays the main menu prompt.
- 6 The supervisor chooses to monitor a single agent from the main menu and enters the directory number (DN) for the agent to monitor.
- 7 The VRU checks with the VLEngine whether that agent can be monitored. The VLEngine checks whether the agent with that DN is logged in, is in talking state, and belongs to the supervisor's team, using previously cached event feed information from the CTI or CTI OS server. If the agent is available to be monitored, the VLEngine replies back to the VRU node.
- 8 The VRU node sends a monitor request to the PhoneSim to monitor the entered DN. For Unified CVP, this request is sent using RSTP protocol. For Unified IP IVR, this request is sent over HTTP.
- 9 The VLEngine works internally using HTTP.
- 10 The VLEngine sends a JTAPI request to Unified Communications Manager to monitor the agent's phone. The VLEngine then receives a JTAPI success response.
- 11 The PhoneSim component receives a SIP-based instruction from Unified Communications Manager to establish a monitoring call with the agent's phone.
- 12 The simulated phone establishes the monitoring call with the agent phone.

- 13 After the monitoring call is established, the Built-In Bridge (BiB) on the agent phone forwards the call conversation to the PhoneSim in the form of RTP packets.
- 14 For Unified IP IVR, the PhoneSim strips the RTP headers and streams this data to the VRU node over HTTP as a response to the request made in step 8. For Unified CVP, the PhoneSim streams the audio back to the CVP Gateway using RTSP and RTP protocols.
- 15 The VRU then plays the data to the supervisor as if it were a streaming audio prompt.

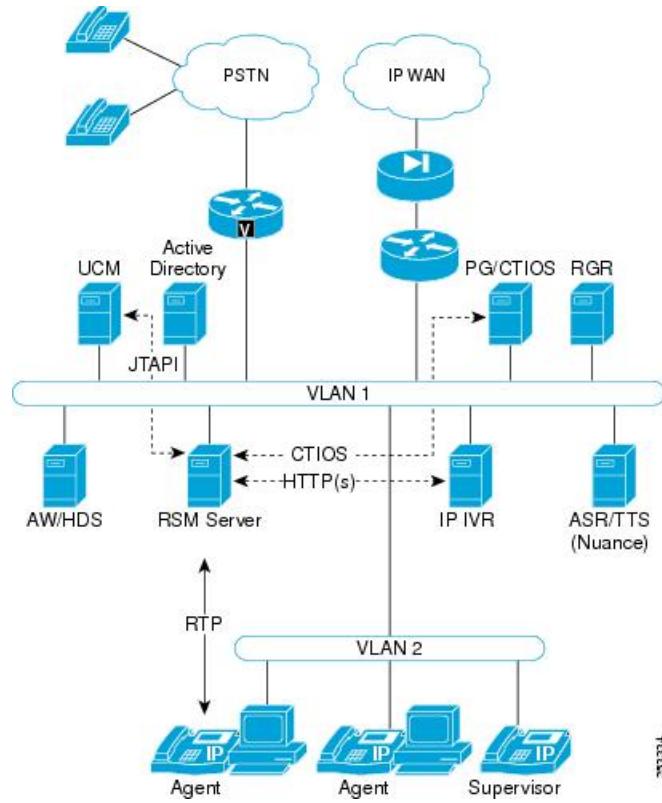
RSM Deployment Models

The following sections describe some basic supported RSM deployments.

Single Site

The following figure illustrates the basic network connectivity of RMS deployed within a typical single-site configuration.

Figure 57: Typical RSM VLAN Configuration



As shown in the preceding figure, supervisors may dial in through a VoIP phone or through the PSTN. The VRU that handles the supervisor's call in this case is Unified IP IVR.

This RSM VLAN configuration also illustrates the various protocol interfaces that RSM has into the rest of the system:

- **HTTPs:** VRU-based requests into the RSM system use HTTP as the carrier protocol. A request takes standard URL form as shown in the following examples:

```
http://rsmserver:8080/vlengine/checkUserCredentials?supervisorID=1101&pin=1234&outputFormat=plain
```

```
http://rsmserver:8080/vlengine/canMonitorAgentID?supervisorID=1101&agentID=1001&outputFormat=vxml
```

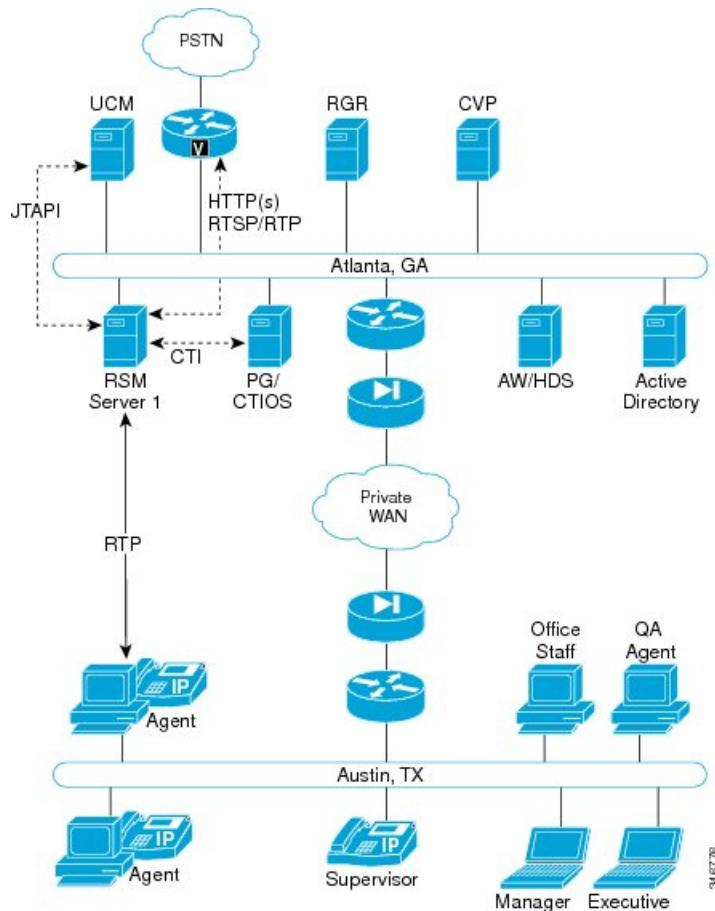
The first request is for the checkUserCredentials API call, while the second is for the canMonitorAgentID API call. Parameters to these requests are passed using the GET method. The return data (as an HTTP response) is either plain text or encapsulated in VoiceXML, depending on the API call used and on the value specified for the outputFormat parameter (if available for that call).

- **CTI or CTI OS:** The RSM server makes several connections to CTI or CTI OS. One connection is for receiving platform events (monitor-mode connection). The other connections (agent-mode connections) are used to authenticate supervisor sign-ins if standard authentication facilities are used.
- **JTAPI:** The request to start monitoring an agent phone is made through JTAPI. A JTAPI application user must be defined on each Unified Communications Manager cluster in the environment and be associated to all agent phones.
- **RTP:** While a dialed-in supervisor monitors an agent, a monitoring call is in place from the BIB of the agent phone to the RSM server. The signaling data for this call runs through Unified Communications Manager while the RTP traffic flows between the agent phone and the RSM server.

Multisite WAN

The following figures illustrate basic supported configurations for RSM in a multisite deployment.

Figure 58: Multisite Deployment with a Single Unified Communications Manager Cluster and Single VRU



In this scenario, the Unified Communications Manager and Unified CCE environment are co-located in Atlanta. The Austin location contains the entire end-user population. The VRU is a VXML Gateway/Voice Gateway in Atlanta, controlled by a Unified CVP Server, which is also in Atlanta.

The supervisor in Austin has two ways of dialing in to the RSM system:

- Through the PSTN — In this case, the supervisor dials an E.164 number and the call is hairpinned through the Voice Gateway. The Unified CVP RSM call flow application handles the call as normal from that point.
- As a VoIP extension — In this case, Unified Communications Manager has a trunk configuration set up to the VRU. The call remains VoIP all the way through. The Unified CVP RSM call flow application handles the call .

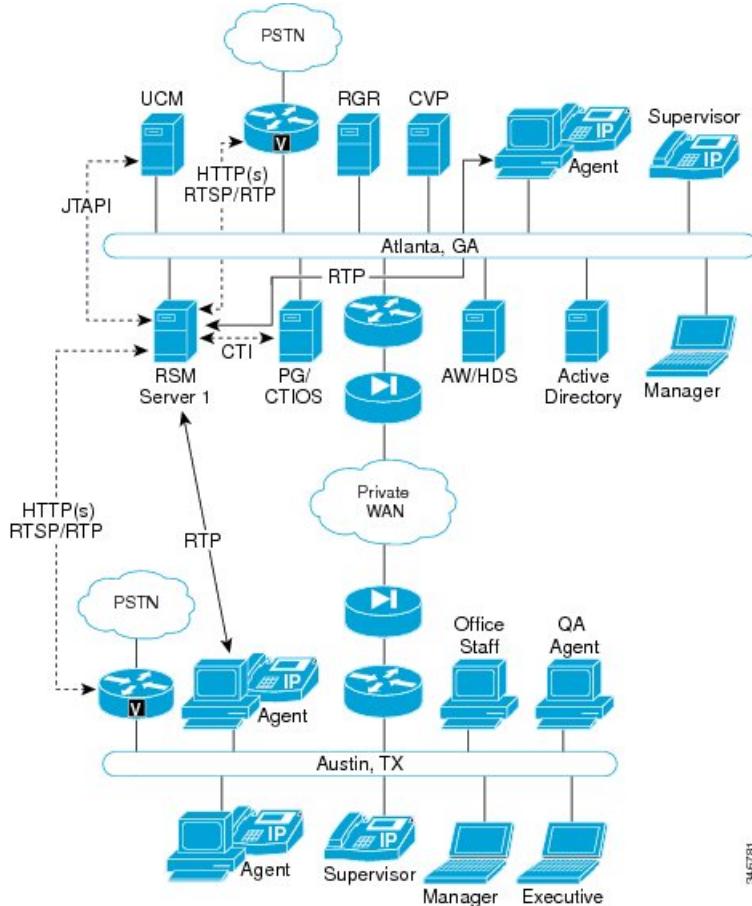
In this scenario, all RSM traffic is confined to the Atlanta site except the following:

- the RTP traffic of the monitored agent

- the actual supervisor call into the platform

The following figure depicts a multisite deployment with a single Unified Communications Manager cluster and multiple VRUs.

Figure 59: Multisite Deployment with a Single Unified Communications Manager Cluster and Multiple VRUs



346781

This scenario is similar to the previous scenario, with the addition of PSTN access at the Austin site. This scenario also adds personnel to the Atlanta site.

With the addition of a PSTN egress point in Austin, a call from a supervisor at the Austin location to the RSM system can be backhauled across the WAN (if VoIP end-to-end) or sent across the PSTN if the Atlanta DID associated with the RSM application was dialed.

In this example, Unified CVP is used, as well as the Unified CVP Server. However, there are two VXML Gateways, one at each site. The environment is configured so that a supervisor dialing RSM from Austin is routed to the RSM call flow application on the Austin VXML Gateway, while a supervisor dialing in from Atlanta is routed to the Atlanta VXML Gateway.

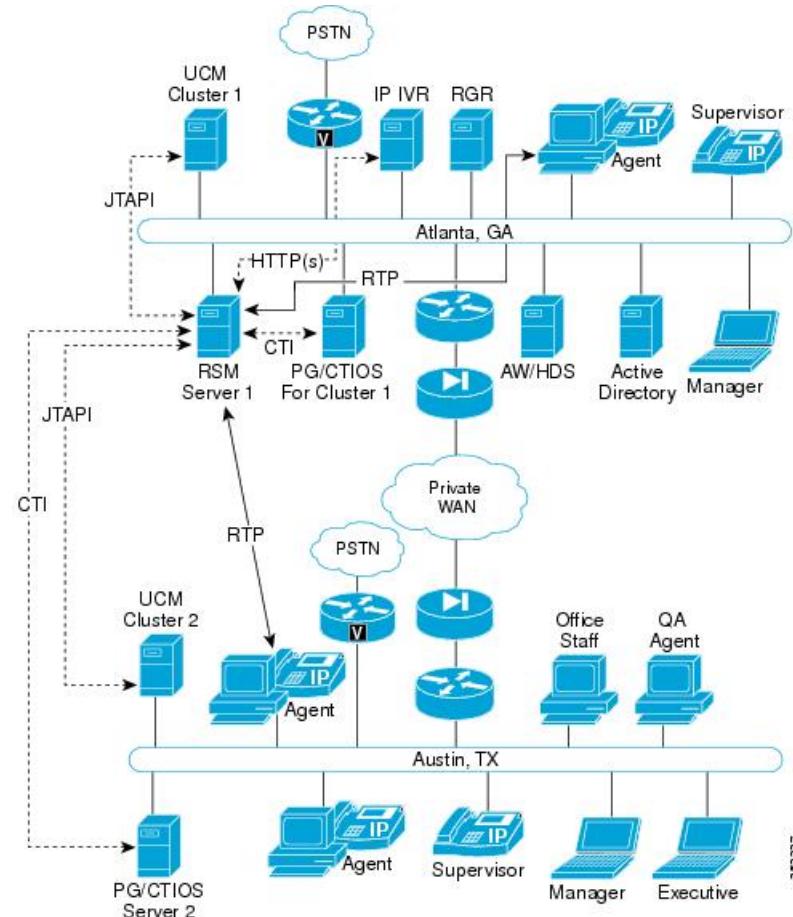
Because the Atlanta site houses the Unified Communications Manager and Unified CCE environment, all RSM-related JTAPI and CTI or CTI OS traffic is still confined there. However, the addition of a VXML Gateway at Austin leads to HTTP-based traffic streamed between the sites over the WAN. This traffic consists

of relatively small requests from the gateway to the RSM server for services, and the RSM server responses. The responses themselves can be sizeable, especially when it is the data for a monitored conversation.

When an agent in Austin is monitored, the RTP data for that conversation is sent over the WAN back to the RSM server as well.

The following figure depicts a multisite deployment with multiple Unified Communications Manager clusters and a single VRU.

Figure 60: Multisite Deployment with Multiple Unified Communications Manager Clusters and a Single VRU



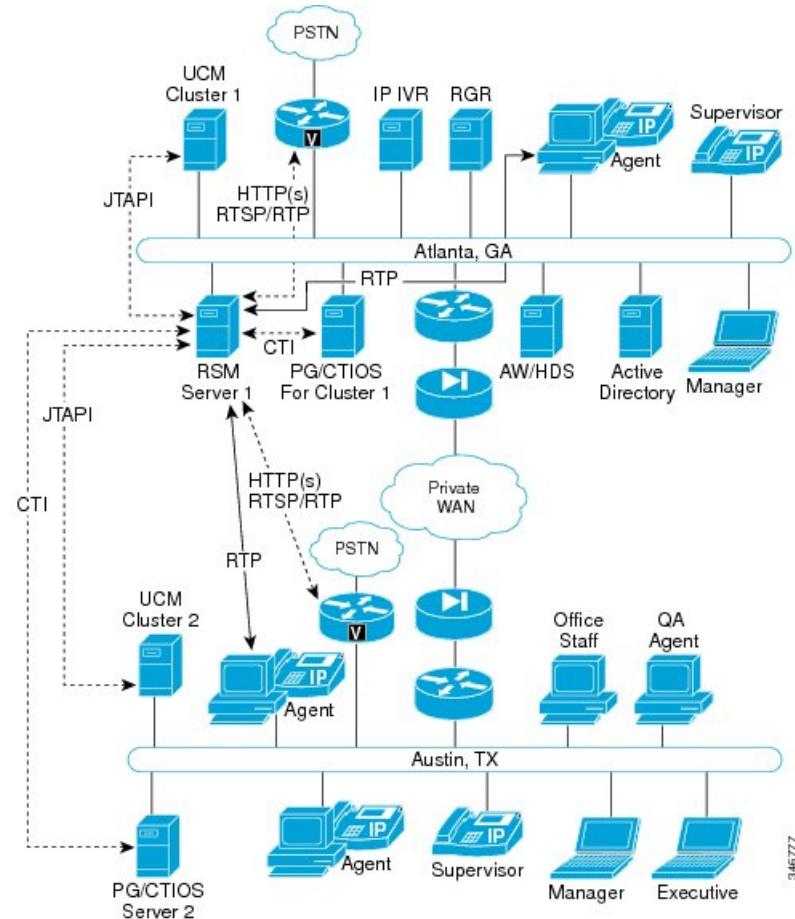
This scenario includes a Unified Communications Manager cluster at both the Atlanta and Austin sites and a single Unified IP IVR VRU in Atlanta. Cluster 1 handles the phone devices at the Atlanta site, while Cluster 2 handles the ones at the Austin site. The RSM server links to the CTI or CTI OS servers of both clusters to track all agents in the enterprise.

As Unified IP IVR is in use, a supervisor call to the RSM call flow is routed to, and media-terminated on, this Unified IP IVR system over either the PSTN or IP WAN. No VXML Gateway is involved in this configuration, and all RSM-related HTTP interaction is confined to the Atlanta site, between the RSM and Unified IP IVR systems.

Because a Unified Communications Manager cluster exists at the Austin site, several classes of data that RSM uses to track environment state and initiate agent monitoring requests (CTI or CTI OS and JTAPI traffic) are sent over the IP WAN.

The following figure depicts a multisite deployment with multiple Unified Communications Manager clusters and multiple VRUs.

Figure 61: Multisite Deployment with Multiple Unified Communications Manager Clusters and Multiple VRUs



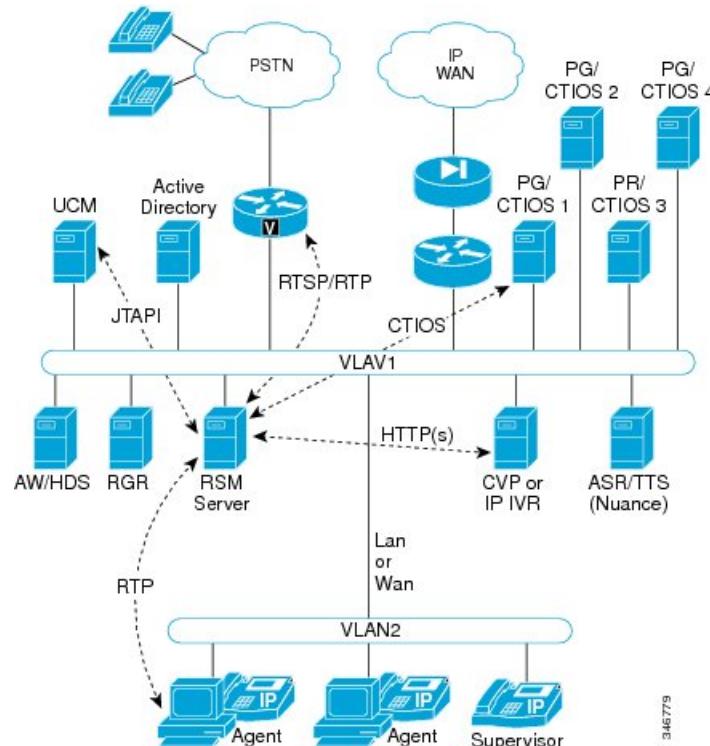
In this scenario, a Unified Communications Manager cluster and a Unified CVP VXML Gateway/Voice Gateway exist at each site. This deployment model is a combination of the previous models, and has the following characteristics:

- The Unified CVP Server controls the VXML Gateway at each site.
- Because there are agent phones at both sites, RTP data can be streamed either within the LAN at Atlanta (if the requested agent to monitor is in Atlanta) or across the WAN (if the requested agent is in Austin).
- As with the previous multisite, multicluster deployment, the RSM tracks the state of the entire enterprise. A supervisor can dial in from either site (or from anywhere in the world through PSTN) and listen to an agent in Atlanta or Austin.

Single Cluster with Multiple PG/CTI OS

The following diagram depicts a setup involving a single Unified Communications Manager cluster and multiple (up to 4) Agent PG/CTI OS servers:

Figure 62: Single Unified Communications Manager Cluster and Multiple (up to four) Agent PG/CTI OS Servers



In this scenario, a separate Agent PG/CTI OS server is deployed for each subscriber node pair (primary and backup).

A separate RSM cluster (in a single RSM instance) should be configured that corresponds to each Agent PG/CTI OS server and its Unified Communications Manager subscriber pair.

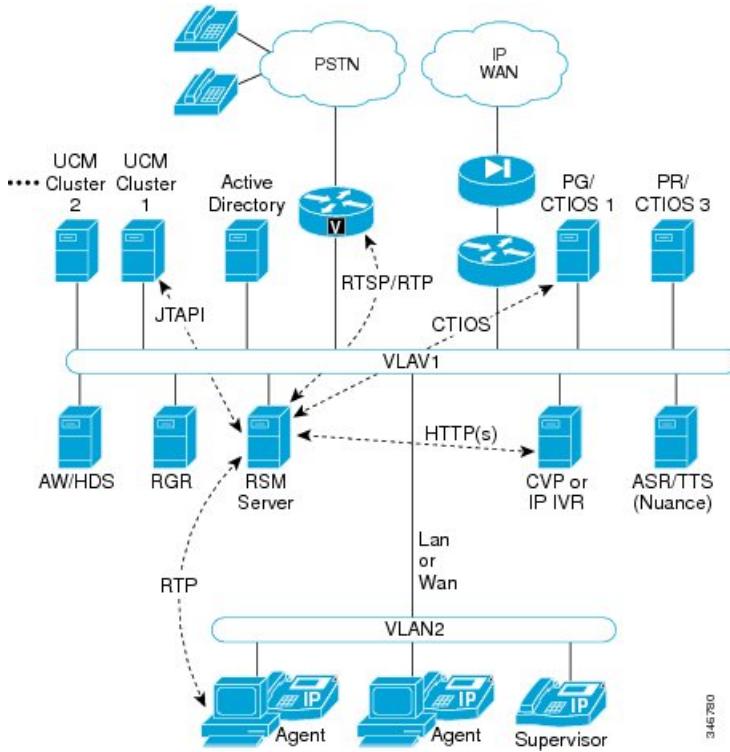
Related Topics

[Deployment of Agent PG in Unified Communications Manager Cluster, on page 235](#)

Multiple Cluster with Multiple PG/CTI OS

The following diagram depicts a setup involving multiple Unified Communications Manager clusters and multiple Agent PG/CTI OS servers.

Figure 63: Multiple Unified Communications Manager Clusters and Multiple Agent PG/CTI OS Servers



In this scenario, a separate Agent PG/CTI OS server is deployed for each Unified Communications Manager cluster.

A separate RSM cluster (in a single RSM instance) should be configured that corresponds to each Agent PG/CTI OS Server and its Unified Communications Manager cluster subscriber pair.

Remote Silent Monitoring Bandwidth Requirements

Before you deploy the RSM solution, verify that your network infrastructure can support the bandwidth requirements of RSM.

The RSM solution connects with multiple components in the larger Cisco environment. The following table lists these components, along with the nature of the data exchanged and the relative bandwidth requirements of that data. If RSM exchanges multiple types of data with a specific component, it is listed multiple times.

Table 14: Bandwidth Requirements

RSM peer	Purpose	Protocols used	Data format	Relative bandwidth requirements	Link latency requirements
VRU	Service requests / responses	TCP (HTTP)	Textual	Minimal	< 500 ms avg.
VRU	Requested voice data from PhoneSim to VRU	TCP (HTTP) TCP (RTSP) UDP (RTP)	For IP IVR - G.711 u-law in WAV format and HTTP chunked transfer encoding format. For CVP - G.711 u-law, G.711 a-law, and G.729 in RTP.	High (about 67 to 87 kbps per session)	< 400 ms avg.
Unified CM	Issuance of agent phone monitoring	TCP (JTAPI)	Binary (JTAPI stream)	Minimal	< 300 ms avg.
CTI or CTI OS server (PG)	Environment events / supervisor logins	TCP (CTI or CTI OS)	Binary (CTI or CTI OS stream)	Minimal (< 1000 agents) Moderate (> 1000 agents) (with 2000 agents, about 100 kbps)	< 300 ms avg.
Agent phones	Simulated phone signaling	TCP or UDP (SIP)	Textual	Minimal	< 400 ms avg.
Agent phones	Monitored phone voice data	UDP (RTP)	Binary (G.711 u-law, G.711 a-law, and G.729)	High (about 67 to 87 kbps per session)	< 400 ms avg.

Agent Phone Bandwidth Figures

The simulated phones on the RSM server support and advertise G.711 u-law, G.711 a-law, and G.729 codecs to establish the monitor call with agent phones.

For information about bandwidth usage, see the [Cisco Voice Over IP - Per Call Bandwidth Consumption Tech Note](#).

Sufficient bandwidth must be available from the agent IP phone to the RSM server to support the monitoring voice stream, in addition to the regular voice streams for the call. This is important for agents who work

remotely, at home, and in small branches on limited bandwidth or WAN connectivity. Regular Call Admission Control (CAC) and bandwidth calculations are applicable for monitoring calls.

Use the [Cisco TAC Voice Bandwidth Codec Calculator](#) for additional bandwidth capacity planning.

RSM Codec Support

The monitoring call established between the RSM simulated phone (simphone) and agent phone is subject to regular call admission control (CAC) procedures. The simphones on the RSM server support and advertise G.711 u-law, G.711 a-law, and G.729 codecs to establish the monitor call with agent phones.

Incoming Calls from Phone (BIB) to RSM

For all incoming calls from phone (BIB) to RSM, the built-in phone transcoding resources transcode the call (if necessary) and send it on to the monitoring call leg. No additional transcoding resources are needed on Unified Communications Manager or Voice Gateway. If the incoming call is G.729, configure the RSM Sim Phones Region to allow a Max Audio Bit Rate of 8 Kbps. If the incoming call is G.711, configure the RSM Sim Phones Region to allow a Max Audio Bit Rate of 64 Kbps.



Note

The Cisco IP phone's BIB does not support dual codec calls. If the Caller-Agent conversation is in G.729, the media forking to RSM is also in G.729.

Outgoing Calls to Unified CVP Systems from RSM

The outgoing monitoring call to the CVP gateway can be G.711 u-law, G.711 a-law, or G.729. The RSM server performs the necessary transcoding. No Voice Gateway transcoding resources are needed.



Note

If the RSM application is configured using a comprehensive flow, only G.711 a-law or G.711 u-law can be configured for the RSM-to-CVP call leg. This requirement is due to other dependencies related to Agent Greeting and other ICM functionality and their inability to handle G.729. To use G.729 for the CVP call leg, configure RSM in a standalone call flow.

Outgoing Calls to Unified IP IVR Systems from RSM

The outgoing monitoring call leg is G.711 u-law. The RSM server encodes to G.711 u-law, if transcoding is necessary.

For more information, see the section on codecs for monitoring and recording calls in the *Features and Services Guide for Cisco Unified Communications Manager* at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-installation-and-configuration-guides-list.html>.

Failover Redundancy and Load Balancing

RSM does not support true failovers or load balancing. RSM does support the deployment of multiple standalone RSM servers within a single Unified CCE environment. This concept is demonstrated in the advanced deployment scenarios described in this document.

The following table indicates how a failure of each of the various components affects a live supervisor call.

Table 15: Impact of Failures on a Supervisor Call

Component that fails	Worst possible impact
VRU node (Unified IP IVR, Unified CVP)	The supervisor's call is terminated as any VRU fail-over occurs (depends). The supervisor may dial back in and log in again once VRU fail-over is complete and/or the original failed VRU is working again.
RSM server (hardware failure)	Callers listening to a voice stream from the failed server have the voice stream terminated and are returned to the main menu. Their next attempt to make a service request to the failed server (or a new caller's first attempt to make such a request) results in a configurable delay of 3 to 5 seconds, as the request times out and an error message is played. Furthermore, any action that attempts to contact the RSM server (for example, logging in or attempting to monitor an agent), fails. The RSM callflow is answered because it is hosted on the VRU node.
VLEngine or PhoneSIM software failure	Service automatically restarted via service wrapper. Supervisors with a request in-progress receive an error message and can retry their last action. During the time either service is not functioning, any action that attempts to contact the RSM server (for example, logging in or attempting to monitor an agent), fails. The RSM callflow is still answered because it is being hosted on the VRU node.
Unified CCE fails (CTI or CTI OS)	RSM loses connectivity to the CTI or CTI OS server when the PG fails or is cycled. If connectivity to both CTI servers on a cluster fails, RSM keeps retrying both, and connects to the first available server. (The CIL failover code is used for all of this.) When connectivity comes back up to a CTI server, the agent and call lists are cleared and refreshed (to avoid stale agents). During this time, no new call events are received, and the system works from an out-of-date agent and call list. Therefore, some monitoring requests fail, saying the agent is not talking when he or she is, and some monitoring requests fail because the system thinks the agent is talking when he or she currently is not. This is believed to be preferable to the scenario where all cached data is deleted when the server goes down, in which case no monitoring works.
Unified Communications Manager fails (JTAPI)	Connectivity to one or more JTAPI providers are lost. RSM can be configured for connectivity to a maximum of 2 JTAPI providers per cluster. If this is the case and connectivity to either of the providers is lost, VLEngine fails over to the other provider if necessary, making it the active one and making its requests through it. If connectivity to both providers is lost, VLEngine periodically retries both and re-establishes connectivity to the first available provider. Attempts to monitor agents (for example, monitorAgent calls) made during this time fail until the JTAPI connection is re-established.

Host-Level Security

You can restrict incoming access to the RSM server to only the necessary components with the host-based Access Control List (ACL) functionality in the Windows Server OS. In the most secure configuration, incoming access to the RSM system is permitted from the VRU systems. You can also employ this host-based access control to allow limited access to other services, such as remote administration mechanisms like Windows Remote Desktop and VNC.

ACL is not required, but an example ACL Configuration for a single-server RSM configuration is as follows:

- Deny incoming access to all
- Permit incoming TCP on port 8080 to each VRU node in the environment (VLEngine HTTP API Access)
- Permit incoming TCP on port 29001 to each VRU node in the environment (PhoneSim HTTP API Access)

Transport or Session-Level Security

Because RSM maintains multiple connections to a number of components in the contact center environment, support for transport or session-level security varies by protocol type. The following notes describe RSM support for transport or session-level security by protocol type:

RSM to VRU (HTTP, RTSP, and RTP)

Encryption of the HTTP-, RTSP-, and RTP-based data exchange between RSM and the VRU node is not supported.

RSM to PG/CTI OS Server (CTI)

Because RSM makes use of the Java CIL, all CTI OS servers that use RSM must be set up with security disabled. CTI OS traffic may be encrypted through the use of IPSec transport mode encryption. For more information, see the section on security settings in the *Cisco Remote Silent Monitoring Installation and Administration* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

RSM to Unified Communications Manager (JTAPI)

Like CTI OS traffic, JTAPI traffic may be encrypted through the use of IPSec transport mode encryption. For more information, see the section on security settings in the *Cisco Remote Silent Monitoring Installation and Administration*.

RSM to Agent Phone (RTP)

Encryption of the RTP stream between agent phones (BIB) and the RSM SimPhone is not supported. RSM SimPhones do not support secure RTP (SRTP).

Support for Mobile Agent, IP Communicator, and Other Endpoints

Unified Communications Manager monitoring functionality does not provide monitoring support for endpoints using any one of the following:

- Cisco Mobile Agent
- Phones that do not have a Built-In Bridge (BIB).

- A media-terminated CTI OS agent desktop
- Monitoring of encrypted phone calls

Therefore, support for these products is also not available through RSM. See the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE for a list of supported phones.

Related Topics

[Silent Monitoring](#), on page 122

Deployment Considerations

Citrix and Microsoft Terminal Services

This section discusses Unified CCE desktops in a Citrix or Microsoft Terminal Services (MTS) environment.



Note

AWSs, Configuration-Only Administration Servers, and Administration Clients can operate only as a single remote instance on a given VM.

Cisco Finesse

Cisco Unified CCE supports running the Cisco Finesse desktop within a Citrix environment. Finesse supports Citrix XenApp and XenDesktop.

For more information about supported versions, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.



Note

You cannot use Cisco Finesse with Microsoft Terminal Services.

CTI OS Toolkit Desktop

Cisco Unified CCE supports running CTI Toolkit Desktop within the Citrix and Microsoft Terminal (MTS) Services environments. If you plan to use Citrix terminal services with the CTI Toolkit Desktop, take the following considerations into account:

- Versions of Citrix MetaFrame Presentation Server prior to Version 4.0 or 4.5 are not supported. Earlier versions have limitations for publishing Microsoft .NET applications.
- CTI OS Java CIL client applications are supported only on Citrix MetaFrame Presentation Server 4.0 and 4.5 for the Windows platform. There is no planned support for Citrix MetaFrame Presentation Server 4.0 or 4.5 on UNIX.
- Silent Monitoring is supported within a Citrix or MTS environment.

- CTI OS Client Desktop sounds such as dial tones and DTMF tones are not audible.

For implementation details, see *Integrating Cisco CTI OS Into a Citrix Metaframe Presentation Server/Microsoft Terminal Services Environment* at <http://www.cisco.com/c/en/us/support/customer-collaboration/computer-telephony-integration-option/products-installation-and-configuration-guides-list.html>.


Note

CTI OS supports virtualized desktop infrastructure from Citrix and VMware. CTI OS also supports Cisco VXI endpoints. When you deploy VDI or VXI, the performance, bandwidth, and timing requirements for CTI OS, as defined in this document, must still be met.

NAT and Firewalls

This section discusses deploying CTI Toolkit Desktop and Cisco Finesse in an environment where two or more disjointed networks are interconnected using Network Address Translation (NAT).

For more information about NAT and firewalls, see the chapter “Securing Cisco Unified Contact Center Enterprise”.

Cisco Finesse and NAT

Cisco Finesse provides limited support for NAT. Finesse does support basic NAT (one-to-one IP address mapping) between Finesse servers and Finesse clients.

The following caveats apply to Finesse and NAT:

- You cannot use PAT/NPAT (one-to-many address mapping that uses ports) between Finesse servers and Finesse clients.
- You cannot use NAT between the Finesse servers and any of the servers to which they connect (such as Unified CCE or Unified Communications Manager servers).

CTI Toolkit Desktop and NAT

When you deploy the Cisco CTI Toolkit Desktop in a network environment where NAT connects two or more disjointed networks, Unified Communications Manager, the physical IP phone, the CTI OS server, CTI Toolkit Desktop, and the CTI OS Unified CCE supervisor desktop must all be on the same network.

Cisco Finesse and CTI OS Agents on the Same PG

Unified CCE deployments can support a mix of Finesse and CTI OS agents on the same PG with the following limitations:

- The maximum number of CTI all events connections supported by the CTI server (seven) is not exceeded.
- The total number of combined Finesse and CTI OS agents does not exceed the capacity of the common PG.

If a mix is deployed, the sizing limitations of Finesse apply.

**Note**

The Finesse supervisor application can monitor only Finesse agents and the CTI OS supervisor application can only monitor CTI OS agents.

IP Phone and IP Communicator Support

Cisco Finesse and the CTI Toolkit Desktop support the use of Cisco IP hardware phones and the Cisco IP Communicator software phone.

For more information about the phone models and IP Communicator versions that each desktop supports, see the *Unified CCE Compatibility Matrix* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.

IP Phones and Silent Monitoring

Silent monitoring supports both IP hardware phones and Cisco IP Communicator.

IP Phones and Mobile Agent

The Mobile Agent feature does not require any specific type of phone. You can even use analog phones with this feature.

IP Phones and Citrix or MTS

Finesse and the CTI Toolkit Desktop support both IP hardware phones and Cisco IP Communicator when using Citrix or MTS. In these environments, you must install Cisco IP Communicator on the agent desktop PC. You cannot deploy Cisco IP Communicator on the Citrix or MTS server.

Cisco Jabber Support

Cisco Finesse supports Cisco Jabber for Windows as a contact center voice endpoint. Finesse supports the following Jabber functionality:

- Voice and Video
- Built-In Bridge (BIB) for silent monitoring
- IM and Presence

**Note**

Agents cannot use Jabber to transfer or conference calls. Agents must use the Finesse desktop for transfer and conference.

To use Jabber with Finesse, you must change the default Jabber configuration as follows:

- Change Maximum number of calls from 6 to 2.
- Change Busy trigger from 2 to 1.

Desktop Latency

You can locate Agent and Supervisor desktops remotely from the Agent PG. In a poorly designed deployment, high time-out values can cause the delay between the desktop server and desktop clients to be too high. Large latency affects the user experience and can become confusing or unacceptable from the user perspective. For example, the phone can start ringing before the desktop updates. Limit the latency between the server and agent desktop to 400-ms round-trip time for Finesse and CTI OS.

Finesse also requires that you limit latency between the Finesse server and the PG to 200-ms round-trip time.



CHAPTER 6

Cisco Outbound Option Description

- [Cisco Outbound Option Feature Description, page 149](#)
- [Cisco Outbound Option Processes, page 150](#)
- [Benefits of Cisco Outbound Option, page 150](#)
- [Cisco Outbound Option Deployment Considerations, page 151](#)
- [Outbound Dialing Modes, page 152](#)
- [Cisco Outbound Option for Unified CCE, page 156](#)

Cisco Outbound Option Feature Description

The Outbound Option Dialer is a software-only process that coresides on the Unified Communications Manager PG. The SIP Dialer process communicates with Voice Gateways, Outbound Option Campaign Manager, CTI Server, and MR PIM. The Dialer communicates with the Campaign Manager to retrieve outbound customer contact records and to report outbound call disposition (including live answer, answering machine, RNA, and busy). The Dialer communicates with the Voice Gateway to place outbound customer calls. The Dialer communicates with the CTI Server to monitor skill group activity and to perform third-party call control for agent phones. The SIP Dialer communicates with the MR PIM to submit route requests to select an available agent.

The Outbound Option Dialer can dial customers on behalf of all agents located on its peripheral. The Dialer is configured with routing scripts that can run in the following modes:

- Full blended mode—An agent can handle inbound and outbound calls
- Scheduled modes—for example, 8:00 AM to 12:00 PM in inbound mode and 12:01 PM to 5:00 PM in outbound mode
- Completely in outbound mode

If blended mode is enabled, the Dialer competes with inbound calls for agents. The Dialer does not reserve more agents than are configured in the administrative script Outbound Percent variable. If all agents are busy, then the Dialer does not attempt to reserve any additional agents.

You can achieve high-availability for SIP Dialer deployment with multiple Voice Gateways and Unified SIP Proxy servers. The redundancy is also achieved with redundant SIP Dialers.

Cisco Outbound Option supports Call Progress Analysis configuration on a campaign basis. When you enable this feature, the SIP Dialer instructs the Voice Gateway to analyze the media stream to determine the nature of the call (such as voice, answering machine, modem, or fax detection).

Campaigns are run as agent-based campaigns or VRU-based campaigns. A VRU is configured in an agent-based campaign to allow for handling of overflow calls when all agents are busy. In a transfer to an VRU-based campaign, all the calls are transferred to a VRU application after the outbound call is answered.


Note

The SCCP Dialer was deprecated in Unified CCE Release 10.0(1). Do not include the SCCP Dialer in new deployments. The SCCP Dialer will be removed in a future release.

Cisco Outbound Option Processes

Cisco Outbound Option for Unified CCE places outbound calls through a Voice Gateway. The Outbound Option Dialer does not require telephony cards to generate tones or to detect tones or voices.

The Cisco Outbound Option involves the following processes:

- Campaign Manager and Import processes manage campaigns.
- Campaign Manager and Import processes are always installed on the Side-A Logger and service only one customer instance.
- The Dialer process dials customers and connects them with properly skilled agents or available VRUs. The Dialer reports the results of all contact attempts back to the Campaign Manager. The central Campaign Manager manages all Dialer processes. The Dialer is installed on the same platform as the Agent PG.
- A Media Routing Peripheral is required for the Dialer to reserve agents for outbound use. It can coreside on other servers in a Unified CCE deployment.


Note

Precision Routing does not support Cisco Outbound Option. Outbound campaigns use skill groups. However, an agent involved in an outbound campaign (through an outbound skill group) can be logged in to a Precision Queue and handle inbound Precision Routing calls.

Related Topics

[Sizing Unified CCE Components and Servers, on page 215](#)

Benefits of Cisco Outbound Option

Cisco Outbound Option provides the following benefits:

- Enterprise-wide dialing, with IP Dialers placed at multiple call center sites. The Campaign Manager server is located at the central site.
- Centralized management and configuration through the Unified CCE Administration & Data Server.
- Call-by-call blending of inbound and outbound calls.

- Flexible outbound mode control. Use the Unified CCE script editor to control the type of outbound mode and percentage of agents within a skill to use for outbound activity.
- Integrated reporting with outbound specific reporting templates.

Cisco Outbound Option Deployment Considerations

Follow these requirements when implementing Cisco Outbound Option:

- Configure abandon to VRU in agent-based campaigns. Telemarketing laws often require this behavior.
- Schedule large imports of the contact list and Do-Not-Call list during off-hours because the Campaign Manager runs on the same system as the Side-A Logger.
- Do not use Cisco IP Communicator soft phone for agents configured for Cisco Outbound Option. IP Communicator can introduce an additional delay in transferring customer calls to the agent.

SIP Dialer Deployment Considerations

Cisco Outbound Option enables an agent to participate in outbound campaigns and take inbound calls through a SIP software dialer.

Follow these requirements when implementing the SIP Dialer:

- The Outbound SIP Dialer supports the T1 PRI and E1 PRI interfaces to the PSTN.
- Use a media routing PG with one Media Routing PIM for redundant SIP Dialers. One SIP Dialer is active while another SIP Dialer is in warm standby mode. One MR PIM is for each SIP Dialer. In a redundant MR PG environment, each PG side has only one PIM that connects to the local dialer when the Dialer becomes active.
- Use the g.711 codec in the dialer peer configuration of the gateway in the cases when the recording is enabled in the campaign configuration in a SIP Dialer deployment.
- Enable SIP Dialer call throttling to prevent overloading the Voice Gateways.
- The Voice Gateway dial peers and CUSP routing policies are used for SIP Dialers to place outbound calls. This enables calls to be placed using gateways that are deployed to leverage toll-bypass and lower local calling rates.
- When the SIP Dialer and Unified CVP share gateways, where the VXML gateway that is selected is the same as the gateway placing the outbound call for transfer to a VRU campaign or abandon to a VRU feature, configure Unified CVP to send the call back to the gateway it comes from to reduce network DSP resource usage and traffic, and to improve media transfer.



Note

Cisco Finesse now supports Progressive, Predictive, and Preview modes.

Related Topics

[SIP Dialer Throttling Considerations, on page 165](#)

Outbound Dialing Modes

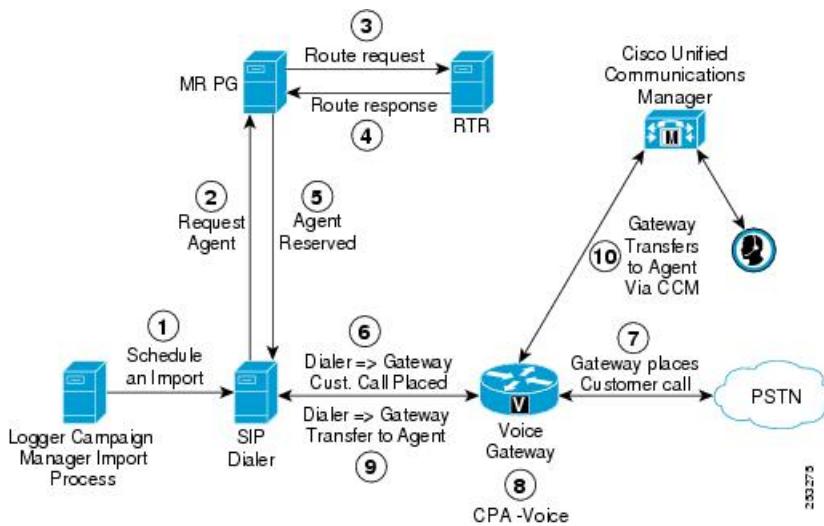
Cisco Outbound Option initiates calls using any of several modes, depending on the skill group:

- Predictive Mode—Dynamically calculates the number of lines to dial per agent to minimize agent idle time between calls.
- Progressive Mode—Uses a fixed number of lines per agent, set by the administrator.
- Preview Mode—Agent manually accepts, rejects, or skips customer calls (through enabled desktop buttons). Dials one line per agent.
- Direct Preview Mode—Allows the agent to hear the call ring-out from the desktop, similar to having the call placed by the agent directly. Dials one line per agent.
- Personal Callback Mode — When the person who is called requests to be called back later, the agent can specify that the callback is directed to the same agent. The system then calls the customer back at a pre-arranged time established between the requested agent and the customer.

Call Flow for Agent-Based Campaign

In an agent-based campaign, completed Dialer calls are routed to a live agent using a Unified IP Phone and desktop. The following figure shows the SIP Dialer call flow for agent-based campaigns with direct VG deployment.

Figure 64: SIP Dialer Call Flow for Agent-Based Campaigns—Direct VG Deployment



The SIP Dialer call flow with direct VG deployment for predictive/progressive dialing proceeds as follows:

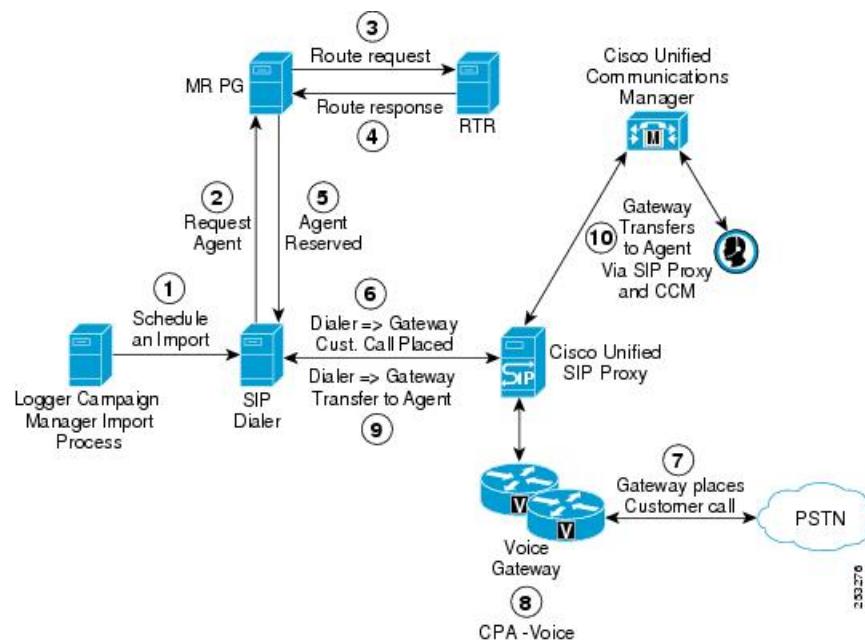
- 1 Import is scheduled and the campaign starts. Records are delivered to Dialer.
- 2 The dialer process continually monitors peripheral skill group statistics from the CTI server for an available agent. Concurrently the campaign manager monitors the database for customer records and forwards active

records to the dialer. When the dialer identifies an available agent for use in an outbound campaign, it sends a route request to the MR PIM.

- 3 The MR PIM forwards the route request to the router.
- 4 The Unified ICM/CCE/CCH CallRouter executes a routing script and selects an available agent. The CallRouter reserves that agent and returns a routing label (phone extension) identifying the reserved agent.
- 5 Media Routing PIM notifies the Dialer that the agent is available. The dialer then sends an agent reservation request to the Agent PG. The Agent PG generates a virtual agent reservation call to the agent desktop. The PG automatically places that virtual reservation call into answered state and then on hold.
- 6 Dialer signals the gateway to place outbound calls to the customers by using a SIP INVITE.
- 7 The VG places outbound calls to the customers, and Dialer is notified the VG is trying.
- 8 Call Progress Analysis is done at the VG. Voice is detected, and Dialer is notified.
- 9 The Dialer asks the VG to transfer the answered outbound call to the reserved agent by its agent extension.
- 10 The VG directs the answered outbound calls to the agents through Unified Communications Manager, using agent extensions and Unified Communications Manager host address. The dialer automatically answers the transferred call for the agent by way of the CTI server. This action quickly establishes the voice path between the customer and the agent.

The following figure shows the SIP Dialer call flow for agent-based campaigns in a Unified SIP Proxy deployment.

Figure 65: SIP Dialer Call Flow for Agent-Based Campaigns – Unified SIP Proxy Deployment



The SIP Dialer call flow in a Unified SIP Proxy deployment for predictive or progressive mode dialing proceeds as follows:

- 1 Import is scheduled and the campaign starts. Customer records are delivered to Dialer.
- 2 Dialer looks for an available agent by using the Media Routing Interface.
- 3 MR PG forwards the request to the Router.
- 4 The Routing Script identifies an agent and responds to the MR PG.
- 5 Media Routing PIM notifies the Dialer that the agent is available.

- 6 Dialer signals the Unified SIP Proxy server to find a gateway and tell it to place outbound calls to the customers through a SIP INVITE.
- 7 The VG places outbound calls to the customer.
- 8 Call Progress Analysis is done at the VG. Voice is detected, and Dialer is notified.
- 9 The Dialer asks the VG to transfer the answered outbound call to the reserved agent by its agent extension.
- 10 The VG begins the transfer to the Unified SIP Proxy server, and the SIP Proxy forwards the invitations onto Unified Communications Manager. Unified Communications Manager forwards the call invitations to the agent phone. The dialer automatically answers the transferred call for the agent by way of the CTI server. This action quickly establishes the voice path between the customer and the agent.

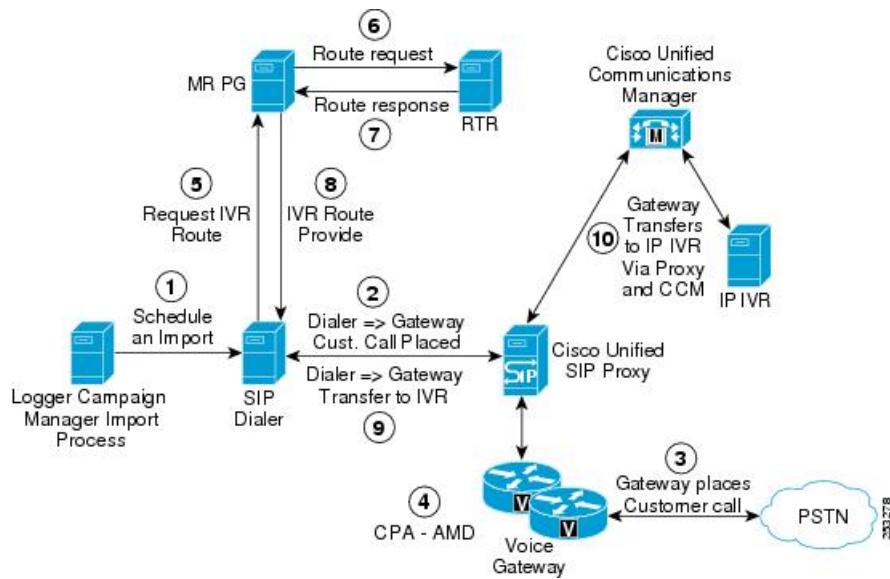
These message flows describe the flow for predictive or progressive mode dialing. The only difference in these two dialing modes is how the dialer determines its dialing rate (dynamic or fixed). For preview mode dialing, the agent receives a customer record screen pop. If the agent wants to call the customer, the agent must click the Accept button on the agent desktop. The button triggers a CTI event, which causes the dialer to call this customer.

Call Flow for Transfer to VRU Campaign

The SIP Dialer does not use CTI RP because the Agent PG does not monitor outbound calls during transfer to the VRU campaign. The SIP Dialer would also lose ECC variables with CTI RP. SIP Dialer uses the MR routing interface instead to request a transferred label from the Router.

The SIP Dialer call flow for VRU-based campaigns with a Unified SIP Proxy server and an Unified IP IVR deployment proceeds as shown in the following figure:

Figure 66: SIP Dialer and Unified IP IVR Call Flow for VRU Campaigns

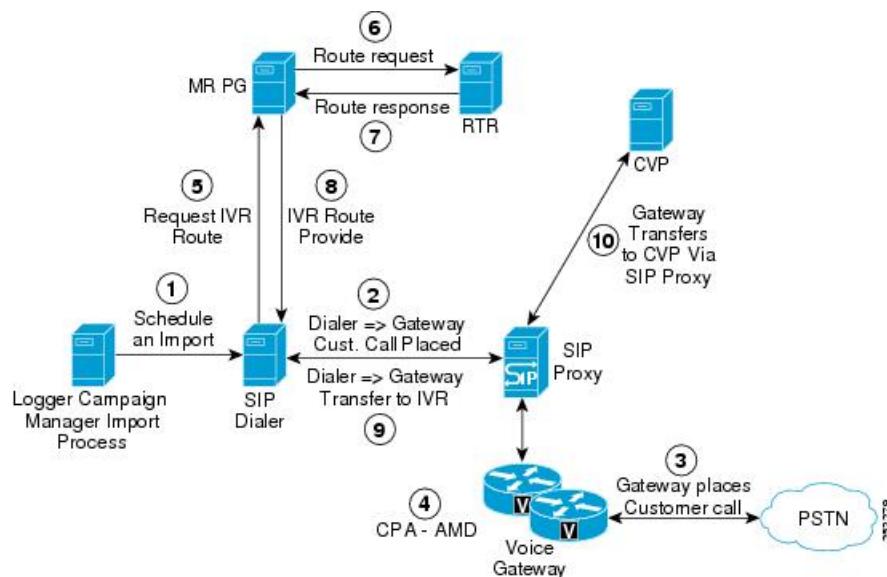


- 1 An unattended VRU campaign starts. Customer records are delivered to the Dialer.
- 2 The Dialer asks the SIP Proxy to forward an invitation to an available gateway to start a call.
- 3 The Voice Gateway (VG) calls the customer.
- 4 VG does Call Progress Analysis and detects live speech. The Dialer is notified.

- 5 The Dialer asks the MR PG where the VRU is.
- 6 MR PG forwards the request to the Router.
- 7 Routing Script identifies the VRU and notifies the MR PG.
- 8 The MR PG forwards the route response to the Dialer.
- 9 The Dialer notifies the VG to transfer the call to the VRU.
- 10 The VG begins the transfer to the SIP Proxy and the SIP Proxy forwards the call invitation to Unified Communications Manager.
- 11 Unified Communications Manager forwards the call invitation to the Unified IP IVR.
- 12 Media is set up between the VG and the Unified IP IVR.

The SIP Dialer call flow VRU-based campaigns with Unified SIP Proxy server and Unified CVP deployment proceeds as follows:

Figure 67: SIP Dialer and Unified CVP Call Flow for VRU Campaigns



- 1 In this example, an unattended VRU campaign starts. Customer records are delivered to the Dialer.
- 2 The Dialer asks the SIP Proxy to forward an invitation to an available Voice Gateway to start a call.
- 3 The VG calls the customer.
- 4 The VG does Call Progress Analysis and detects live speech. The Dialer is notified.
- 5 The Dialer asks the MR PG where the VRU is.
- 6 MR PG forwards the request to the Router.
- 7 Routing Script identifies the VRU and notifies the MR PG.
- 8 The MR PG forwards the route response to the Dialer.
- 9 The Dialer notifies the VG to transfer the call to the VRU.
- 10 The VG sends its invitation to the SIP Proxy, which forwards it to Unified CVP. The transfer is completed and media is set up between Unified CVP and the VG.

Cisco Outbound Option for Unified CCE

Enterprise Deployments

Run Cisco Outbound Option on a VM that meets the minimum requirements specified for the latest version in the *Virtualization for Unified CCE DocWiki* at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE.

The SIP Dialer is preferred for new deployments due to its high scalability by offloading call process resources and call progress analysis to the gateway. Furthermore, the SIP Dialer has no Unified CM or gateway proximity requirements.

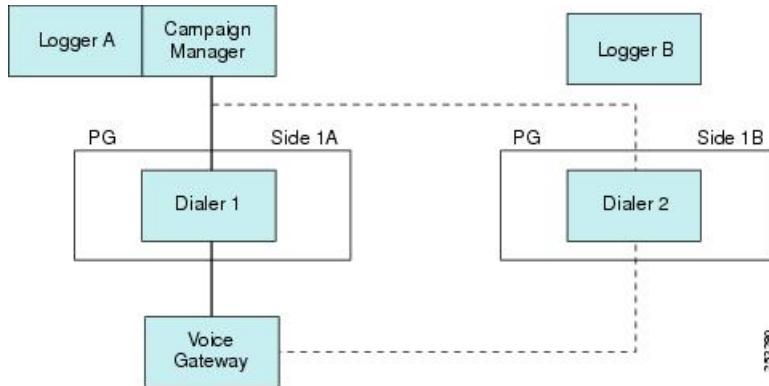
You can deploy the SIP dialer either on a VM with just the MR-PG or on a VM with the MR PG and the Agent PG. Redundant MR-PGs and Agent PGs are required.

The redundant Agent PG supports only redundant SIP Dialers; one dialer is active and another dialer is in warm-standby mode. For redundant SIP Dialer installations, each SIP Dialer connects to the MR PIM on the same MR PG side (Side_A or Side_B).

Single Gateway Deployment for SIP Dialer

The following figure shows the installation of redundant SIP Dialers with a single Gateway. The Dialers are shown to be installed on Side A and Side B of the redundant PGs. The port capacity depends on the type of Cisco Voice Gateway deployed. This deployment model is used when scaling and high availability are not factors.

Figure 68: Single Gateway Deployment for SIP Dialer



The SIP Dialer architecture supports only one active SIP Dialer per peripheral. Only one SIP Dialer needs to be configured. Two Dialers are installed on separate PG platforms, but each Dialer is installed using the same Dialer Name.

For Unified CCE deployments, the SIP Dialer and Media Routing PG processes can run on a separate VM or on the same VM as the Agent PG. For a deployment with redundant SIP Dialers and MR PGs on the Agent PGs, each MR PG has one MR PIMs that connects to the coresident SIP Dialer.

With Unified Communications Manager in single gateway deployments, the SIP Dialer uses the local static route file to place and transfer outbound calls when **Sip Server Type** is set to **Voice Gateway** in the Dialer setup dialog. These outbound calls are transferred to Unified CVP, Unified IP IVR, or outbound agents. Make sure the SIP Dialer uses the local static route file for single gateway deployments.

With Unified Communications Manager in single gateway deployments, the SIP Dialer uses the Unified SIP Proxy server to place and transfer outbound calls when **Sip Server Type** is set to **CUSP Server** in the Dialer setup dialog. These calls are placed or transferred to Unified CVP, Unified IP IVR, or outbound agents.

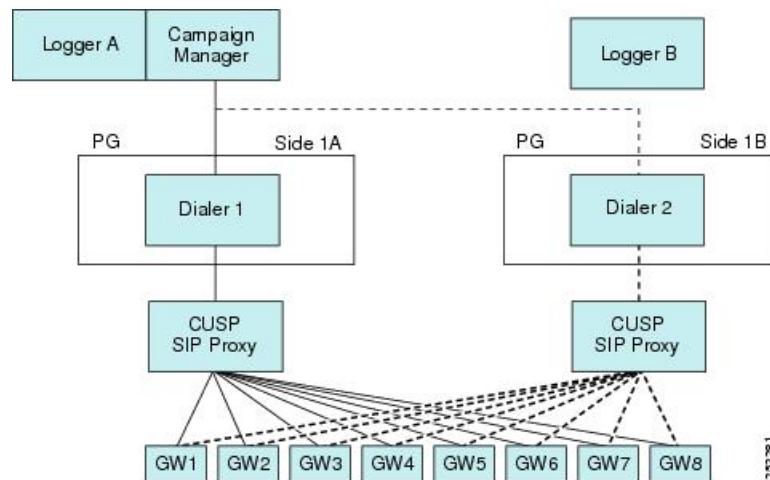
**Note**

Codec configuration (g.729 versus g.711) impacts port capacity and CPU utilization of gateways. Configuring g.729 requires more DSP and CPU resources for gateways.

Multiple Gateway Deployment for SIP Dialer

The following figure shows the deployment model for Unified SIP Proxy and eight Voice Gateways. The active Dialer points to the Unified SIP Proxy server. The proxy handles load balancing and fail-over. The SIP Dialer supports Unified SIP Proxy on the Cisco 3845 Integrated Services Router.

Figure 69: Multiple Gateway Deployment for SIP Dialer



In a multiple gateway deployment, the SIP Dialer requires Server Group and Route Table configurations on Unified SIP Proxy servers to identify the gateways, as well as numbers so that the gateways can determine where to send calls to Unified CVP, Unified IP IVR, or agents when the Dialer asks the gateway to transfer customer calls. Setting the **Sip Server Type** radio button to **SIP Proxy** in the Dialer setup dialog is required for multiple gateway deployment.

Related Topics

[High Availability Design for SIP Dialer, on page 167](#)

Clustering Over the WAN

The deployment model for clustering Unified CCE over the WAN allows for improved high availability by deploying redundant components on the other end of the WAN (see [Deployments, on page 27](#)). The Cisco Outbound Option high-availability model differs from the model that is used in clustering over the WAN; therefore, when deploying clustering over the WAN, keep in mind that its benefits are for inbound traffic only.

Distributed Deployments

A distributed deployment model involves a central Unified CCE system and Unified Communications Manager cluster located at one site, with the Campaign Manager installed on the logger at this site, and a second site reachable over a WAN, which consists of the dialer, a PG, and a second cluster with Cisco Outbound Option.

For SIP Dialer deployment, a Unified SIP Proxy server is installed for one SIP Dialer on each PG side, and the Side A/Side B Dialer is targeting the same set of Voice Gateways through its own Unified SIP Proxy server. Multiple Voice Gateways can be installed locally to customer phones, or each Voice Gateway can be installed locally to an area so that tolls are not encountered if leased circuits or IP MPLS WAN circuits are available.

The Campaign Manager sends dialer records over the WAN, and the dialer places calls to local customers. The second site would support inbound agents as well.

The following bandwidth options are available between India and the US in customer environments:

- 1 Terrestrial P2P leased 2 Mbps circuits
- 2 Terrestrial P2P DS3 (44 Mbps) leased circuits
- 3 IP MPLS WAN circuits. Varying speeds are available from the service provider depending on customer needs. Typical usage is 44 Mbps.
- 4 The service provider hands off PRI (E1) trunks to India. The WAN cloud is usually built on SIP by the service provider. The service provider converts TDM to IP at the ingress/egress point in the United States and converts IP to TDM in India.

Options 1 and 2 above are the most common. Option 3 is becoming more popular with outsourcers because the MPLS cloud can connect to several of their customers. For example, the diagrams in the following sections show that the Outbound Contact Center System is deployed across multiple sites in the United States and India for various agent-based campaigns or transfer to a VRU campaign. The customers are in one country; for example, in the United States.

Distributed Deployment for Agent-Based Campaign

In this distributed deployment example for an agent-based campaign:

- The Voice Gateway and Router/Logger A servers are distributed between two sites (Site 1 and Site 3) in the United States.
- The Unified Communications Manager cluster is located at Site 2 in India along with the Agent PG.
- The redundant MRPG/Dialer and redundant Agent PGs are installed on the same VM at Site 2 in India.
- The SIP Dialer uses the Voice Gateways that are located at Site 3 in the United States.

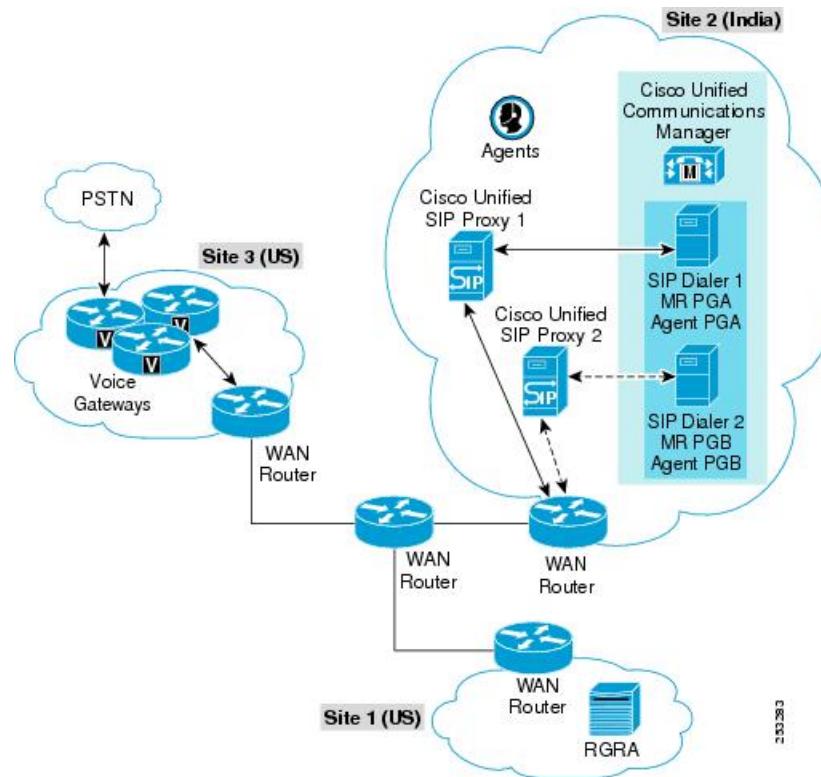
- The Voice Gateways are included in the diagram with CT3 interface at Site 3 in the United States. These routers provide 1:1 redundancy for Dialer calls.
- The Unified SIP Proxy servers are locally redundant at Site 2 to avoid the WAN SIP signaling traffic that is needed to transfer live outbound calls.
- Each SIP Dialer connects to its own Unified SIP Proxy server at Site 2.
- Each Unified SIP Proxy server controls the set of Voice Gateways at Site 3 in the United States.
- Each Unified SIP Proxy server controls the set of Voice Gateways at Site 3 in the United States.

If recording is enabled at the SIP Dialer, the bandwidth requirements are as follows:

- Answered outbound calls require the following bandwidth for each agent call:
 - g.711 Codec calls require a WAN bandwidth of 80 kbps
 - g.729 Codec calls require a WAN bandwidth of 26 kbps
- Alerting outbound calls require the following bandwidth for each agent call:
 - g.711 Codec calls require a WAN bandwidth of 80 kbps
 - g.729 Codec calls require a WAN bandwidth of 26 kbps

The following figure provides an example of a distributed deployment for an agent-based campaign.

Figure 70: Distributed Deployment Example for Agent-Based Campaign



Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

Distributed Deployment for Transfer to Unified CVP Campaign

In this distributed deployment example for an agent-based campaign:

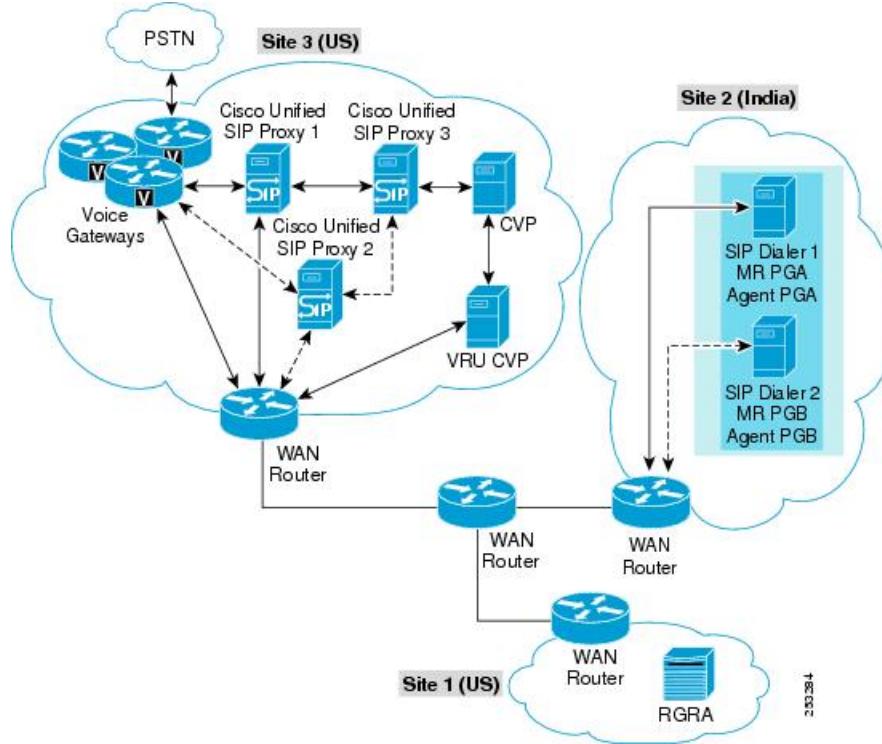
- The Voice Gateway and Router/Logger A servers are distributed between two sites (Site 1 and 3) in the United States.
- The redundant MRPG/Dialer and redundant Agent PGs are installed on the same VM at Site 2 in India.
- Unified CVP with local redundancy is included at Site 3 (United States). Unified CVP has its own Unified SIP Proxy servers for load balancing and redundancy.
- The VRU PGs are locally redundant at Site 3 (United States).
- The SIP Dialer uses the Voice Gateways located at Site 3 (United States).
- The Voice Gateways are included in the diagram with CT3 interface at Site 3 (United States). These routers will provide 1:1 redundancy for Dialer calls.
- The Unified SIP Proxy servers are locally redundant at Site 3 to avoid the WAN SIP signaling traffic to transfer live outbound calls.
- Each SIP Dialer connects to its own Unified SIP Proxy server at Site 3. Each Unified SIP Proxy server controls the set of Voice Gateways at Site 3 (United States).
- The Unified SIP Proxy servers provide $(N + 1)$ redundancy.

If recording is enabled at the SIP Dialer, the bandwidth requirements are as follows:

- Answered outbound calls require the following bandwidth for each agent call:
 - g.711 Codec calls require a WAN bandwidth of 80 kbps
 - g.729 Codec calls require a WAN bandwidth of 26 kbps
- Alerting outbound calls require the following bandwidth for each agent call:
 - g.711 Codec calls require a WAN bandwidth of 80 kbps
 - g.729 Codec calls require a WAN bandwidth of 26 kbps

- Outbound calls being queued or self-serviced at Unified IP IVR do not require WAN bandwidth.

Figure 71: Distributed Deployment Example for Transfer-to-VRU Campaign with Unified CVP



Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

Distributed Deployment for Transfer to IP IVR Campaign

In this distributed deployment example for a transfer-to-VRU campaign with IP IVR:

- The Voice Gateway and Router/Logger A servers are distributed between two sites (Site 1 and 3) in the United States.
- The Unified Communications Manager cluster is located at Site 3 (United States) along with the VRU PG.
- The redundant VRU PGs are at Site 3 (United States).
- IP IVR is included at Site 3 (United States).
- The redundant MRPG/Dialer and redundant Agent PGs are installed on the same VM at Site 2.
- The SIP Dialer uses the Voice Gateways located at Site 3 (United States).
- The Voice Gateways are included in the diagram with CT3 interface at Site 3 (United States). These routers will provide 1:1 redundancy for Dialer calls.

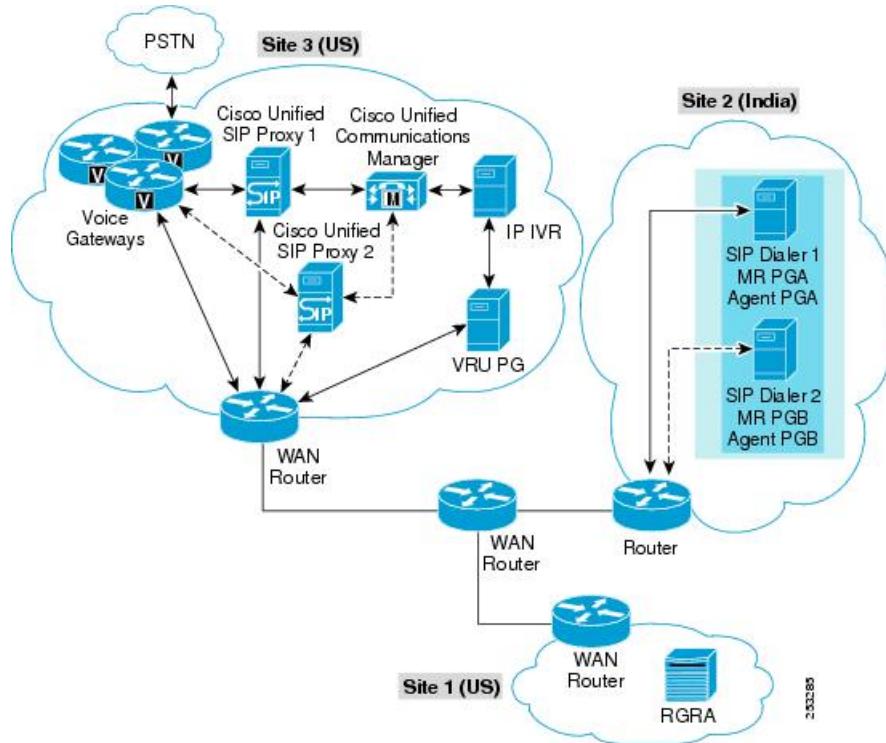
- The redundant Unified SIP Proxy servers are at Site 2 to avoid the WAN SIP signaling traffic to transfer live outbound calls.
- Each SIP Dialer connects to its own Unified SIP Proxy server at Site 2. Each Unified SIP Proxy server controls the set of Voice Gateways at Site 3 (United States).
- The Unified SIP Proxy servers provide (N+1) redundancy.

If recording is enabled at the SIP Dialer, the bandwidth requirements are as follows:

- Answered outbound calls require the following bandwidth for each agent call:
 - g.711 Codec calls require a WAN bandwidth of 80 kbps for the Call Progress Analysis time period.
 - g.729 Codec calls require a WAN bandwidth of 26 kbps for the Call Progress Analysis time period.
- Alerting outbound calls require the following bandwidth for each agent call:
 - g.711 Codec calls require a WAN bandwidth of 80 kbps
 - g.729 Codec calls require a WAN bandwidth of 26 kbps
- Outbound calls being queued or self-serviced at the IP IVR do not require WAN bandwidth.

The following figure provides an example of a distributed deployment for transfer-to-VRU campaign for IP IVR.

Figure 72: Distributed Deployment Example for Transfer-to-VRU Campaign with IP IVR



Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

Unified Contact Center Enterprise Deployments

Configure Cisco Outbound Option for Unified CCE

Cisco Outbound Option can run fully blended campaigns in which agents can handle inbound and outbound calls alternately.

When sizing your deployment, do not use the maximum number of outbound agents allowed on a PG without also looking at expected hit rate, lines dialed per agent, and average handle times. Use the [Cisco Unified Communication Sizing Tool](#) to size your Cisco Outbound Option deployment.

SIP Dialer targets the support of 1000 outbound agents for one PIM per PG. The number of supported agents is smaller when deploying mobile agents. To support this number of agents, the deployment must have at least five high-end gateways dedicated to outbound dialing.

SIP Dialer can support 1500 ports and 30 calls per second (cps). To achieve the rate of 30 cps, the SIP Dialer has to support from 1000 through 2000 ports, depending on hit rates and handle times.

Each port can dial two calls per minute, assuming an average 30 seconds per call attempt, so 30 ports can handle one call per second for the Dialer. If the time to get all ports busy exceeds the average port busy time, then some ports are always idle.

Calculate Number of Dialer Ports

The following formula can be used to calculate the number of dialer ports that are required to achieve targeted call rate:

$$\text{Number of Ports} = [\text{target call rate} * \text{average call duration} * (1 + \text{hit rate \%})]$$

For example, given an estimated average of 30-seconds per outbound call and given an estimated 20% hit rate, the following table shows the number of ports that are required to achieve targeted outbound call rates:

Table 16: Ports Required to Achieve Targeted Outbound Call Rates

Targeted outbound calls per second	Number of ports required
10	360
20	720
30	1080

Voice Gateway Considerations

The most powerful Voice Gateway supports about 12 calls per second, even under the most favorable conditions. Five gateways can support an aggregate spike of up to 60 calls per second when evenly distributed. However, even distribution does not account for occasions when ports are tied up with agent or VRU calls after the

transfer. So assuming a 50% transfer rate and using a conservative estimate, eight Voice Gateways are required to support a spike of up to 60 calls per second.

For the most current information about Voice Gateway models and releases that are supported by a Unified CCE SIP Dialer, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.

For gateway sizing considerations, see the published Cisco gateway performance data and Unified CCE sizing tool.

Agent PG Considerations

The Unified Communications Manager PIM can support up to 15 calls per second.

If the voice hit rate for the campaign is 15%, then the PG can sustain dialing at a rate of 100 calls per second.

Unified Communications Manager Considerations

The Unified Communications Manager subscriber can support a certain number of outbound calls per second. If the Dialer attempts to transfer a large numbers of live outbound calls per second at the agent PG, then it must be distributed across multiple subscribers using a Unified SIP Proxy server.

Cisco Unified SIP Proxy Considerations

A typical outbound call requires two transactions, if the call is transferred to an agent or VRU. A typical outbound call requires one transaction, if the call is not transferred to an agent or VRU.

The following table shows Unified SIP Proxy Sizing.

Table 17: Unified SIP Proxy Sizing

Hardware Model	Maximum Transaction Rate Per Second
NME-CUSP-522	100

Unified CVP Considerations

Calls can be distributed to Unified CVP using translation routes. Any load balancing across Unified CVPs happens in the routing script.

Since four SIP Proxy transactions are required for some outbound call scenarios with Unified CVP, give Unified CVP its own Unified SIP Proxy server in large-scale deployments.

Unified IP IVR Considerations

If Unified IP IVR is deployed, then front-end all calls through Unified Communications Manager. This deployment results in a higher call load on the subscribers. Because the subscriber supports only five calls per second, distribute calls transferred to agents and the VRU across multiple subscribers using the Unified SIP Proxy server.

Unified Mobile Agent Considerations

The SIP Dialer supports 500 unified mobile agents per Agent PG. With the SIP Dialer solution, the outbound calls have the same impact on Unified Communications Manager as inbound calls. Maintain a 2:1 ratio for number of inbound agents versus outbound agents. Since the SIP Dialer solution supports 1000 outbound regular agents per Agent PG, 500 outbound mobile agents per Agent PG is supported by the SIP Dialer.

For sizing the Cisco Outbound Option for SIP Dialer, use the [Cisco Solution Sizing Tool](#).

SIP Dialer Throttling Considerations

SIP Dialer Throttling is controlled by the field **Port Throttle** in the dialer configuration. Port Throttle indicates the number of ports to throttle per second. Setting the value to Port Throttle = 5 will allow SIP Dialer to dial outbound calls at a rate of five calls per second per Dialer.

When the SIP Dialer connects to the Voice Gateway directly in the deployment, limit the dialer port throttle by the maximum dialer call setup rate listed on the gateway sizing table.

When the SIP Dialer connects through the CUSP in the deployment, the port throttle setting on the dialer must not exceed the total gateway capacity under assumption. Calls is load-balanced through CUSP and each gateway will reach its maximum available capacity. Limit the port throttle by the CUSP maximum transaction. Currently, the dialer maximum throttle setting is 60 calls per second. Under normal transfer rate, calls through CUSP will not exceed maximum CUSP transaction rate given that CUSP is exclusively used by outbound deployments.

In a single or multiple gateway deployment, the SIP Dialer raises an alarm if any gateway is overloaded, and it automatically throttles the dialing rate down to ten percent of the configured port throttle value per 5000 customer attempts until fifty percent of the correction is met. Fifty percent of the correction means the SIP Dialer stops auto-throttling when it reaches fifty percent of the configured port throttle value.

SIP Dialer provides the option to disable the auto-throttle mechanism by setting the value of registry key EnableThrottleDown to 0. The auto-throttle mechanism is enabled by default. SIP Dialer still raises an alarm even though the auto-throttle mechanism is disabled.

Set the port throttle value to 5 for Cisco 2800 Series Integrated Services Routers, set the port throttle value to 15 for Cisco 3800 Series Integrated Services Routers, and set this value to 20 for Cisco Access Servers and Universal Gateways.

Single Gateway Deployment

Use the following formula to calculate the Port Throttle if the gateway is dedicated 100% for outbound campaigns:

$$\text{Port Throttle} = (\text{Value for Gateway})$$

Use the following formula to calculate the Port Throttle if the gateway is shared by multiple SIP Dialers for outbound campaigns:

$$\text{Port Throttle} = (\text{Value for Gateway}) / (\text{Number of SIP Dialers})$$

Use the following formula to calculate the Port Throttle if the gateway is shared by multiple Unified CCE components (Unified Communications Manager, Unified CVP, and SIP Dialer) for inbound/outbound calls:

$$\text{Port Throttle} = (\text{Value for Gateway}) * (\text{Percentage of outbound calls}) * (1 - \text{Hit Rate})$$

Multiple Gateway Deployment

Use the following formula to calculate the Port Throttle if the gateways are dedicated 100% for outbound campaigns:

$$\text{Port Throttle} = \text{Total Values for Gateways}$$

Use the following formula to calculate the Port Throttle if the gateways are shared by multiple SIP Dialers for outbound campaigns:

$$\text{Port Throttle} = (\text{Total Values for Gateways}) / (\text{Number of SIP Dialers})$$

Use the following formula to calculate the Port Throttle, if the gateways are shared by multiple Unified CCE components (Unified Communications Manager, Unified CVP, and SIP Dialer) for inbound/outbound calls:

$$\text{Port Throttle} = (\text{Total Values for Gateways}) * (\text{Percentage of outbound calls}) * (1 - \text{Hit Rate})$$

The throttling mechanism in the SIP Dialer process is not aware of which gateway the Unified SIP Proxy server selects to place outbound calls, so the appropriate weight for each gateway in the Server Group configuration of the Unified SIP Proxy server must be calculated for the load balance.

$$\text{Weight} = (\text{Value for Gateway}) / (\text{Port Throttle}) * 100$$

For example, if a Cisco 3800 Series Gateway (192.168.10.3) and a Cisco 2800 Series Gateway (192.168.10.4) are used in a multiple gateway deployment, the following configuration allows that 3800 Series gateway in the cucm.example.com server group to receive 75 percent of the traffic and the 2800 Series gateway to receive 25 percent.

```
netmod(cusp-config)> server-group sip group cucm.example.com enterprise
netmod(cusp-config-sg)> element ip-address 192.168.10.3 5060 tls q-value 1.0 weight 75
netmod(cusp-config-sg)> element ip-address 192.168.10.4 5060 tls q-value 1.0 weight 25
netmod(cusp-config-sg)> lbttype weight
netmod(cusp-config-sg)> end server-group
```

SIP Dialer Recording

The SIP Dialer can record ("Recording") or enable the recording of Call Progress Analysis by third-party applications ("Media Termination") to be used for CPA troubleshooting. Note that it does not record the full conversation.

There usually is no media stream between the SIP Dialer and the Voice Gateway. But when the recording or media termination is enabled in the Campaign configuration, the SIP Dialer requests the Voice Gateways to send the media stream to the SIP Dialer. The media stream is in g.711 or g.729 codec, depending on the dial peer configuration on the Voice Gateway. The SIP Dialer can record the media stream only with g.711 codec, but it can receive media streams for both g.711 and g.729 codecs to allow a third recording server to perform SPAN-based recording for outbound calls.

When "Recording" is enabled in the Campaign configuration, the SIP Dialer receives media streams, decodes RTP packets in g.711 codec, and writes them into a recording file. The SIP Dialer will send an alarm if the media stream is g.729 codec. The SIP Dialer has been tested to be able to support a maximum of 100 recording sessions per Dialer server due to CPU resource and disk I/O limitations.

When "Media Termination" is enabled in the Campaign configuration, the SIP Dialer will only receive the media stream to allow a third-party recording server to perform SPAN-based recording.

There is a limit for Media Termination Sessions because of a thread resource limitation per process. The SIP Dialer has to create a thread to listen on the media stream. The current limit for Media Termination Sessions is 200.

The SIP Dialer uses the following Registry keys to allow users to manage recording sessions and disk space:

Table 18: SIP Dialer Registry Keys

Name	Data Type	Description	Default Value
MaxRecordingSessions	DWORD	The maximum recording sessions per SIP Dialer, if the recording is enabled in the Campaign configuration.	100
MaxMediaTerminationSessions	DWORD	The maximum media termination sessions per SIP Dialer, if the recording is enabled in the Campaign configuration.	200
MaxAllRecordFiles	DWORD	The maximum recording file size (bytes) per SIP Dialer.	500,000,000
MaxPurgeRecordFiles	DWORD	The maximum recording file size (bytes) that SIP Dialer will delete when the total recording file size, MaxAllRecordFiles, is reached.	100,000,000

Call Transfer Timelines

The length of time required to complete a call transfer of a customer call to an agent is highly dependent on the telephony environment. The following factors can add to transfer times:

- Improperly configured Cisco Unified Communications infrastructure—Port speed mismatches between servers or inadequate bandwidth.
- WAN—WAN unreliable or not configured properly.
- IP Communicator—Media termination running on a desktop does not have the same system priority as software running on its own hardware platform, such as a hard phone (use hard phones instead of soft phones when using Outbound Option).
- Call Progress Analysis—When you enable Call Progress Analysis for the campaign, it takes approximately half a second to differentiate between voice and an answering machine if the voice quality is good. When calling cell phones, the voice quality is quite often less than optimal, so it might take the dialer or Voice Gateway a bit longer to differentiate.

High Availability Design for SIP Dialer

The Cisco Outbound Option with SIP Dialer provides high availability through fault tolerant design in SIP Dialer, Agent PG and Unified SIP Proxy server. Many components in the Cisco Outbound Option with SIP Dialer are duplicated for redundancy.

Campaign Manager and Import

The Campaign Manager and Import process components of Outbound Option are simplex components and must be co-located with the Logger (Side A).

The Campaign Manager supports a single active dialer per peripheral. Only one SIP Dialer needs to be configured. Install two SIP Dialers on separate PG platforms, but install each using the same Dialer Name.

The peripheral setup program allows users to input the dialer name in the setup page for each SIP Dialer.

When the SIP Dialer starts, it will attempt to register with the Campaign Manager. The Campaign Manager checks if the SIP Dialer is configured based on the dialer name from the registration message. It will reject the registration if it cannot find the configured SIP Dialer with that name. A maximum of two SIP Dialers can register with the same name; the Campaign Manager will reject the registration if that limit is exceeded.

The Campaign Manager activates only one SIP Dialer in the ready state from its registered SIP Dialer pool. If the activated SIP Dialer changes state from ready to not ready due to a failed CTI link to CTI Server or a failed heartbeat to SIP Server, the Campaign Manager activates the standby SIP Dialer.

If the Campaign Manager detects that the connection has failed from the activated SIP Dialer, it will activate the standby SIP Dialer. The Campaign Manager marks all outstanding records with an Unknown status and return them to pending status after a certain time-out period.

SIP Dialer

The SIP Dialer is considered in ready state after it has successfully registered with Campaign Manager, has been configured successfully, has established a CTI connection to CTI Server/Agent PG, and has successfully sent a heartbeat to the SIP Server. The SIP Server can be a gateway or Unified SIP Proxy server to which the SIP Dialer is connected.

In the case of a CTI link or heartbeat failure, the SIP Dialer sends all active and pending customer records to the Campaign Manager (dialer flush), or closes them internally if the link to the Campaign Manager is not available. The SIP Dialer cancels alerting calls, abandons the connected calls that have not yet transferred to outbound agents or VRU , and leaves the outbound calls that were transferred.

The Dialer sends a heartbeat to the gateway in a single gateway deployment or to the Unified SIP Proxy Server in a multiple gateway deployment. The Dialer transitions to the ready state only when the heartbeat is enabled and the initial heartbeat is successful.

The heartbeat can be disabled by setting the Dialer Registry, EnableHeartBeat=0.

If the heartbeat fails in several attempts defined by the Dialer registry HBNumTries, the SIP Dialer changes the state to not ready and updates the status to the Campaign Manager to trigger the warm standby mechanism.

The gateway or Unified SIP Proxy server does not play any role in warm standby behavior for the SIP Dialer.

An alarm is raised when the SIP Dialer detects SIP Server heartbeat failure.

CTI Server and Agent PG

Both the activated and standby SIP Dialers maintain active connections to the CTI Server at same time.

If the CTI Server or Agent PG fails to cause the CTI link failure, the SIP Dialer changes the state to not ready and updates the status to the Campaign Manager to trigger the warm standby mechanism.

An alarm is raised when the SIP Dialer detects the CTI link failure.

Cisco Unified SIP Proxy Server

The Unified SIP Proxy server provides weighted load balancing and redundancy in a multiple gateway deployment by configuring each gateway as an element in the Server group configuration. In the following configuration, one gateway in the cucm.example.com server group receives 50 percent of the traffic and the

other two elements receive 25 percent each. You can change the weights and q-values to configure a different priority or load-balancing scheme.

```
server-group sip group cucm.example.com enterprise
element ip-address 192.168.10.4 5060 tls q-value 1.0 weight 50
element ip-address 192.168.10.5 5060 tls q-value 1.0 weight 50
element ip-address 192.168.10.3 5060 tls q-value 1.0 weight 100
fail-over-resp-codes 503lbtype weightpingend server-group
```

If one gateway is overloaded or loses its WAN link to the PSTN network, the Unified SIP Proxy server receives a SIP 503 response message. The “fail-over-resp-codes 503” configuration in the Server Group allows the Unified SIP Proxy server to select the next available gateway to resend an outbound call.

The Unified SIP Proxy server supports the Hot Swappable Router Protocol (HSRP). This protocol can build redundancy into your network by allowing two Unified SIP Proxy servers to continuously test each other for connectivity. The other server takes over if one Unified SIP Proxy server fails.

Do not use the HSRP configuration for the Unified SIP Proxy servers dedicated for Cisco Outbound Option. The Campaign Manager and SIP Dialer have a built-in warm standby feature. Also, configuring HSRP for the Unified SIP Proxy server adds undesirable complexity for Cisco Outbound Option.

Server Group and Route Table configurations are duplicated for two redundant Unified SIP Proxy servers.

Cisco Outbound Option for Unified Mobile Agents

Mobiles agents are supported only with a nailed connection for outbound campaigns.

Related Topics

[Cisco Unified Mobile Agent, on page 171](#)

References

For more information, see the [Cisco Outbound Option documentation](#).



Cisco Unified Mobile Agent

- [Cisco Unified Mobile Agent Architecture, page 171](#)
- [Unified Mobile Agent with Cisco Outbound Option, page 183](#)
- [Cisco Unified Mobile Agent Fault Tolerance, page 184](#)
- [Unified Mobile Agent Sizing, page 184](#)

Cisco Unified Mobile Agent Architecture

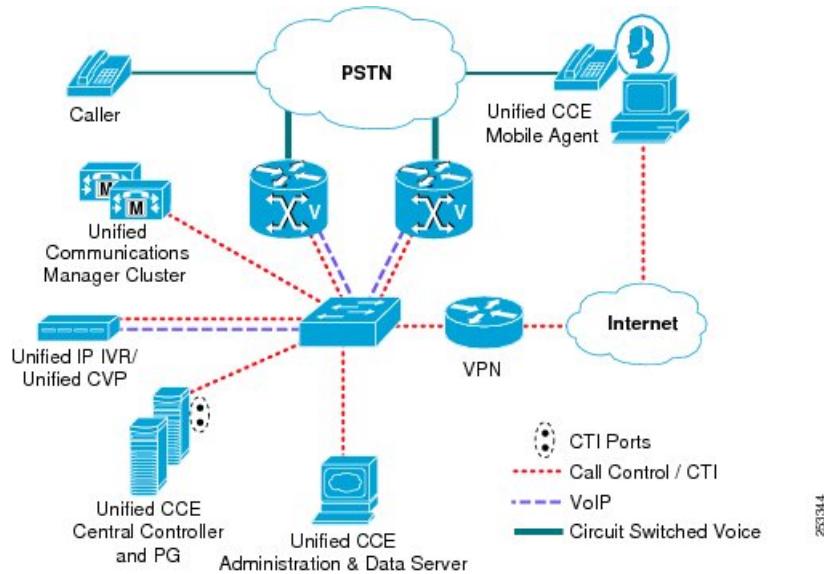
The Cisco Unified Mobile Agent feature enables an agent using any PSTN phone and a broadband VPN connection (for agent desktop communications) to function just like a Unified CCE agent sitting in a formal call center and using a Cisco IP Phone monitored and controlled by Cisco Unified Communications Manager JTAPI.

Cisco Unified Mobile Agent uses a pair of CTI ports that function as proxies for the mobile agent phone (or endpoint) and the caller phone (or endpoint). Two CTI ports (local and remote) are required for every logged-in mobile agent, and the two CTI ports take the place of the Cisco IP Phone monitored and controlled by Unified CM JTAPI. The local CTI port DN is used by the agent at login and is where callers are routed when this agent is selected. The remote CTI port calls the agent either at login for a nailed connection or upon being selected for a call-by-call connection. Then, by using media redirection, the CTI ports signal for the two VoIP endpoints to stream RTP packets directly, with no further involvement from the CTI ports until further call control (transfer, conference, hold, retrieve, or release) is required. Any subsequent call control must be performed from the agent desktop application. The PG transmits the necessary subsequent call control via

Connection Modes

JTAPI to Unified CM for the two CTI ports to do whatever is needed to the media of the call (see the figure below).

Figure 73: Cisco Unified Mobile Agent Architecture



The two CTI ports (local and remote) are logically and statically linked within the PG software by using the documented naming convention required. The CTI Ports are registered at PG initialization. Call observers are added for these two CTI Ports when a mobile agent logs in using these CTI Ports. Call control for the CTI Ports (and thus the call) is provided by the PG. As mentioned earlier, the voice path is between the two Voice Gateways.

When a mobile agent is in the office, the agent can log in as a non-mobile agent from a JTAPI monitored and controlled phone, using the same agent ID. (This document refers to these non-mobile agents as local agents.) Historical call reporting does not distinguish between calls handled as a mobile agent and those handled as a local agent.

Queuing calls to mobile agents is supported with both Unified CVP and Cisco Unified IP IVR.

Connection Modes

With Cisco Unified Mobile Agent, administrators can configure agents to use either call-by-call dialing or a nailed connection, or the administrator can configure agents to choose the connection mode at login time.

Call-by-Call Connection Mode

In call-by-call dialing, the agent's remote phone is dialed for each incoming call. When the call ends, the agent's phone is disconnected before the agent is made ready for the next call.

A basic call flow for this type of dialing is as follows:

- At login, a mobile agent specifies their login name or agent ID, password, a local CTI port DN as the instrument (CTI OS) or extension (Cisco Agent Desktop), and a phone number at which to call them. This CTI port DN must be selected carefully by an administrator based on the agent's location.

- 2 A customer call arrives in the system and is queued for a skill group or an agent through normal Unified CCE configuration and scripting. This processing is the same as for local agents.
- 3 When an agent is selected for the call, and if the agent happens to be a mobile agent, then the new processing for mobile agent begins. The Unified CCE CallRouter uses the directory number for the agent's local CTI port as the routing label.
- 4 The incoming call rings at the agent's local CTI port. The Agent PG is notified that the local CTI port is ringing but does not answer the call immediately. The caller will hear ringing at this point.
- 5 Simultaneously, a call to the agent is initiated from the remote CTI port for the selected agent. This process might take a while to complete, depending on connection time. If the agent does not answer within the configured time, RONA processing is initiated.
- 6 When the agent answers their phone by going off-hook, this second call is temporarily placed on hold. At that time, the original customer call is answered and directed to the agent call media address. The agent call is then taken off hold and directed to the customer call media address. The result is an RTP stream directly between the two VoIP endpoints.
- 7 When the call ends, both connections are disconnected and the agent is set to ready, not ready, or wrap-up, depending on agent configuration and agent desktop input.

If the agent phone is configured with voicemail, disable voicemail to allow RONA call processing to occur. With call-by-call connection, an agent must answer the phone by going off hook. The Answer button on the agent desktop will not be enabled.

Auto-answer is not possible with call-by-call connections because there is no call control mechanism to make the mobile agent phone go off hook.

Related Topics

[Agent Location and Call Admission Control Design, on page 175](#)

Nailed Connection Mode

In nailed connection mode, the agent is called once at login, and the line stays connected through multiple customer calls.

A basic call flow for this type of connection is as follows:

- 1 At login, a mobile agent specifies their agent ID, password, a local CTI port DN as the instrument (CTI OS) or extension, and a phone number at which to call them. The administrator must preselect the CTI port DN based on the agent's location.
- 2 The remote CTI port statically associated with the local CTI port used at login initiates a call to the phone number supplied at mobile agent login. When the agent answers, the call is immediately placed on hold. The agent is not considered logged in and ready until this process completes.
- 3 A customer call arrives in the system and is queued for a skill group or an agent through normal Unified CCE configuration and scripting. This process is the same as for local agents.
- 4 When an agent is selected for the call, and if the agent is a mobile agent, the new processing for a mobile agent begins.
- 5 The incoming call rings at the local CTI port that the agent uses at login. The JTAPI gateway detects that the CTI port is ringing, but does not immediately answer the call. The caller hears ringing at this point.

- 6 The agent's desktop indicates that a call is ringing, but the agent phone does not ring because it is already off hook. If the agent does not answer within the configured time, RONA processing is initiated.
- 7 When the agent presses the Answer button to accept the call, the customer call is answered and directed to the agent call media address. The agent call is then taken off hold and directed to the customer call media address.
- 8 When the call ends, the customer connection disconnects and the agent connection is placed back on hold. The agent is set to ready, not ready, or wrap-up, depending on agent configuration and agent desktop input.

A nailed connection mobile agent can log off by using the desktop or by just hanging up the phone.

Auto-answer is allowed with a nailed connection.

The following two Unified Communications Manager timers can terminate a mobile agent nailed connection call:

- Maximum Call Duration timer (the default value is 720 minutes)
- Maximum Call Hold timer (the default value is 360 minutes)

This termination can log out a nailed connection mobile agent. To keep the mobile agent logged in, set the values for both of these timers to 0 so these timers never expire. To configure these timers, use the Unified Communications Manager Administration web page for service parameters using Unified Communications Service.

In a deployment with a firewall, if an agent in a nailed connection mode is idle longer than the firewall idle timeout value, the firewall can block the media stream when the firewall idle timeout expires. To prevent the firewall from blocking the media stream, increase the firewall idle timeout value.

Mobile Agent Connect Tone for Nailed Connection Mobile Agent

The Cisco Unified Mobile Agent connect tone provides an audible indication when a call is delivered to the nailed connection mobile agent. The connection tone is two beeps, which the nailed connection mobile agent will hear upon answering a call. This feature is turned off by default; for information about how to enable the Mobile Agent connect tone, see the *Cisco Unified Contact Center Enterprise Features Guide*.

Supported Mobile Agent and Caller VoIP Endpoints

Cisco Unified Mobile Agents can log in to Unified CCE using any PSTN phone that gets routed to a Cisco Voice Gateway. You can register that Voice Gateway with the same cluster as the associated Agent PG or with another cluster. In addition to using a phone, a Cisco Unified Mobile Agent must use an agent desktop application.

Any Voice Gateway supported by Unified CM and Unified CCE is supported for mobile agents. You can configure caller (ingress) and mobile agent (egress) Voice Gateways with either MGCP or SIP. A combination of Voice Gateway types is also supported. If supervisory Silent Monitoring is not required, the ingress and egress Voice Gateways can be the same Voice Gateway.

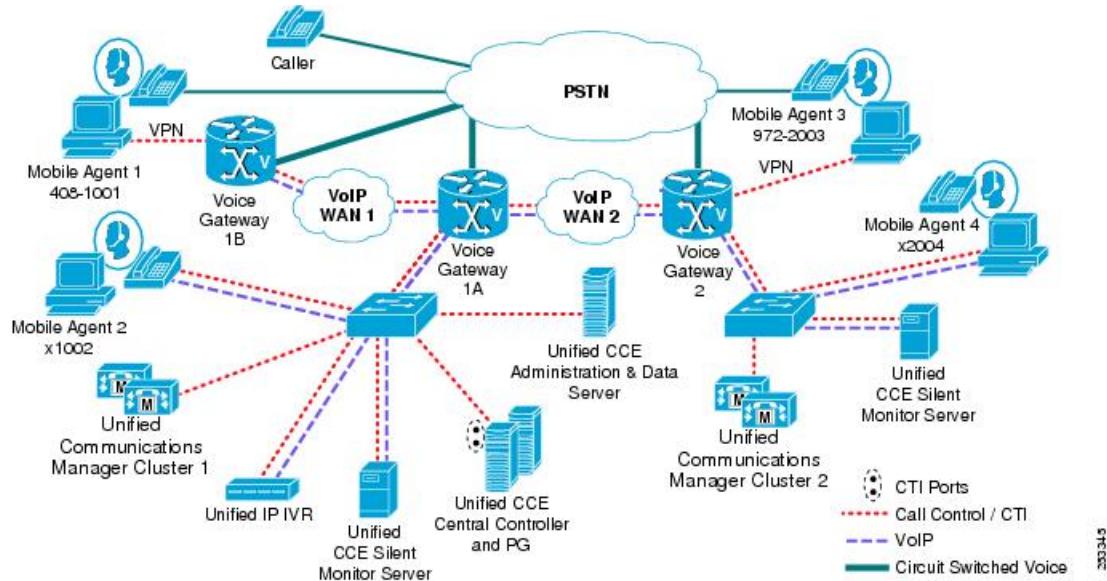
Cisco Unified Mobile Agents can also log in using a Cisco IP Phone. The IP Phone can be configured for SIP or SCCP, and a mixture is also allowed. You can register this IP Phone with the same cluster as the associated Agent PG or with another cluster. Calls to mobile agents can also originate from SIP or SCCP IP Phones.

For improved Unified Communications Manager performance, configure agents with IP Phones on the same cluster as the associated Agent PG to use Extension Mobility instead of the Unified Mobile Agent feature.

Because the IP Phone device is associated with the JTAPI user, there is a small performance hit on Unified Communications Manager for making that association.

In the following figure, Voice Gateways 1A and 1B both register with cluster 1, and Voice Gateway 2 registers with cluster 2. The call arrives into ingress Voice Gateway 1A and can be routed to any of the four agents. Mobile agent 4's IP phone (not monitored and controlled by JTAPI) registers with cluster 2, and there is no PG for cluster 2. If Silent Monitoring of mobile agent 3 is required, then a Silent Monitoring server must be deployed for agents connecting through Voice Gateway 2.

Figure 74: Mobile Agent Call Scenarios



Consider the following factors when designing an Unified Mobile Agent solution:

- If you use SIP trunks, configure Media Termination Points (MTPs). This requirement also applies if you use TDM trunks to interface with service providers. For detailed information, see *Cisco Unified Contact Center Enterprise Features Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-feature-guides-list.html>.
- Enabling the use of an MTP on a trunk affects all calls that traverse that trunk, even noncontact-center calls. Ensure that the number of available MTPs can support the number of calls traversing the trunk.

Agent Location and Call Admission Control Design

The pair of CTI ports being used by a mobile agent must be configured in Unified CM with the same location as the agent's VoIP endpoint. Because a CTI Port is a virtual type of endpoint, it can be located anywhere. System administrators need to be careful to set the proper location for the mobile agent CTI ports. Call center supervisors also must ensure that the CTI port pair assigned to a mobile agent is in the same location with the Voice Gateway (or VoIP endpoint) that will call the agent. If the location for the CTI ports is set incorrectly or if a mobile agent is assigned a CTI port pair with a different location than the Voice Gateway that will call the mobile agent, then call admission will not be accounted for correctly.

For example, assume Mobile Agent 3 in [Figure 74: Mobile Agent Call Scenarios , on page 175](#) wants to be called at 972-2003, and the dial plans for Unified CM clusters 1 and 2 are configured to route calls to 972-2003

through Voice Gateway 2. Under normal operations, Agent 3 must log in using a CTI Port pair configured with the same location as the inter-cluster trunk from Cluster 1 to Cluster 2. This configuration would allow for call admission control to properly account for calls to this mobile agent across VoIP WAN 2. If Agent 3 were to log in using a CTI Port pair with the same location as Voice Gateway 1B, then call admission control would incorrectly assume that the call was traversing VoIP WAN 1 instead of VoIP WAN 2.

Call admission control sees this mobile agent call as two completely separate calls. Call leg 1 is the call from the caller to the agent's local CTI port, and call leg 2 is the call from the remote CTI port to the agent. Because the CTI ports are in the same location as the agent endpoint, call admission control counts only the call from the caller location to the agent location (just like a normal call). This is why it is important for an agent to use CTI ports for their current location.

From the perspective of call admission control locations for the mobile agent CTI ports, there are three deployment scenarios. In the indicated figure, Agent 1 needs to use CTI ports configured in the same location as the egress Voice Gateway (Voice Gateway 1B) that will call the agent. Agent 2 needs to use CTI ports configured in the same location as the ingress Voice Gateway (Voice Gateway 1A). Agents 3 and 4 both need to use CTI ports in the same location as the inter-cluster trunk from Cluster 1 to Cluster 2. For each location possibly used by mobile agents, there must be a pool of local and remote CTI ports. The three pools of CTI ports shown in the indicated figure are co-located with the VoIP endpoint type for the agent (Voice Gateway or IP phone).

Callers and agents can also use VoIP endpoints on another Unified CM cluster. As shown in the indicated figure, this configuration would allow agents in remote locations to be called from local Voice Gateways that are associated with a different Unified CM cluster. However, a monitoring server is required at the remote site with the agent (egress) Voice Gateway if Silent Monitoring were required.

For additional information about call admission control design, see the call admission control information in the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Related Topics

[CTI OS Silent Monitoring, on page 181](#)

Dial Plan Design

As mentioned in the previous section, the Unified CM dial plan must be configured in such a way to ensure that, when the remote CTI port calls the phone number supplied by the mobile agent at login, it routes to a Voice Gateway in the same location as the mobile agent CTI ports. Otherwise, call admission control accounting will not work correctly.

Another possible design for the Unified CM dial plan is to configure it so that all calls from the CTI ports go through a specific gateway regardless of what phone number is being called. This configuration is desirable if you want a dedicated gateway for mobile agents to use. It is more easily managed, but it is not necessarily the most efficient configuration from the perspective of PSTN trunk utilization.

For additional information about dial plan design, see the dial plan information in the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Music on Hold Design

If you want a caller to hear music when a mobile agent places the caller on hold, assign Music on Hold (MoH) resources to the ingress Voice Gateway or trunk that is connected to the caller, as you would do with traditional agents. The user or network audio source is specified on the local CTI port configuration. Likewise, if you want a mobile agent to hear music when the agent is put on hold, assign MoH resources to the egress Voice

Gateway or trunk that is connected to the mobile agent. In that case, the user or network audio source is specified on the remote CTI port configuration.

**Note**

Do *not* assign MoH resources to local and remote CTI ports because it is unnecessary and might have some performance impact on the system.

A Mobile Agent remote call over a nailed connection is put on hold when there is no active call to the agent. In general, enable MoH to the mobile agent phone for nailed connection calls. If MoH resources are an issue, consider multicast MoH services.

If MoH is disabled for the nailed connection mobile agent remote phone device associated to the call, it is possible that hold tone is played to the agent phone during the hold time, depending on the call processing agent that controls the mobile agent remote phone. For Unified CM, the hold tone is enabled by default and is very similar to the Mobile Agent connect tone. With the Unified CM hold tone enabled, it is very difficult for the agent to identify if a call has arrived by listening for the Mobile Agent connect tone. Therefore, disable the hold tone for Unified CM by changing the setting of the Tone on Hold Timer service parameter on Unified CM. For details on setting this parameter, see the Unified CM product documentation available at cisco.com.

For additional information about MoH design, see the MoH information in the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Codec Design

Media streams between the ingress and egress Voice Gateways can be g.711 or g.729, but not a mix, because all CTI ports for a PG must advertise the same codec type. This requirement could result in g.711 (instead of g.729) calls being sent across the WAN. If most calls are routed to agents in the same location as the ingress Voice Gateway, then sending a few g.711 calls over the WAN might not be an issue. The alternative is to make all mobile agent calls be g.729. If a very large portion of all Unified CCE calls will always cross a WAN segment, then it probably makes sense to have all CTI ports configured for g.729. However, it is not possible to have g.711 for some mobile agent calls and g.729 for others. A dedicated region is required for the CTI ports to ensure that all calls to and from this region will use the same encoding format.

From the perspective of Silent Monitoring, the CTI OS Supervisor Desktop can silently monitor g.711 or g.729. All mobile agents would have to use the same codec, but local agents on the supervisor's team could use a mix of codecs.

For additional information about codec design considerations, see the media resources information in the *Cisco Collaboration System Solution Reference Network Designs*.

Related Topics

[CTI OS Silent Monitoring, on page 181](#)

DTMF Considerations with Mobile Agent

MTP resources might be required for mobile agents who is consulting a VRU or other network component that requires DTMF to navigate. The Mobile Agent feature relies on Cisco Unified CM CTI ports, which do not support in-band DTMF (RFC 2833). If the endpoints being used by mobile agents supports only in-band DTMF (or if they are configured to use in-band DTMF per RFC 2833), then Unified CM automatically inserts MTP resources because of the capabilities mismatch. If the mobile agent call flow requires in-band DTMF (RFC 2833), make a sufficient amount of MTP resources available.

Cisco Unified Border Element Considerations with Mobile Agent

Some SIP devices such as the Cisco Unified Border Element or other Session Border Controllers could dynamically change the media port during the call. In this case, if the Mobile Agent feature is used, MTP resources are required on the SIP trunk connecting to the agent endpoint.

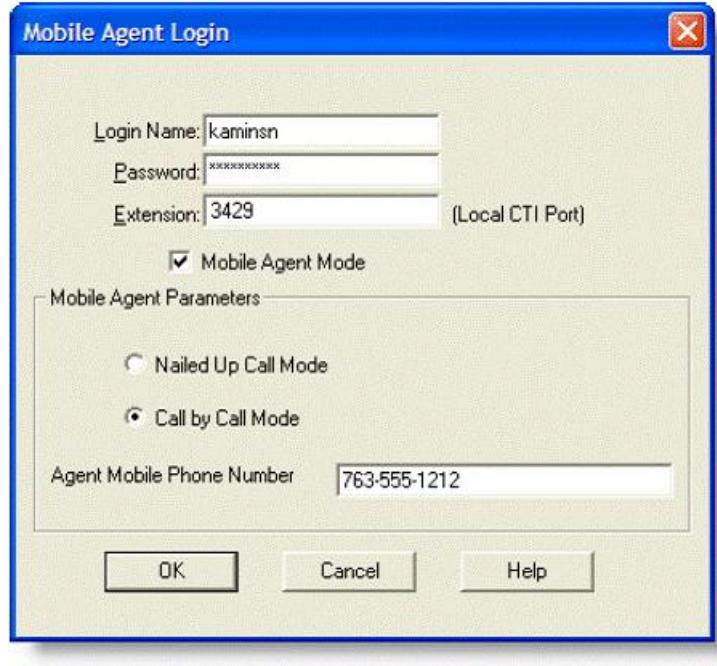
Cisco Unified Mobile Agent Interfaces

CAD IP Phone Agent (IPPA) is not an applicable agent interface for mobile agents. CAD IPPA is available only from JTAPI monitored and controlled phones that support XML applications.

Cisco Agent Desktop

The latest release of Cisco Agent Desktop supports mobile agents. At agent login, if the mobile agent mode is selected, the mobile agent login dialog box is presented to the agent. The mobile agent must provide the local CTI port extension, a call mode, and a dialable phone number.

Figure 75: Mobile Agent Login



253348

The phone number supplied must route to a VoIP endpoint (Voice Gateway, IP phone, or inter-cluster trunk) in the same location as the CTI port pair used by the agent. Otherwise, call admission control will not work correctly.

A supervisor using Cisco Supervisor Desktop (CSD) can view the state and real-time statistics for a mobile agent using Cisco Agent Desktop (CAD). A supervisor using Cisco Supervisor Desktop can also barge-in and intercept calls of mobile agents using Cisco Agent Desktop. A supervisor using CSD cannot manage agents (view statistics, Silent Monitor, record, barge-in, or intercept) using CTI-OS Toolkit applications.

Cisco Agent Desktop Silent Monitoring and Recording

The latest release of Cisco Supervisor Desktop (CSD) can silently monitor and record mobile agents using Cisco Agent Desktop SPAN Port Monitoring of the mobile agent Voice Gateway. However, Cisco Unified Mobile Agent does not support the use of Unified Communications Manager Silent Monitoring.

The Cisco Agent Desktop SPAN port monitor server provides a mechanism to access an agent's RTP stream when desktop monitoring is not possible (primarily for Cisco Agent Desktop mobile agents, Cisco Agent Desktop IP Phone Agents, or agents using lower-end IP phones without a data port for connection to the agent workstation). When a supervisor clicks the Silent Monitor button on the CSD application, the CSD application requests the SPAN port monitor server for that agent to forward a copy of both RTP streams for that agent to the CSD application. The CSD application then blends the two RTP streams and plays the resulting audio stream to the supervisor through the supervisor workstation speakers. Silent Monitoring uses two one-way RTP streams flowing from the SPAN port monitor server to the CSD workstation.

If the supervisor using CSD wants to record an agent using Cisco Agent Desktop, then the supervisor clicks the record button and the CSD application requests the recording server to request the appropriate SPAN port monitor server to forward a copy of both RTP streams to the Cisco Agent Desktop recording server to be saved onto disk. An agent can also request for a call to be recorded by clicking the Record button (if enabled) on their Cisco Agent Desktop application. Clicking this button also sends a request to the recording server to request the appropriate SPAN port monitor server to forward a copy of both RTP streams to the recording server to be saved onto disk. When recording, there are two one-way RTP streams flowing from the SPAN port monitor server to the CAD recording server.

Cisco Agent Desktop SPAN Port Monitoring of the agent Voice Gateway is somewhat different than Cisco Agent Desktop SPAN Port Monitoring of local agent Cisco IP Phones. When SPANning a LAN segment with JTAPI monitored and controlled Cisco IP Phones being used by Unified CCE local agents, the Cisco Agent Desktop SPAN Port Monitoring software is searching for RTP packets with the MAC address of the local agent's Cisco IP Phone. When SPANning a LAN segment with mobile agent Voice Gateways, the Cisco Agent Desktop SPAN Port Monitoring software is searching for RTP packets to and from the agent Voice Gateway IP address and port.

A single Cisco Agent Desktop SPAN port monitor server can SPAN a network segment with both local agent Cisco IP Phones and multiple mobile agent Voice Gateways. The Cisco Agent Desktop SPAN port monitor server is intelligent enough to find an agent's RTP stream, whether it is a local agent using a Cisco IP Phone or a mobile agent connected through an agent Voice Gateway. With Cisco Agent Desktop, a single deployment for a PG instance can support up to five Cisco Agent Desktop SPAN port monitor servers. Voice gateways are statically mapped to a specific SPAN port monitor server, and multiple agent Voice Gateways can be mapped to the same SPAN port monitor server (assuming the network SPAN is set up accordingly). Unlike local Cisco Agent Desktop agents (which are statically associated in Cisco Agent Desktop administration to a SPAN port monitor server), mobile Cisco Agent Desktop agents are not mapped to a specific SPAN Port Monitoring server. Therefore, when a Cisco Agent Desktop agent (who is not using desktop monitoring) is a local agent, they must be using an IP phone on the appropriate LAN segment that is being SPANned by their associated SPAN port monitor server. However, when that same agent is logging in as a mobile agent, there is no need to worry about which Voice Gateway or SPAN port monitor server is used to gain access to the RTP streams.

The Cisco Agent Desktop SPAN port monitor server must run separately from the agent PG, and one virtual NIC must be connected to the SPAN port of a Cisco Catalyst switch to capture the RTP streams. A second virtual NIC interface on the SPAN port monitor server is also required to communicate with other Unified CCE components such as the CSD and the Cisco Agent Desktop recording server. There is no redundancy for SPAN port monitor servers.

**Note**

Virtualization of Cisco Agent Desktop Silent Monitoring was tested only on UCS-C series servers.

The Cisco Agent Desktop SPAN port monitor server supports both g.711 and g.729 RTP streams, but it cannot support encrypted RTP streams.

Cisco Agent Desktop SPAN Port Monitoring of the ingress (or customer) Voice Gateway is not supported. Cisco Agent Desktop SPAN Port Monitoring of mobile agents using Cisco IP Phones is also not supported. For SPAN Port Monitoring to work, calls must pass through an egress (or agent) Voice Gateway, and the egress Voice Gateway must be a different Voice Gateway than the ingress Voice Gateway.

Related Topics

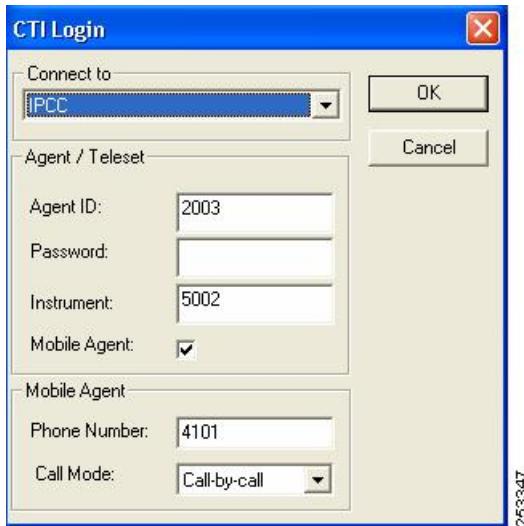
[CAD Silent Monitoring and Recording, on page 357](#)

CTI OS

The latest CTI OS releases support mobile agents. To use the Mobile Agent feature, the system administrator must enable the mobile agent while running the CTI OS setup program during or after installation. The CTI OS agent desktop will contain the Mobile Agent checkbox only after the mobile agent is enabled.

At agent login, if the mobile agent mode is selected, the CTI login dialog box is presented to the agent. The mobile agent must provide the local CTI port extension as the instrument, select a call mode, and provide a dialable phone number.

Figure 76: CTI OS Login



The phone number supplied must route to a VoIP endpoint (Voice Gateway, IP phone, or inter-cluster trunk) in the same location as the CTI port pair used by the agent. Otherwise, call admission control will not work correctly.

A supervisor using the CTI OS supervisor desktop can view the state and real-time statistics for a mobile agent using CTI OS agent desktop. A supervisor using the CTI OS supervisor desktop can also barge-in, intercept, and Silent Monitor calls of mobile agents using the CTI OS agent desktop. CTI OS does not provide agent call recording.

CTI OS Silent Monitoring

Supervisors can use silent monitoring on mobile agents with CTI OS desktops. The CTI OS Silent Monitoring service runs on a separate virtual machine for performance reasons. The virtual machine requires a dedicated virtual NIC connection to a SPAN port on a Cisco Catalyst switch and a second virtual NIC connection to carry the Unified CCE public network traffic. The Catalyst switch can SPAN a VLAN segment with either multiple ingress Voice Gateways (VG) or multiple egress VGs. The VLAN segment cannot include both types of VGs. A second virtual machine runs the CTI OS Server with the standard two virtual NICs, one for the Unified CCE public network and the other for the Unified CCE private network.

CTI OS provides a method for a supervisor to silently monitor a mobile agent using the CTI OS agent desktop. CTI OS includes a Silent Monitoring service that runs on a separate VM. The Silent Monitoring service for mobile agents requires a NIC interface on the physical CTI OS Silent Monitor server to be connected to a SPAN port on a Cisco Catalyst switch. The Catalyst switch can SPAN a VLAN segment with multiple ingress or egress Voice Gateways, but not both. For more information about SPAN-based Silent Monitoring for CTI OS, see the *CTI OS System Manager Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

Most Cisco Catalyst switches allow the destination port of a SPAN configuration to act as a normal network connection. However, some Cisco Catalyst switches do not support network traffic on the SPAN destination port. On those switches, because the server NIC interface connected to the SPAN port cannot be used for communications with supervisor desktops and other Unified CCE components, a NIC interface must be dedicated for connection to the SPAN port. In redundant Unified CCE installations, the second server NIC interface is used for the private WAN connection and thus is not available for Silent Monitoring. Therefore, in redundant Unified CCE installations and as shown in [Figure 74: Mobile Agent Call Scenarios](#), on page 175, a separate VM must be deployed with the Silent Monitor service running. One NIC interface communicates with supervisor desktops, and the other NIC interface is used to connect to the SPAN port on the Cisco Catalyst switch.



Note

For information about the Cisco Catalyst switch types that do not support outgoing traffic on a SPAN destination port, see the “Network Traffic Restrictions” section in <http://www.cisco.com/c/en/us/support/docs/voice-unified-communications/icm-cti-desktop/28804-VOIP.html>.

A Silent Monitoring service can monitor multiple ingress or egress Voice Gateways (but not both), and a CTI OS instance may have only two monitoring services. However, a Unified CM cluster can support multiple PGs if more monitoring servers are needed.

Mobile agents using IP phones can use desktop monitoring to obtain the RTP stream.

The CTI OS supervisor desktop supports Silent Monitoring of both g.711 and g.729 media streams. The supervisor desktop is sent copies of whichever encoding format is used by the agent call. Note that there are two unidirectional media streams from the monitoring server to the supervisor desktop, which represent the bidirectional media streams of the agent call. The supervisor desktop blends those media streams and plays the resulting blended media stream through the sound resources on the supervisor workstation.

The CTI OS supervisor desktop enables a supervisor to silently monitor mobile CTI OS agents connected to any Voice Gateway that is being SPANned by a CTI OS Silent Monitoring service on the same CTI OS instance. The CTI OS supervisor desktop also allows a supervisor to silently monitor local CTI OS agents by using desktop monitoring.

Unlike CAD SPAN Port Monitoring, CTI OS SPAN Port Monitoring does not statically associate an agent with a specific SPAN Port Monitoring service.

Related Topics

[Silent Monitoring, on page 122](#)

Cisco Finesse

Cisco Finesse supports mobile agents. The Cisco Finesse server does not need to be reconfigured to enable the Mobile Agent feature. To use the Mobile Agent feature, the system administrator must follow all configurations as outlined in *Cisco Unified Contact Center Enterprise Features Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-feature-guides-list.html>.

Figure 77: Cisco Finesse Agent Desktop Sign-In

The screenshot shows a blue-themed sign-in form. At the top, there are three text input fields labeled 'ID*', 'Password*', and 'Extension*'. Below these is a checkbox labeled 'Sign in as a Mobile Agent' which is checked. At the bottom is a 'Sign In' button.

On the Finesse sign-in page, if you select the mobile agent check box, the mobile agent options are presented to the agent. The mobile agent must provide the local CTI port extension in the Extension field, select a mode (Call by Call or Nailed Connection), and provide a dial number for the agent's phone.

Figure 78: Cisco Finesse Mobile Agent Sign-In

This screenshot shows the same sign-in form as Figure 77, but with additional fields. Below the 'Sign in as a Mobile Agent' checkbox, there is a 'Mode*' dropdown menu set to 'Call by Call' with a question mark icon next to it. Below that is a 'Dial Number*' text input field. The overall layout is identical to Figure 77, with the same blue header and footer.

The phone number the agent supplies must route to a VoIP endpoint (Voice Gateway, IP phone, or inter-cluster trunk) in the same location as the CTI port pair used by the agent for call admission control to work properly.

A Finesse mobile supervisor can perform all of the functions that a non-mobile supervisor can perform, except for Silent Monitoring.

Cisco Finesse Mobile Agent Silent Monitoring

Cisco Finesse does not support Silent Monitoring of mobile agents.

Customer Relationship Management Integrations

You can integrate Customer Relationship Management (CRM) applications with Unified CCE through CTI OS. The integration allows an agent to sign in through their CRM application. You can enhance the CRM interface to support using mobile agents. As part of the enhancement, provide a mobile agent check box option and to supply a call mode and phone number.

Alternately, a mobile agent might be able to sign in through the CTI OS agent desktop. The agent could then continue to use the integrated CRM agent interface as usual for call control and any further agent state control. However, verify this capability for each CRM-integrated offering.

Related Topics

[Unified CCE Desktop Deployment Scenarios, on page 109](#)

Unified Mobile Agent with Cisco Outbound Option

Mobile agents can participate in outbound campaigns, but they must use nailed connection mode for all outbound dialing modes.

The call flow for predictive or progressive dialing is as follows:

- 1 Mobile agents log in using the local CTI port DN as their agent phone number.
- 2 Without knowing whether the agents to be selected are local or mobile agents, the dialer process continually monitors peripheral skill group statistics from the CTI server for an available agent. Concurrently, the campaign manager monitors the database for customer records and forwards active records to the dialer. When the dialer identifies an available agent for use in an outbound campaign, it sends a route request to the media routing (MR) PIM.
- 3 The MR PIM forwards the route request to the Unified ICM/CCE/CCH CallRouter.
- 4 The Unified ICM/CCE/CCH CallRouter executes a routing script, selects an available agent, reserves that agent, and then returns a routing label (phone extension) identifying the reserved agent.
- 5 The MR PG returns the label (local CTI port DN) for an available agent to the dialer.
- 6 The dialer then places a reservation phone call to the local CTI port DN. The dialer auto-answers this reservation call for the agent via the CTI Server and then automatically places that reservation call on hold. At this point, a mobile agent has been reserved by having the dialer port call the local CTI port, and the CTI port has placed that call on hold.
- 7 The dialer initiates customer calls through Unified Communications Manager at whatever rate is configured for the campaign.
- 8 When a live answer is detected, the dialer immediately initiates a transfer of the call (along with call context for screen pop) to the next reserved agent extension from the list maintained by the dialer. If a mobile agent is selected, then that agent extension is the local CTI port used by that mobile agent at login.
- 9 The dialer auto-answers the transferred call for the agent through the CTI server so that the voice path between the customer and the agent can be established quickly, thus releasing the dialer port used to call

the customer. The dialer then hangs up the reservation call to this agent. The dialer also updates the Campaign Manager to indicate a live answer was detected for this call. After the agent completes handling the outbound call, the agent can be reserved for another outbound call via the same message flow.

Related Topics

[Cisco Unified Mobile Agent](#), on page 171

Cisco Unified Mobile Agent Fault Tolerance

Because the RTP stream for a mobile agent call is between the ingress and egress Voice Gateways, a failure of Unified CM or Unified CCE does not impact call survivability. However, subsequent call control (transfer, conference, or hold) may not be possible after a fail-over. A mobile agent is notified of a fail-over on their agent desktop, but they must log in again after a Unified CM or Unified CCE fail-over occurs.

Unified Mobile Agent Sizing

Mobile agent call processing uses significantly more server resources and therefore reduces the maximum number of supported agents on both Unified Communications Manager and the Agent PG. Unified Mobile Agent uses conference bridge resources for Agent Greeting. Be sure to use the sizing calculator, but indicate a conference on every call in place of agent greeting for each of the Mobile Agents.

Related Topics

[Sizing Unified CCE Components and Servers](#), on page 215

[Sizing Cisco Unified Communications Manager Servers](#), on page 229



CHAPTER 8

Video Contact Center

- Video Contact Center Overview, page 185

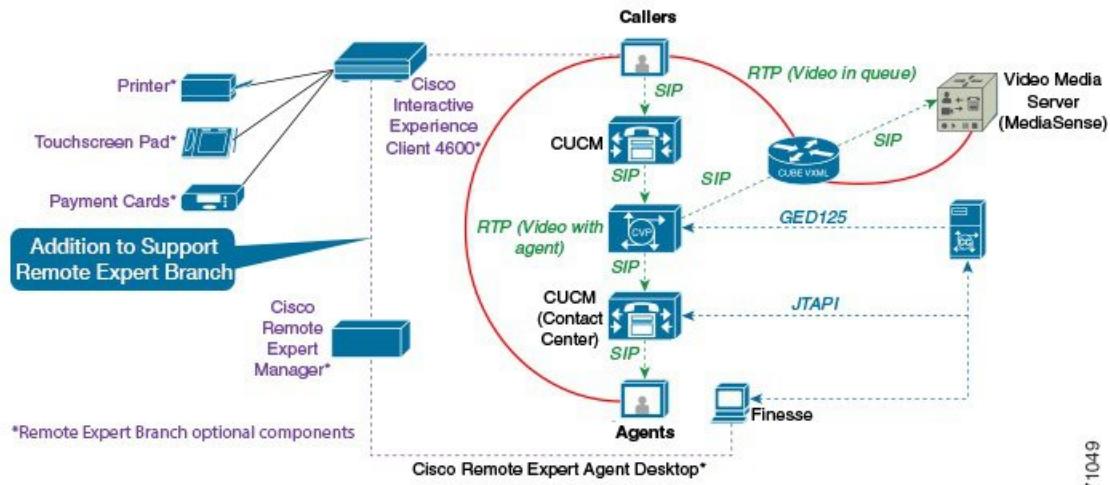
Video Contact Center Overview

With Cisco Video Remote Expert, customers and agents can have a face-to-face conversation over the network and collaborate like never before.

Through a combination of technologies and design that allows the contact center caller and remote agent to feel as if they are in the same room, the Cisco Contact Center Enterprise portfolio has the potential to provide great productivity benefits and transform your business. Organizations use it to control costs, make decisions faster, improve customer intimacy, and scale scarce resources.

Video Remote Expert allows video callers to be queued. Optionally, with CVP Video In Queue (ViQ), the caller can interact through high definition video prompt or navigate a video menu using DTMF keys. You can deploy Video Remote Expert in Immersive/Kiosk mode.

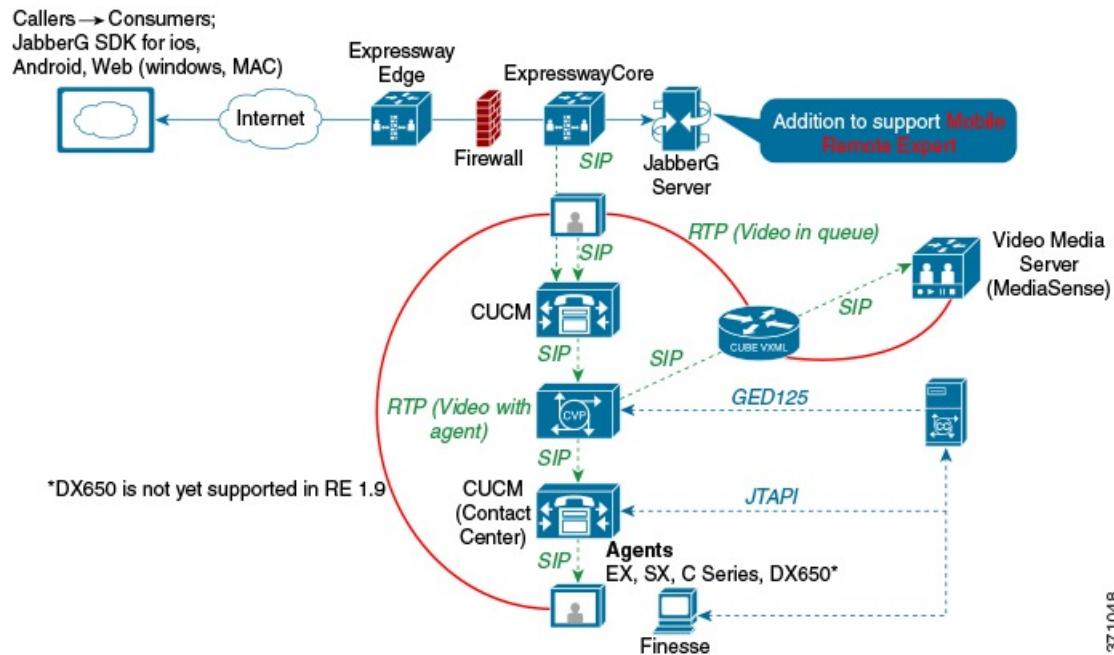
Figure 79: Immersive/Kiosk Remote Expert



371049

You can also deploy Video Remote Expert in Consumer mode.

Figure 80: Consumer Remote Expert



Video Contact Center Features

The following table summarizes the key features for the Cisco Video Remote Expert.

Feature Summary	Notes
Video Endpoints	<p>Video Callers and Agents</p> <ul style="list-style-type: none"> • HD: EX, SX, C Series, DX650* • SD: 8941, 8945, 9951, 9971

Feature Summary	Notes
User Experience	<ul style="list-style-type: none"> • EX, SX, C Series, DX650* can now be used as a standard Enterprise Contact Center agent on UCM. • Show and share capability from agent with screen sharing using dual desktop support from selected agent endpoints (EX, SX, C Series, DX650*). • Finesse and CTIOS agent desktop support for agent login, Agent State (Ready, Not Ready), Dial, Answer, Release and CTI data. • Supplementary service (Hold, Retrieve, Alternate, Reconnect, Blind/Consult, Transfer/Conference), support from Agent Desktop.
Optional: CVP Video in queue (ViQ)	Enables caller interaction through high definition video prompt or navigation through a video menu using DTMF keys

Video Remote Expert

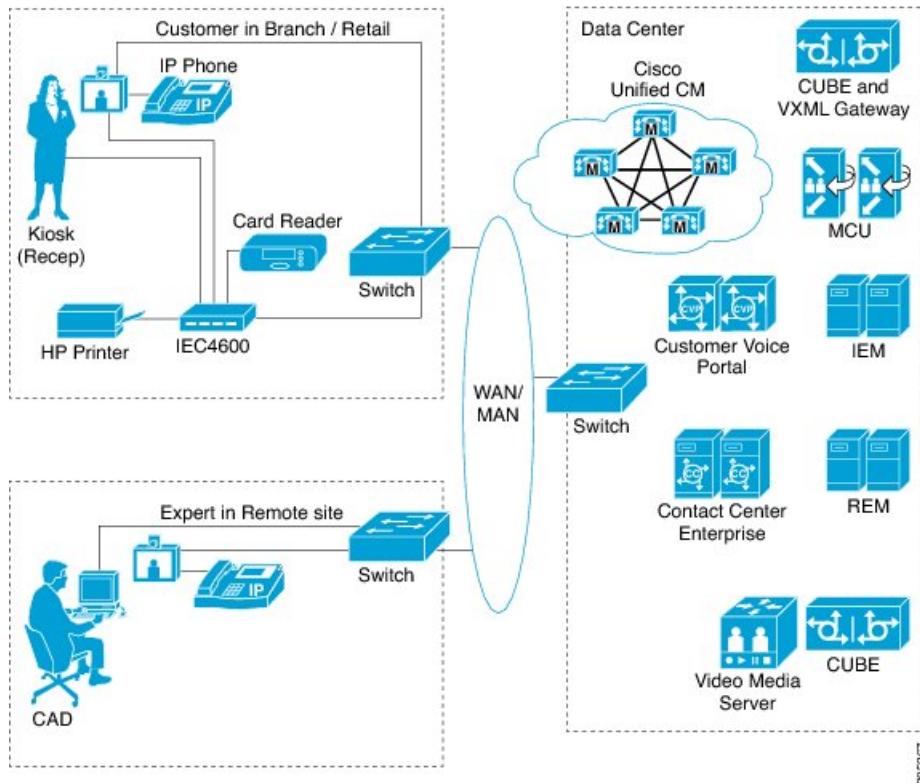
Video Remote Expert is a subset of Video Contact Center functionality and addresses a specific use case with a video kiosk that can connect a customer from kiosk to a remote expert agent with the press of a single button. The customer and the agent are then connected with a high definition video feed and can share documents back and forth as well as perform financial transactions. The remote agent can also move the camera so that the agent can show the customer how to perform a task. The customer from the kiosk can be offered a video while the call is in queue waiting for a remote expert. Optionally, with CVP Video In Queue (ViQ), the caller can interact through high definition video prompt or navigate a video menu using DTMF keys.

Video Remote Expert builds on the Unified CCE. There are many deployment models, components, and features that comprise the total Video Remote Expert solution. This chapter addresses only those that are relevant to the contact center implementation for Video Remote Expert.

Video Topologies

The following diagram illustrates topology call flow.

Figure 81: Topology Call Flow

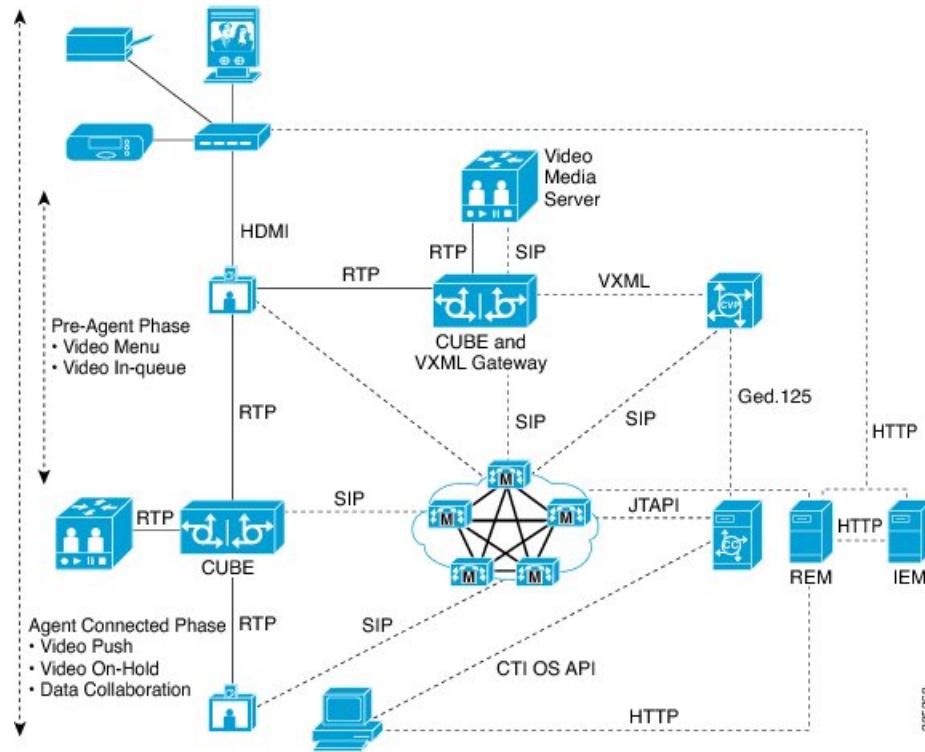


Call flow:

- 1 Customer submits video call into CCE-CVP data center from branch kiosk.
- 2 CCE script invokes CVP Server "Call Studio" application.
- 3 Call is connected to VXML gateway and video playback is invoked from Media Server.
- 4 Video RTP steamed from Gateway to Phone at branch.
- 5 Customer navigates VRU video via DTMF digits.
- 6 Customer submits DTMF for digit collection and stored in Call Context via CCE.
- 7 When customer selects to talk to agent and agent becomes available, CVP transfers the call from the VXML gateway to the Unified CM managed Video Remote Expert.
- 8 Customer is connected to Agent and video RTP is streaming from customer video phone to the agent video phone.
- 9 Via phones, agent can move the video camera around to pan the video if desired.
- 10 Desktop sharing is also available feature if required.

The following figure shows a logical view of the topology.

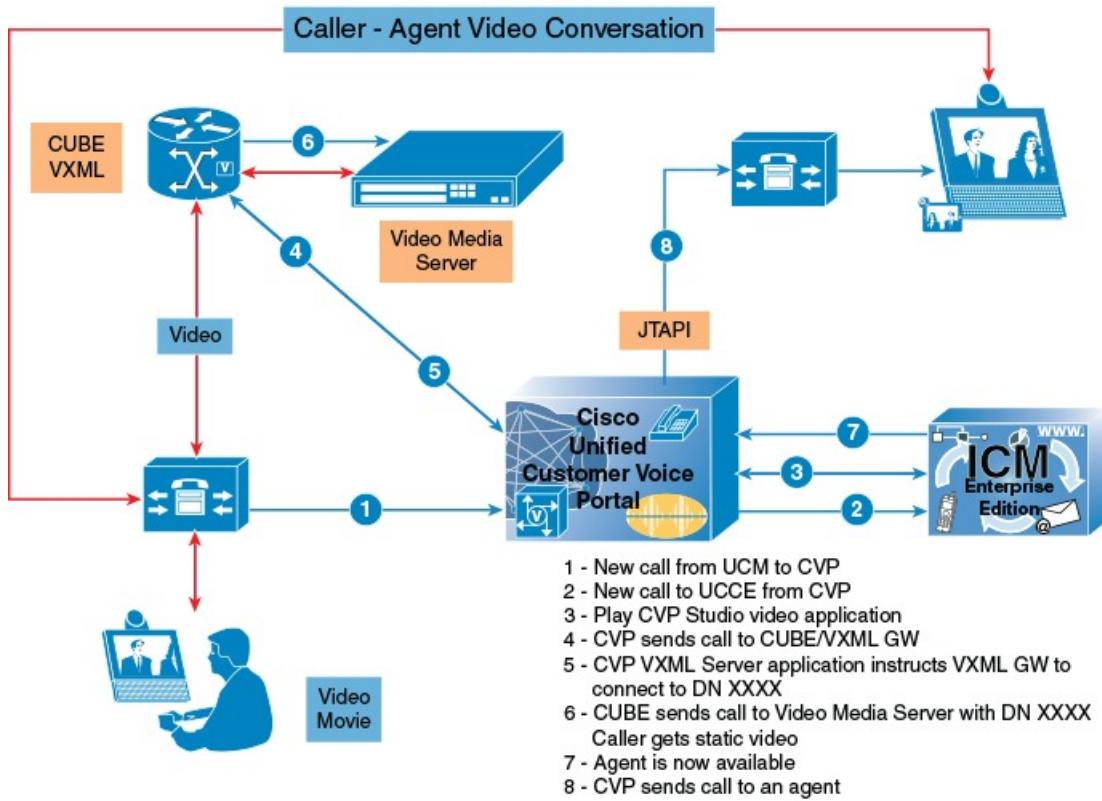
Figure 82: Logical View



Video Call Flows

The following figure illustrates the video call flow.

Figure 83: Call Flows



371047

Video Remote Expert Support for Contact Center Features

This table lists support for features with Video Remote Expert.

Feature	Video Remote Expert
Video Silent Monitor	Not Supported
Video Recording	Not Supported
Video Supervisor Barge-in	Not Supported

Video Infrastructure

The following table lists the video infrastructure:

	Video Remote Expert
Video Conference Bridge	<ul style="list-style-type: none">• ISR-G2 with PVDM3• Codian MCU
Video Media Server	Cisco MediaSense Note ViQ videos play back using G.711, AAC-LD, or G.722, whichever the endpoint prefers. However, there is no video resolution scaling in MediaSense. A 320p video plays at 320p on every device, and a 1080p video plays at 1080p on every device. Most devices properly handle any necessary correction to the scaling, but the 8941, 8945, 9951, 9971 devices do not. Upload as 480p or less any video that is intended to play on a 8941, 8945, 9951, 9971 device. MediaSense does not support the G.711 A-Law codec for video playback.



CHAPTER 9

Securing Unified CCE

- [Introduction to Security, page 193](#)

Introduction to Security

Achieving Unified CCE system security requires an effective security policy that accurately defines access, connection requirements, and systems management within your contact center. A good security policy enables you to use many state-of-the-art Cisco technologies to protect your data center resources from internal and external threats. Security measures ensure data privacy, integrity, and system availability.

The security considerations for Unified CCE at a high level are similar to the considerations for the other applications in a Cisco Unified Communications solution. Deployments of Unified CCE vary greatly and often call for complex network designs. These deployments require competence in Layer 2 and Layer 3 networking as well as voice, VPN, QoS, Microsoft Windows Active Directory, and other networking issues. This chapter provides some guidance that touches on these areas. But, this chapter is not an all-inclusive guide for deploying a secure Unified CCE network.

Along with the Unified Communications Security Solution portal, use [other Cisco solution reference network design guides \(SRNDs\)](#) in addition to this document to answer many design and deployment questions. These documents provide information on properly building a network infrastructure for Cisco Unified Communications. In particular, consult the following relevant documents about security and Cisco Unified Communications:

- *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*
- *Data Center Networking: Server Farm Security SRNDv2*
- *Site-to-Site IPSec VPN SRND*
- *Voice and Video Enabled IPSec VPN (V3PN) SRND*
- *Business Ready Teleworker SRND*

Updates and additions to these documents are posted periodically, so visit the SRND web site frequently.

This chapter provides limited guidance on the intricacies of designing and deploying a Windows Active Directory. More information is available from Microsoft on the following topics:

- Designing a new Active Directory logical structure

- Deploying Active Directory for the first time
- Upgrading an existing Windows environment to Microsoft Windows Server 2008 R2 Active Directory
- Restructuring your current environment to a Windows Active Directory environment

In particular, see the Designing and Deploying Directory and Security Services section of the Microsoft Windows Server 2008 R2 Deployment Kit. That section can assist you in meeting all the Active Directory design and deployment goals for your organization. This [development kit and its related documentation](#) are available from Microsoft.

Security Layers

An adequately secure Unified CCE deployment requires a multilayered approach to protecting systems and networks from targeted attacks and the propagation of viruses, among other threats. The goal of this chapter is to stress the various areas pertinent to securing a Unified CCE deployment, but it does not delve into the details of each area. Specific details can be found in the relevant product documentation.

Implement the following security layers and establish policies around them:

- Physical Security

You must ensure that the servers hosting the Cisco contact center applications are physically secure. They must be located in data centers to which only authorized personnel have access. The cabling plant, routers, and switches also have controlled access. Implementing a strong physical-layer network security plan also includes utilizing such things as port security on data switches.

- Perimeter Security

While this document does not delve into the details on how to design and deploy a secure data network, it does provide references to resources that can aid in establishing an effective secure environment for your contact center applications.

- Data Security

To ensure an increased level of protection from eavesdropping for customer-sensitive information, Unified CCE provides support for Transport Layer Security (TLS) on the CTI OS and Cisco Agent Desktops. It also supports IPSec to secure communication channels between servers.

- Host-Based Firewall

You can use the Windows Firewall to protect from malicious users and programs that use unsolicited incoming traffic to attack servers. Use the Windows Firewall Configuration Utility on VMs or the Agent Desktop Installers to integrate with the firewall component of Windows Server 2008 R2.



Note

While Unified CCE does not support Windows XP, Cisco Finesse 10.0 does support Windows XP clients.

- Virus Protection

All VMs must be running antivirus applications with the latest virus definition files (scheduled for daily updates). The *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE contains a list of all the tested and supported antivirus applications.

- Patch Management

A system is typically not connected to a live network until all security updates have been applied. It is important for all hosts to be kept up-to-date with Microsoft (Windows, SQL Server, Internet Explorer, and so forth) and other third-party security patches.

For most of these security layers, the Unified CCE solution supports a number of capabilities. However, what Cisco cannot control or enforce is your enterprise policies and procedures for deploying and maintaining a secure Unified CCE solution.

Platform Differences

Before discussing how to design the various security layers required for a Unified CCE network, this section introduces the differences that are inherent in the applications making up the Unified CCE solution.

The Unified CCE solution consists of a number of application servers that are managed differently. The primary servers, those with the most focus in this document, are the Routers, Loggers (also known as Central Controllers), Peripheral Gateways, Administration & Data Servers, and so forth. These application servers can be installed only on a standard (default) operating system installation. For Unified CCE components that you install on Windows Server 2008 R2, use only a default retail version of the Windows Server software. The maintenance of this operating system in terms of device drivers, security updates, and so forth, is the responsibility of the customer, as is acquiring the necessary software from the appropriate vendors. This category of application servers is the primary focus of this topic.

The secondary group of servers, those running applications that are part of the solution but that are deployed differently, are Cisco Unified Communications Manager (Unified CM), Cisco Unified IP IVR, and so forth. Customers are required to obtain all relevant patches and updates to this operating system from Cisco. The security hardening specifications for this operating system can be found in the *Cisco Collaboration System Solution Reference Network Designs* and other Unified CM product documentation at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/tsd-products-support-series-home.html>.

The approach to securing the Unified CCE solution as it pertains to the various layers listed above differs from one group of servers to another. It is useful to keep this in mind as you design, deploy, and maintain these servers in your environment. Cisco is constantly enhancing its Unified Communications products with the eventual goal of having them all support the same customized operating system, antivirus applications, and security path management techniques.

Security Design Elements

Cisco has a security guide for the primary group of servers. The guide covers details of security implementation along with general guidance for securing a Unified CCE deployment. The security guide includes the following topics:

- Encryption Support
- IPSec and NAT Support
- Windows Firewall Configuration
- Automated Security Hardening
- Updating Microsoft Windows
- SQL Server Hardening

- SSL Encryption
- Microsoft Baseline Security Analysis
- Auditing
- Antivirus Guidelines
- Secure Remote Administration

For a more in-depth discussion of security, see the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

The guidelines are based in part on hardening guidelines published by Microsoft and other third-party vendors. The guide also serves as a reference point for most of the security functionality in the product. The guide also covers installation for the Automated OS and SQL Security Hardening bundled with the application installer, Windows Firewall Configuration Utility, the SSL Configuration Utility, the Network Isolation IPSec Utility, and the Unified CC Security Wizard.

Other Security Guides

Other documents containing security guidance include, but are not limited to the following:

Table 19: Other Security Documentation

Security Topic	Document and URL
Server staging and Active Directory deployment	<i>Staging Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html
CTI OS encryption	<i>CTI OS System Manager Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html and <i>Cisco Agent Desktop Installation Guide</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/products-installation-guides-list.html
SNMPv3 authentication and encryption	<i>SNMP Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html
Feature Control (Software access control)	<i>Configuration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html
Validating real-time clients	<i>Setup and Configuration Guide for Cisco Unified ICM Hosted</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html

Network Firewalls

There are several important factors to consider when deploying firewalls in a Unified CCE network. The application servers making up a Unified CCE solution are not meant to reside in a demilitarized zone (DMZ) and must be segmented from any externally visible networks and internal corporate networks. The VMs must be placed in data centers, and the applicable firewalls or routers must be configured with access control lists (ACL) to control the traffic that is targeted to the VMs, thereby allowing only designated network traffic to pass through.

Deploying the application in an environment in which firewalls are in place requires the network administrator to be knowledgeable about which TCP/UDP IP ports are used, firewall deployment and topology considerations, and impact of Network Address Translation (NAT).

TCP/IP Ports

For an inventory of the ports used across the contact center suite of applications, see the following documentation:

- *Port Utilization Guide for Cisco Unified Intelligent Contact Management Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.
- *Cisco Unified Contact Center Express Port Utilization Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-express/products-installation-and-configuration-guides-list.html>
- *TCP and UDP Port Usage Guide for Cisco Unified Communications Manager* at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-maintenance-guides-list.html>

To aid in firewall configuration, these guides list the protocols and ports used for agent desktop-to-server communication, application administration, and reporting. They also provide a listing of the ports used for intra-server communication.

Network Firewall Topology

The deployment shown in the section on AD Administrator created OUs represents the placement of firewalls and other network infrastructure components in a Unified CCE deployment. The design model incorporates a parent Unified ICM system with legacy peripheral hosts and a child Cisco Unified Contact Center Enterprise with a Unified CM cluster. For this deployment type, do the following:

- Block the following ports at the enterprise perimeter firewall:
 - UDP ports 135, 137, 138, and 445
 - TCP ports 135, 139, 445, and 593
- Deploy Layer-3 and Layer-4 ACLs that are configured as described in the port guides.
- Isolate database and web services by installing dedicated historical data servers.
- Minimize the number of Administration & Data Servers (ADS) and use Administration Clients (no database required) and internet script editor clients.

- Use the same deployment guidelines when the parent Unified ICM or child Unified CCE central controllers are geographically distributed.
- Deploy Windows IPSec (ESP) to encrypt intraserver communications.
- Use Cisco IOS IPSec for site-to-site VPNs between geographically distributed sites, remote branch sites, or outsourced sites.

Related Topics

- [AD Administrator-Created OUs, on page 199](#)
[IPSec Deployment, on page 201](#)

Network Address Translation

Network Address Translation (NAT) is a feature that resides on a network router and permits the use of private IP addressing. A private IP address is an IP address that cannot be routed on the Internet. When NAT is enabled, users on the private IP network can access devices on the public network through the NAT router.

When an IP packet reaches the NAT-enabled router, the router replaces the private IP address with a public IP address. For applications such as HTTP or Telnet, NAT does not cause problems. However, applications that exchange IP addresses in the payload of an IP packet experience problems because the IP address that is transmitted in the payload of the IP packet is not replaced. Only the IP address in the IP header is replaced.

To overcome this problem, Cisco IOS-based routers and PIX/ASA firewalls implement *fix-ups* for a variety of protocols and applications including SCCP and CTIQBE (TAPI/JTAPI). The fix-up allows the router to look at the entire packet and replace the necessary addresses when performing the NAT operation. For this process to work, the version of Cisco IOS or PIX/ASA must be compatible with the Unified CM version.

Unified CCE supports connectivity through a NAT except when CTI OS desktop monitoring/recording is in use. The IP address of the agent phone is seen as the NAT IP address, which causes the agent desktop to filter the IP packets improperly. For more information, consult the “IPSec and NAT Support” section of the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

Active Directory Deployment

This section describes the Active Directory and Firewall Deployment topology. For more detailed Active Directory (AD) deployment guidance, consult the *Staging Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

While Unified ICM and Unified CCE systems may still be deployed in a dedicated Windows Active Directory domain, it is not a requirement. What makes this possible is the capability of the software security principals to be installed in Organizational Units. This closer integration with AD and the power of security delegation means that corporate AD directories can be used to house application servers (for domain membership), user and service accounts, and groups.

Related Topics

- [AD Administrator-Created OUs, on page 199](#)

AD Site Topology

In a geographically distributed deployment of Unified ICM or Unified CCE, redundant domain controllers must be located at each of the sites, and properly configured Inter-Site Replication Connections must be established with a Global Catalog at each site. The Unified CCE application is designed to communicate with the AD servers that are in their site, but this requires an adequately implemented site topology in accordance with Microsoft guidelines.

Organizational Units

Application-Created OUs

The installation of Unified ICM or Unified CCE software requires that the AD Domain in which the VMs are members must be in Native Mode. The installation will add a number of OU objects, containers, users, and groups that are necessary for the operation of the software. Adding these objects can be done only in an Organizational Unit in AD over which the user running the install program has been delegated control. The OU can be located anywhere in the domain hierarchy, and the AD Administrator determines how deeply nested the Unified ICM/Unified CCE OU hierarchy is created and populated.

**Note**

Local Server Accounts and groups are not created on the application servers. All created groups are Domain Local Security Groups, and all user accounts are domain accounts. The Service Logon domain account is added to the Local Administrators' group of the application servers.

Unified ICM and Unified CCE software installation is integrated with a Domain Manager tool that can be used standalone for pre-installing the OU hierarchies and objects required by the software, or can be used when the Setup program is invoked to create the same objects in AD. The AD/OU creation can be done on the domain in which the running VM is a member or on a trusted domain.

Do not confuse the creation of AD objects with Group Policy Objects (GPO). The Automated Security Hardening, which is provided and follows the standard Microsoft Security Template format, is not added to AD as part of the software installation through the configuration of a GPO. The security policy provided by this customized template (for Unified ICM/Unified CCE applications) is applied locally when a user chooses to apply hardening, or it can be pushed down through a GPO through manual AD configuration using the provided policy file, `CiscoICM_Security_Template.inf`.

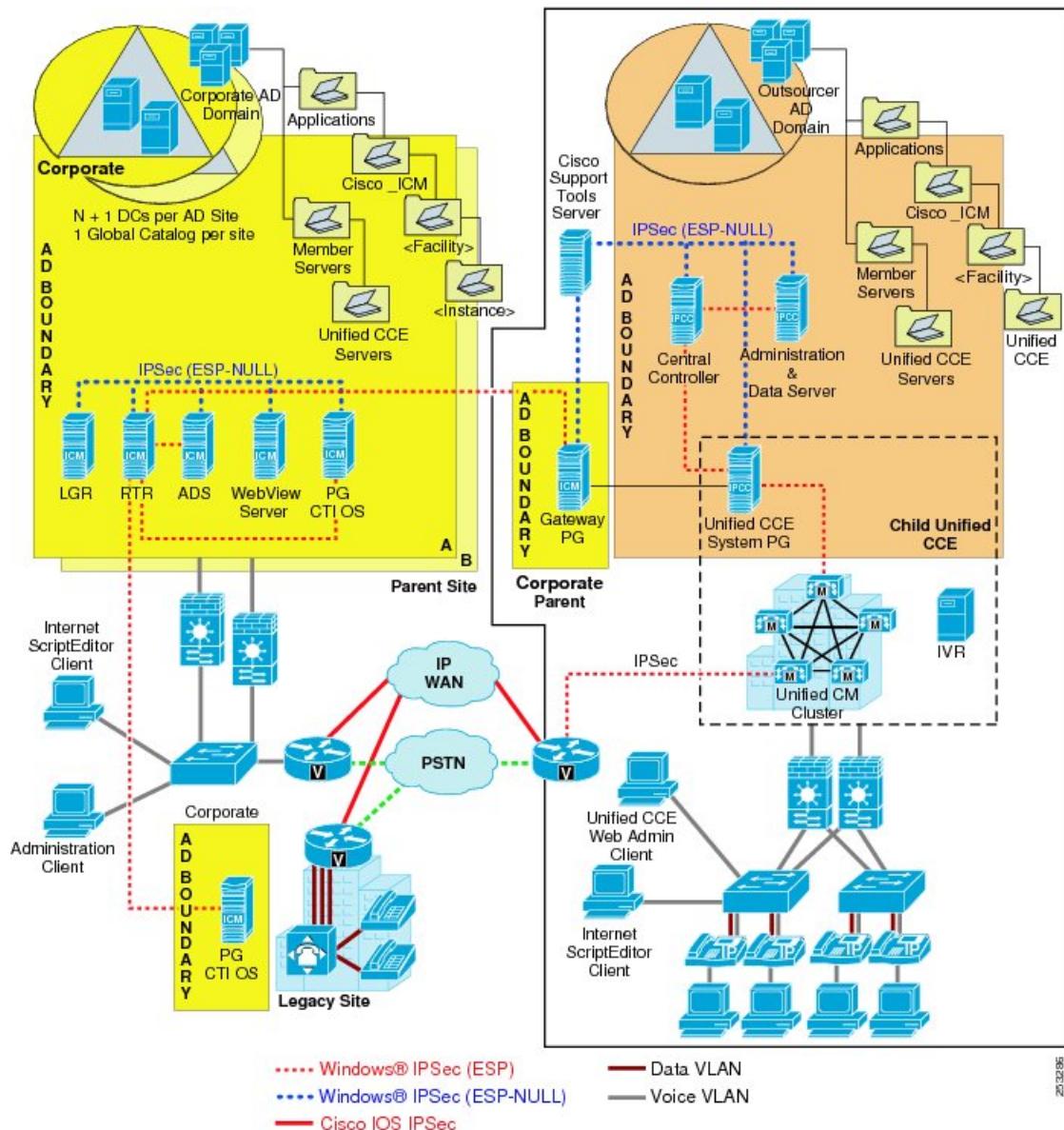
AD Administrator-Created OUs

An administrator can create certain AD objects. A prime example is the OU container for Unified CCE Servers. This OU container is manually added to contain the VMs that are members of a given domain. You move these VMs to this OU once they are joined to the domain. This segregation controls who can or cannot administer the servers (delegation of control). Most importantly, the segregation controls the AD Domain Security Policies that the application servers in the OU can or cannot inherit.

As noted before, Unified ICM/Unified CCE servers ship with a customized security policy. You can apply this policy at this server OU level through a Group Policy Object (GPO). Block any differing policies from being inherited at the Unified ICM/Unified CCE Servers' OU. Remember that someone can override blocking inheritance, a configuration option at the OU object level, by selecting the Enforced/No Override option at a higher hierarchy level. The application of group policies must follow a well-planned design. Start with the most common denominator, and restrict those policies only at the appropriate level in the hierarchy. For a

more in-depth explanation on how to deploy group policies properly, see the pages in the *Microsoft Security Compliance Manager* site at <http://technet.microsoft.com/en-us/library/cc677002.aspx>.

Figure 84: Active Directory and Firewall Deployment Topology



The following notes apply to the preceding figure:

- The application setup creates the Cisco_ICM organizational unit object hierarchies.
- The AD administrators create Unified ICM Servers and Unified CCE Servers organizational unit objects to separately apply custom Cisco Unified ICM Security Policies through a GPO if necessary.
- Flexible Single Master Operation servers must be distributed across Domain Controllers in the appropriate sites according to Microsoft guidelines.

IPSec Deployment

The Unified CCE solution relies on one or both of Microsoft Windows IPSec and Cisco IOS IPSec to secure critical links between VMs and sites. You can secure the solution in the following ways:

- By deploying peer-to-peer IPSec tunnels between the VMs and sites
- By deploying a more restrictive and preconfigured Network Isolation IPSec policy
- Using a combination of both

The peer-to-peer IPSec deployment requires manual configuration for each communication path that must be secured, using the tools provided by Microsoft. However, you can automatically deploy the Network Isolation IPSec policy on each VM by using the Network Isolation IPSec utility. The utility secures all communication paths to or from that VM unless an exception is made. The Network Isolation IPsec utility is installed by default on all Unified CCE servers.

For more details, see the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise-products-installation-and-configuration-guides-list.html>.

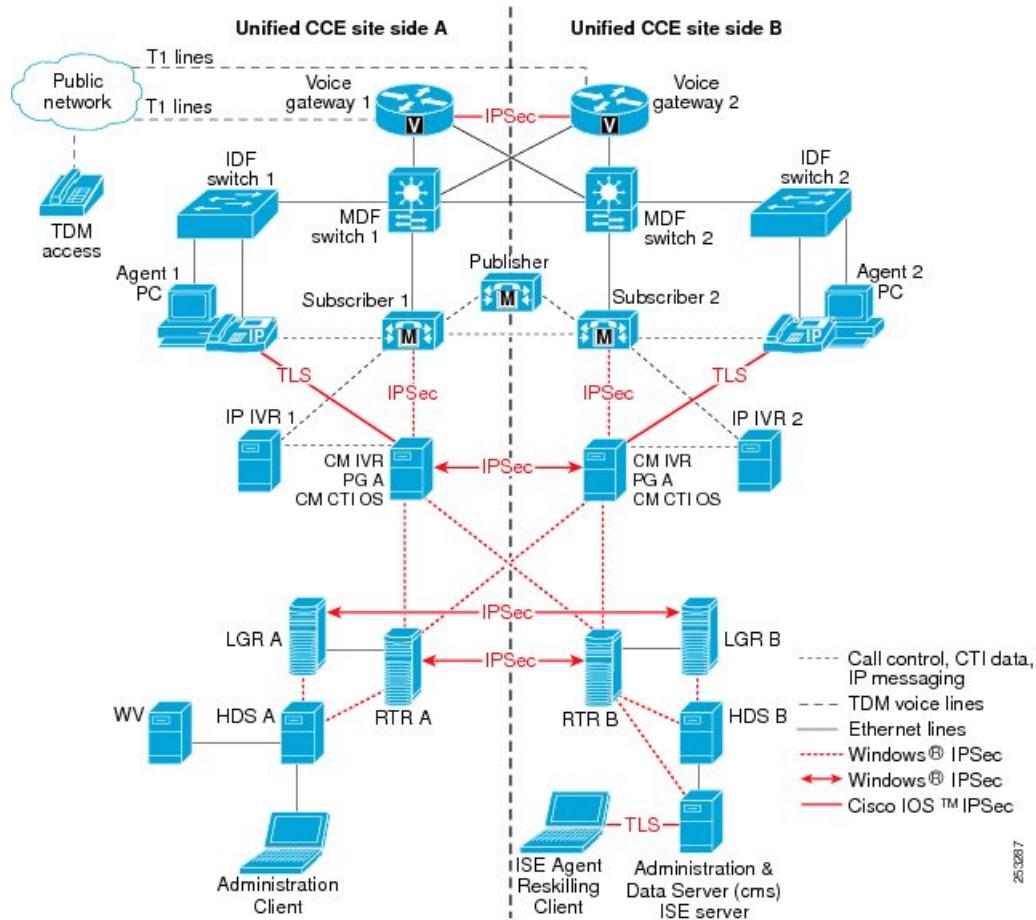
This guide not only lists the supported paths, but also information to help users deploy Windows IPSec, including appropriate settings and much more.

**Note**

Enabling IPSec affects scalability in several key areas.

Several connection paths in Unified CCE support IPSec. The following figure illustrates these guidelines. The figure shows the various server interconnections that must be secured with either Windows IPSec or Cisco IOS IPSec. The diagram also shows several paths that support TLS.

Figure 85: IPSec Deployment Example



Related Topics

[Endpoint Security, on page 205](#)

[Scalability Impacts of Components and Features, on page 323](#)

Host-Based Firewall

By providing host firewall protection on the innermost layer of your network, Windows Firewall can be an effective part of your defense-in-depth security strategy. Unified CCE supports the deployment of Windows Firewall on the VMs. The *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* contains a chapter on the implementation and configuration of this feature.

The configuration of the exceptions and the opening of the ports required by the application will still be done locally using the Windows Firewall Configuration Utility, which is included with the Unified CCE application.

The Windows Firewall is set up during Unified CCE installation, during which required ports are opened.

For more information about the Windows Firewall, see the *Windows Firewall Operations Guide* at [http://technet.microsoft.com/en-us/library/cc739696\(v=WS.10\).aspx](http://technet.microsoft.com/en-us/library/cc739696(v=WS.10).aspx).

Configuring Server Security

Unified Contact Center Security Wizard

The Unified Contact Center Security Wizard allows easy configuration of the security features defined above, namely, SQL Server Hardening, Windows Firewall configuration, and Network Isolation IPSec policy deployment. The Security Wizard encapsulates the functionality of these four utilities in an easy-to-use wizard-like interface that guides the user with the steps involved in configuring the security feature. (This is particularly helpful when deploying the Network Isolation IPSec policy.) The Security Wizard is installed by default with Unified CCE. The *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-and-configuration-guides-list.html> contains a chapter explaining the Security Wizard in detail.

Virus Protection

Antivirus Applications

A number of third-party antivirus applications are supported for the Unified CCE system. For a list of applications and versions supported on your particular release of the Unified CCE software, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE DocWiki and the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal (Unified CVP)*, as well as the Cisco Unified CCX and Unified Communications Manager product documentation for the applications supported. These documents are available on cisco.com.

Deploy only the supported applications for your environment to avoid a software conflict.

Configuration Guidelines

Antivirus applications have numerous configuration options that allow granular control of what and how data must be scanned on a VM.

With any antivirus product, configuration is a balance of scanning versus the performance of the VM. The more you choose to scan, the greater the potential performance overhead. The role of the system administrator is to determine what the optimal configuration requirements for installing an antivirus application within a particular environment. See the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-and-configuration-guides-list.html> and your particular antivirus product documentation for more detailed configuration information about a Unified ICM environment.

The following list highlights some general rules:

- Upgrade to the latest supported version of the third-party antivirus application. Newer versions improve scanning speed over previous versions, resulting in lower overhead on VMs.

- Avoid scanning of any files accessed from remote drives (such as network mappings or UNC connections). Where possible, each of these remote machines must have its own antivirus software installed, thus keeping all scanning local. With a multitiered antivirus strategy, scanning across the network and adding to the network load is not required generally.
- Heuristics scanning has a higher overhead than traditional antivirus scanning. Use this advanced scanning option only at key points of data entry from untrusted networks (such as email and internet gateways).
- You can enable real-time or on-access scanning, but only on incoming files (when writing to disk). This setting is the default for most antivirus applications. On-access scanning of file reads yields a higher than necessary impact on system resources in a high-performance application environment.
- On-demand and real-time scanning of all files gives optimum protection. But, this configuration imposes the unnecessary overhead of scanning those files that cannot support malicious code (for example, ASCII text files). Exclude files or directories of files in all scanning modes that are known to present no risk to the system. Also, follow the guidelines for which specific Unified CCE files to exclude in Unified CCE implementation, as provided in the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*.
- Schedule regular disk scans only during low usage times and at times when application activity is lowest. To determine when application purge activity is scheduled, see the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* listed in the previous item.

Intrusion Prevention

Cisco does not test or support intrusion prevention products by vendors such as Sygate, McAfee, and so on. Such products are capable of blocking legitimate application functionality if they incorrectly identify that application as a security threat. These products must be configured to allow legitimate operations to execute.

Patch Management

Security Patches

The [security updates qualification process for Contact Center products](#) is documented. This process applies to the VMs running the standard Windows Operating System.

Follow Microsoft guidelines regarding when and how to apply updates. All Contact Center customers must separately assess all security patches released by Microsoft and install those deemed appropriate for their environments.

Automated Patch Management

Unified CCE servers (except for the applications installed on VOS) support integration with Microsoft's Windows Server Update Services, whereby customers control which patches can be deployed to those VMs and when the patches can be deployed.

Selectively approve updates and determine when they get deployed on production VMs. The Windows Automatic Update Client (installed by default on all Windows hosts) can be configured to retrieve updates by polling a VM that is running Microsoft Window Update Services in place of the default Windows Update Web site.

For more configuration and deployment information, see the [Microsoft Deployment Guide and other step-by-step guides](#).

More information is also available on this topic in the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

The Cisco Unified Communications Operating System configuration and patch process does not currently allow for an automated patch management process.

Endpoint Security

Agent Desktops

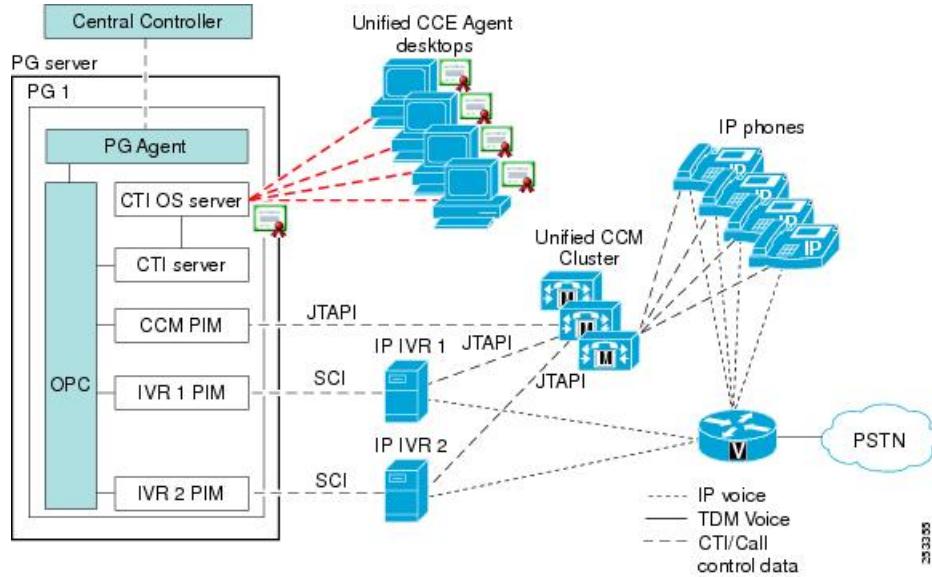
The CTI OS (C++/COM toolkit) and CAD agent desktop servers both support TLS encryption to the server. This encryption protects agent sign-in and CTI data from snooping. A mutual authentication mechanism enables the CTI OS Server and client to agree on a cipher suite for authentication, key exchange, and stream encryption. The cipher suite used is as follows:

- Protocol: SSLv3
- Key exchange: DH
- Authentication: RSA
- Encryption: AES (128)
- Message digest algorithm: SHA1

The following figure shows the encryption implementation's use of X.509 certificates on the agent desktops and on the servers. The implementation supports the integration with a Public Key Infrastructure (PKI) for the most secure deployment. By default, the application installs and relies on a self-signed certificate authority (CA) to sign client and server requests. However, Cisco supports integration with a third-party CA. This

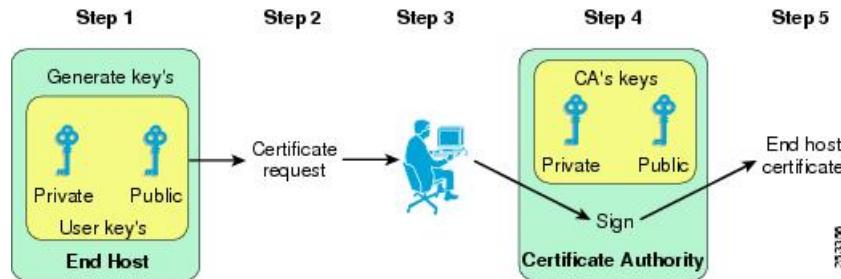
mechanism is the preferred method because of the increased security provided by a corporate-managed CA or external authority such as VeriSign.

Figure 86: Secure Agent Desktops (Certificate-Based Mutual Authentication)



The following figure shows the Certificate Authority enrollment procedure to generate certificates used by the agent and the servers. The agent desktop certificate enrollment process is manual. The process requires the creation of certificate signing requests (CSRs) at each endpoint. The CSRs are then transferred to the certificate authority responsible for signing and generating the certificates.

Figure 87: Certificate Authority Enrollment Procedure



Cisco Finesse supports HTTPS for the Administration Console and Agent and Supervisor Desktops. HTTPS is not supported for Agent and Supervisor Desktops in large deployments (over 1000 agents).

Unified IP Phone Device Authentication

When designing a Unified CCE solution based on Unified Communications Manager, customers may choose to implement device authentication for the Cisco Unified IP Phones. Unified CCE supports Unified Communications Manager's Authenticated Device Security Mode, which ensures the following:

- Device Identity — Mutual authentication using X.509 certificates

- Signaling Integrity — SCCP/SIP messages authenticated using HMAC-SHA-1
- Signaling Privacy — SCCP/SIP message content encrypted using AES-128-CBC

Media Encryption (SRTP) Considerations

Certain IP phones support Secure Real-Time Transport Protocol (SRTP). Before enabling SRTP in your deployment, consider the following points:

- The Unified CVP VXML Browser does not support SRTP.
- Deployments that use span-based silent monitoring do not support SRTP.
- Mobile Agents cannot use SRTP.
- The Cisco Outbound Option Dialers do not support SRTP. While calls are connected to the Dialer, the calls cannot use SRTP. But, calls can negotiate SRTP once the call is no longer connected to the Dialer.

IP Phone Hardening

The IP phone device configuration in Unified CM provides the ability to disable a number of phone features to harden the phones, such as disabling the phone's PC port or restricting a PC from accessing the voice VLAN. Changing some of these settings can disable the monitoring/recording feature of the Unified CCE solution. The settings are defined as follows:

- PC Voice VLAN Access
 - Indicates whether the phone will allow a device attached to the PC port to access the Voice VLAN. Disabling Voice VLAN Access will prevent the attached PC from sending and receiving data on the Voice VLAN. It will also prevent the PC from receiving data sent and received by the phone. Disabling this feature will disable desktop-based monitoring and recording.
 - Setting: Enabled (default)
- Span to PC Port
 - Indicates whether the phone will forward packets transmitted and received on the Phone Port to the PC Port. To use this feature, PC Voice VLAN access must be enabled. Disabling this feature will disable desktop-based monitoring and recording.
 - Setting: Enabled

Disable the following setting to prevent man-in-the-middle (MITM) attacks unless the third-party monitoring and/or recording application deployed uses this mechanism for capturing voice streams. The CTI OS Silent Monitoring feature and CAD Silent Monitoring and Recording do not depend on Gratuitous ARP.

- Gratuitous ARP
 - Indicates whether the phone will learn MAC addresses from Gratuitous ARP responses.
 - Setting: Disabled



CHAPTER 10

Sizing Contact Center Resources

- [Sizing Contact Center Resources, page 209](#)
- [Contact Center Basic Traffic Terminology, page 209](#)

Sizing Contact Center Resources

Central to designing a Cisco Unified Contact Center (or any contact center) is the proper sizing of its resources. This chapter discusses the tools and methodologies needed to determine the required number of contact center agents (based on customer requirements such as call volume and service level desired), the number of Unified IP IVR ports required for various call scenarios (such as call treatment, prompt and collect, queuing, and self-service applications), and the number of Voice Gateway ports required to carry the traffic volume coming from the PSTN or other TDM source such as PBXs and TDM VRUs.

The methodologies and tools presented in this chapter are based on traffic engineering principles using the Erlang-B and Erlang-C models applied to the various resources in a Unified CCE deployment. Examples are provided to illustrate how resources can be impacted under various call scenarios such as call treatment (prompt and collect) in the Unified IP IVR and agent wrap-up time. These tools and methodologies are intended as building blocks for sizing contact center resources and for any telephony applications in general.

Contact Center Basic Traffic Terminology

It is important to be familiar with, and to be consistent in the use of, common contact center terminology. Improper use of these terms in the tools used to size contact center resources can lead to inaccurate sizing results.

The terms listed in this section are the most common terms used in the industry for sizing contact center resources. There are also other resources available on the internet for defining contact center terms.

Busy Hour or Busy Interval

A busy interval can be one hour or less (such as 30 minutes or 15 minutes, if sizing is desired for such smaller intervals). The busy interval occurs when the most traffic is offered during this period of the day. The busy hour or interval varies over days, weeks, and months. There are weekly busy hours and seasonal busy hours. There is one busiest hour in the year. Common practice is to design for the average busy hour (the average of the 10 busiest hours in one year). This average is not always applied, however, when staffing is required

to accommodate a marketing campaign or a seasonal busy hour such as an annual holiday peak. In a contact center, staffing for the maximum number of agents is determined using peak periods, but staffing requirements for the rest of the day are calculated separately for each period (usually every hour) for proper scheduling of agents to answer calls versus scheduling agents for offline activities such as training or coaching. For trunks or VRU ports, in most cases it is not practical to add or remove trunks or ports daily, so these resources are sized for the peak periods. In some retail environments, additional trunks can be added during the peak season and disconnected afterwards.

Busy Hour Call Attempts (BHCA)

The BHCA is the total number of calls during the peak traffic hour (or interval) that are attempted or received in the contact center. For the sake of simplicity, we assume that all calls offered to the Voice Gateway are received and serviced by the contact center resources (agents and Unified IP IVR ports). Calls normally originate from the PSTN, although calls to a contact center can also be generated internally, such as by a help-desk application.

Calls Per Second as reported by Call Router (CPS)

These are the number of call routing requests received by the Unified CCE Call Router per second. Every call will generate one call routing request in a simple call flow where the call comes in from an ingress gateway, receives some VRU treatment and is then sent to an Agent; however, there are conditions under which a single call will need more than one routing request to be made to the Unified CCE Call Router to finally get to the right agent.

An example of this is when the first agent who receives the call wants to transfer/conference to another agent by using a post route. This will generate an additional routing request resulting in the same call generating two routing requests to the Unified CCE Call Router. A routing request is made to the Unified CCE Call Router whenever a resource is required for a call/task. These requests also include multimedia requests for Email, Chat, Blended Collaboration, Callback and certain Outbound Calls. Call center administrators must take into account these additional call routing requests when they size their contact center.

The maximum supported call rate is the call rate reported by the Unified CCE Call Router and not the BHCA at the ingress gateway. These additional routing requests need to be factored into the calculation of BHCA at the ingress gateway. In general, the BHCA at the ingress gateway is lower than or equal to the corresponding CPS rate reported by the Unified CCE Call Router.

For example, consider the following situation. If the BHCA at the ingress gateway is 36,000, then the call rate at the ingress gateway is 10 CPS. If we assume that 10% of the calls are transferred through the Call Router, the CPS reported by Call Router is equal to 11 CPS. In this case, the Unified CCE platform needs a capacity of 11 CPS.

Servers

Servers are resources that handle traffic loads or calls. There are many types of servers in a contact center, such as PSTN trunks and gateway ports, agents, voicemail ports, and VRU ports.

Talk Time

Talk time is the amount of time an agent spends talking to a caller, including the time an agent places a caller on hold and the time spent during consultative conferences.

Wrap-Up Time (After-Call Work Time)

After the call is terminated (the caller finishes talking to an agent and hangs up), the wrap-up time is the time it takes an agent to wrap up the call by performing such tasks as updating a database, recording notes from

the call, or any other activity performed until an agent becomes available to answer another call. The Unified CCE term for this concept is *after-call work time*.

Average Handle Time (AHT)

AHT is the mean (or average) call duration during a specified time period. It is a commonly used term that refers to the sum of several types of handling time, such as call treatment time, talk time, and queuing time. In its most common definition, AHT is the sum of agent talk time and agent wrap-up time.

Erlang

Erlang is a measurement of traffic load during the busy hour. The Erlang is based on having 3600 seconds (60 minutes, or 1 hour) of calls on the same circuit, trunk, or port. (One circuit is busy for one hour regardless of the number of calls or how long the average call lasts.) If a contact center receives 30 calls in the busy hour and each call lasts for six minutes, this equates to 180 minutes of traffic in the busy hour, or 3 Erlangs (180 min/60 min). If the contact center receives 100 calls averaging 36 seconds each in the busy hour, then total traffic received is 3600 seconds, or 1 Erlang (3600 sec/3600 sec).

Use the following formula to calculate the Erlang value:

$$\text{Traffic in Erlangs} = (\text{Number of calls in the busy hour} * \text{AHT in sec}) / 3600 \text{ sec}$$

The term is named after the Danish telephone engineer A. K. Erlang, the originator of queuing theory used in traffic engineering.

Busy Hour Traffic (BHT) in Erlangs

BHT is the traffic load during the busy hour and is calculated as the product of the BHCA and the AHT normalized to one hour:

$$\text{BHT} = (\text{BHCA} * \text{AHT seconds}) / 3600, \text{ or } \text{BHT} = (\text{BHCA} * \text{AHT minutes}) / 60$$

For example, if the contact center receives 600 calls in the busy hour, averaging 2 minutes each, then the busy hour traffic load is $(600 * 2/60) = 20$ Erlangs.

BHT is typically used in Erlang-B models to calculate resources such as PSTN trunks or self-service VRU ports. Some calculators perform this calculation transparently using the BHCA and AHT for ease of use and convenience.

Grade of Service (Percent Blockage)

This measurement is the probability that a resource or server is busy during the busy hour. All resources might be occupied when a user places a call. In that case, the call is lost or blocked. This blockage typically applies to resources such as Voice Gateway ports, VRU ports, PBX lines, and trunks. In the case of a Voice Gateway, grade of service is the percentage of calls that are blocked or that receive busy tone (no trunks available) out of the total BHCA. For example, a grade of service of 0.01 means that 1% of calls in the busy hour is blocked. A 1% blockage is a typical value to use for PSTN trunks, but different applications might require different grades of service.

Blocked Calls

A blocked call is a call that is not serviced immediately. Callers are considered blocked if they are rerouted to another route or trunk group, if they are delayed and put in a queue, or if they hear a tone (such as a busy tone) or announcement. The nature of the blocked call determines the model used for sizing the particular resources.

Service Level

This term is a standard in the contact center industry, and it refers to the percentage of the offered call volume (received from the Voice Gateway and other sources) that are answered within x seconds, where x is a variable. A typical value for a sales contact center is 90% of all calls answered in less than 10 seconds (some calls are delayed in a queue). A support-oriented contact center might have a different service level goal, such as 80% of all calls answered within 30 seconds in the busy hour. Your contact center's service level goal determines the number of agents needed, the percentage of calls that are queued, the average time calls spend in queue, and the number of PSTN trunks and Unified IP IVR ports needed. For an additional definition of service level within Unified CCE products, see the Unified CCE glossary, which is available in the Configuration Manager's online help.

Queuing

When agents are busy with other callers or are unavailable (after call wrap-up mode), subsequent callers must be placed in a queue until an agent becomes available. The percentage of calls queued and the average time spent in the queue are determined by the service level desired and by agent staffing. Cisco's Unified CCE solution uses a Unified IP IVR to place callers in queue and play announcements. It can also be used to handle all calls initially (call treatment, prompt and collect such as DTMF input or account numbers or any other information gathering) and for self-service applications where the caller is serviced without needing to talk to an agent (such as obtaining a bank account balance, airline arrival/departure times, and so forth). Each of these scenarios requires a different number of Unified IP IVR ports to handle the different applications because each has a different average handle time and possibly a different call load. The number of trunks or gateway ports needed for each of these applications will also differ accordingly.

Contact Center Resources and the Call Timeline

The focus of this topic is on sizing the following main resources in a contact center:

- Agents
- Gateway ports (PSTN trunks)
- Unified IP IVR ports

It is helpful first to understand the anatomy of an inbound contact center call as it relates to the various resources used and the holding time for each resource. The following figure shows the main resources and the occupancy (hold/handle time) for each of these resources.

Figure 88: Inbound Call Timeline



Ring delay time (network ring) must be included, if calls are not answered immediately. This delay can be a few seconds on average, and it must be added to the trunk average handle time.

Erlang Calculators as Design Tools

Many traffic models are available for sizing telephony systems and resources. Choosing the right model depends on three main factors:

- Traffic source characteristics (finite or infinite)
- How lost calls are handled (cleared, held, delayed)
- Call arrival patterns (random, smooth, peaked)

For purposes of this document, there are mainly two traffic models that are commonly used in sizing contact center resources, Erlang-B and Erlang-C. There are many other resources on the internet that give detailed explanations of the various models (search using traffic engineering).

Erlang calculators are designed to help answer the following questions:

- How many PSTN trunks do I need?
- How many agents do I need?
- How many VRU ports do I need?

Before you can answer these basic questions, you must have the following minimum set of information that is used as input to these calculators:

- The busy hour call attempts (BHCA)
- Average handle time (AHT) for each of the resources
- Service level (percentage of calls that are answered within x seconds)
- Grade of service, or percent blockage, desired for PSTN trunks and Unified IP IVR ports

The next two sections of this chapter present a brief description of the generic Erlang models in simple terms. Also described are the input/output of the Erlang models and which model to use for sizing the specific contact center resource (agents, gateway ports, and Unified IP IVR ports). There are various web sites that provide contact center sizing tools free of charge (some offer feature-rich versions for purchase), but they all use the two basic traffic models, Erlang-B and Erlang-C. Cisco does not endorse any particular vendor product; it is up to the customer to choose which tool suits their needs. The input required for any of the tools, and the methodology used, are the same regardless of the tool itself.

Erlang-C

The Erlang-C model is used to size agents in contact centers that queue calls before presenting them to agents. This model assumes:

- Call arrival is random.
- If all agents are busy, new calls are queued and not blocked.

The input parameters required for this model are:

- The number of calls in the busy hour (BHCA) to be answered by agents
- The average talk time and wrap-up time

- The delay or service level desired, expressed as the percentage of calls answered within a specified number of seconds

The output of the Erlang-C model lists the number of agents required, the percentage of calls delayed or queued when no agents are available, and the average queue time for these calls.

Erlang-B

The Erlang-B model is used to size PSTN trunks, gateway ports, or Unified IP IVR ports. It assumes the following:

- Call arrival is random.
- If all trunks/ports are occupied, new calls are lost or blocked (receive busy tone) and not queued.

The input and output for the Erlang B model consists of the following three factors. You need to know any two of these factors, and the model will calculate the third:

- Busy Hour Traffic (BHT), or the number of hours of call traffic (in Erlangs) during the busiest hour of operation. BHT is the product of the number of calls in the busy hour (BHCA) and the average handle time (AHT).
- Grade of Service, or the percentage of calls that are blocked because not enough ports are available.
- Ports (lines), or the number of Unified IP IVR or gateway ports.



11

CHAPTER

Sizing Unified CCE Components and Servers

- Sizing Considerations for Unified CCE, page 215
- Peripheral Gateway and Server Options, page 223
- Agent Greeting Sizing Considerations, page 224
- Whisper Announcement Sizing Considerations, page 225
- Throttling During Precision Queue Changes , page 225
- System Performance Monitoring, page 226
- Summary, page 227

Sizing Considerations for Unified CCE

Proper sizing of your Cisco Unified Contact Center Enterprise (Unified CCE) solution is important for optimum system performance and scalability. Sizing considerations include the number of agents the solution can support, the maximum busy hour call attempts (BHCA), and other variables that affect the number, type, and configuration of application servers required to support the deployment. Regardless of the deployment model chosen, Unified CCE is based on a highly distributed architecture, and questions about capacity, performance, and scalability apply to each element within the solution as well as to the overall solution.

This chapter presents design practices focusing on scalability and capacity for Unified CCE deployments. The design considerations and capacities presented in this chapter are derived primarily from testing and, in other cases, extrapolated test data. This information is intended to enable you to size and provision Unified CCE solutions appropriately.

This chapter refers to sizing tools that are available [online](#). The sizing tools are available to Cisco internal employees and Cisco partners only, and proper login authentication is required.

Core Unified CCE Components

When sizing Unified CCE deployments, Cisco Unified Communications components are a critical factor in capacity planning. Good design, including multiple Cisco Unified Communications Managers and clusters, must be utilized to support significant call loads. For additional information about Cisco Unified Communications Manager capacity and sizing of Cisco Unified Communications components, see the latest

version of the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Additionally, because of varying agent and skill group capacities, consider proper sizing of the Agent PG, including Finesse and CTI OS Desktop servers, together with the Cisco Unified Communications components.

Finally, the remaining Unified CCE components, while able to scale extremely well, are affected by specific configuration element sizing variables that also have an impact on the system resources. These factors, discussed in this section, must be considered and included in the planning of any deployment.


Note

Unless otherwise explicitly noted, the capacity information presented in *Operating Conditions* specifies capacity for inbound calls only and does not apply equally to all implementations of Unified CCE. The data is based on testing in particular scenarios, and it represents the maximum allowed configuration. This data, along with the sizing variables information in this chapter, serves only as a guide. As always, be conservative when sizing and plan for growth.


Note

Sizing considerations are based on capacity and scalability test data. Major Unified CCE software processes were run on individual VMs to measure their specific CPU and memory usage and other internal system resources. Reasonable extrapolations were used to derive capacities for co-resident software processes and multiple vCPU VMs. This information is meant as a guide for determining when Unified CCE software processes can be co-resident within a single VM and when certain processes need their own dedicated VM. [Table 20: Sizing Information for Unified CCE Components and Servers, on page 218](#) assumes that the deployment scenario includes two fully redundant VMs.

Related Topics

[Operating Conditions, on page 216](#)

[Sizing Cisco Unified Communications Manager Servers, on page 229](#)

Operating Conditions

The sizing information presented in this chapter is based on the following operating conditions:

- Maximum of 30 busy hour call attempts (BHCA) per agent.
- Five skill groups or precision queues per agent.
- The total number of agents indicated in the following tables and figures consists of 90% agents and 10% supervisors. For example, if a table or figure indicates 100 agents, the assumption is that there are 90 agents and 10 supervisors.
- Supervisors do not handle calls.
- Total number of teams is equal to 10% of total number of agents.
- Team members consist of 90% agents and 10% supervisors.
- Call types consist of 85% straight calls, 10% consultative transfers, and 5% consultative conferences.
- The default refresh rate for skill group updates is 10 seconds.
- The default number of skill group statistics columns configured at the CTI OS server is 17 columns.

- Agent Statistics is turned ON.
- The default number of agent statistics columns configured at the CTI OS server is 6 columns.
- Average of five Voice Response Unit (VRU) scripts, running consecutively in the Unified CCE script, per VRU call.
- Five Expanded Call Context (ECC) scalars.
- Transport Layer Security (TLS) for CTI OS is turned OFF.
- No mobile agents.
- Outbound hit rate is averaged at 30%.

The following notes apply to all figures and tables in this topic:

- The number of agents indicates the number of logged-in agents
- Server types:
 - APG = Agent Peripheral Gateway
 - PGR = Lab deployment
 - RGR = Rogger

Figure 89: Minimum Servers Required for Unified CCE Deployments with CTI OS Desktop

Maximum Agent Count	450*	2,000*	4,000	8,000	12,000
Central Controller					
Peripheral Gateways Agent Services					

* Deployment supported in Unified System CCE

© 2014 Cisco

The following notes apply to the figure above:

- Sizing is based on the information listed above.
- Voice Response Unit (VRU), Administration & Data Server, and Unified Communications Manager components are not shown.

**Note**

The terms Rogger and Central Controller are used interchangeably throughout this chapter.

Table 20: Sizing Information for Unified CCE Components and Servers

Component	Notes
Administration & Data Server	For the current specifications for a VM running the Administration & Data Server, see the <i>Virtualization for Unified CCE DocWiki</i> at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE .
Voice Response Unit (VRU) PG	Use the number of ports instead of agent count. Average of 5 Run VRU Script Nodes per call.
Agent PG with Outbound Voice (Includes Dialer and Media Routing PG)	<p>To determine the maximum inbound agent capacity, see the Inbound Agent PG entry in this table. The capacity depends on your Unified CCE software release, hardware server class, and agent desktop type.</p> <p>Impact of Outbound Option on agent capacity with the SIP Dialer: (Maximum PG agent capacity) – (1.33 x [number of SIP Dialer ports])</p> <p>The formula is an indicator of platform capacity. The formula does not indicate outbound resources in terms of how many agents can be kept busy by the number of dialer ports in the deployment. A quick but inexact estimate is that two ports are required for each outbound agent, but your outbound resources can vary depending on hit rate, abandon limit, and talk time for the campaigns in the deployment. Use the sizing tool to determine outbound resources required for your campaigns.</p> <p>Example: Agent PG with Cisco Finesse and 30 SIP Dialer ports.</p> <p>Available inbound Finesse agents = 2000 - (1.33*30) = 1960.</p> <p>Note The Cisco Media Blender is not supported when installed on a PG system.</p>
Cisco Unified Web and E-Mail Interaction Manager	For the most current server specifications and sizing guidelines for Cisco Unified Web and E-Mail Interaction Manager, see the latest documentation at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-email-interaction-manager/products-implementation-design-guides-list.html .
Cisco Unified Customer Voice Portal (CVP) Application Server And Voice Browser	For the most current server specifications for Unified CVP, see the latest version of the <i>Hardware and System Software Specification for Cisco Unified CVP</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-technical-reference-list.html .
Unified IP IVR Server	For the most current Unified IP IVR server specifications, see the documentation available through valid Cisco Employee or Partner login.
Cisco Unified Intelligence Center (Unified Intelligence Center)	For the most current server specifications for Unified Intelligence Center, see the latest version of the <i>Hardware and System Software Specification (Bill of Materials)</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-intelligence-center/products-implementation-design-guides-list.html .

For further details on sizing VMs, see the [Unified Communications in a Virtualized Environment](#) DocWiki pages.

Related Topics

- [Peripheral Gateway and Server Options, on page 223](#)
- [Cisco Agent Desktop, on page 351](#)
- [System Requirements and Constraints, on page 287](#)

Additional Sizing Factors

Many variables in the Unified CCE configuration and deployment options can affect the server requirements and capacities. This section describes the major sizing variables and how they affect the capacity of the various Unified CCE components.

Busy Hour Call Attempts (BHCA)

The number of calls attempted during a busy hour is an important metric. As BHCA increases, there is an increase in the load on all Unified CCE components, most notably on Unified CM, Unified IP IVR, and the Unified CM PG. The capacity numbers for agents assume up to 30 calls per hour per agent. If a deployment requires more than 30 calls per hour per agent, it decreases the maximum number of supported agents for the agent PG. Handle such occurrences on a case-by-case basis.

Agents

The number of agents is another important metric that impacts the performance of most Unified CCE server components, including Unified CM clusters.

Average Skill Groups or Precision Queues Per Agent

The number of skill groups or precision queues per agent (which is independent of the total number of skills per system) significantly affects the following:

- CTI OS servers
- Finesse servers
- Agent PG
- Call Router
- Logger

Limit the number of skill groups and precision queues per agent to 5 or fewer, when possible. Periodically remove unused skill groups or precision queues so that they do not affect the system's performance. You can also manage the effects on the CTI OS Server by increasing the value for the frequency of statistical updates.

The Finesse server does not display statistics for unused skill groups. Therefore, the number of skill groups that are assigned to agents affects the performance of the Finesse server more than the total number of skill groups configured.

Queue (skill group) statistics are updated on the Finesse Desktop at 10-second intervals. The Finesse Desktop also supports a fixed number of queue statistics fields. These fields cannot be changed.

Additional Sizing Factors

The first table shows examples of the number of skill groups or precision queues (PQ) per agent affecting the capacity of the Unified CCE system. The table shows the capacity for each CTI OS instance. The Finesse server supports the same number of agents and skill groups as CTI OS.

Unified CCE supports a maximum of 50 unique skill groups across all agents on a supervisor's team, including the supervisor's own skill groups. If this number is exceeded, all skill groups monitored by the supervisor still appear on the supervisor desktop. However, exceeding this number can cause performance issues and is not supported.

**Note**

Each precision queue that you configure creates a skill group per Agent PG and counts toward the supported number of skill groups per PG.

The numbers in this table are subject to specific hardware and software requirements.

Table 21: Sizing Effects of Skill Groups or Precision Queues for Each Agent (12,000 Agents)

Avg Configured PQ or SG for each Agent	System	Generic PG Limits				
		Max Concurrent Agent for each System	Max Concurrent Agent for each PG	Max Configured PQ or SG for each PG	Max Configured VRU Ports for each PG	Max Configured VRU PIMs for each PG
5	12000	2000	4000	1000	4	
10	11038	1832	4000	1000	4	
15	10078	1663	4000	1000	4	
20	9116	1495	4000	1000	4	
25	8156	1326	4000	1000	4	
30	7194	1158	4000	1000	4	
35	6234	989	4000	1000	4	
40	5272	820	4000	1000	4	
45	4312	652	4000	1000	4	
50	3350	484	4000	1000	4	

**Note**

CTI OS monitor mode applications are supported only at 20 or lower skill groups per agent.

Supervisors and Teams

The number of supervisors and team members can also be a factor impacting the CTI OS Server performance. Distribute your agents and supervisors across multiple teams and have each supervisor monitor only a few agents.

**Note**

Supervisors can monitor only agents within their own team, and all of the agents must be configured on the same peripheral.

**Note**

You can add a maximum of 50 agents per team. You can add a maximum of ten supervisors per team.

A Unified CCE system can support a maximum of 50 agents per supervisor with the assumptions below. If a particular environment requires more than 50 agents per supervisor, then use the following formula to ensure that there is no impact to the CTI OS Server and Supervisor desktop. The most important factor in this calculation is the number of updates per second.

$$X = (Y * (N + 1) / R) + ((Z * N * A) / 3600), \text{ rounded up to the next integer}$$

Where:

X = Number of updates per second received by the CTI OS Supervisor desktop.

Y = Weighted Average of Number of Skill Groups or Precision Queues per Agents. For example, if total of 10 agents have the following skill group distribution: 9 have 1 skill group and 1 agent has 12 Skill Groups. The number of skills per agent ('Y') is, $Y = 90\% * 1 + 10\% * 12 = 2.1$. (The number of configured statistics in the CTI OS server is 17.)

Z = Calls per hour per agent.

A = Number of agent states. (Varies based on call flow; average = 10.)

N = Number of agents per supervisor.

R = The skill group or precision queue refresh rate configured on the CTI OS Server. (Default = 10 seconds.)

$(Y * (N + 1) / R)$ = Number of updates per second, based on skill groups.

$(Z * N * A) / 3600$ = Number of updates per second, based on calls.

The CTI OS Supervisor desktop is not impacted as long as there are fewer than 31 updates per second. This threshold value is derived by using the above formula to calculate the update rate for 50 agents per supervisor ($N = 50$), as follows:

$$X = (5 * (50 + 1) / 10) + ((30 * 50 * 10) / 3600) = 25.5 + 5 = 31 \text{ updates per second}$$

The maximum number of agents per supervisor must not exceed 200 for any given configuration, still holding updates per second to a maximum of 31 with the above formula.

CTI OS Monitor Mode Applications

A CTI OS Monitor Mode application can affect the performance of the CTI OS Server. CTI OS supports only two such applications per server pair. Depending on the filter specified, the impact on the CPU utilization might degrade the performance of the Agent PG.

Unified CM Silent Monitor

Each silently monitored call adds more processing for the PG and Unified CM. Each silently monitored call is equivalent to two unmonitored calls to an agent. Make sure that the percentage of the monitored calls is within the capabilities of PG scalability.

CTI OS Skill Group Statistics Refresh Rate

The skill group statistics refresh rate can also affect the performance of CTI OS Server. Cisco requires that you do not lower the refresh rate below the default value of 10 seconds.

Call Types

The call type is also an important metric that affects the performance of most Unified CCE server components. An increase in the number of transfers and conferences increase the load on the system which decreases the total capacity.

Queuing

The Unified IP IVR and Unified Customer Voice Portal (CVP) place calls in a queue and play announcements until an agent answers the call. For sizing purposes, it is important to know whether:

- The VRU will handle all calls initially (call treatment), and direct the callers to agents after a short queuing period.
- The agents will handle calls immediately, and the VRU queues only unanswered calls when all agents are busy.

The answer to this question determines very different VRU sizing requirements and affects the performance of the Call Router/Logger and Voice Response Unit (VRU) PG.

Translation Route Pool

Sizing the translation route pool depends on the expected call arrival rate. Use the following formula to size the translation route pool:

$$\text{Translation route pool} = 20 * (\text{Calls per second})$$

This calculation is specific to Unified CCE. For more general Unified ICM deployments, consult your Cisco Account Team or Partner.

Unified CCE Script Complexity

As the complexity and/or number of Unified CCE scripts increase, the processor and memory overhead on the Call Router and VRU PG increases significantly. The delay time between replaying Run VRU scripts also has an impact.

Reporting

Real-time reporting can have a significant effect on Logger and Rogger processing due to database access. A separate VM is required for an Administration & Data Server to off-load reporting overhead from the Logger and Rogger.

VRU Script Complexity

As VRU script complexity increases with features such as database queries, the load placed on the IP IVR Server and the Router also increases. There is no good rule or benchmark to characterize the Unified IP IVR performance when used for complex scripting, complex database queries, or transaction-based usage. Test complex VRU configurations in a lab or pilot deployment to determine the response time of database queries under various BHCA and how they affect the processor and memory for the VRU server, PG, and Router.

Unified IP IVR Self-Service Applications

In deployments where the Unified IP IVR is also used for self-service applications, the self-service applications are in addition to the Unified CCE load. Factor self-service applications into the sizing requirements as stated in the sizing tables above.

Third-Party Database and Cisco Resource Manager Connectivity

Carefully examine connectivity of any Unified CCE solution component to an external device and/or software to determine the overall effect on the solution. Cisco Unified CCE solutions are flexible and customizable, but they can also be complex. Contact centers are often mission-critical, revenue-generating, and customer-facing operations. Therefore, engage a Cisco Partner (or Cisco Advanced Services) with the appropriate experience and certifications to help you design your Unified CCE solution.

Expanded Call Context (ECC)

The ECC usage impacts PG, Router, Logger, and network bandwidth. There are many ways that ECC can be configured and used. The capacity impact varies based on ECC configuration, handled on a case-by-case basis.

Related Topics

[Operating Conditions, on page 216](#)

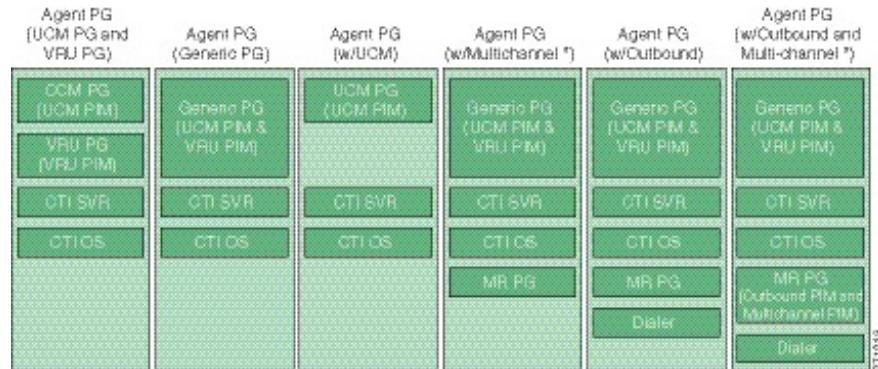
[Sizing Cisco Unified Communications Manager Servers, on page 229](#)

Peripheral Gateway and Server Options

A Unified CCE Peripheral Gateway (PG) translates messages coming from the Unified Communications Manager servers, the Unified IP IVR, Unified CVP, or voice response units (VRUs) into common internally formatted messages that are then sent to and understood by Unified CCE. In the reverse, it also translates Unified CCE messages so that they can be sent to and understood by the peripheral devices.

The figures below illustrate various configuration options for the Agent PG with CTI OS.

Figure 90: Agent PG Configuration Options with CTI OS



* with CCE, the MR PG for outbound and multi-channel is co-resident with the Agent PG. With System CCE, the MR PG for outbound (outbound controller) is co-resident with the Agent PG, but not the MR PG for multi-channel (multi-channel controller).

The table below gives sizing guidelines for PGs and PIMs.

Table 22: PG and PIM Sizing Guidelines

Sizing variable	Guidelines based on Unified CCE Release 10
Maximum number of PGs per Unified CCE	150
Maximum number of PG types per VM	Up to two PG types are permitted per VM, but each VM must meet the maximum agent and VRU port limitations.
Maximum number of Unified Communications Manager PGs per VM	Only one Unified Communications Manager PG, Generic PG, or System PG is allowed per VM.
Maximum number of Unified Communications Manager PIMs per PG	1
Can PGs be remote from Unified CCE?	Yes
Can PGs be remote from Unified Communications Manager?	No
Maximum number of VRUs controlled by one Unified Communications Manager	See the <i>Cisco Collaboration System Solution Reference Network Designs</i> at http://www.cisco.com/go/ucsrnd .
Maximum number of CTI servers per PG	1
Can PG be co-resident with Cisco Unified Communications Manager?	No

Agent Greeting Sizing Considerations

Agent Greeting invokes conference resources to bring the greeting into the call. With supported hard phones, the Built in Bridge on the phone is used. For Mobile Agent, conference resources are used. This adds a short but additional call leg to every call, which has impacts on all components.

Central Controller

Agent Greeting has an impact of up to 1.5 regular calls on the Router and Logger. This implies that the maximum call rate on Unified CCE is reduced from 60 calls per second to 40 calls per second, as measured by new calls that originate from the service provider. As each Agent Greeting involves an additional route request, the Router PerfMon counter displays 80 calls per second under a full supported load. The number of agents supported per System is dependent upon the overall call rate. For a specific scenario, see the [Unified CCE Sizing Tool](#).

Peripheral Gateway

Agent Greeting does have an impact on the PG resource utilization, but it is not enough impact to require reducing the supported agent capacity per PG. Other factors like additional skill groups per agent or total configured skill groups also play a factor in PG sizing. When sizing the PG, the sizing calculator factors in Agent Greeting.

Communications Manager

When Agent Greeting and/or Mobile Agent and Unified IP IVR are in use, the number of agents supported by a Unified Communications Manager subscriber is impacted.

The [Cisco Unified Collaboration Sizing Tool](#) takes call rate and the other factors into account to determine the capacity for a specific scenario.

Mobile Agent

Agent Greeting with Mobile Agent uses additional Conference Bridge and MTP resources. The agent greeting calls are relatively short and they need not be factored into sizing considerations. To properly size Conference Bridge and UCM resources, indicate a conference in place of an Agent Greeting for each Mobile Agent (when Agent Greeting is enabled) for each inbound call.

CVP and VXML Gateway

Agent Greeting also utilizes CVP and VXML gateway resources, so it is important to consider the call rate when sizing. The *Design Guide for Cisco Unified Customer Voice Portal* includes information about how to size based on call rate; however, most deployments are sized based on the number of ports. The Agent Greeting utilization has a profile of short calls but at a high call rate, so do not overlook this consideration.

Whisper Announcement Sizing Considerations

The impact of Whisper Announcement on solution component sizing is not as significant as the impact caused by Agent Greeting. To factor in Whisper Announcement, run the [Cisco Unified Collaboration Sizing Tool](#).

Throttling During Precision Queue Changes

A configuration update on a precision queue (made through the API or the Unified CCE Administration tool) can result in many agents with changed precision queue associations. These updates could overload the system if done all at once. Therefore, when an update affects many agents, the system moves the agents into and out of their respective precision queues gradually, based on available system resources.

You can submit another precision queue configuration update before an earlier update completes. If you submit the updates too quickly, the new update can cause the pending configuration updates to queue in the system. To avoid a backlog, the system rejects new precision queue configuration updates after reaching five

concurrent pending updates. Once the pending precision queue updates fall below the threshold, the system accepts new configuration updates.

To mitigate possible overload conditions on the agent peripheral during these operations, the system limits the number of calls to the peripheral during an overload condition. When an overload occurs, the system stops sending Precision Routing calls to that peripheral for a short time.

System Performance Monitoring

Supporting and maintaining an enterprise solution requires many steps and procedures. Depending on the customer environment, the support procedures vary. System performance monitoring is one procedure that helps maintain the system. This section provides a guide for monitoring Unified CCE to ensure that the system is performing within system tolerances. System monitoring is especially critical for customers as they expand or upgrade their system. Monitor the system during times of heavy activity. For more information about monitoring, see the *Serviceability Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-configuration-examples-list.html>.

The following system components are critical to monitor:

- CPU
- Memory
- Disk
- Network

The following list highlights some of the important counters for the critical system components, along with their threshold values:

- Monitoring the CPU
 - %Processor Time; the threshold of this counter is 60%.
 - ProcessorQueueLength; this value must not go above ($2 * [\text{the total number of CPUs on the system}]$).
- Monitoring Memory
 - % Committed Bytes; this value must remain less than ($0.8 * [\text{the total amount of physical memory}]$).
 - Memory\Available MByte; this value must not be less than 16 MB.
 - Page File \%usage; the threshold for this counter is 80%.
- Monitoring the Disk Resources
 - AverageDiskQueueLength; this value must remain less than ($1.5 * [\text{the total number of disks in the array}]$).
 - %Disktime; this value must remain less than 60%.
- Monitoring Network Resources
 - NIC\bytes total/sec; this value must remain less than ($0.3 * [\text{the bandwidth of the NIC}]$).
 - NIC\Output Queue Length; the threshold for this counter is 1.

- Monitoring Unified CCE application
 - Cisco Call Router(_Total)\Agents Logged On
 - Cisco Call Router(_Total)\Calls in Progress
 - Cisco Call Router(_Total)\calls /sec

**Note**

The above performance counters for CPU, memory, disk, and network are applicable to all VMs within the deployment. The preferred sample rate is 15 seconds.

Summary

Proper sizing of Unified CCE components requires analysis beyond the number of agents and busy hour call attempts. Configurations with multiple skill groups per agent, significant call queuing, and other factors contribute to the total capacity of any individual component. Careful planning and discovery in the pre-sales process uncovers critical sizing variables; apply these considerations to the final design and hardware selection.

Correct sizing and design can ensure stable deployments for large systems up to 8000 agents and 216,000 BHCA. For smaller deployments, cost savings can be achieved with careful planning and co-resident Unified CCE components (for example, Rogger and Agent PG).

Pay careful attention to the sizing variables that impact sizing capacities such as skill groups per agent. While it is often difficult to determine these variables in the pre-sales phase, it is critical to consider them during the initial design, especially when deploying co-resident PGs.



CHAPTER 12

Sizing Cisco Unified Communications Manager Servers

- [Sizing Unified Communications Manager Clusters for Unified CCE, page 229](#)
- [Cluster Sizing Concepts, page 230](#)
- [Cisco Unified Collaboration Sizing Tool, page 231](#)
- [Cluster Guidelines and Considerations, page 231](#)
- [Unified Communications Manager Redundancy, page 234](#)
- [Load Balancing for Unified Communications Manager, page 235](#)
- [Deployment of Agent PG in Unified Communications Manager Cluster, page 235](#)
- [Sizing Considerations for Unified Mobile Agent, page 236](#)

Sizing Unified Communications Manager Clusters for Unified CCE

This chapter discusses the concepts, provisioning, and configuration of Cisco Unified Communications Manager clusters when used in a Unified CCE deployment. Clusters provide a mechanism for distributing call processing across a converged IP network infrastructure to support Cisco Unified Communications. This mechanism also facilitates redundancy and provides feature transparency and scalability.

This chapter covers only the Unified CCE operation with clusters and proposes reference designs for implementation.

The information in this chapter builds on the concepts presented in the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>. Some duplication of information is necessary to clarify concepts relating to Unified CCE as an application supported by the Unified Communications Manager call processing architecture. However, the foundational concepts are not duplicated here; become familiar with them before continuing with this chapter.

This chapter documents general requirements and scalability considerations for sizing the subscribers used with your Unified CCE deployments. Within the context of this document, scalability refers to the subscriber and cluster capacity when used in a Unified CCE deployment. For information about sizing and choosing a

gateway, see the gateway information in the latest version of the *Cisco Collaboration System Solution Reference Network Designs*.

Cluster Sizing Concepts

Before attempting to size a Unified Communications Manager cluster for a Unified CCE deployment, perform the following design tasks:

- Determine the different types of call flows.
- Determine the required deployment model (single site, centralized, distributed, clustering over the WAN, or remote branches within centralized or distributed deployments).
- Determine whether Unified CVP or IP IVR is used for call treatment, self-service, and queuing.
- Determine the protocols to be used.
- Determine redundancy requirements.
- Determine all other customer requirements for Cisco Unified Communications that share a cluster with a Unified CCE deployment. For example, consider Cisco Unified IP Phones, applications that are not part of Unified CCE, route patterns, and so forth.

After you complete these tasks, you can begin to accurately size the necessary clusters. Many factors affect the sizing of a cluster, and the following list mentions some of those factors:

- Number of office phones and the busy hour call attempt (BHCA) rate per phone
- Number of inbound agent phones and the BHCA rate per phone
- Number of CTI ports and the BHCA rate on those VoIP endpoints. (If you use Unified CVP for call treatment, self-service, and queuing, these factors might not apply.)
- Number of Voice Gateway ports and the BHCA rate on those VoIP endpoints
- Number of outbound agent phones, outbound dialing mode, and BHCA rate per phone
- Number of outbound dialer ports, number of VRU ports for outbound campaigns, and the BHCA rate per port for both
- Number of mobile agents and the BHCA rate per mobile agent
- Number of voicemail ports and the BHCA rate to those VoIP endpoints
- Signaling protocols used by the VoIP endpoints
- Percent of agent call transfers and conferences
- Dial plan size and complexity, including the number of dialed numbers, lines, partitions, calling search spaces, locations, regions, route patterns, translations, route groups, hunt groups, pickup groups, and route lists
- Amount of media resources needed for functions such as transcoding, conferences, encryption, and so forth
- Coresident applications and services such as CTI Manager, E-911, and Music on Hold
- Unified Communications Manager release (sizing varies per release)

- Type of Unified Communications Manager OVA

Other factors can affect cluster sizing, but the preceding list shows the most significant factors in terms of resource consumption.

In general, sizing a cluster involves estimating the resource consumption (CPU, memory, and I/O) for each of these factors. You then choose VMs that satisfy the resource requirements. Gather information about these factors before you can size a cluster with any accuracy.

Cisco Unified Collaboration Sizing Tool

To size Cisco Unified Contact Center Enterprise servers, use the Cisco Unified Collaboration Sizing Tool (Unified CST) available at <http://tools.cisco.com/cust/faces>.

You can use the Unified CST to size large and complex Unified Communications Systems. This tool supports the sizing of many Unified Communications components, such as Unified Communications Manager, Unified CCE, Unified IP IVR, Unified CVP, and gateways.

The Unified CST is available to Cisco internal employees and Cisco partners, and proper login authentication is required. For detailed instructions, see the documentation for this tool.

Cluster Guidelines and Considerations

The following guidelines apply to all Unified Communications Manager clusters with Unified CCE:

- All primary and backup subscribers must use the same OVF template. All subscribers in the cluster must run the same Unified Communications Manager software release and service pack.
- Within a cluster, you can enable a maximum of eight subscribers (four primary and four backup subscribers) with the Cisco Call Manager Service. You can use more VMs for dedicated functions such as TFTP, publisher, and music on hold.
- In Unified CCE 4,000 Agent deployments with only Unified CVP, a Unified CM cluster can support about 4,000 Unified CCE agents. In Unified CCE 12,000 Agent deployments, a Unified CM cluster with four primary and four backup subscribers can support about 8,000 Unified CCE agents. These limits assume that the BHCA call load and all configured devices are spread equally among the eight call processing subscribers with 1:1 redundancy. These capacities can vary, depending on your specific deployment. All deployments must be sized by using the Cisco Unified Communications Manager Capacity Tool or the Unified Collaboration Sizing Tool.

A subscriber can support a maximum of 1,000 agents. In a fail-over scenario, the primary subscriber supports a maximum of 2,000 agents.



Note

In a Unified CCE 4,000 Agent deployment, a cluster with four subscribers (two primary and two backup) can support the maximum load. If you create clusters with more subscribers, do not exceed the maximum of 4,000 agents for the cluster.

- In Unified CCE deployments with only Unified IP IVR, a Unified CM cluster can support about 2,000 Unified CCE agents. These limits assume that the BHCA call load and all configured devices are spread equally among the eight call processing subscribers with 1:1 redundancy. These capacities can vary,

depending on your specific deployment. All deployments must be sized by using the Cisco Unified Communications Manager Capacity Tool or the Unified Collaboration Sizing Tool.

A subscriber can support a maximum of 250 agents. In a fail-over scenario, the primary subscriber supports a maximum of 500 agents.

When sizing the cluster to support contact center solutions for the appropriate number of CTI resources, remember to account for the following:

- Configured phones from agents who are not signed in
- Applications which remotely control the device like Call Recording, Attendant Console, and PC-clients
- Other 3rd-party applications which consume CTI resources

Unified Communications Manager can support multiple concurrent CTI resources, for example, when multiple lines, the contact center, and recording are used concurrently. Those CTI resources follow the same CTI rules as described in the *Cisco Collaboration System Solution Reference Network Designs*. Size all deployments with the Cisco Unified Collaboration Sizing Tool:

- Devices (including phones, music on hold, route points, gateway ports, CTI ports, JTAPI Users, and CTI Manager) must never reside or be registered on the publisher. If there are any devices registered with the publisher, any administrative work on Unified Communications Manager impacts call processing and CTI Manager activities.
- Do not use a publisher as a fail-over or backup call processing subscriber in production deployments. Any deviations require review by Cisco Bid Assurance on a case-by-case basis.
- Any deployment with more than 150 agent phones requires a minimum of two subscribers and a combined TFTP and publisher. The load-balancing option is not available when the publisher is a backup call processing subscriber.
- If you require more than one primary subscriber to support your configuration, then distribute all agents equally among the subscriber nodes. This configuration assumes that the BHCA rate is uniform across all agents.
- Similarly, distribute all gateway ports and Unified IP IVR CTI ports equally among the cluster nodes.
- Some deployments require more than one Unified CCE JTAPI user (CTI Manager) and more than one primary subscriber. In these deployments, if possible, group and configure all devices monitored by the same Unified CCE JTAPI User (third-party application provider), such as Unified CCE route points and agent devices, on the same VM.
- Enable CTI Manager only on call processing subscribers, thus allowing for a maximum of eight CTI Managers in a cluster. To provide maximum resilience, performance, and redundancy, load-balance CTI applications across the various CTI Managers in the cluster. For more CTI Manager considerations, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.
- If you have a mixed cluster with Unified CCE and general office IP phones, if possible, group and configure each type on a separate VM (unless you need only one subscriber). For example, all Unified CCE agents and their associated devices and resources are on one or more Unified Communications Manager servers. Then, all general office IP phones and their associated devices (such as gateway ports) are on other Unified Communications Manager servers, as long as cluster capacity allows. If you use the Cisco Unified Communications Manager Capacity Tool, run the tool separately with the specific device configuration for each primary Unified Communications Manager server. You need to run it multiple times because the tool assumes that all devices are equally balanced in a cluster. Remember that with Unified CCE, you must use the 1:1 redundancy scheme. If you use the Unified Collaboration

Sizing Tool instead, you can run the tool once because this tool supports deployments with dedicated Unified Communications Manager servers for agent phones or with a mixed cluster.

- Use hardware-based conference resources whenever possible. Hardware conference resources provide a more cost-effective solution and allow better scalability within a cluster.
- Have all CTI route points that are associated with the Unified CCE Peripheral Gateway (PG) JTAPI user register with the subscriber node running the CTI Manager instance that communicates with that Unified CCE PG.
- The Cisco Unified Communications Manager Capacity Tool and the Unified Collaboration Sizing Tool do not currently measure CTI Manager impact on each VM separately. However, CTI Manager does place an extra burden on the subscriber running that process. The tools report the resource consumption based on these subscribers. The actual resource consumption on the other Unified Communications Manager subscribers can be slightly lower.
- Count as an agent device all devices that are associated with a Unified CCE PG JTAPI user, but that a call center agent does not use. The PG is still notified of all device state changes for that phone, even though an agent does not use the phone. To increase cluster scalability, if a device is not used regularly by an agent, do not associate the device with the Unified CCE PG JTAPI user.
- For deployments requiring large numbers of VRU ports, use Unified CVP instead of Unified IP IVR. Unified IP IVR ports place a significant call processing burden on Unified Communications Manager, while Unified CVP does not. Thus, Unified CCE deployments with Unified CVP allow more agents and higher BHCA rates per cluster. Size all deployments by using the Unified Collaboration Sizing Tool.
- In deployments with multiple Unified IP IVRs, associate those servers with different CTI Managers on different subscribers to better balance call processing across the cluster.
- CPU resource consumption by Unified Communications Manager varies, depending on the trace level enabled. Changing the trace level from Default to Full on Unified Communications Manager can increase CPU consumption significantly under high loads. The Cisco Technical Assistance Center does not support changing the tracing level from Default to No tracing.
- Under normal circumstances, place all subscribers from the cluster within the same LAN or MAN. Do not place all members of a cluster on the same VLAN or switch.
- If the cluster spans an IP WAN, follow the specific guidelines for clustering over the IP WAN as described in *Geographically distributed data centers* in this guide and *Clustering over the IP WAN* in the *Cisco Collaboration System Solution Reference Network Designs*.

For the most current information about Unified Communications Manager and Unified CCE supported releases, see the latest version of the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.

For more Unified Communications Manager clustering guidelines, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Related Topics

[Geographically Redundant Data Centers, on page 55](#)

[Unified Communications Manager Redundancy, on page 234](#)

Unified Communications Manager Redundancy

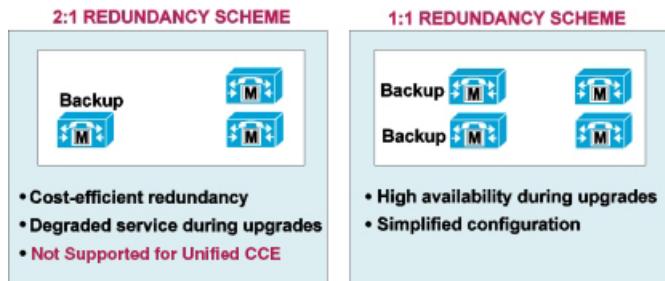
With Unified Communications Manager, you can choose from the following redundancy schemes:

- 2:1—For every two primary subscribers, there is one shared backup subscriber.
- 1:1—For every primary subscriber, there is a backup subscriber.

Due to the higher phone usage in contact centers and the increased downtime required during upgrades, do not use the 2:1 redundancy scheme for Unified Communications Manager deployments with Unified CCE.

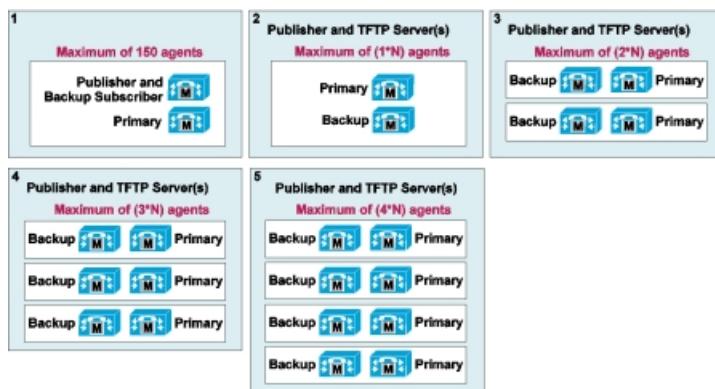
The following figure illustrates these options. This illustration shows only call processing subscribers and does not show publisher, TFTP, music on hold (MoH), or other servers. For details on additional cluster deployment and redundancy options, see the latest version of the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Figure 91: Basic Redundancy Schemes



In the following figure, the options shown all provide 1:1 subscriber redundancy. Option 1 is used for clusters supporting fewer than 150 Unified CCE agents on any supported version of Unified CM. Options 2 through 5 illustrate increasingly larger clusters. In this figure, for deployments with Unified Communications Manager 8.x and Unified CVP (not Unified IP IVR), N is equal to 2000. For deployments with Unified IP IVR, N is equal to 500. For other types of deployments, use the Cisco Unified Communications Manager Capacity Tool or the Cisco Unified Collaboration Sizing Tool.

Figure 92: Redundancy Configuration Options



Load Balancing for Unified Communications Manager

An additional benefit of using the 1:1 redundancy scheme is that it enables you to balance the devices over the primary and backup subscriber pairs.

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

To plan for 50/50 load balancing, calculate the capacity of a cluster without load balancing and then distribute the load across the primary and backup subscribers based on devices and call volume. To allow for failure of the primary or the backup, the total load on the primary and secondary subscribers must not exceed that of a single subscriber. In a 1:1 redundancy pair, you can split the load between the two subscribers, configuring each subscriber with half of the agents. To provide for system fault tolerance, make sure that all capacity limits are observed so that Unified CCE agent phones, Unified IP phones, CTI limits, and so on, for the subscriber pair do not exceed the limits allowed for a subscriber's VM.

Distribute all devices and call volumes as equally as possible across all active subscribers. For instance, distributing the Unified CCE agents, CTI ports, gateways, trunks, voicemail ports, and other users and devices among all subscribers equally, minimizes the impact of any outage.

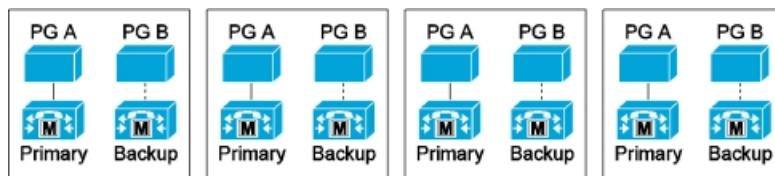
For additional information about general call processing topics such as secondary TFTP servers and gatekeeper considerations, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/go/ucsrnd>.

Deployment of Agent PG in Unified Communications Manager Cluster

You can deploy Agent PGs in a Unified Communications Manager cluster in either of the following ways:

- Deploy an Agent PG for each pair of subscribers. In this case, each subscriber runs the CTI Manager service, and each Agent PG connects to a CTI Manager running on its corresponding subscriber pair. The following diagram shows an example where four primary subscribers are required and four backup subscribers are deployed to provide 1:1 redundancy.

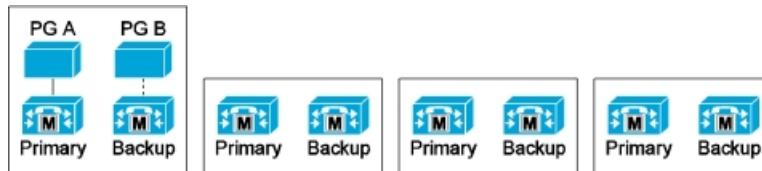
Figure 93: Deploy Agent PG for Each Pair of Subscribers



- Deploy a single Agent PG for the entire cluster. This type of deployment requires a single pair of subscribers running CTI Manager. Spread agent phone registration among all the subscribers, including

the subscribers running the CTI Manager service. The following diagram shows an example where four primary subscribers are required and four backup subscribers are deployed to provide 1:1 redundancy.

Figure 94: Deploy Single Agent PG for Entire Cluster



This model reduces the server count for the PG. Another benefit is that there is a single PIM for the entire cluster. So, you can create teams that span across many subscribers. This allows supervisors, for example, to monitor agent phones registered to any subscriber in the cluster. However, this deployment can have slightly higher resource utilization on the cluster. Use the *Cisco Unified Communications Manager Capacity Tool* or the *Cisco Unified Collaboration Sizing Tool* to size the Unified CM servers for your solution.

Sizing Considerations for Unified Mobile Agent

Unified Mobile Agent requires the use of two CTI ports per contact center call. One CTI port controls the caller endpoint, and the other CTI port controls the selected agent endpoint. The actual RTP stream is between the two endpoints and is not bridged through these two CTI ports. However, there is additional call processing activity on Unified CM when setting up calls to mobile agents through these two CTI ports (when compared with setting up calls to local Unified CCE agents).

While mobile agents may essentially log in from any location (by using the agent desktop) where they have a high-quality broadband connection and a PSTN phone, they will still be associated logically with a particular Unified CCE Peripheral and Unified Communications Manager cluster, even if the Voice Gateway used to call the mobile agent is registered with a different cluster. The agent is associated with a particular peripheral and cannot migrate freely to other peripherals without some custom modifications.

For specific subscriber and cluster sizing for Unified CCE deployments, the Cisco Unified Communications Manager Capacity Tool or the Unified Collaboration Sizing Tool must be used. When sizing the cluster, input the maximum number of simultaneously logged-in mobile agents. In cases where the number of configured mobile agents is higher than the maximum number of simultaneous logged-in mobile agents, consider the pairs of CTI ports configured for mobile agents who are not logged in the Cisco Unified Communications Manager Capacity Tool by entering CTI ports type 1 with a BHCA and BHT of 0. This is similar to the method for taking into account local agent phones that are not logged in by using the CTI third-party controlled lines in the Cisco Unified Communications Manager Capacity Tool. As an alternative, or when using the Cisco Unified Collaboration Sizing Tool, you can input all mobile agents (logged-in and not logged-in) into the tool and adjust the BHCA and BHT per mobile agent accordingly. The total BHCA and BHT must remain the same as when considering simultaneous logged-in mobile agents with their actual BHCA and BHT.

Related Topics

[Cisco Unified Mobile Agent, on page 171](#)



CHAPTER 13

Bandwidth Provisioning and QoS Considerations

- [Bandwidth Provisioning and QoS Considerations for Unified CCE, page 237](#)
- [Unified CCE Network Architecture Overview, page 238](#)
- [Bandwidth and Latency Requirements, page 245](#)
- [Quality of Service, page 245](#)
- [Bandwidth Provisioning, page 252](#)
- [Outbound Option Bandwidth Provisioning and QoS Considerations, page 259](#)
- [Bandwidth Requirements and QoS for Agent and Supervisor Desktops, page 264](#)
- [Bandwidth Requirements for an Administration and Data Server and Reporting, page 267](#)
- [Bandwidth Requirements for Cisco EIM/WIM, page 267](#)
- [Bandwidth and Latency Requirements for the User List Tool, page 267](#)

Bandwidth Provisioning and QoS Considerations for Unified CCE

This chapter presents an overview of the Unified CCE network architecture, deployment characteristics of the network, and provisioning requirements of the Unified CCE network. Essential network architecture concepts are introduced, including network segments, keep-alive (heartbeat) traffic, flow categorization, IP-based prioritization and segmentation, and bandwidth and latency requirements. Provisioning guidelines are presented for network traffic flows over the WAN, including how to apply proper Quality of Service (QoS) to WAN traffic flows.

Generally, you deploy Unified CCE with private, point-to-point leased-line network connections for both its Private and Visible (public) WAN network structure. For optimal network performance characteristics (and route diversity for the fault-tolerant fail-overs), Unified CCE requires dedicated private facilities, redundant IP routers, and appropriate priority queuing.

Enterprises deploying networks that share multiple traffic classes prefer to maintain their existing infrastructure rather than revert to an incremental, dedicated network. Convergent networks offer both cost and operational efficiency, and such support is a key aspect of Cisco Powered Networks.

If your deployment meets the required latency and bandwidth requirements, you can deploy Unified CCE with a convergent QoS-aware public network and a convergent QoS-aware private network environment. This chapter presents QoS marking, queuing, and shaping guidelines for both the Unified CCE public and private network traffic.

Unified Contact Center Enterprise use the Differentiated Services (DiffServ) model for QoS. DiffServ categorizes traffic into different classes and applies specific forwarding treatments to the traffic class at each network node.

Adequate bandwidth provisioning and implementation of QoS are critical components in the success of Unified CCE deployments. Bandwidth guidelines and examples are provided in this chapter to help with provisioning the required bandwidth.

Related Topics

[Architecture Overview, on page 1](#)

Unified CCE Network Architecture Overview

Unified CCE is a distributed, resilient, and fault-tolerant network application that relies on its network infrastructure meeting real-time data transfer requirements. A properly designed Unified CCE network requires proper bandwidth, low latency, and a prioritization scheme favoring specific UDP and TCP application traffic. These design requirements ensure both the fault-tolerant message synchronization between redundant Unified CCE nodes. The requirements also ensure the delivery of time-sensitive system status data (routing messages, agent states, call statistics, trunk information, and so forth) across the system. Unified CCE requires prompt delivery of PG data to the Central Controller for accurate state updates and real-time reporting data.

In a Cisco Unified Communications deployment, WAN and LAN traffic can be grouped into the following categories:

- Voice and video traffic
Voice calls (voice carrier stream) consist of Real-Time Transport Protocol (RTP) packets that contain the actual voice samples between various endpoints such as PSTN gateway ports, Unified IP IVR Q-points (ports), and IP phones. This traffic includes voice streams of silently monitored and recorded agent calls.
- Call control traffic
Call control consists of packets belonging to one of several protocols (MGCP, SCCP, or TAPI/JTAPI), according to the endpoints involved in the call. Call control includes functions to set up, maintain, tear down, or redirect calls. For Unified CCE, control traffic includes routing and service control messages that route voice calls to peripheral targets (such as agents or services) and other media termination resources (such as Unified IP IVR ports). Control traffic also includes the real-time updates of peripheral resource status.
- Data traffic
Data traffic can include email, web activity, and CTI database application traffic sent to the agent desktops. Unified CCE priority data includes data for non-real-time system states, such as reporting and configuration update events.

This topic focuses primarily on the types of data flows and bandwidth used between the following:

- A remote Peripheral Gateway (PG) and the Unified CCE Central Controller (CC)
- Sides A and B of a PG or a CC

- The desktop application and the Finesse, CTI OS, or Cisco Agent Desktop servers

Guidelines and examples are presented to help estimate required bandwidth and to help implement a prioritization scheme for these WAN segments.

The flows discussed in this chapter encapsulate call control and data traffic. Because media (voice and video) streams are maintained primarily between Cisco Unified Communications Manager and its endpoints, voice and video provisioning are not addressed here.

For bandwidth estimates for the voice RTP stream generated by the calls to Unified CCE agents and the associated call control traffic generated by the various protocols, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/tsd-products-support-series-home.html>.

Data traffic and other mission-critical traffic varies according to the specific integration and deployment model used. For information about proper network design for data traffic, see the Network Infrastructure and Quality of Service (QoS) documentation at http://www.cisco.com/en/US/netsol/ns742/networking_solutions_program_category_home.html.

Network Segments

The fault-tolerant architecture employed by Unified CCE requires two independent communication networks. The private network (using a separate path) carries traffic necessary to maintain and restore synchronization between the systems and to allow clients of the Message Delivery Subsystem (MDS) to communicate. The public network carries traffic between each side of the synchronized system and foreign systems. The public network is also used as an alternate network by the fault-tolerance software to distinguish between node failures and network failures.

**Note**

The terms public network and visible network are used interchangeably throughout this document.

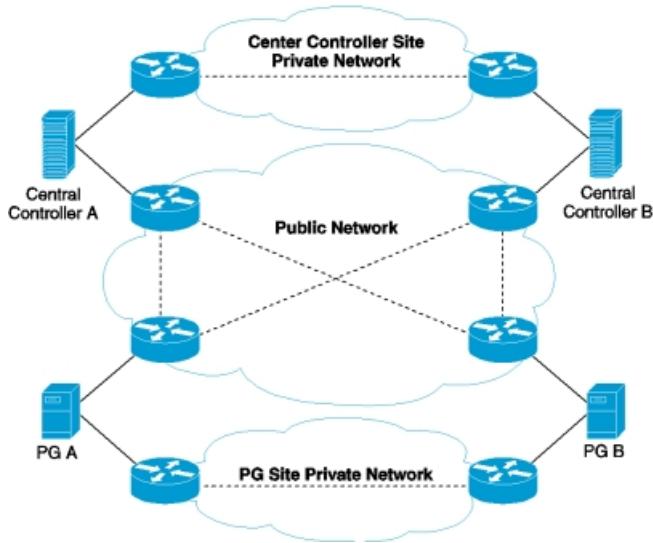
A third network, the signaling access network, may be deployed in Unified CCE systems that also interface directly with the carrier network (PSTN) and that deploy the Unified CCE architecture. The signaling access network is not addressed in this chapter.

**Note**

Cisco Unified CCH is deprecated. Use Cisco HCS for Contact Center instead.

The figure below illustrates the fundamental network segments for a Unified CCE system with a duplexed PG and a duplexed Central Controller (with Sides A and B geographically separated).

Figure 95: Example of Public and Private Network Segments for a Unified CCE System



The following notes apply to the figure above:

- The private network carries Unified CCE traffic between duplexed sides of the Central Controller or a Peripheral Gateway. This traffic consists primarily of synchronized data and control messages, and it also conveys the state transfer necessary to re-synchronize duplexed sides when recovering from an isolated state. When deployed over a WAN, the private network is critical to the overall responsiveness of Cisco Unified CCE. It must meet aggressive latency requirements and, therefore, either IP-based priority queuing or QoS must be used on the private network links.
- The public network carries traffic between the Central Controller and call centers (PGs and Administration & Data Servers). The public network can also serve as a Central Controller alternate path, used to determine which side of the Central Controller retains control if the two sides become isolated from one another. The public network is never used to carry synchronization control traffic. Public network WAN links must also have adequate bandwidth to support the PGs and Administration & Data Servers at the call center. The IP routers in the public network must use either IP-based priority queuing or QoS to ensure that Unified CCE traffic classes are processed within acceptable tolerances for both latency and jitter.
- Call centers (PGs and Administration & Data Servers) local to one side of the Central Controller connect to the local Central Controller side through the public Ethernet and to the remote Central Controller side over public WAN links. This arrangement requires that the public WAN network must provide connectivity between Side A and Side B. Bridges may optionally be deployed to isolate PGs and Administration & Data Servers from the Central Controller LAN segment to enhance protection against LAN outages.
- To achieve the required fault tolerance, the private WAN link must be fully independent from the public WAN links (separate IP routers, network segments or paths, and so forth). Independent WAN links ensure that a single point of failure is truly isolated between the public and the private networks. Deploy public network WAN segments that traverse a routed network so that you maintain PG-to-CC (Central Controller) connectivity.

Controller) route diversity throughout the network. Avoid routes that result in common path selection (and, thus, a common point of failure) for the multiple PG-to-CC sessions.

IP-Based Prioritization and QoS

For each of the WAN links in the public and private network segments of a Unified CCE deployment, a prioritization scheme is required. Two such prioritization schemes are supported: IP-based prioritization and QoS. Traffic prioritization is needed because it is possible for large amounts of low-priority traffic to get in front of high-priority traffic, thereby delaying delivery of high-priority packets to the receiving end. In a slow network flow, the amount of time a single large (for example, 1500-byte) packet consumes on the network (and delays subsequent packets) can exceed 100 ms. This delay would cause the apparent loss of one or more heartbeats. To avoid this situation, a smaller Maximum Transmission Unit (MTU) size is used by the application for low-priority traffic, thereby allowing a high-priority packet to get on the wire sooner. (MTU size for a circuit is calculated from within the application as a function of the circuit bandwidth, as configured at PG setup.)

A network that is not prioritized correctly almost always leads to call time-outs and problems from loss of heartbeats as the application load increases or (worse) as shared traffic is placed on the network. A secondary effect often seen is application buffer pool exhaustion on the sending side, due to extreme latency conditions.

Unified CCE applications use three priorities: high, medium, and low. However, prior to QoS, the network effectively recognized only two priorities identified by source and destination IP address (high-priority traffic was sent to a separate IP destination address) and, in the case of UDP heartbeats, by specific UDP port range in the network. The approach with IP-based prioritization is to configure IP routers with priority queuing in a way that gives preference to TCP packets with a high-priority IP address and to UDP heartbeats over the other traffic. When using this prioritization scheme, 90% of the total available bandwidth is granted to the high-priority queue.

A QoS-enabled network applies prioritized processing (queuing, scheduling, and policing) to packets based on QoS markings as opposed to IP addresses. Unified CCE provides a marking capability of Layer-3 DSCP for private and public network traffic. Traffic marking in Unified CCE implies that configuring dual IP addresses on each Network Interface Controller (NIC) is no longer necessary because the network is QoS-aware. However, if the traffic is marked at the network edge instead, dual-IP configuration is still required to differentiate packets by using access control lists based on IP addresses.

**Note**

**Note**

Layer-2 802.1p marking is also possible if Microsoft Windows Packet Scheduler is enabled (for PG/Central Controller traffic only). However, this is not supported. Microsoft Windows Packet Scheduler is not well supported or suited to Unified CCE, and support is removed in future versions. 802.1p markings are not widely used, nor are they required when DSCP markings are available.

Related Topics

[Where to Mark Traffic, on page 246](#)

UDP Heartbeat and TCP Keep-Alive

The primary purpose of the UDP heartbeat design is to detect if a circuit has failed. Detection can be made from either end of the connection, based on the direction of heartbeat loss. Both ends of a connection send heartbeats at periodic intervals (typically every 100 or 400 milliseconds) to the opposite end. Each end looks for analogous heartbeats from the other. If either end does not receive a heartbeat after five times the heartbeat period, that end assumes that something is wrong and the application closes the socket connection. At that point, a TCP Reset message is typically generated from the closing side. Various factors can cause loss of heartbeats, such as:

- The network failed.
- The process sending the heartbeats failed.
- The VM with the sending process is shut down.
- The UDP packets are not properly prioritized.

There are several parameters associated with heartbeats. In general, leave these parameters set to their system default values. Some of these values are specified when a connection is established, while others can be specified by setting values in the Microsoft Windows 2008 registry. The two values of most interest are:

- The amount of time between heartbeats
- The number of missed heartbeats (currently hard-coded as 5) that the system uses to determine whether a circuit has apparently failed

The default value for the heartbeat interval between redundant components is 100 milliseconds. One side can detect the failure of the circuit or the other side within 500 ms. The default heartbeat interval between a central site and a peripheral gateway is 400 ms. In this case, it takes 2 seconds to reach the circuit failure threshold.

As part of the Unified CCE QoS implementation, a TCP keep-alive message in the public network connecting a Central Controller to a Peripheral Gateway replaces the UDP heartbeat. A consistent heartbeat or keep-alive mechanism is enforced on public network interface whereas keep-alive mechanism is enforced on private network interface. When QoS is enabled on the network interface, a TCP keep-alive message is sent; otherwise UDP heartbeats are retained.

The TCP keep-alive feature, provided in the TCP stack, detects inactivity and then causes the server or client side to terminate. The TCP keep-alive feature sends probe packets (namely, keep-alive packets) across a connection after the connection has been idle for a certain period. The connection is considered down if a keep-alive response from the other side is not heard. Microsoft Windows 2012 allow you to specify keep-alive parameters on a per-connection basis. For Unified CCE public connections, the keep-alive timeout is set to 5 * 400 ms, matching the failure detection time of 2 seconds with the UDP heartbeat.

The reasons for moving to TCP keep-alive with QoS enabled are as follows:

- In a converged network, algorithms used by routers to handle network congestion conditions can have different effects on TCP and UDP. As a result, delays and congestion experienced by UDP heartbeat traffic can result in connection failures from timeouts.
- The use of UDP heartbeats creates deployment complexities in a firewall environment. With the dynamic port allocation for heartbeat communications, you open a large range of port numbers which weakens the security of your firewall.

HSRP-Enabled Network

In a network where Hot Standby Router Protocol (HSRP) is deployed on the default gateways that are configured on the Unified CCE servers, follow these requirements:

- Set the HSRP hold time and its associated processing delay lower than five times the heartbeat interval (100 ms on the private network and 400 ms on the public network). This level avoids Unified CCE private network communication outage during HSRP active router switch-over.

**Note**

With convergence delays that exceed private or public network outage notification, HSRP fail-over times can exceed the threshold for network outage detection which results in a fail-over. If the HSRP configuration has primary and secondary designations and the primary path router fails over, HSRP reinstates the primary path when possible. That reinstatement can lead to a second private network outage detection.

For this reason, do not use primary and secondary designations with HSRP convergence delays that approach 500 ms for the private network and 2 seconds for the public network. On the other hand, convergence delays below the detected threshold (which result in HSRP fail-overs that are transparent to the application) do not mandate a preferred path configuration. This approach is preferable. Keep enabled routers symmetrical if path values and costs are identical. However, if available bandwidth and cost favor one path (and the path transition is transparent), then designation of a primary path and router is advised.

- The Unified CCE fault-tolerant design requires the private network to be physically separate from the public network. Therefore, do not configure HSRP to fail-over one type of network traffic to the other network link.
- The bandwidth requirement for Unified CCE must be guaranteed at all times with HSRP, otherwise the system behavior is unpredictable. For example, if HSRP is initially configured for load sharing, ensure that sufficient bandwidth for Unified CCE remains on the surviving links in the worst-case failure situations.

RSVP

Cisco Unified Communications Manager provides support for Resource Reservation Protocol (RSVP) between endpoints within a cluster. As a protocol for call admission control, RSVP is used by the routers in the network to reserve bandwidth for calls.

RSVP traces the path between two RSVP agents that reside on the same LAN as the phones. The RSVP agent is a software media termination point (MTP) that runs on Cisco IOS routers. The RSVP agents are controlled by Unified Communications Manager and are inserted into the media stream between the two phones when a call is made. The RSVP agent of the originating phone will traverse the network to the RSVP agent of the destination phone, and reserve bandwidth. Since the network routers keep track of bandwidth usage instead of Unified Communications Manager, multiple phone calls can traverse the same RSVP controlled link even if the calls are controlled by multiple clusters.

You might use RSVP in a scenario in which two different clusters provide service to phones at the same remote site. This may occur if a cluster is assigned to handle an IP call center. In the scenario, two users at the same office are serviced by different clusters. RSVP offloads the bandwidth calculation responsibilities of Unified Communications Manager to the network routers.

For more information about Unified Communications Manager RSVP, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://cisco.com/go/ucsrnd>.

Traffic Flow

This section briefly describes the traffic flows for the public and private networks.

Public Network Traffic Flow

The active PG continuously updates the Central Controller call routers with state information related to agents, calls, queues, and so forth, at the respective call center sites. This type of PG-to-Central Controller traffic is real-time traffic. The PGs also send up historical data at each 15-minute or half-hour interval based on configuration of the PG. The historical data is low priority, but it must complete its journey to the central site before the start of the next interval (to get ready for the next half hour of data).

When a PG starts, its configuration data is supplied from the central site so that it can know which agents, trunks, and so forth, it has to monitor. This configuration download can be a significant network bandwidth transient.

In summary, traffic flows from PG to Central Controller can be classified into the following distinct flows:

- High-priority traffic — Includes routing and Device Management Protocol (DMP) control traffic. It is sent in TCP with the public high-priority IP address.
- Heartbeat traffic — UDP messages with the public high-priority IP address and in the port range of 39500 to 39999. Heartbeats are transmitted at 400-ms intervals bi-directionally between the PG and the Central Controller. The UDP heartbeat is replaced with TCP keep-alive if QoS is enabled on the public network interface through the Unified CCE setup.
- Medium-priority traffic — Includes real-time traffic and configuration requests from the PG to the Central Controller. The medium-priority traffic is sent in TCP with the public high-priority IP address.
- Low-priority traffic — Includes historical data traffic, configuration traffic from the Central Controller, and call close notifications. The low-priority traffic is sent in TCP with the public non-high-priority IP address.

Private Network Traffic Flow

Traffic destined for the critical Message Delivery Service (MDS) client (Router or OPC) is copied to the other side over the private link.

The private traffic can be summarized as follows:

- High-priority traffic — Includes routing, MDS control traffic, and other traffic from MDS client processes such as the PIM CTI Server, Logger, and so forth. It is sent in TCP with the private high-priority IP address.
- Heartbeat traffic — UDP messages with the private high-priority IP address and in the port range of 39500 to 39999. Heartbeats are transmitted at 100-ms intervals bi-directionally between the duplexed sides. The UDP heartbeat is replaced with TCP keep-alive if QoS is enabled on the private network interface through the Unified CCE setup.

- Medium-priority and low-priority traffic — For the Central Controller, this traffic includes shared data sourced from routing clients as well as (non-route control) call router messages, including call router state transfer (independent session). For the OPC (PG), this traffic includes shared non-route control peripheral and reporting traffic. This class of traffic is sent in TCP sessions designated as medium priority and low priority, respectively, with the private non-high priority IP address.
- State transfer traffic — State synchronization messages for the Router, OPC, and other synchronized processes. It is sent in TCP with a private non-high-priority IP address.

Bandwidth and Latency Requirements

The amount of traffic sent between the Central Controllers (call routers) and Peripheral Gateways is largely a function of the call load at that site, although transient boundary conditions (for example, startup configuration load) and specific configuration sizes also affect the amount of traffic. Bandwidth calculators and sizing formulas can project bandwidth requirements far more accurately. See the section on private and visible network bandwidth requirements for more details.

A site that has an ACD as well as a VRU has two peripherals, and the bandwidth requirement calculations need to take both peripherals into account. As an example, a site that has four peripherals, each taking 10 calls per second, will generally be configured to have 320 kbps of bandwidth. The 1000 bytes per call is a rule of thumb, but monitor the actual behavior once the system is operational to ensure that enough bandwidth exists. (Unified CCE meters data transmission statistics at both the Central Controller and PG sides of each path.)

As with bandwidth, specific latency requirements must be guaranteed for Unified CCE to function as designed. The side-to-side private network of duplexed Central Controller and PG nodes has a maximum one-way latency of 100 ms (50 ms preferred). The PG-to-CC path has a maximum one-way latency of 200 ms to perform as designed. Meeting or exceeding these latency requirements is particularly important in an environment using Unified CCE post-routing and/or translation routes.

As discussed previously, Unified CCE bandwidth and latency design is fully dependent on an underlying IP prioritization scheme. Without proper prioritization in place, WAN connections will fail. The Cisco Unified CCE support team has custom tools (for example, Client/Server) that can be used to demonstrate proper prioritization and to perform some level of bandwidth utilization modeling for deployment certification.

Depending on the final network design, an IP queuing strategy is required in a shared network environment to achieve Unified CCE traffic prioritization concurrent with other non-DNP traffic flows. This queuing strategy is fully dependent on traffic profiles and bandwidth availability, and success in a shared network cannot be guaranteed unless the stringent bandwidth, latency, and prioritization requirements of the product are met.

In general, Agent Greeting feature requires shorter latency cross system. For example, the PG-to-CC path has a maximum one-way latency of 50 ms to support Agent Greeting feature as designed.

Quality of Service

This section covers the planning and configuration issues to consider when moving to a Unified CCE QoS solution.

Where to Mark Traffic

In planning QoS, a question often arises about whether to mark traffic in Unified CCE or at the network edge. Each option has its pros and cons. Marking traffic in Unified CCE saves the access lists for classifying traffic in IP routers and switches.


Note

While Cisco allows Microsoft Packet Scheduler with Unified CCE 8.5, it is not supported and future releases will remove this option.

There are several disadvantages to marking traffic in Unified CCE. First, it is hard to make changes. For instance, to change the marking values for the public network traffic, you have to make changes on all the PGs. For a system with more than 30 PGs, for example, all those changes would require quite a lot of work. Second, QoS trust has to be enabled on access-layer routers and switches, which could open the network to malicious packets with inflated marking levels.


Note

In Windows, you can use the Group Policy Editor to apply a QoS policy to apply DSCP Level 3 markings to packets. You can also administer these policies through the Active Directory Domain Controller. This may simplify the administration issue. For more information, see appropriate Microsoft documentation.

In contrast, marking traffic at the network edge allows for centralized and secured marking policy management, and there is no need to enable trust on access-layer devices. A little overhead is needed to define access lists to recognize Unified CCE packets. For access-list definition criteria on edge routers or switches, see [Table 23: Public Network Traffic Markings \(Default\) and Latency Requirements, on page 247](#), [Table 24: Router Private Network Traffic Markings \(Default\) and Latency Requirements, on page 248](#), and [Table 25: PG Private Network Traffic Markings \(Default\) and Latency Requirements, on page 248](#). Do not use port numbers in the access lists for recognizing Unified CCE traffic (although they are provided in the tables for reference purposes) because port numbers make the access lists extremely complex and you would have to modify the access lists every time a new customer instance is added to the system.


Note

A typical Unified CCE deployment has three IP addresses configured on each virtual NIC, and the Unified CCE application uses two of them. For remote monitoring using PCAnywhere or VNC, because the port numbers are not used in the access lists, use the third IP address to prevent the remote monitoring traffic from being marked as the real Unified CCE traffic.

Related Topics

[How to Mark Traffic, on page 246](#)

How to Mark Traffic

The default Unified CCE QoS markings can be overwritten if necessary. The tables below show the default markings, latency requirement, IP address, and port associated with each priority flow for the public and private network traffic respectively, where $i\#$ stands for the customer instance number. Notice that in the public network the medium-priority traffic is sent with the high-priority public IP address and marked the

same as the high-priority traffic, while in the private network it is sent with the non-high-priority private IP address and marked the same as the low-priority traffic.

For details about Cisco Unified Communications packet classifications, see the *Cisco Collaboration System Solution Reference Network Designs* at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.


Note

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

Table 23: Public Network Traffic Markings (Default) and Latency Requirements

Priority	Server-Side IP Address and Port	One-Way Latency Requirement	DSCP / 802.1p Marking
High	IP address: Router's high-priority public IP address TCP port: <ul style="list-style-type: none">• 40003 + (i# * 40) for DMP high-priority connection on A• 41003 + (i# * 40) for DMP high-priority connection on B UDP port: 39500 to 39999 for UDP heartbeats if QoS is not enabled on Unified CCE	200 ms	AF31 / 3
Medium	IP address: Router's high-priority public IP address TCP port: <ul style="list-style-type: none">• 40017 + (i# * 40) for DMP high-priority connection on A• 41017 + (i# * 40) for DMP high-priority connection on B	1000 ms	AF31 / 3
Low	IP address: Router's non-high-priority public IP address TCP port: <ul style="list-style-type: none">• 40002 + (i# * 40) for DMP low-priority connection on A• 41002 + (i# * 40) for DMP low-priority connection on B	5 seconds	AF11 / 1

Table 24: Router Private Network Traffic Markings (Default) and Latency Requirements

Priority	Server-Side IP Address and Port	One-Way Latency Requirement	DSCP / 802.1p Marking
High	IP address: Router's high-priority private IP address TCP port: $41005 + (i\# * 40)$ for MDS high-priority connection	100 ms (50 ms preferred)	AF31 / 3
Medium	IP address: Router's non-high-priority private IP address TCP port: $41016 + (i\# * 40)$ for MDS medium-priority connection	1000 ms	AF11/1
Low	IP address: Router's non-high-priority private IP address TCP port: <ul style="list-style-type: none">• $41004 + (i\# * 40)$ for MDS low-priority connection• $41022 + (i\# * 40)$ for CIC StateXfer connection• $41021 + (i\# * 40)$ for CLGR StateXfer connection• $41023 + (i\# * 40)$ for HLGR StateXfer connection• $41020 + (i\# * 40)$ for RTR StateXfer connection	1000 ms	AF11/1

Table 25: PG Private Network Traffic Markings (Default) and Latency Requirements

Priority	Server-Side IP Address and Port	One-Way Latency Requirement	DSCP / 802.1p Marking
High	IP address: PG high-priority private IP address TCP port: <ul style="list-style-type: none">• $43005 + (i\# * 40)$ for MDS high-priority connection of PG no.1• $45005 + (i\# * 40)$ for MDS high-priority connection of PG no.2	100 ms (50 ms preferred)	AF31/3
Medium	IP address: PG's non-high-priority private IP address TCP port: <ul style="list-style-type: none">• $43016 + (i\# * 40)$ for MDS medium-priority connection of PG no.1• $45016 + (i\# * 40)$ for MDS medium-priority connection of PG no.2	1000 ms	AF11/1

Priority	Server-Side IP Address and Port	One-Way Latency Requirement	DSCP / 802.1p Marking
Low	IP address: PG's non-high-priority private IP address TCP port: • $43004 + (i\# * 40)$ for MDS low-priority connection of PG no.1 • $45004 + (i\# * 40)$ for MDS low-priority connection of PG no.2 • $3023 + (i\# * 40)$ for OPC StateXfer of PG no.1 • $45023 + (i\# * 40)$ for OPC StateXfer of PG no.2	1000 ms	AF11/1

QoS Configuration

This section presents some QoS configuration examples for the various devices in a Unified CCE system.

QoS Enablement in Unified CCE

QoS is enabled by default on private network traffic.

- Disable QoS for the Visible (public) network traffic. For most deployments, disabling QoS for the Visible network traffic ensures timely failover handling.
You can add QoS markings outside the contact center applications with a Windows Group Policy or by enabling marking on the IP edge routers.

For information about enabling QoS on the router during install, see the *Cisco Unified Contact Center Enterprise Installation and Upgrade Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

QoS Configuration on Cisco IOS Devices

This section presents some representative QoS configuration examples. For details about campus network design, switch selection, and QoS configuration commands, see the [Enterprise QoS Solution Reference Network Design \(SRND\)](#).



Note

The marking value, bandwidth data, and queuing policy in the examples below are provided for demonstration purpose only. Do not copy and paste the examples without making corresponding changes in the real working system.

Configuring 802.1q Trunks on IP Switches

If 802.1p is an intended feature and the 802.1p tagging is enabled on the NIC for the visible network, the switch port into which the Unified CCE server plugs must be configured as an 802.1q trunk, as illustrated in the following configuration example:

```
switchport mode trunk
switchport trunk encapsulation dot1q
switchport trunk native vlan [data/native VLAN #]
switchport voice vlan [voice VLAN #]
switchport priority-extend trust
spanning-tree portfast
```

Configuring QoS Trust

Assuming Unified CCE DSCP markings are trusted, the following commands enable trust on an IP switch port:

```
mls qos
  interface mod/port
    mls qos trust dscp
```

Configuring Queuing Policy to Act on Marked Traffic

Using the public (visible) network as an example, the class map below identifies two marking levels, AF31 for high-priority traffic (which actually includes medium-priority public network traffic because it is marked the same as the high-priority traffic by default) and AF11 for low-priority traffic:

```
class-map match-all Unified ICM_Public_High
  match ip dscp af31
class-map match-all ICM_Public_Low
  match ip dscp af11
```

If the link is dedicated to Unified CCE Public traffic only, the policy map puts ICM_Public_High traffic into the priority queue with the minimum and maximum bandwidth guarantee of 500 kbps, and it puts ICM_Public_Low traffic into the normal queue with a minimum bandwidth of 250 kbps:

```
policy-map ICM_Public_Queueing
  class ICM_Public_High
    priority 500
  class ICM_Public_Low
    bandwidth 250
```

You can also use the commands **priority percent** and **bandwidth percent** to assign bandwidth on a percentage basis. Assign 90% of the link bandwidth to the priority queue.

If it is a shared link, then use the sizing tools introduced in the section on Bandwidth Provisioning, to calculate the bandwidth requirement at each priority level and add it to the allocation for non-CCE traffic in the same queue. For example, if the link is shared with Unified CM ICCS traffic and RTP traffic and they respectively require 600 kbps and 400 kbps, and if the link also carries the private traffic in case of fail-over and the high-priority and low-priority private Unified CCE traffic respectively require 200 kbps and 100 kbps, the configuration is:

```
policy-map Converged_Link_Queueing
  class RTP
    priority 400
```

```

class ICCS
  bandwidth 600
class ICM_Public_High
  bandwidth 500
class ICM_Public_Low
  bandwidth 250
class ICM_Private_High
  bandwidth 200
class ICM_Private_Low
  bandwidth 100

```

You can also use the commands `priority percent` and `bandwidth percent` to assign bandwidth on a percentage basis. If the link is dedicated to Unified CCE traffic only, assign 90% of the link bandwidth to the priority queue. If it is a shared link, use the sizing tools to calculate the bandwidth requirement at each priority level and add it to the allocation for non-CCE traffic in the same queue.

Finally, the queuing policy is applied to the outgoing interface:

```

interface mod/port
  service-policy output ICM_Public_Queueing

```

Configuring Marking Policy to Mark Traffic

As discussed earlier, rather than marking traffic in Unified CCE, another option is to mark traffic at the network edge. First, define access lists to recognize Unified CCE traffic flows:

```

access-list 100 permit tcp host Public_High_IP any
access-list 100 permit tcp any host Public_High_IP
access-list 101 permit tcp host Public_NonHigh_IP any
access-list 101 permit tcp any host Public_NonHigh_IP

```

Second, classify the traffic using a class map:

```

class-map match-all ICM_Public_High
  match access-group 100
class-map match-all ICM_Public_Low
  match access-group 101

```

Third, define the marking policy using a policy map:

```

policy-map ICM_Public_Marking
  class ICM_Public_High
    set ip dscp af31
  class ICM_Public_Low
    set ip dscp af11

```

Finally, apply the marking policy to the incoming interface:

```

interface mod/port
  service-policy input ICM_Public_Marking

```

Related Topics

[Bandwidth Provisioning, on page 252](#)

QoS Performance Monitoring

Once the QoS-enabled processes are up and running, the Microsoft Windows Performance Monitor (PerfMon) can be used to track the performance counters associated with the underlying links. For details on using PerfMon, see the Microsoft documentation.

Bandwidth Provisioning

This section discusses bandwidth provisioning considerations for the Unified CCE system.

Bandwidth Requirements for Unified CCE Public and Private Networks

This section briefly describes bandwidth sizing for the public (visible) and private networks.

Public Network Bandwidth

Special tools are available to help calculate the bandwidth needed for the following public network links:

- Unified CCE Central Controller to Unified CM PG
A tool is accessible to Cisco partners and Cisco employees for computing the bandwidth needed between the Unified CCE Central Controller and Unified CM. This tool is called the ACD/CallManager Peripheral Gateway to Unified CCE Central Controller Bandwidth Calculator, and it is available (with proper login authentication) through the [Steps to Success Portal](#).
- Unified CCE Central Controller to Unified IP IVR or Unified CVP PG
A tool is accessible to Cisco partners and Cisco employees for computing the bandwidth needed between the Unified CCE Central Controller and the IP IVR PG. This tool is called the VRU Peripheral Gateway to Unified Central Controller Bandwidth Calculator, and it is also available through the [Steps to Success Portal](#).

At this time, no tool exists that specifically addresses communications between the Unified CCE Central Controller and the Cisco Unified Customer Voice Portal (Unified CVP) PG. Testing has shown, however, that the tool for calculating bandwidth needed between the Unified CCE Central Controller and the Unified IP IVR PG will also produce accurate measurements for Unified CVP if you perform the following substitution in one field:

For the field labeled **Average number of RUN VRU script nodes**, substitute the number of Unified CCE script nodes that interact with Unified CVP.

Private Network Bandwidth

The following table is a worksheet to assist with computing the link and queue sizes for the private network. Definitions and examples follow the table.



Note Minimum link size in all cases is 1.5 Mbps (T1).

Table 26: Worksheet for Calculating Private Network Bandwidth

Component	Effective BHCA	Multiplication Factor	Calculated Link	Multiplication Factor	Calculated Queue	
Router + Logger		* 30		* 0.8		Total Router + Logger High-Priority Queue Bandwidth
Unified CM PG		* 100		* 0.9		Add these numbers together and total in the box below to get the PG High-Priority Queue Bandwidth
Unified IP IVR PG		* 60		* 0.9		
Unified CVP PG		* 120		* 0.9		
Unified IP IVR or Unified CVP Variables		* ((Number of Variables * Average Variable Length)/40)		* 0.9		
		Total Link Size				Total PG High-Priority Queue Bandwidth

If one dedicated link is used between sites for private communications, add all link sizes together and use the Total Link Size at the bottom of the table above. If separate links are used, one for Router/Logger Private and one for PG Private, use the first row for Router/Logger requirements and the bottom three (out of four) rows added together for PG Private requirements.

Effective BHCA (effective load) on all similar components that are split across the WAN is defined as follows:

Router + Logger

This value is the total BHCA on the call center, including conferences and transfers. For example, 10,000 BHCA ingress with 10% conferences or transfers are 11,000 effective BHCA.

Unified CM PG

This value includes all calls that come through Unified CCE Route Points controlled by Unified CM and/or that are ultimately transferred to agents. This assumes that each call comes into a route point and is eventually sent to an agent. For example, 10,000 BHCA ingress calls coming into a route point and being transferred to agents, with 10% conferences or transfers, are 11,000 effective BHCA.

Unified IP IVR PG

This value is the total BHCA for call treatment and queuing. For example, 10,000 BHCA ingress calls, with all of them receiving treatment and 40% being queued, are 14,000 effective BHCA.

Unified CVP PG

This value is the total BHCA for call treatment and queuing coming through a Unified CVP. 100% treatment is assumed in the calculation. For example, 10,000 BHCA ingress calls, with all of them receiving treatment and 40% being queued, are 14,000 effective BHCA.

Unified IP IVR or Unified CVP Variables

This value represents the number of Call and ECC variables and the variable lengths associated with all calls routed through the Unified IP IVR or Unified CVP, whichever technology is used in the implementation.

Example of a Private Bandwidth Calculation

The table below shows an example calculation for a combined dedicated private link with the following characteristics:

- BHCA coming into the contact center is 10,000.
- 100% of calls are treated by Unified IP IVR and 40% are queued.
- All calls are sent to agents unless abandoned. 10% of calls to agents are transfers or conferences.
- There are four Unified IP IVRs used to treat and queue the calls, with one PG pair supporting them.
- There is one Unified CM PG pair for a total of 900 agents.
- Calls have ten 40-byte Call Variables and ten 40-byte ECC variables.

Table 27: Example Calculation for a Combined Dedicated Private Link

Component	Effective BHCA	Multiplication Factor	Calculated Link	Multiplication Factor	Calculated Queue	
Router + Logger	11,000	* 30	330,000	* 0.8	264,000	Total Router + Logger High-Priority Queue Bandwidth
Unified CM PG	11,000	* 100	1,100,000	* 0.9	990,000	Add these three numbers together and total in the box below to get the PG High-Priority Queue Bandwidth
Unified IP IVR PG	14,000	* 60	840,000	* 0.9	756,000	
Unified CVP PG	0	* 120	0	* 0.9	0	
Unified IP IVR or Unified CVP Variables	14,000	* ((Number of Variables * Average Variable Length)/40)	280,000	* 0.9	252,000	

Component	Effective BHCA	Multiplication Factor	Calculated Link	Multiplication Factor	Calculated Queue	
		Total Link Size	2,550,000		1,998,000	Total PG High-Priority Queue Bandwidth

For the combined dedicated link in this example, the results are as follows:

- Total Link Size = 2,550,000 bps
- Router/Logger high-priority bandwidth queue of 264,000 bps
- PG high-priority queue bandwidth of 1,998,000 bps

If this example were implemented with two separate links, Router/Logger private and PG private, the link sizes and queues are as follows:

- Router/Logger link of 330,000 bps (actual minimum link is 1.5 Mb, as defined earlier), with high-priority bandwidth queue of 264,000 bps
- PG link of 2,220,000 bps, with high-priority bandwidth queue of 1,998,000 bps

When using Multilink Point-to-Point Protocol (MLPPP) for private networks, set the following attributes for the MLPPP link:

- Use per-destination load balancing instead of per-packet load balancing.



Note You must have two separate multilinks with one link each for per-destination load balancing.

- Enable Point-to-Point Protocol (PPP) fragmentation to reduce serialization delay.

Bandwidth Requirements for Clustering over WAN

Bandwidth must be guaranteed across the highly available (HA) WAN for all Unified CCE private, public, CTI, and Unified Communications Manager intracluster communication signaling (ICCS). Moreover, bandwidth must be guaranteed for any calls going across the highly available WAN. Minimum total bandwidth required across the highly available WAN for all Unified CCE signaling is 2 Mbps.

This section covers the bandwidth requirements for the private and public networks. This section also discusses bandwidth analysis for the connections from Unified IP IVR or Unified CVP PG to Unified IP IVR or Unified CVP, CTI Server to CTI OS, and Unified Communications Manager intracluster communication signaling (ICCS).

Unified IP IVR or Unified CVP PG to Unified IP IVR or Unified CVP

Currently, no tool exists that specifically addresses communication between the Unified IP IVR or Unified CVP PG and the Unified IP IVR or Unified CVP. However, the tool mentioned in the previous section produces a fairly accurate measurement of the needed bandwidth. The bandwidth consumed between the Unified CCE

Bandwidth Requirements for Clustering over WAN

Central Controller and Unified IP IVR or Unified CVP PG is similar to the bandwidth consumed between the Unified IP IVR or Unified CVP PG and the Unified IP IVR or Unified CVP.

The *VRU Peripheral Gateway to Unified CCE Central Controller Bandwidth Calculator* tool is available (with proper logon authentication) through the [Cisco Steps to Success Portal](#).

If the Unified IP IVR or Unified CVP PGs are split across the WAN, the total bandwidth required is double what the tool reports: once for the Central Controller to the PG and once for the PG to Unified IP IVR or Unified CVP.

CTI Server to CTI OS

The worst case for bandwidth utilization across the WAN link between the CTI OS and CTI Server occurs when the CTI OS is remote from the CTI Server. To guarantee availability for this worst case, use a bandwidth queue.

For this model, the following simple formula can be used to compute worst-case bandwidth requirements:

- With no Expanded Call Context (ECC) or Call Variables:
BHCA * 20 = bps
- With ECC and Call Variables
 $BHCA * (20 + ((\text{Number of Variables} * \text{Average Variable Length}) / 40)) = \text{bps}$

Example: With 10,000 BHCA and 20 ECC variables with average length of 40 bits:
 $10,000 * (20 + ((20 * 40) / 40)) = 10,000 * 40 = 400,000 \text{ bps} = 400 \text{ kbps}$

CTI Server to Cisco Finesse

To determine the bandwidth required where Cisco Finesse connects to the CTI server over a WAN link, use the Finesse Bandwidth Calculator at <http://www.cisco.com/c/en/us/support/customer-collaboration/finesse-products-technical-reference-list.html>.

Unified Communications Manager Intracluster Communication Signaling (ICCS)

The bandwidth required for Intracluster Communication Signaling (ICCS) between subscribers is higher when Unified CCE is deployed. Unified CCE requires more call redirects and extra CTI/JTAPI communications for the intracluster communications. Use the following formulae to calculate the required bandwidth for the ICCS and database traffic between subscribers in Unified CCE:

- **Intracluster Communications Signaling (ICCS)**

Total Bandwidth (Mbps) = $(\text{Total BHCA}) / 10,000 * [1 + (0.006 * \text{Delay})]$, where Delay = Round-trip-time delay in msec

This value is the bandwidth required between each Unified CM subscriber that is connected to Voice Gateways, agent phones, and Agent PGs. The minimum value for this link is 1.544 Mbps.



Note This formula assumes a BHCA of 10,000 or more. For a BHCA of less than 10,000, use the minimum of 1.544 Mbps.

- **Database and other communications**

1.544 Mbps for each subscriber remote from the publisher

The BHCA value to use for this ICCS formula is the total BHCA for all calls coming into the contact center.

- **CTI ICCS**

$$\text{Bandwidth (Mbps)} = (\text{Total BHCA}/10,000) * 0.53$$

These bandwidth requirements assume proper design and deployment. Inefficient design (for example, if ingress calls to Site 1 are treated in Site 2) causes more intracluster communications, possibly exceeding the defined bandwidth requirements.

For more information on Unified Communications Manager Signaling, see the section on the local failover deployment model in the *Cisco Collaboration System Solution Reference Network Designs* at http://www.cisco.com/c/en/us/td/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

Related Topics

[Geographically Redundant Data Centers, on page 55](#)

Bandwidth Requirements for Finesse Client to Finesse Server

The most expensive operation from a network perspective is the agent or supervisor login. This operation involves the web page load and includes the CTI login and the display of the initial agent state. After the desktop web page loads, the required bandwidth is significantly less.

Because Cisco Finesse is a web application, caching can significantly impact the required bandwidth. For example, the first time an agent logs in, the number of bytes transmitted is approximately 2 megabytes. If caching is enabled in the browser, during subsequent logins, the number of bytes transmitted is 138 kilobytes.

Because of the additional gadgets on the supervisor desktop (Team Performance, Queue Statistics), this number is higher for a supervisor login – approximately 2.5 megabytes without caching and 333 kilobytes with caching. To minimize the amount of bandwidth required for login, make sure that caching is enabled in the browser.

Cisco does not mandate a minimum bandwidth for the login operations. You must determine how long you want the login to take and determine the required bandwidth accordingly. To help you with this calculation, Cisco Finesse provides a bandwidth calculator (at <http://www.cisco.com/c/en/us/support/customer-collaboration/finesse/products-technical-reference-list.html>) to estimate the bandwidth required to accommodate the client login time.

Note that during failover, agents are redirected to the alternate Finesse server and required to log in again. For example, if you configure your bandwidth so that login takes 5 minutes and a client failover event occurs, agents will take 5 minutes to successfully log in to the alternate Finesse server.

After login is complete, the most intensive operation for both an agent and a supervisor is making an outbound call to a route point. For the supervisor, updates to the Team Performance and Queue Statistics gadgets may be occurring concurrently. You can use the Cisco Finesse bandwidth calculator to calculate the total bandwidth required for connections between all Finesse clients and the Finesse server.



Note

The Cisco Finesse bandwidth calculator does not include the bandwidth required for any third-party gadgets in the Finesse container or any other applications running on the agent desktop client.

Other applications at the remote client location may compete for total bandwidth to that remote client.

The bandwidth listed in the bandwidth calculator must be available for Finesse after you account for the bandwidth used by other applications, including voice traffic that may share this bandwidth. The performance

of the Finesse interface, and potentially the quality of voice sharing this bandwidth, may degrade if sufficient bandwidth is not continuously available.

Auto Configuration

If auto configuration is used, the child PG might send the entire agent, skill group, and route-point configuration to the parent PG. If not much bandwidth is available, it could take considerable time for this data to be sent.

This table lists the approximate number of bytes (worst case) that are sent for each of the data entities. If you know the size of the configuration on a child PG, you can calculate the total number of bytes of configuration data that is sent. The values in the table are worse-case estimates that assume sending only one item per record, with each field having the maximum size (which is unlikely).

Table 28: Bytes Sent Per Data Item Under Worst-Case Conditions

Data Item sent	Size
Agent	500 bytes
Call type	250 bytes
Skill group	625 bytes
Device (route point, and so forth)	315 bytes

For example, if the child PG has 100 agents, 10 call types, 5 skill groups, and 20 route points, then you can estimate the amount of configuration data sent as follows:

$$100 \text{ agents} * 500 \text{ bytes} = 50,000 \text{ bytes}$$

$$10 \text{ call types} * 250 \text{ bytes} = 2500 \text{ bytes}$$

$$5 \text{ skill groups} * 625 \text{ bytes} = 3125 \text{ bytes}$$

$$20 \text{ route points} * 315 \text{ bytes} = 6300 \text{ bytes}$$

$$50,000 + 2500 + 3125 + 6300 = 61,925 \text{ bytes}$$

The total amount of data (approximate maximum) sent for this configuration is 61,925 bytes.

Options for Gateway PG and Unified CCE

To mitigate the bandwidth demands, use any combination of the following options:

- Use fewer call and ECC variables on the child PG.

Certain messages transmit call data from the child Unified CCE system to the parent. Reducing the size and quantity of variables used will reduce the data transmitted for these events.

**Note**

Call variables used on the child PG are transmitted to the parent PG regardless of their use or the setting of the MAPVAR parameter. For example, if call variables 1 through 8 are used on the child PG but are never referenced on the parent PG (and assume MAPVAR = EEEEEEEEEE, meaning Export all but Import nothing), they will still be transmitted to the PG where the filtering takes place, therefore bandwidth is still required. For the reverse situation, bandwidth is spared. For example, if the map setting is MAPVAR = IIIIIIIIII (Import all but Export nothing), then bandwidth is spared. Call variable data will not be transmitted to the child PG on a ROUTE_SELECT response.

- Use the MAPVAR = IIIIIIIIII and MAPECC = IIIIIIIIII peripheral configuration parameters. If you do not use the MAPVAR and MAPECC option (which means that the settings default to MAPVAR = BBBB BBBB BBBB and MAPECC = BBBB BBBB BBBB), then for every ROUTE_SELECT sent to the child, all Call and ECC variables used on the parent are also sent to the child. If you use the I (Import) or N (None) option for MAPVAR, MAPECC, or both, then the Gateway PG does not send these variables over the line to the child system. If a lot of call variables and/or ECC variables are used on the parent, these parameter settings can save some bandwidth.

**Note**

Eliminating Import (I or B setting) of data does not save any bandwidth because, even though the Gateway PG does not import the data, the child Unified CCE system still transmits it.

Outbound Option Bandwidth Provisioning and QoS Considerations

In many Outbound Option deployments, all components are centralized; therefore, there is no WAN network traffic to consider.

For some deployments, if the outbound call center is in one country (for example, India) and the customers are in another country (for example, US), then the WAN network structure must be considered in a Unified CCE environment under the following conditions:

- In a distributed Outbound Option deployment, when the Voice Gateways are separated from the Outbound Option Dialer servers by a WAN.
- When using Unified CVP deployments for transfer to a VRU campaign, and the Unified CVP servers are separated from the Outbound Option Dialer servers by a WAN. Provide Unified CVP with its own Cisco Unified SIP Proxy server in the local cluster to reduce the WAN traffic.
- When using Unified IP IVR deployments for transfer to a VRU campaign, and the Unified IP IVR is separated from the Outbound Option Dialer servers by a WAN. Provide Unified IP IVR with its own Unified CM cluster to reduce the WAN traffic.
- When deploying a SIP Dialer solution for transfer to a VRU campaign, and the Cisco Unified SIP Proxy servers for the SIP Dialers are separated from the Outbound Option Dialer servers by a WAN. Configure the recording server local to the Voice Gateways.
- When the third-party recording server is separated from the Outbound Option Dialer servers by a WAN. Configure the recording server local to the Voice Gateways.

Adequate bandwidth provisioning is an important component in the success of the Outbound Option deployments.

Distributed SIP Dialer Deployment

SIP is a text-based protocol; therefore, the packets used are larger than some protocols. The typical SIP outbound call flow uses an average of 12,500 bytes per call that is transferred to an outbound agent. The average hit call signaling bandwidth usage is:

$$\text{Hit Call Signaling Bandwidth} = (12,500 \text{ bytes/call}) (8 \text{ bits/byte}) = 100,000 \text{ bits per call} = 100 \text{ Kb per call}$$

The typical SIP outbound call flow uses about 6,200 bytes per call that is disconnected by the outbound dialer. Those outbound calls can be the result of a busy ring no-answer, an invalid number, and so forth. The average non-hit call signaling bandwidth usage is:

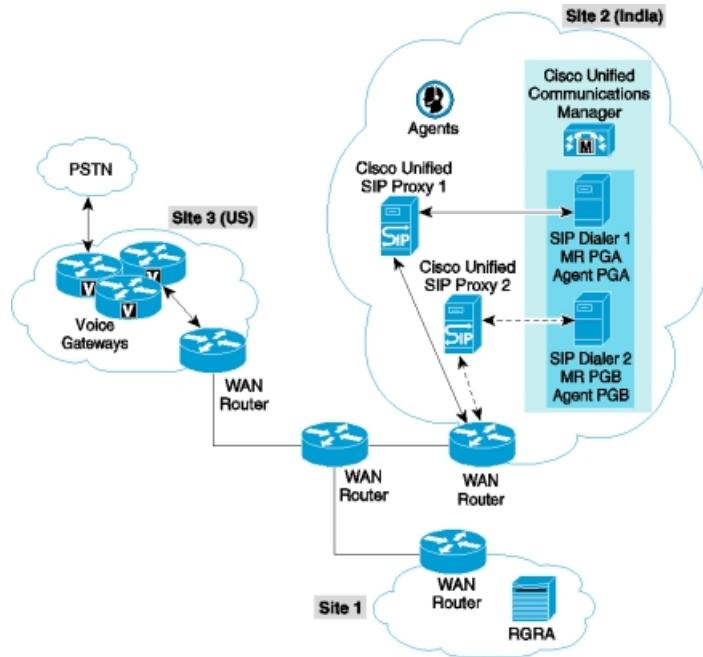
$$\text{Non-Hit Signaling Call Bandwidth} = (6,200 \text{ bytes/call}) (8 \text{ bits/byte}) = 49,600 \text{ bits per call} = 49.6 \text{ Kb per call}$$

Codec Bandwidth = 80 Kbps per call for g.711 Codec, or 26 Kbps per call for g.729 Codec

Agent-Based Campaign – No SIP Dialer Recording

This figure shows an example of the distributed Outbound SIP Dialer deployment for an agent-based campaign.

Figure 96: Distributed Outbound SIP Dialer Deployment for an Agent-Based Campaign



The average WAN bandwidth usage in this case is:

$$\text{WAN Bandwidth} = \text{Calls Per Second} * (\text{Hit Rate} * (\text{Codec Bandwidth} * \text{Average Call Duration} + \text{Hit Call Signaling Bandwidth}) + (1 - \text{Hit Rate}) * \text{Non-Hit Call Signaling Bandwidth}) = \text{Kbps}$$

Example 1

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the agent-based campaign, and a WAN link with g.711 codec and average call duration of 40 seconds, the bandwidth usage is:

$$60 * (20\% * (80 * 40 + 100) + (1 - 20\%) * 49.6) = 41980.8 \text{ kbps} = 41.98 \text{ Mbps}$$

Example 2

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the agent-based campaign, and a WAN link with g.729 codec and average call duration of 40 seconds, the bandwidth usage is:

$$60 * (20\% * (26 * 40 + 100) + (1 - 20\%) * 49.6) = 16060.8 \text{ kbps} = 16.06 \text{ Mbps}$$

Agent-Based Campaign – SIP Dialer Recording

The average WAN bandwidth usage in this case is:

$$\text{WAN Bandwidth} = \text{Calls Per Second} * (\text{Codec Bandwidth} * \text{Average Call Duration} + \text{Hit Rate} * \text{Hit Call Signaling Bandwidth} + (1 - \text{Hit Rate}) * \text{Non-Hit Call Signaling Bandwidth}) = \text{Kbps}$$

Example 3

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the agent campaign, and a WAN link with average g.711 codec and average call duration of 40 seconds, the bandwidth usage is:

$$60 * (80 * 40 + 20\% * 100 + (1 - 20\%) * 49.6) = 199180.8 \text{ kbps} = 199.18 \text{ Mbps}$$

Example 4

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the agent campaign, and a WAN link with average g.729 codec and average call duration of 40 seconds, the bandwidth usage is:

$$60 * (26 * 40 + 20\% * 100 + (1 - 20\%) * 49.6) = 67660.8 \text{ kbps} = 67.66 \text{ Mbps}$$

Transfer-To-VRU Campaign – No SIP Dialer Recording

The following figures show examples of the distributed Outbound SIP Dialer deployment for transfer to a VRU campaign.

Figure 97: Distributed Outbound SIP Dialer Deployment for Transfer to a VRU Campaign Using Cisco Unified CVP

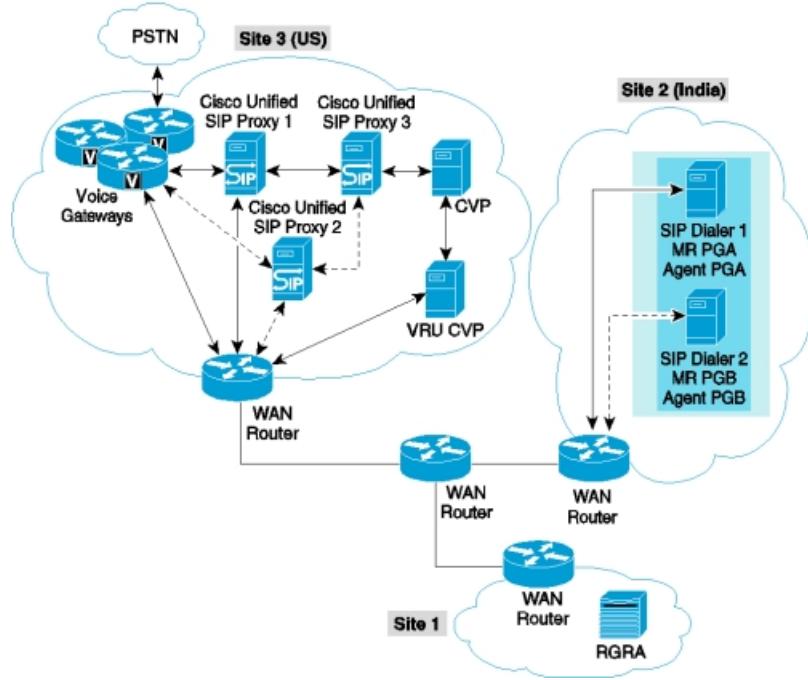
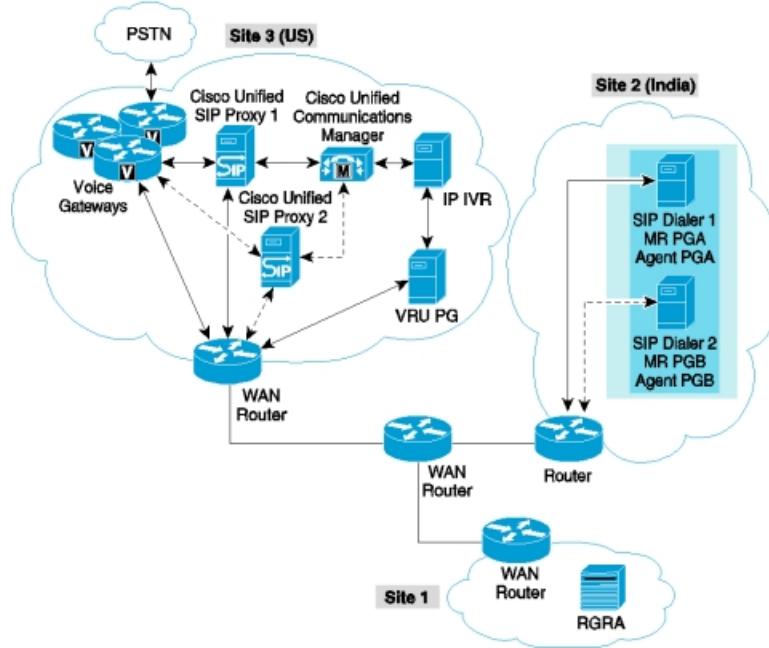


Figure 98: Distributed Outbound SIP Dialer Deployment for Transfer to a VRU Campaign Using Cisco Unified IP IVR



The average WAN bandwidth usage in this case is:

$$\text{WAN Bandwidth} = \text{Calls Per Second} * \text{Hit Rate} * \text{Hit Call Signaling Bandwidth} + \text{Calls Per Second} * (1 - \text{Hit Rate}) * \text{Non-Hit Call Signaling Bandwidth} = \text{Kbps}$$

Example 5

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the transfer-to-IVR campaign, and a WAN link with g.711 codec, the bandwidth usage is:

$$60 * 20\% * 100 + 60 * (1 - 20\%) * 49.6 = 3600 \text{ kbps} = 3.6 \text{ Mbps}$$

Transfer-To-VRU Campaign – SIP Dialer Recording

The average WAN bandwidth usage in this case is:

$$\text{WAN Bandwidth} = \text{Calls Per Second} * (\text{Codec Bandwidth} * \text{Average Call Duration} + \text{Hit Rate} * \text{Hit Call Signaling Bandwidth} + (1 - \text{Hit Rate}) * \text{Non-Hit Call Signaling Bandwidth}) = \text{Kbps}$$

Example 6

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the agent campaign, and a WAN link with g.711 codec and average call duration of 40 seconds, the bandwidth usage is:

$$60 * (80 * 40 + 20\% * 100 + (1 - 20\%) * 49.6) = 199180.8 \text{ kbps} = 199.18 \text{ Mbps}$$

Example 7

With call throttling of 60 cps on the SIP Dialer, a 20% hit rate for the transfer-to-VRU campaign, and a WAN link with g.729 codec and average call duration of 40 seconds, the bandwidth usage is:

$$60 * (26 * 40 + 20\% * 100 + (1 - 20\%) * 49.6) = 67660.8 \text{ kbps} = 67.66 \text{ Mbps}$$

Bandwidth Requirements and QoS for Agent and Supervisor Desktops

There are many factors to consider when assessing the traffic and bandwidth requirements for Agent and Supervisor Desktops in a Unified CCE environment. While the VoIP packet stream bandwidth is the predominant contributing factor to bandwidth usage, other factors such as call control, agent state signaling, Silent Monitoring, recording, and statistics must also be considered.

VoIP packet stream bandwidth requirements are derived directly from the voice codec deployed (G.729, G.711, and so forth), and can range from 4 kbps to 64 kbps per voice stream. Therefore, the contact center's call profile must be well understood because it defines the number of straight calls (incoming or outgoing), consultative transfers, and conference calls, and consequently the number of VoIP packet streams, that are active on the network. In general, the number of VoIP packet streams typically is slightly greater than one per agent, to account for held calls, Silent Monitoring sessions, active recordings, consultative transfers, and conference calls.

Call control, agent state signaling, Silent Monitoring, recording, and statistics bandwidth requirements can collectively represent as much as 25% to 50% of total bandwidth utilization. While VoIP packet stream bandwidth calculations are fairly straightforward, these other factors depend heavily on implementation and deployment details and are therefore discussed further in the sections below.

Because WAN links are usually the lowest-speed circuits in a Cisco Unified Communications network, attention must be given not only to bandwidth, but also to reducing packet loss, delay, and jitter where voice traffic is sent across these links. G.729 is the preferred codec for use over the WAN because the G.729 method for sampling audio introduces the least latency (only 30 ms) in addition to any other delays caused by the network. The G.729 codec also provides good voice quality with good compression characteristics, resulting in a relatively low (8 kbps) bandwidth utilization per stream.

Consider the following QoS factors:

- Total delay budget for latency, taking into account WAN latency, serialization delays for any local area network traversed, and any forwarding latency present in the network devices.
- Impact of routing protocols. For example, Enhanced Interior Gateway Routing Protocol (EIGRP) uses quick convergence times and conservative use of bandwidth. EIGRP convergence also has a negligible impact on call processing and Unified CCE agent logins.
- Method used for silently Monitoring and Recording agent calls. The method used dictates the bandwidth load on a given network link.
- Cisco Unified Mobile Agent deployments that use QoS mechanisms optimize WAN bandwidth utilization.
- Use advanced queuing and scheduling techniques in distribution and core areas as well.

Bandwidth Requirements for CTI OS Agent Desktop

This section addresses the traffic and bandwidth requirements between CTI OS Agent Desktop and the CTI OS server. These requirements are important in provisioning the network bandwidth and QoS required between the agents and the CTI OS server, especially when the agents are remote over a WAN link. Even if the agents are local over Layer 2, it is important to account for the bursty traffic that occurs periodically because this traffic presents a challenge to bandwidth and QoS allocation schemes and can impact other mission-critical traffic traversing the network.

CTI-OS Client/Server Traffic Flows and Bandwidth Requirements

The network bandwidth requirements increase linearly as a function of agent skill group membership. The skill group statistics are the most significant sizing criterion for network capacity, while the effect of system call control traffic is a relatively small component of the overall network load. CTI OS Security affects the network load as well. When CTI OS Security is enabled (turned on), the bandwidth requirement increases significantly due to the OpenSSL overhead.

The following table shows the type of messaging of each CTI OS application.

Table 29: Messaging Type By CTI OS Application

Application Name	Message Types
CTI OS Agent Desktop	Agent state changes
	Call Control
	Call status information
	Chat messages
	Agent and skill-group statistics
CTI OS Supervisor Desktop	Agent state changes
	Call Control
	Call status information
	Monitoring agent states
	Silent Monitoring
	Chat messages
	Agent and skill-group statistics
All Agents Monitor Application	Agent state changes for all agents

Silent Monitoring Bandwidth Usage

Silent Monitoring provides supervisors with a means of listening in on agent calls in Unified CCE call centers that use CTI OS. Voice packets sent to and received by the monitored agent's IP hardware phone are captured from the network and sent to the supervisor desktop. At the supervisor desktop, these voice packets are decoded and played on the supervisor's system sound card.

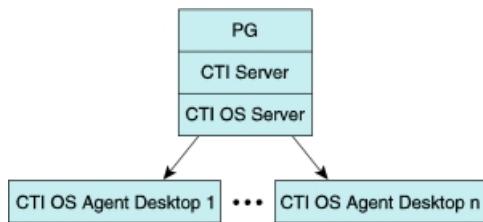
Silent Monitoring of an agent consumes roughly the same network bandwidth as an additional voice call. If a single agent requires bandwidth for one voice call, then the same agent being silently monitored would require bandwidth for two concurrent voice calls.

To calculate the total network bandwidth required for your call load, you would then multiply the number of calls by the per-call bandwidth figure for your particular codec and network protocol.

CTI OS Server Bandwidth Calculator

CTI OS provides a bandwidth calculator (at <http://www.cisco.com/c/en/us/support/customer-collaboration/computer-telephony-integration-option/products-technical-reference-list.html>) that examines the CTI OS Server-to-CTI OS Desktop bandwidth, as illustrated in the figure below. It calculates Total bandwidth, agent bandwidth, and supervisor bandwidth requirements with CTI OS Security turned on or off.

Figure 99: CTI OS Server-to-CTI OS Desktop Communication



Bandwidth reductions for CTI OS Server and CTI OS Agent Desktop

To mitigate the bandwidth demands, use any combination of the following options.

Configure fewer statistics

CTI OS allows the system administrator to specify, in the registry, the statistics items that are sent to all CTI OS clients. The choice of statistics affects the size of each statistics packet and, therefore, the network traffic. Configuring fewer statistics decreases the traffic sent to the agents. The statistics cannot be specified on a per-agent basis currently. For more information about agent statistics, see the *CTI OS System Manager Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-guides-list.html>.

Turn off statistics on a per-agent basis

You can turn off statistics on a per-agent basis by using different connection profiles. For example, if Unified Mobile Agents use a connection profile with statistics turned off, these client connections have no statistics traffic between the CTI OS Server and the Agent or Supervisor Desktop. This option could eliminate the need for a separate CTI OS Server in remote locations.

If more limited statistics traffic is acceptable for the remote site, a remote supervisor or selected agents can still log statistics through a different connection profile with statistics enabled.

If Unified Mobile Agents have their skill group statistics turned off, but the supervisor needs to see the agent skill group statistics, the supervisor could use a different connection profile with statistics turned on. In this case, the volume of traffic sent to the supervisor is considerably less. For each skill group and agent (or supervisor), the packet size for a skill-group statistics message is fixed. So an agent in two skill groups would get two packets, and a supervisor observing five skill groups would get five packets. Assume there are 10 agents at a remote site and one supervisor, all with the same two skill groups configured. In Unified CCE, the supervisor sees all the statistics for the skill groups to which any agent in the agent team belongs. If only the supervisor has statistics turned on to observe the two skill groups and agents have statistics turned off, then this approach reduces skill-group statistics traffic by 90%.

Also, at the main location, if agents want to have their skill-group statistics turned on, they could do so without impacting the traffic to the remote location if the supervisor uses a different connection profile. Again, in this case no additional CTI OS servers are required.

In the case where there are multiple remote locations, assuming only supervisors must see the statistics, it is sufficient to have only one connection profile for all remote supervisors.

Turn off all skill group statistics in CTI OS

If skill group statistics are not required, turn them all off. Doing so would remove the connections between the CTI OS Server and the Agent or Supervisor Desktop and would eliminate all statistics traffic.

Bandwidth Requirements for an Administration and Data Server and Reporting

For more information about the bandwidth requirement for Cisco Unified Intelligence Center, see the latest version of the *Hardware and System Software Specification (Bill of Materials)* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-intelligence-center/products-implementation-design-guides-list.html>.

Bandwidth Requirements for Cisco EIM/WIM

The bandwidth requirements for Cisco EIM/WIM integrations are documented in *Cisco Unified Web and E-Mail Interaction Manager Solution Reference Network Design Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-email-interaction-manager/products-implementation-design-guides-list.html>.

Bandwidth and Latency Requirements for the User List Tool

In deployments in which an Administration Client is remote (connected over a WAN) from the domain controller and Administration & Data Server, specific network bandwidths and latencies are required to achieve reasonable performance in the User List Tool. Reasonable performance is defined as less than 30 seconds to retrieve users. This information is provided in an effort to set expectations and to encourage upgrading to later Cisco Unified CCE later releases. In this version, changes were made to enhance the performance of the tool under these conditions.

There are a number of other things that can be done to improve performance of the User List Tool. Moving an Administration & Data Server and a domain controller local to the Administration Client can greatly enhance performance, as shown by the LAN row in the table below. Improving the latency in your WAN connection will improve performance, and increasing the bandwidth of your WAN connection will also improve performance.

The following data points describe scenarios in which the User List Tool can retrieve users within 30 seconds in Unified CCE 7.2(3) or later releases. Additionally, laboratory testing has determined that the tool cannot perform reasonably for any number of users on networks with a one-way latency greater than 50 ms.

Table 30: Latency and Bandwidth Requirements for the User List Tool

Maximum One-Way Latency (ms)	Available Bandwidth	Number of Users Supported
Negligible	LAN	8000
15	3.4 Mbits and higher	4000
15	2 Mbits	500
15	256 Kbits	500
50	64 Kbits and higher	25



CHAPTER 14

Cisco Unified Contact Center Management Portal

- [Unified Contact Center Management Portal, page 269](#)
- [Unified CCMP Architecture, page 270](#)
- [Portal Interfaces, page 270](#)
- [Deployment Modes, page 271](#)
- [Software Compatibility, page 273](#)
- [Reporting, page 273](#)
- [Bandwidth Requirements, page 274](#)
- [References, page 274](#)

Unified Contact Center Management Portal

Cisco Unified Contact Center Management Portal (Unified CCMP) is a browser-based management application designed for use by contact center system administrators, business users, and supervisors. It is a dense multi-tenant provisioning platform that overlays the Cisco Unified Contact Center Enterprise (Unified CCE), Cisco Unified Communications Manager, and Cisco Unified Customer Voice Portal (Unified CVP) equipment.

From a Unified CCMP perspective, the underlying Unified CCE equipment is viewed as configuration items, generally known as resources, such as agents or IP phones. Unified CCMP partitions the resources in the equipment using a familiar folder paradigm, and these folders are then secured using a sophisticated security structure that allows administrators to specify which users can perform which actions within the specified folders.

The Unified CCMP focus on supplying dense multi-tenancy functionality helps support the business plans of large enterprises because it allows the distributed or disparate contact center equipment to be partitioned or segmented to satisfy the following business goals:

- Unified CCMP abstracts and virtualizes the underlying contact center equipment, thereby allowing centralized deployment and decentralized control, which in turn provides economies of scale while supporting multilevel user command and control.
- Unified CCMP allows the powerful and flexible native Unified CCE provisioning operations to be abstracted into simple high-level tasks that enable business users to rapidly add and maintain contact center services across the virtualized enterprise (or a portion thereof).

- Unified CCMP users see only the resources in the platform that they are entitled to see, thereby providing true multi-tenancy.
- Unified CCMP users may manipulate only those resources visible to them by using Unified CCMP tools and features they have been authorized to use, thereby providing role-based task control.

The Unified CCMP Web interface allows for the concurrent provisioning activities of hundreds of end-users, thus avoiding the surge of activity at the Administration & Data Server (formerly known as Admin Workstation, or AW) sometimes experienced in Unified CCE deployments where provisioning requests can stack up during busy periods. This surge of activity is smoothed by Unified CCMP, so that the central site is not overloaded with provisioning requests.

Unified CCMP Architecture

Unified CCMP is a multitier architecture consisting of a web server, application server, and database. This architecture maintains a complete data model of the contact center equipment to which it is connected, and the data model is periodically synchronized with the underlying Unified CCE equipment. The Unified CCMP data model and synchronization activity allow for resources to be provisioned either through the Unified CCMP Web interfaces or from the standard equipment-specific user interfaces (the so-called closed loop provisioning cycle).

All provisioning operations entered through the Unified CCMP Web interfaces are checked for capacity (Is there room on Unified CCE?) and concurrency (Has another user already modified or deleted the resource?) before the request is committed to Unified CCMP. Unified CCMP then executes the provisioning request through the relevant Unified CCE APIs and checks until the action has successfully passed through the Unified CCE servers (the confirmation). At all stages, the process is audited to allow the business users to run audit reports to determine who changed what and when.

Unified CCMP back-end components connect to the Unified CCE interfaces with a *preferred* connection and a backup. This applies more to the dual-sided Unified CCE than to the Unified CM cluster, but typically Unified CCMP connects to the local Administration & Data Server (the preferred connection) and switches to the backup connection if the preferred connection fails. Unified CCMP switches back to the preferred connection when its monitoring software detects the return to normal service.

Portal Interfaces

Users connect to Unified CCMP through an HTTP/S connection. This is a standard Internet Explorer browser connection to the Unified CCMP web server.

Unified CCMP uses three interface points with the rest of Unified CCE:

- The Configuration Management Service (CMS) server, which runs on an Administration & Data Server, acts as the provisioning interface for Unified CCE. It uses the Java RMI protocol, and the CMS server option must be selected as part of the Administration & Data Server installation.
- The Administration & Data Server “AWDB” database catalog acts as the read-only configuration confirmation interface for Unified CCE. This is an OLEDB protocol interface that uses either Integrated Security or SQL Server integration. Integrated Security means that either Unified CCMP must be in the same Active Directory domain as the Administration & Data Server, or that suitable permissions between the domains must be set up.

- The Unified CM AXL interface acts as both the provisioning interface and the confirmation interface for Unified CM. This is the standard web service using HTTP and XML SOAP protocol.

Deployment Modes

Unified CCMP supports all Unified CCE deployment modes, including parent/child. This section explains the deployment modes and guidelines that pertain to them.

**Note**

For all deployments, each Unified CCE instance connected to a Unified CCMP system requires a separate VM configured as an Administration & Data Server.

Lab Deployment

In lab environments only, Unified CCMP software can be installed on the Unified CCE Administration & Data Server. This co-located model can be used only in labs due to the high processing requirements of the Administration & Data Server and the maximum configuration of 200 Named Agents.

Standard Deployments

In dedicated server mode, two deployments are supported:

- Single Server—In this simplex mode, all Unified CCMP components are installed on a single VM. Most Unified CCE customers use this deployment because it represents the lowest cost of deployment and ongoing cost of ownership. This mode supports a maximum configuration of 1500 Concurrent Agents.
- Dual Server—In this mode, the front-end Unified CCMP components are installed on one VM (the Web Server) and back-end components on another VM (the Database Server). This allows the use of a firewall between the Web and Database Servers, which creates a DMZ for Internet connectivity and provides for higher capacity and performance throughout the system. This mode supports a maximum configuration of 8000 Concurrent Agents.

**Note**

Both of the above deployments are non-resilient in nature. The workaround in the event of Portal failure is to revert to provisioning on Unified CCE or Unified Communications Manager until Unified CCMP is returned to service, at which time an automatic resynchronization occurs.

Resilient Deployments

Either of the two standard deployment modes can be enhanced to a resilient configuration using a duplicate set of hardware with Unified CCMP integrated data replication facilities to provide a geographically dispersed solution.

Unified CCMP uses SQL Server replication to keep the two sides synchronized. Use the resilient forms of these deployments if you require fault tolerance. The system capacity limits remain unchanged from the equivalent standard deployments.

**Note**

If a load balancing solution is to be provided to the front end (for example, Cisco Local/Remote Director), then it must support 'sticky' connections.

Parent/Child Deployment

In parent/child deployments, a single Unified CCMP instance connects to each of the child Unified CCE Administration & Data servers, which must be configured as physically separate Primary Administration & Data Servers. Each child instance appears as a *tenant* within Unified CCMP. Resources added via Unified CCMP are linked to a tenant, and the added resource is replicated from the Unified CCE child to its parent using the standard replication process.

Unified CCE Administration and Data Server

Beginning with Unified CCE 8.0(1), the Distributor AW (with or without HDS) is renamed as an Administration & Data Server. Multiple Administration & Data Server deployments with different roles are available, based on the functionality and amount of reporting data that it can handle.

Roles

The Administration & Data Servers are classified into roles based on the system configuration and the call load that they can handle. The Administration & Data Server role, known as a Configuration-Only Administration Server, was designed for use with Unified CCMP when a lightweight Administration & Data Server running on VMs is desirable.

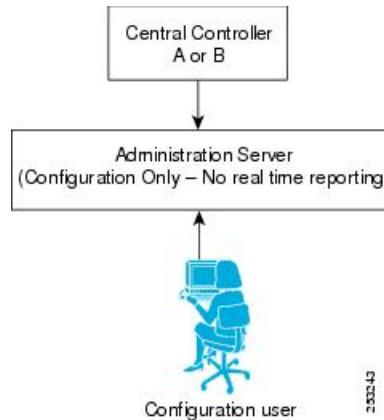
Related Topics

[Deployments, on page 27](#)

Administration Server (Configuration-Only Administration Server)

In this role as a Configuration-Only Administration Server, the HDS is not enabled and real-time reporting is turned off. This distributor deployment provides support for configuration changes only. No real-time or historical reporting is supported, as shown in the following figure.

Figure 100: Configuration-Only Administration Server



Systems That Exceed Published Limits

If your requirements exceed the capacity limits outlined in this document and detailed in the *Virtualization DocWiki*, you must contact Cisco Systems to confirm that your Unified CCMP deployment plan is suitable and will not cause performance issues. Unified CCMP users must pay particular attention to the limits on provisioning operations per hour.

For details on additional resource capacity criteria and required networking connectivity within geographically dispersed deployments, see the *Virtualization for Unified CCE DocWiki* at http://docwiki.cisco.com/wiki/Unified_Communications_in_a_Virtualized_Environment.

Software Compatibility

For information about Unified CCMP compatibility, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE

Reporting

The provisioning audit information collected by Unified CCMP can be viewed by the end-user using the Unified CCMP multi-tenanted and partitioned reporting engine. For reporting of Unified CCE call data, use Cisco Unified Intelligence Center (Unified IC), Unified CCE's advanced reporting platform.

Bandwidth Requirements

Unified CCMP does not have any voice data or call signaling paths; therefore, it does not have any QoS requirements. Very low bandwidth or the use of congested network links will either increase the latency of the requests or cause application time-outs to be returned to the user.

Use the following bandwidths:

- A minimum of a 256 kbps link between Unified CCMP and Unified CCE /AXL.



Note AXL is particularly sensitive to slow networks due to the relatively verbose SOAP packets returned during the import phase.

- A minimum of 2 Mbps links between the client browsers and Unified CCMP Web Servers, and 2 Mbps between the Unified CCMP Web Servers and Unified CCMP Database Servers if quad deployment mode is used.

References

For further information, see the Unified CCMP product documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/tsd-products-support-series-home.html>.



APPENDIX A

Acronym List

Numerics

3DES

Triple Data Encryption Standard

A

ACD

Automatic call distribution

AD

Active Directory

ADSL

Asymmetric digital subscriber line

AHT

Average handle time

ANI

Automatic Number Identification

APG

Agent Peripheral Gateway

AQT

Average queue time

ARM

Agent Reporting and Management

ASA

Average speed of answer

ASR

Automatic speech recognition

AVVID

Cisco Architecture for Voice, Video, and Integrated Data

AW

Administrative Workstation

AWDB

Administrative Workstation Database

B**BBWC**

Battery-backed write cache

BHCA

Busy hour call attempts

BHCC

Busy hour call completions

BHT

Busy hour traffic

BOM

Bill of materials

bps

Bits per second

Bps

Bytes per second

C**CAD**

Cisco Agent Desktop

CC

Contact Center

CCE

Contact Center Enterprise

CG

CTI gateway

CIPT OS
Cisco Unified Communications Operating System
CIR
Cisco Independent Reporting
CMS
Configuration Management Service
CPE
Customer premises equipment
CPI
Cisco Product Identification tool
CRM
Customer Relationship Management
CRS
Cisco Customer Response Solution
CSD
Cisco Supervisor Desktop
CSS
Cisco Content Services Switch
CSV
Comma-separated values
CTI
Computer telephony integration
CTI OS
CTI Object Server
CVP
Cisco Unified Customer Voice Portal
D
DCS
Data Collaboration Server
DES
Data Encryption Standard
DHCP
Dynamic Host Configuration Protocol

DID

Direct inward dial

DiffServ

Differentiated Services

DMP

Device Management Protocol

DMZ

Demilitarized zone

DN

Directory number

DNP

Dialed Number Plan

DNS

Domain Name System

DSCP

Differentiated Services Code Point

DSL

Digital subscriber line

DSP

Digital signal processor

DTMF

Dual Tone Multi Frequency

E**ECC**

Expanded Call Context

H**HA WAN**

Highly available WAN

HDS

Historical Data Server

HSRP

Hot Standby Router Protocol

I

ICCS

Intra-Cluster Communication Signaling

ICM

Cisco Unified Intelligent Contact Management

IDF

Intermediate distribution frame

IDS

Intrusion detection system

IntServ

Integrated services

IP

Internet Protocol

IPM

Internetwork Performance Monitor

IPPA

Unified IP Phone Agent

ISN

Internet service node

IVR

Interactive voice response

IXC

Inter-exchange carrier

J

JTAPI

Java Telephony Application Programming Interface

K

kb

Kilobits

kB

Kilobytes

kbps

Kilobits per second

kBps

Kilobytes per second

L**LAMBDA**

Load Adaptive Message-Base Data Archive

LAN

Local area network

LCC

Logical contact center

LDAP

Lightweight Directory Access Protocol

LEC

Local exchange carrier

LAA

Longest available agent

LSPAN

Local switched port analyzer

M**MAC**

Media access control

Mbps

Megabits per second

MC

Management center

MDF

Main distribution frame

MDS

Message delivery Subsystem

MED

Minimum expected delay

MGCP

Media Gateway Control Protocol

MoH

Music on hold

MR

Media routing

MRCP

Media Resource Control Protocol

MTU

Maximum transmission unit

N

NAT

Network Address Translation

NDIS

Network driver interface specification

NIC

Network interface controller

O

OAMP

Operations, Administration, Maintenance, and Provisioning

OPC

Open peripheral controller

OS

Object server

OU

Organizational unit

P

PAT

Port address translation

PerfMon

Microsoft Windows Performance Monitor

PG

Peripheral gateway

PHB

Per-hop behavior

PIM

Peripheral interface manager

PLAR

Private line automatic ringdown

PPPoE

Point-to-Point Protocol over Ethernet

PSPAN

Port switched port analyzer

PSTN

Public switched telephone network

PVC

Permanent virtual circuit

Q**QoS**

Quality of Service

R**RAID**

Redundant array of inexpensive disks

RIS

Real-time information server

Roger

Router and Logger

ROI

Return on investment

RONA

Reroute On No Answer

RSPAN

Remote switched port analyzer

RSVP

Resource Reservation Protocol

RTD

Real-Time Distributor

RTMT

Real-Time Monitoring Tool

RTP

Real-Time Transport Protocol

S

SAA

Service assurance agent

SCCP

Skinny Client Control Protocol

SCI

Service control interface

SCSI

Small computer system Interface

SDL

Signal distribution layer

SE

Systems engineer

SIP

Session Initiation Protocol

SLG

Service level goal

SNMP

Simple Network Management Protocol

SPAN

Switched port analyzer

SRND

Solution Reference Network Design

SRST

Survivable remote site telephony

SSL

Secure Socket Layer

SUS

Microsoft Software Update Services

T**TAC**

Cisco Technical Assistance Center

TAPI

Telephony application programming interface

TCD

Telephony Call Dispatcher

TCP

Transmission Control Protocol

TDM

Time-division multiplexing

TES

Task event services

TFTP

Trivial File Transfer Protocol

TLS

Transport Layer Security

TNT

Takeback N Transfer

TOS

Test other side

TTS

Text-to-speech

U**UCS**

Unified Computing System

UDP

User Datagram Protocol

UI

User interface

URL

Uniform resource locator

V

V3PN

Cisco Voice and Video Enabled Virtual Private Network

VLAN

Virtual local area network

VMS

CiscoWorks VPN/Security Management Solution

VoIP

Voice over IP

VPN

Virtual private network

VPNSM

Virtual Private Network Services Module

VRU

Voice response unit

VSPAN

Virtual LAN switched port analyzer

VXML

Voice XML (Extensible Markup Language)

W

WAN

Wide area network

WUS

Microsoft Windows Update Services

X

XML

Extensible markup language



APPENDIX **B**

System Requirements and Constraints

- [Introduction to the Unified CCE Reference Designs, page 287](#)
- [Configuration Limits and Scalability Constraints for Non-Reference Designs, page 309](#)
- [Administration and Data Server Limits by Deployment Type, page 314](#)
- [Standard Operating Conditions, page 315](#)
- [Data Store Configurations, page 318](#)
- [Workstation Specifications, page 318](#)
- [All-Event Client and Monitor-Mode Connection Limits, page 318](#)
- [G.711 Audio Codecs Support, page 319](#)
- [Solution Component and Feature Availability by Deployment Type, page 321](#)
- [Congestion Control Limits by Deployment Type, page 322](#)
- [Scalability Impacts of Components and Features, page 323](#)
- [Notes on Unified ICM/Unified CCE Components, page 324](#)
- [Unified Contact Center Management Portal, page 327](#)
- [E.164 Dial Plan Design Considerations, page 328](#)

Introduction to the Unified CCE Reference Designs

The Cisco contact center product suite has grown with the changes in computing and telecommunications technology. As a result, the product suite includes multiple possibilities to accomplish certain tasks. In response to the increasing scope and complexity of the contact center products, Cisco created the Unified CCE Reference Designs.

The Unified CCE Reference Designs standardize some of your choices to reduce the possible design permutations. Some decisions are based on changes in real-world infrastructure; others are based on the future development plans for certain product features. These designs provide you with a set of common models for contact centers.

You can simplify the design and acceptance processes by basing your deployment on these designs. The designs also help insulate your deployment against future problems by avoiding technology that is nearing its EOL.

The Unified CCE Reference Designs require the use of the following:

- Installation on virtualized servers
- Cisco Unified Customer Voice Portal (Unified CVP) over Cisco Unified IP Interactive Voice Response (Unified IP IVR)
- SIP over SCCP dialers
- Cisco Finesse for the agent desktop



Note

Unified CCE still supports the Cisco Computer Telephony Integration Option (CTI OS) desktop. But, you cannot use CTI OS in a deployment that follows the Unified CCE Reference Designs.

Unified CCE Reference Designs

The Unified CCE Reference Designs define models that match common requirements for a contact center. The Unified CCE Reference Designs contain the following models:

Packaged CCE

The Packaged CCE product is its own model. It includes a prebuilt set of VM templates for deploying a contact center that supports up to 1000 agents. If this predefined contact center can answer your business needs, it greatly simplifies deploying a new contact center. For more information on Cisco Packaged Contact Center Enterprise, see the Packaged CCE documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/packaged-contact-center-enterprise/tsd-products-support-series-home.html>.

Unified CCE

Unified CCE is flexible enough to support several Reference Design models. Unified CCE can support a maximum of 12,000 agents and implement the preferred components in the Unified CCE product suite. Unified CCE works as a distributed environment with components installed on multiple VMs. The distributed environment offers flexibility to design deployments for a range of performance, capacity, and network topology requirements. Unified CCE includes the following Reference Design models:

- **Unified CCE–4,000 Agents.** This model supports up to 4,000 pure Unified CCE Agents with CVP queueing.
- **Unified CCE–12,000 Agents.** This model supports up to 12,000 pure Unified CCE Agents with CVP queueing.
- **Large Unified CCE.** This model is a two-tier system to support call volumes beyond the Unified CCE–12,000 Agents model. The first layer is an IVR ICM system that acts as a self-service layer. There are no agents at this layer. The IVR ICM layer passes opt-out traffic to an agent node in the second layer. Use the Unified CCE–12,000 Agents model for the agent node.

Unsupported Configurations in the Unified CCE Reference Designs

The following list details some of the components, options, and configurations that the Unified CCE Reference Designs do not support:

- Multi-CUCM PIM configuration
- Simplex configurations for components that are normally deployed in redundant pairs
- Progger deployment type
- 300+ total CPS or 90+ Agent CPS
- Mobile Agent with Silent Monitoring
- Desktop Monitoring (Used for phones that do not include BIB.)
- CAD and its IPPA agent desktops
- CAD VoIP Monitor Server
- Remote Silent Monitoring
- Third-party ACDs
- Database integration with DB Lookup (SQL Gateway). (HCS for CC does support this integration with limits. HCS for CC also does not support the Application Gateway.)
- SCCP Dialer for Cisco Outbound Option. (SCCP Dialer is deprecated.)
- Unified Communications Manager Multi-cast Music on Hold
- Unified Communications Manager software-based conference bridges and Unified Communications Manager software-based transcoder as media resources
- H.323 and MGCP protocols
- G.722, iSAC, and iLBC codecs
- Unified IP IVR
- KPML for DTMF
- Tomcat as the media server
- CUBE-SP (Service Provider)
- SS7
- Pre-route ingress call flows

Configuration Limits and Scalability Constraints for Unified CCE Reference Designs

This table gives a brief description of the Unified CCE Reference Designs Models.

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	Large Unified CCE
Model Description	Pure Unified CCE Agents with CVP	Pure Unified CCE Agents with CVP	A two-tiered deployment that supports traffic volumes beyond the capacity of the Unified CCE –12,000 Agents model. Use IVR ICM as the self-service layer. For opt-out traffic, use the Unified CCE –12,000 Agents model for the agent node.

The following tables list configuration limits and scalability constraints for various Unified CCE deployments.

Configuration Limits

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Configured agents	22,000	72,000	NA	For Unified CCE, you can have a maximum of 12,000 configured agent on a PG.
Configured dialed numbers	240,000	240,000	240,000	
Configured labels	100,000	160,000	160,000	Use agent targeting rules.

Concurrent Users

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Concurrent agents	4,000	12,000	12,000 (at the agent Node, only one agent node in each deployment)	
Concurrent agents on each UCM cluster	4,000	8,000		
Active Administrators (Users)	350	1000	1000	Includes setup, configuration, and scripting users. Each Distributor can support up to 64 users.

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Real-Time Only Distributors (for configuration only)	4 (2 on each side)	10 (5 on each side)	10 (5 on each side)	
Concurrent WIM/EIM agents	200 (lite) or 1250 distributed EIM/WIM	200 (lite) or 1250 distributed EIM/WIM	NA	
Concurrent Outbound Option agents	Sizing is required	Sizing is required	NA	

Unified Communications Manager

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Maximum number of clusters	4	12	NA	The maximum number of PGs for each CUCM Cluster is 4.

Run Time

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Maximum queued calls/tasks	6,000	10,000	10,000	
CPS	30	90	90 for Unified CCE agent node 300 for IVR ICM	Reference Designs do not support greater than 300 total CPS or greater than 90 agent CPS.

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Provisioning operations on each hour	120	120	120	<p>This number represents the maximum number of save operations across all ADS in the solution in a 1-hour period.</p> <p>For Configuration Manager, CCMP, or AAS, each reskilling operation can include up to 200 changes (such as, Add an attribute, Remove a skill).</p> <p>For the cceadmin gadget and associated APIs, each reskilling operation can include up to 3500 changes.</p>
Maximum agents tracked in Agent State Trace	100	100	NA	<p>For more information on Agent State Trace, see the Configuration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-installation-and-configuration-guides-list.html</p>

WAN

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Round-trip time for Contact Center with Geographic Resiliency WAN (not clustering over WAN)	200 msec	200 msec	200 msec	
Round-trip time for Unified Communications Manager cluster over the WAN	80 msec	80 msec	80 msec	

Variables

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
ECC (Extended Context Call) and User variables size (bytes)	2000	2000	2000	Unified CVP and Outbound Option rely on a subset of this maximum limit for integration with Unified ICM. The maximum indicated is independent from the number of ECC and user variables used, with each representing approximately 50-bytes extra storage per record. The maximum includes both persistent and nonpersistent variables.
Number of Peripheral Variables (Call Variables)	10	10	10	
Peripheral Variable length (characters)	40	40	40	40 characters, excluding terminating NULL.

Peripheral Gateways and PIMs

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Maximum PGs	30, but only 4 Agent PGs System capacity limits maximum Agent PGs and VRU PGs. (See system limits.)	150, but only 12 Agent PGs System capacity limits maximum Agent PGs and VRU PGs. (See system limits.)	Maximum VRU PGs at the IVR ICM are 40.	Agent PGs for Unified CCE are either Generic PGs or Unified CM PGs with up to 2,000 agents. For deployments with <= 450 agents and Cisco Outbound Option, there is another MR PG. See UCM PIM limit.

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Maximum VRU ports	12,000	36,000	72,000	The Reference Designs do not support multiple Unified CM PIM configurations.
CTI OS or Finesse on each Agent PIM	1	1	1	Unified ICM does not support Finesse.
CTI OS or Finesse servers on each Generic PG	1	1	NA	
VRU PIMs on each Generic PG	NA	NA	NA	A Generic PG with Cisco Unified Call Manager (CUCM) PIM only supports 1000 Ports total.
VRU PIMs on each System PG	NA	NA	NA	IP-IVR PIMs only.
TDM PIMs on each PG	NA	NA	NA	Multiple PIMs on a PG affect performance. Compared to a single PIM on each PG, multiple PIMs lower the total number of agents, VRU ports, and supported call volume. There is a maximum of one PIM on each TDM PG with CTI OS coresident.
MR PIMs on each MR PG	2	2	2	
PIMs on each system (total)	18	52	40	

Cisco Outbound Option

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Platform	C series	B series	NA	
Maximum outbound agents	2,000	4,000	NA	
Maximum Calls/Second	40	120	NA	

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Maximum Calls/Second per Dialer	30	30	NA	
Maximum dialer ports on each SIP dialer	1500	1500	NA	
SIP dialers on each PG pair (Side A + Side B)	1 Dialer pair	1 Dialer pair	NA	
Dialer ports on each system (total)	4,000	4,000	NA	
Dialers on each system (total)	4	6	NA	
Campaigns on each system	300	300	NA	
Campaigns skill groups on each system	300	300	NA	Total skill groups from all campaigns.
Campaigns skill groups on each Dialer	100	100	NA	
Campaign skill groups on each campaign	1 for each campaign on each Dialer	1 for each campaign on each Dialer	NA	

Skill Groups, Call Types, and Precision Queues

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Skill groups on each PG	4,000	4,000	NA	Configuration of precision queues creates a skill group per agent PG which counts toward the supported number of skill groups on each PG.

	Unified CCE–4,000 Agents	Unified CCE–12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Skill groups on each system	16,000	27,000	NA	Configuration of precision queues creates a skill group for each agent PG which counts toward the supported number of skill groups on each system.
Call type skill groups on each interval	30,000	30,000	NA	Total call type skill group records.
Configured call types	10,000	10,000	10,000	Total call types configured.
Active call types	8,000	8,000	8,000	
Precision Routing (PR) Attributes on each system	10,000	10,000	NA	
PR Attributes for each Agent	50	50	NA	
PR Precision Queues (PQ) on each system	4,000	4,000	NA	
PR PQ Steps on each system	10,000	10,000	NA	
PR Terms for each PQ Step	10	10	NA	
PR Steps for each PQ	10	10	NA	
PR Unique attributes for each PQ	10	10	NA	
Unique skill groups and PQs in a supervisor team	50	50	NA	

Component and Feature Support for Unified CCE Reference Designs

This table gives a brief description of the Unified CCE Reference Designs Models.

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	Large Unified CCE
Model Description	Pure Unified CCE Agents with CVP	Pure Unified CCE Agents with CVP	A two-tiered deployment that supports traffic volumes beyond the capacity of the Unified CCE – 12,000 Agents model. Use IVR ICM as the self-service layer. For opt-out traffic, use the Unified CCE – 12,000 Agents model for the agent nodes.

The following tables list component and feature support for various Unified CCE deployments.

Allowed Server Types

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Computer Platform	C series B series (required for more than 8,000 agents)	B series (required for more than 8,000 agents)	IVR ICM layer can use: C series B series	You can also use comparable spec-based hardware.
Packaged CCE Core Single server	No	No	No	Reference Designs do not support the Progger deployment type.
Rogger	Yes	No	No	
Router/ Logger	Yes	Yes	Yes	
Historical & RT servers	AW-HDS-DDS	HDS-DDS, HDS-AW, AWD	HDS-DDS, HDS-AW, AWD	

Fault Tolerance

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Unified CCE Redundant (Side A & B)	Yes	Yes	ICM Side A & Side B	
Simplex installs of redundant components	No	No	No	
N+1 (Centralized Contact Center)	CVP N+1	CVP N+1	CVP N+1	
1:1 (Contact Center with Geographic Resiliency)	CVP 1:1	CVP 1:1	CVP 1:1	
Unified CCE Geographic Redundancy	Yes	Yes	Yes	
Unified Communications Manager Cluster over the WAN	Yes	Yes	NA	
SIP Proxy	CUSP 1:1 (Required for either more than 4 CVP servers or more than 2 Unified CM clusters.)	Required	CUSP 1:1	

Reporting

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Reporting subsystem component	CUIC	CUIC	CUIC	
Live Data	No	No	No	
Third-party Reporting sub system	Yes	Yes	Yes	

Desktop

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Finesse and Finesse IPPA	Yes	Yes	NA	
CTI OS	No	No	NA	
CAD and its IPPA	No	No	NA	
Desktop Customization				
Finesse REST API	Yes	Yes	NA	
CTI OS Toolkit	No	No	NA	
CRM Integration				
Solution Plus based CRMs	Yes	Yes	NA	

Agent locations

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Local Agents	Yes	Yes	NA	
Mobile Agent (Call-by-Call)	No	No	NA	
Mobile Agent (Nailed Connection)	Yes	Yes	NA	The Unified CCE Reference Designs do not support Silent Monitor with Mobile Agent.
At home agents (Broadband)	Yes	Yes	NA	

Silent Monitoring and Recording

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Silent Monitoring				
Unified CM-based (BIB)	Yes	Yes	NA	
SPAN/CTI OS Monitor Server (for Mobile Agent)	No	No	NA	
Desktop Monitoring	No	No	NA	Only used for phones without BIB.
CAD VoIP Monitor Service	No	No	NA	
Remote Silent Monitoring	No	No	NA	
Method				
CUCM-based (BiB)	Yes	Yes	NA	
SPAN/CTI OS Monitor Server (for Mobile Agent)	No	No	NA	
CUBE Forking	Yes	Yes	NA	
Recording Server				
MediaSense	Yes	Yes	Yes	
Stand-alone third-party recording server	Yes	Yes	Yes	

Third-party Integrations

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Recording				
CTI OS all events feed	No	No	NA	
CTI Server (GED-188)	Yes	Yes	NA	

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Wallboards				
CTI Server (GED-188), ODBC to AW real-time tables	Either CTI Server or Finesse	Either CTI Server or Finesse	NA	
Workforce Management				
Advanced Services Workforce Management Extensions	Yes	Yes	NA	
Database Integration				
CVP Server	Yes	Yes	Yes	
App Gateway	Yes	Yes	Yes	
DB Lookup	No	No	No	
Third-party ACD				
ACD	None	None	NA	
Third-party VRU				
VRU	None	Yes	Yes	

Cisco Feature Options

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Social Media Customer Care	Yes	Yes	NA	SocialMiner required for Agent Request API
Web Interaction Manager	Yes	Yes	NA	
E-Mail Interaction Manager	Yes	Yes	NA	

Voice Infrastructure

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Music on Hold				
Unicast with Unified Communications Manager Subscriber source only	Yes	Yes	NA	
Multi-cast with Unified Communications Manager Subscriber source only	No	No	No	
Multi-cast using VGW	Yes	Yes	NA	
Media Resources				
Conference bridges using VGWs or SRST	Yes	Yes	Yes	In many cases, IVR ICM does not need media resources.
Unified CM software-based conference bridges	No	No	No	Where possible, requires careful configuration.
Transcoders and Universal Transcoders using VGWs or SRST	Yes	Yes	Yes	
Unified CM software-based transcoder	No	No	No	Where possible, requires careful configuration.
IOS Software MTPs	Yes	Yes	Yes	
Proxy				
SIP Proxy	CUSP	CUSP	CUSP	

PSTN Support

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
ISDN Trunk	Yes	Yes	Yes	
Cisco UBE-ENT (Enterprise)	Yes, with some limitations	Yes, with some limitations	Yes, with some limitations	For more information on Cisco UBE, see the <i>Design Guide for Cisco Unified Customer Voice Portal</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-implementation-design-guides-list.html .
Cisco UBE-SP (Service Provider, usually ASR.)	No	No	No	
Third-party Session Border Controllers	No	See notes	See notes	See the Design Guide for Cisco Unified Customer Voice Portal at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-implementation-design-guides-list.html .
SS7	No	No	No	

End Points and Protocols

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
End Points				
Ingress Gateways	See <i>Compatibility Matrix for Unified CCE</i>	<i>Compatibility Matrix for Unified CCE</i>	<i>Compatibility Matrix for Unified CCE</i>	Note large deployments could also have AS54xx.
Phones	<i>Compatibility Matrix for Unified CCE</i>	<i>Compatibility Matrix for Unified CCE</i>	NA	
VXI	<i>Compatibility Matrix for Unified CCE</i>	<i>Compatibility Matrix for Unified CCE</i>	NA	
Protocol				

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
SIP over TCP	Yes	Yes	Yes	
SIP over UDP	No	No	No	Except for Dialer
H.323	No	No	No	
MGCP	No	No	No	
Codec	G.711 ulaw, alaw, G.729a	G.711 ulaw, alaw, G.729a	G.711 ulaw, alaw, G.729a	The Reference Designs do not support G.722, iSAC, and iLBC.
Phone Signaling				
SIP	Yes	Yes	NA	
SCCP	Yes	No	NA	
Secure RTP	No	No	No	
Load balancer				
Load balancer	Yes	Yes	Yes	

VRU and Queuing

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Queuing				
CVP	CVP Comprehensive Type 10	CVP Comprehensive Type 10	NA	
VRU ONLY (Type 3)	No	No	VRU Type 3 is only used in combination with INAP.	

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Unified IP IVR	No	No	No	
Self service				
CVP Comprehensive Type 10	Yes	Yes	Yes (for pure CVP deployment), Call Director and Third-party VRU	
CVP Standalone w/ ICM lookup	No	Yes	NA	
CVP Call Director	No	Yes, wrapped around 3rd party VRU	Yes, wrapped around 3rd party VRU	
Third-party	No	Yes	Yes For VRU-VRU: 36,000 Self-service Ports	
Scale	Sizing is required.	Sizing is required.	For VRU-VRU: 36,000 Self-service Ports	
Caller Input	DTMF, ASR/TTS as an option	DTMF, ASR/TTS as an option	DTMF, ASR/TTS as an option	
DTMF				
RFC2833	Yes	Yes	Yes	
KPML	No	No	No	
Video				
CVP and Video Basic	Yes	Yes	No	
CVP Video in Queue	Yes	Yes	Yes	

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Media Server				
IIS	Yes	Yes	Yes	Media Server for Reference Design could be dedicated.
Tomcat	No	No	No	

Call Flows

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Ingress Point				
Pre-route	No	No	No	
Post-route by CVP	Yes	Yes	Yes	
Post-route by Unified Communications Manager	No	No	NA	
Call Flows				
System Initiated	VRU co-browse Post call survey Courtesy call back Warm Transfers Blind transfers Router requery Post route using CVP	VRU co-browse Post call survey Courtesy call back Warm Transfers Blind transfers Router requery Post route using CVP	Translation Route to CVP at Agent Node	
Translation Route	Yes	Yes	Yes	

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Agent Initiated	All transfers, conferences, and direct agent calls by ICM script	All transfers, conferences, and direct agent calls by ICM script	NA	
Agent Greeting	Yes, sizing is required	Yes, sizing is required	NA	
Whisper Announcement	Yes, sizing is required	Yes, sizing is required	NA	
Courtesy Callback	Yes, sizing is required	Yes, sizing is required	NA	

Administration

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Provisioning				
ICM Configuration tools, CVP Operations Console	Yes	Yes	Yes	
Unified CCE Web Admin	No	No	No	
CCMP	Yes	Yes	Yes	
VIM	Yes	Yes	Yes	
Service creation				
Script Editor	Yes	Yes	Yes	
Call Studio	Yes	Yes	Yes	
Serviceability				
RTMT Analysis Manager Diagnosis	Yes	Yes	Yes	
System CLI	Yes	Yes	Yes	

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
CUOM	No	No	No	
RTMTAnalyze Call Path	No	No	NA	
Support Tools	No	No	No	

APIs

	Unified CCE – 4,000 Agents	Unified CCE – 12,000 Agents	IVR ICM component of Large Unified CCE	Notes
Administration				
Finesse	Yes	Yes	Yes	
WebAdmin API	Yes, partial	Yes, partial	No	Unified CCE supports only PQ, Attribute Creation, Attribute Assignment, and Agent Reskilling
Desktop				
GED-188 (Wallboard & recording)	Yes	Yes	NA	
CTI OS Tool Kit	No	No	NA	
MediaSense	Yes	Yes	Yes	
VRU				
GED-125	No	Yes	Yes	
GED-145 (Application Gateway)	No	Yes	Yes	

Configuration Limits and Scalability Constraints for Non-Reference Designs

The following tables specify the configuration limits and scalability constraints for the Unified ICM/CCE products. These configuration limits are part of the Unified ICM/CCE product design constraints and were used for system sizing characteristics as tested by Cisco. Most of these system parameters (or combinations of these system parameters) form contribution factors that affect system capacity. When you design your contact center, ensure your design is deployed within these limits. (See the comments in the following table for more information.) Consult Cisco if you have special configuration requirements that might exceed specific parameters.


Important

This information serves as a quick reference. Check the *Unified Communications in a Virtualized Environment* at http://docwiki.cisco.com/wiki/Unified_Communications_in_a_Virtualized_Environment and the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE pages on the DocWiki for more information on system constraints.

The compatibility matrix specifies all supported configurations and versions for Cisco Unified Contact Center Enterprise Release 10.0. The information in the compatibility matrix supersedes compatibility information in any other Cisco Unified Contact Center Enterprise documentation. If a configuration or version is not stated in the compatibility matrix, that configuration or version is not supported.

The check mark in the table indicates that a given parameter is applicable to the indicated Unified ICM/CCE product edition. See the notes at the end of this table.

Table 31: Configuration Limits and Scalability Constraints

Maximum Limit	Limit Value		Applies to				Comments
	<=450 agents (Unified CCE only)	>450 - <=12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
Administrators (Users)	100	000	✓	✓	—	—	Includes setup, config, and scripting users.

Maximum Limit	Limit Value		Applies to				Comments
	<=450 agents (Unified CCE only)	>450 - <=12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
ECC (Extended Context Call) and User variables size (bytes)	2000	2000	✓	✓	✓	✓	Unified CVP and Outbound Option rely on a subset of this maximum limit for integration with Unified ICM. The maximum indicated is independent from the number of ECC and user variables used, with each representing approximately 50-bytes extra storage per record. The maximum includes both persistent and nonpersistent variables.
Number of Peripheral Variables (Call Variables)	10	10	✓	✓	✓	✓	
Peripheral Variable length (characters)	40	40	✓	✓	✓	✓	40 characters, excluding terminating NULL.
VRU PIMs on each VRU PG	N/A	10	✓	✓	✓	✓	Each CVP PIM supports 900 ports, which means that the total maximum ports on a VRU PG is 9000.
VRU PIMs on each Generic PG	2	4	✓	✓	✓	—	A Generic PG with Cisco Unified Call Manager (CUCM) PIM only supports 1000 Ports total.
VRU PIMs on each System PG	N/A	5	✓	—	—	—	IP-IVR PIMs only.

Maximum Limit	Limit Value		Applies to				Comments
	<=450 agents (Unified CCE only)	>450 - <=12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
TDM PIMs on each PG	N/A	5	—	✓	—	✓	Multiple PIMs on a PG affect performance. Compared to a single PIM on each PG, multiple PIMs lower the total number of agents, VRU ports, and supported call volume. There is a maximum of one PIM on each TDM PG with CTI OS coresident.
MR PIMs on each MR PG	1	2	✓	✓	✓	✓	
UCM PIMs	1	12	✓	✓	✓	—	
Maximum Number of PGs for each CUCM Cluster	4	4	✓	✓	✓	✓	
Duplex PGs on each ICM instance	1	150	✓	✓	✓	✓	For deployments with <= 450 agents and Cisco Outbound Option, there is another MR PG. See UCM limit above.
PIMs on each system (total)	4	150	✓	✓	✓	✓	One agent PIM, two VRU PIMs, and one MR PIM (applies to >= 450 agents only).
Configured agents on each system (total)	N/A	65,000	—	✓	✓	✓	
Configured agents on each system (total)	3000	76,000	✓	—	—	—	
Configured agents on each peripheral	3000	12,000	✓	✓	✓	✓	

Configuration Limits and Scalability Constraints for Non-Reference Designs

Maximum Limit	Limit Value		Applies to				Comments
	<=450 agents (Unified CCE only)	>450 - <=12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
Skill groups on each peripheral gateway	1500	4000	✓	✓	✓	✓	Configuration of precision queues creates a skill group per agent PG which counts toward the supported number of skill groups on each PG.
Skill groups on each system	1500	27,000	✓				Configuration of precision queues creates a skill group per agent PG which counts toward the supported number of skill groups on each system.
CTI OS on each Agent PIM	1	1	✓	✓	✓	✓	
Provisioning operations on each hour	30	120	✓	✓	✓	✓	For Configuration Manager, web reskilling, CCMP or AAS – maximum number of save operations across all ADSS in the solution in a 1-hour period. 200 changes per provisioning operation.
Maximum agents tracked in Agent State Trace		100	✓	✓	✓	✓	For more information on Agent State Trace, see the Configuration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/professional-migration/guide.html
SIP dialer ports on each dialer	N/A	1500	✓	—	✓	—	This limit assumes the model is distributed, numbers vary based on deployment.
SIP dialers on each PG pair (Side A + Side B)	N/A	1	✓	—	✓	—	Only one dialer type can be installed per PG.

Maximum Limit	Limit Value		Applies to				Comments
	<=450 agents (Unified CCE only)	>450 - <=12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
SIP dialer Ports on each server (total)	N/A	1500	—	—	✓	—	In multi-instance deployment.
Dialer ports on each system (total)	N/A	4000	✓	—	✓	—	
Dialers on each system (total)	N/A	32	✓	—	✓	—	
Campaigns on each system	N/A	300	✓	—	✓	—	Sizing is required. Smaller deployments cannot support the full 300 campaigns.
Campaigns skill groups on each system	N/A	100	✓	—	✓	—	Total skill groups from all campaigns.
Campaign skill groups on each campaign	N/A	20	✓	—	✓	—	Limitation on skill groups for any given campaign (as long as the maximum 100 campaign skill groups per system not exceeded).
Dialed numbers on each system	1500	240,000	✓				
Labels configured on each system	500	160,000	✓				
Call type skill groups on each interval	1000	30,000	✓	✓	✓	✓	Total call type skill group records.
Configured call types	500	10,000	✓	✓	✓	✓	Total call types configured.
Active call types	250	8000	✓	✓	✓	✓	
Precision Routing (PR) Attributes on each system	N/A	10,000	✓	—	—	—	
PR Attributes for each Agent	N/A	50	✓	—	—	—	

Maximum Limit	Limit Value		Applies to				Comments
	<=450 agents (Unified CCE only)	>450 - <=12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
PR Precision Queues (PQ) on each system	N/A	4000	✓	—	—	—	
PR PQ Steps on each system	N/A	10,000	✓	—	—	—	
PR Terms for each PQ Step	N/A	10	✓	—	—	—	
PR Steps for each PQ	N/A	10	✓	—	—	—	
PR Unique attributes for each PQ	N/A	10	✓	—	—	—	
Unique skill groups and PQs in a supervisor team	50	50	✓	✓	✓	✓	

**Note**

Deployments close to the maximum number of configured agents on a system can show performance degradation and failed call routing, especially if contending capacity limitations also approach maximum thresholds. Expert assistance from partner or professional services is generally necessary for capacity-related system planning. Parameters that most impact performance with large numbers of configured agents include total number of system peripherals, routes, number of active agents, and overall call load. The points at highest risk for degradation are busy hours and the half-hour update period, during which the PG sends report data to the Central Controller. System administrators can lessen the impact of these issues by purging unused configured agents, retiring inactive peripherals, and maintaining systems at current maintenance release levels.

Administration and Data Server Limits by Deployment Type

You can deploy only so many Administration and Data Servers for each Logger. You can only deploy more AWs by removing an equal number of AW-HDSs.

Table 32: Administration and Data Server Deployment Limits

Component on each Logger side	Limits by agent count		
	450	4000	12,000
AW only	—	1	2
AW-HDS-DDS	1	1	—
AW-HDS	—	—	3
HDS-DDS	—	—	1

Limits for Active Administration and Data Server Users

The following table lists the limits for the types of active Administration and Data Server users.

User	Unified CCE 4,000 Agent Deployment	Unified CCE 12,000 Agent Deployment	Notes
Administrators	350	1000	Includes setup, configuration, and scripting users. Each Distributor can support up to 64 users.
Real-Time Only Distributors (for configuration only)	4 (2 on each side)	10 (5 on each side)	

Standard Operating Conditions

Except when explicitly specified, the Unified ICM/CCE hardware selection described in this section is based on the following operating conditions. When you size a Unified CCE Reference Design, Unified ICM, or other Unified CCE software implementation, consider the factors listed here. Many variables affect system capacity, but Cisco chose values for a representative subset. The sizing limitations in this table are based on that subset of variables.

Table 33: Operating Conditions for Unified ICM/Unified CCE

Operating Condition	Value		Applies to				Comments
	<=450 agents	>450 - <12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
Maximum number of CTI OS Servers per PG	1	1	✓	✓	✓	✓	Standalone CTI OS system (not for production systems)
Maximum number of Cisco Finesse servers per PG	2	2	✓	✓	✓	✓	Each PG pair can support 2000 agents.
Average skill groups per agent per team	5	5	✓	✓	✓	✓	Does not include default skill group; assumes 17 statistics for each skill group enabled
Number of supervisors	10%	10%	✓	—	✓	—	10% of total agent population
Number of teams	Agents/10	10%	✓	—	✓	—	10% of total agent population (9 agents and 1 supervisor per team)
Monitor mode applications (CTI OS)	2	2 ¹	✓	✓	—	✓	You can increase the number of monitor mode connections to 5 by reducing the All-Event clients by an equal amount.
	NA	10	—	—	1 each	—	

Operating Condition	Value		Applies to				Comments
	<=450 agents	>450 - <12,000 agents	Unified CCE	Unified ICM	Unified CCH	Unified ICMH	
All-Event clients (CTI Server)	1	5 ²	✓	✓	✓	✓	For 2 vCPU VMs built from the Small Agent PG OVA. These clients are in addition to the 2 CTI OS connections. You can reduce the number of All-Event clients to support more monitor mode connections.
	—	20 ³	✓	✓	✓	✓	For 4 vCPU VMs built from the Large Agent PG OVA. These clients are in addition to the 2 CTI OS connections. You can reduce the number of All-Event clients to support more monitor mode connections.
ECC variables	5 scalars	5 scalars	✓	✓	✓	✓	40 bytes each
Call flow traffic on straight calls	85%	85%	✓	✓	✓	✓	
Call flow traffic on transfer calls	10%	10%					
Call flow traffic on conference calls	5%	5%					
Outbound dialer call transfer rate	N/A	30%	✓	—	✓	—	Percentage of the calls transferred to agent or VRU

¹ See the section on All-Event Clients and monitor-mode connections for more details.

² See the section on All-Event Clients and monitor-mode connections for more details.

³ See the section on All-Event Clients and monitor-mode connections for more details.

Data Store Configurations

If you are setting up Tested Reference Configuration (TRC) or spec-based servers, consult the *UC Virtualization Supported Hardware* page at http://docwiki.cisco.com/wiki/UC_Virtualization_Supported_Hardware.

Workstation Specifications

The following list gives the minimum system requirements for agent and administrator workstations:

- **CPU**—Intel®Pentium® dual-core processor (2.5 GHz)
- **Memory**—4-GB RAM
- **Hard drive**—250-GB SATA
- **Sound card**—Windows-compatible full-duplex sound card required if the contact center uses Cisco IP Communicator, Silent Monitoring, or Jabber.

Administration & Data Servers and Administration Clients have these extra requirements:

- **Graphics card**—Support for 1024 x 768 x 64 K color or better
- **Display monitor**—17 inch or larger

Related Topics

[Cisco Finesse Desktop Solution](#), on page 115

All-Event Client and Monitor-Mode Connection Limits

The CTI server uses All-Event clients and the CTI OS server uses monitor-mode connections.

Maximums for Two vCPU Small Agent PG OVAs

On this OVA, the numbers of each type of connection are linked. If you use more of one, you can have less of the other. On each Agent PG, you can have the following:

- A maximum of seven All-Event clients on the CTI server.
- A maximum of five monitor-mode connections on the CTI OS server.
- A combined maximum of nine All-Event clients and monitor-mode connections.

**Important**

On this OVA, the Agent PG does not support having both seven All-Event clients on the CTI server and five monitor-mode connections on the CTI OS server. Your design must trade off one for the other to stay within the combined maximum of nine.

Maximums for Four vCPU Large Agent PG OVAs

VMs built from the large OVA with four vCPU can support more All-Event Clients and monitor-mode connections. On each Agent PG, you can have the following:

- A maximum of 20 All-Event clients on the CTI server.
- A maximum of five monitor-mode connections on the CTI OS server.

Unlike VMs built from the smaller OVA, these limits are not linked. The number of All-Event Clients does not limit the number of monitor-mode connections that you can have on the large VMs. You can use the maximum amount of each type, if your system can support the load.

All-Event Clients

Each of the agent desktop solutions (Cisco Finesse, CTI OS Desktop, and Cisco Agent Desktop) uses two of the available All-Event clients. Some of the possible consumers of the clients are:

- CAD IP Phone Agent (2)
- Real-Time Adherence (2)
- Some third-party recording vendors (2)
- Unified WIM and EIM (2)
- Cisco Media Blender
- B&S MCAL
- Remote Silent Monitor

Monitor-Mode Connections

If your deployment uses the CTI OS server, it begins with two monitor-mode connections on each side of the redundant pair. With the Small Agent PG OVA, you cannot configure those connections to fail over to the other side in a failure. In a failure on Side A, the resources connected to the CTI OS server on Side A cannot reconnect to the Side B server. Those extra connections would push Side B past the combined maximum of nine.

With the Large Agent PG OVA, you can configure those connections to fail over if that does not exceed five monitor-mode connections on the other server.

G.711 Audio Codecs Support

Unified CCE negotiates the audio codec when a call first connects between two points. For example, the codec is negotiated when:

- A call arrives in the system and media is established between the ingress gateway and a VXML browser.

- A call in the queue connects to an agent and media is established between the ingress gateway and the agent phone.
- An agent conferences a call and media is established between all the parties and a conference resource.

Codec selection can change on each leg of the call, depending on what codecs each component supports.

There are two types of audio codecs for G.711:

- **G.711 muLaw**—Used in North America and Japan
- **G.711 A-Law**—Used everywhere else

Unified CCE supports both audio codecs.

In previous releases, the TDM gateway could convert the A-Law that PCM used in the ISDN trunk to G.711 muLaw on the SIP dial peer. This conversion enabled the voice traffic inside the contact center to run on G.711 muLaw, while the EMEA PSTN was on A-Law.

Service providers now have a SIP interface that bypasses the TDM VGW; this requires a new architecture with internal support for A-Law.

Unified Communications Manager allows configuring codec preferences at the region level; so you can advertise both A-Law and muLaw.

Codec Support in CVP

CVP Default Prompts

CVP has default prompts for both mu-law and A-Law. You can only use one of the codecs in your deployment.

Agent Greeting

You can record agent greetings in either mu-law or A-Law.

ASR/TTS

CVP supports ASR/TTS with either mu-law or A-Law.

Codec Support in Unified CCE

Mobile Agent

Mobile Agent can advertise either G.711 muLaw or A-Law. But, Mobile Agent can only advertise one codec at a time. All the Mobile Agents on a peripheral must use the same codec. The Unified Communications Manager must insert a transcoder in order to send a Mobile Agent a call in a different codec.

Cisco Outbound Option Dialers

SIP Dialers with Cisco UBE can support A-Law with specific design considerations. The SIP Dialer does not advertise A-Law. So, the deployment needs DSP resources on Cisco UBE during the initial negotiation (no media) between the SIP Dialer and the SIP service provider. During a REFER from the Dialer to the agent,

Cisco UBE renegotiates the code with the agent to use A-Law. Cisco UBE can then release the DSP resource (Transcoder).

Silent Monitor Support

The following silent monitoring solutions support both muLaw and A-Law:

- Unified Communications Manager Silent Monitoring
- CTI OS-based Silent Monitoring
- Cisco Remote Silent Monitoring
- Cisco Agent Desktop Silent Monitoring

Mixed Environments Not Supported

Mixed mode A-Law and muLaw deployments are not supported.

Design consideration is needed for a site that is transitioning from muLaw to A-Law.

Mobile Agent

All Mobile Agents on a peripheral are required to use the same codec.

CVP Prompts

CVP prompts must all use the same codec.

Solution Component and Feature Availability by Deployment Type

This table lists the availability of certain components and features in various deployment types.

Solution Component and Feature	Availability by Deployment Type					
	1—NAM 3—NAM Rogger	2—IVR ICM	4—ICM Router/Logger 8—ICM Rogger	5—UCCE 8000 Agents Router/Logger 6—UCCE 12000 Agents Router/Logger 9—UCCE 4000 Agents Rogger 13—UCCE 450 Agents Progger	7—Packaged CCE : CCE-PAC-M1 10—Packaged CCE : CCE-PAC-M1 Lab Only	11—HCS-CC 1000 Agents 12—HCS-CC 500 Agents 14—HCS-CC 4000 Agents 15—HCS-CC 12000 Agents
Unified CCE/Unified Communications Manager Agents	N	N	Y	Y	Y	Y

Congestion Control Limits by Deployment Type

Solution Component and Feature	Availability by Deployment Type					
	1—NAM 3—NAM Rogger	2—IVR ICM	4—ICM Router/Logger 8—ICM Rogger	5—UCCE 8000 Agents 6—UCCE 12000 Agents 9—UCCE 4000 Agents Rogger 13—UCCE 450 Agents Progger	7—Packaged CCE : CCE-PAC-M1 10—Packaged CCE : CCE-PAC-M1 Lab Only	11—HCS-CC 1000 Agents 12—HCS-CC 500 Agents 14—HCS-CC 4000 Agents 15—HCS-CC 12000 Agents
Third-party ACD Agent PG	N	N	Y	N	N	N
Third-party IVR PG	N	Y	Y	Y	N	N
Precision Routing	N	N	N	Y	Y	Y
ICM-to-ICM Gateway	Y	Y	Y	Y	N	N
Cisco Finesse	N	N	Y ⁴	Y	Y	Y
Agent Request API	N	N	N	Y	Y	N
Unified CCMP	N	N	Y	Y	N	Unified CCDM ⁵
Web Admin and REST APIs	N	N	N	Partial ⁶	Y	Unified CCDM ⁷
Parent or Child	N	N	Y	N	N	N
CICM	N	N	Y	N	N	N

⁴ Finesse is only available for Unified CCE/Unified Communications Manager agents. Finesse is not available for Third-party ACD agents.

⁵ Unified CCDM is the management tool for Unified HCS for Contact Center.

⁶ Only Precision Routing configuration is exposed.

⁷ Unified CCDM is the management tool for Unified HCS for Contact Center.

Congestion Control Limits by Deployment Type

Congestion Control provides protection to the Central Control Router from overload conditions caused by high call rates. When faced with extreme overload, congestion control keeps the system running close to its rated capacity.

Congestion Control provides satisfactory service during an overloaded condition to a smaller percentage of calls, rather than a highly degraded service to all calls. The feature keeps the system within its capacity by rejecting calls by the Routing Clients at the call entry point. Throttling the capacities ensures the service of the routed calls is successful without timeouts. This throttling prevents overloading the Router and ensures the designed call processing throughput under overload conditions. The following table lists the supported deployment types with the maximum supported calls per second (CPS).

Table 34: Deployment Types

Deployment type	Maximum calls per second
Unified CCE 12,000 Agents Router/Logger	105
Unified CCE 8000 Agents Router/Logger	69
Unified CCE 4000 Agents Rogger	35
Unified CCE 450 Agents Progger	4
HCS-CC 4000 Agents	32
HCS-CC 1000 Agents	8
HCS-CC 500 Agents	5
Unified ICM Rogger	58
Unified ICM Router/Logger	115
NAM	300
NAM Rogger	150

Scalability Impacts of Components and Features

When your contact center use certain optional components and features, they have impacts on Unified ICM/Unified CCE scalability and capacity calculations. This table lists some of these impacts.

Component or feature	Impact
CTI OS Security	When you enable CTI OS Security, agent capacity decreases by 25%.

Component or feature	Impact
IPSec	<p>When you enable IPSec:</p> <ul style="list-style-type: none"> The maximum supported operational capacities for peripheral gateways in a CCE deployment decrease by 25%. This capacity reduction applies to agents, VRU ports, SIP Dialer ports, and call rate. The maximum call rate (calls per second) that the CCE deployment supports decreases by 25%.
Mobile agents	<p>Unified CCE does not directly control the phones of mobile agents. The two delivery modes, Call-by-Call and Nailed Connection, use resources differently. Run the Cisco Unified Collaboration Sizing Tool to determine the exact impact of mobile agents.</p>
Cisco Outbound Option	<p>This feature reduces the agent capacity of the Agent PGs, as follows:</p> $\text{Max agents} = (\text{Maximum PG agent capacity}) - (1.33 \times (\text{number of SIP dialer ports}))$ <p>These formulas indicate platform capacity; they do not indicate how many agents can be kept busy by the number of dialer ports in the deployment. A quick, but inexact, estimate is that you require two ports for each outbound agent. Your outbound resources can vary based on hit rate, abandon limit, and talk time for the campaigns. Use the sizing tool to determine outbound resources required for your campaigns.</p>
Agent Greeting	<p>The Agent greeting feature has an impact the Router, Logger, and Unified Communications Manager.</p> <ul style="list-style-type: none"> On the Router and Logger, the feature increases route requests made which effectively decreases the maximum call rate by about one third. Run the Cisco Unified Collaboration Sizing Tool to determine the exact impact on the Unified Communications Manager.
Precision Queues and Skill Groups	<p>As the average number of precision queues or skill groups for each agent increases, the maximum concurrent agents for each PG and for the whole system decreases.</p>
Extended Call Context (ECC)	<p>Extended Call Context (ECC) usage greater than that in the Operating Conditions has a performance and scalability impact on critical components of Unified CCE. The capacity impact varies based on ECC configuration, which requires professional guidance on a case-by-case basis.</p>

Related Topics

[Additional Sizing Factors](#), [on page 219](#)

Notes on Unified ICM/Unified CCE Components

This section contains notes on other Unified ICM/Unified CCE server requirements.

Administration & Data Server Deployment Capacities and Requirements

The Administration & Data Server now offers several roles (also known as deployments) based on the functionality and amount of reporting data that it can handle. This section specifies the hardware requirements for an Administration & Data Server that are used with a reporting server (Cisco Unified Intelligence Center) including servers with the following roles:

- Administration Server, Real-Time and Historical Data Server, and Detail Data Server (AW-HDS-DDS)
- Administration Server and Real-Time and Historical Data Server (AW-HDS)
- Historical Data Server and Detail Data Server (HDS-DDS)

Do not run more than ten concurrent reports on any client machine. This is a combined limit for reports that run on the Unified Intelligence Center User Interface, Permalinks, and Dashboards on the client machine.

However, you cannot run ten concurrent reports for the 200 maximum reporting users on each node. Our capacity testing shows that 200 reporting users can each have the following running reports:

Capacity testing uses the following profile:

- 200 Intelligence Center users per Intelligence Center server with:
 - 4 real-time reports with 100 rows and 10 fields, refreshing every 15 seconds
 - 2 historical reports with 2000 rows with 10 fields each, refreshing every 30 minutes


Note

Each reporting user is the equivalent of one Script Editor monitoring user (using Internet Script Editor or Administration Client).

Virtual Machine Capacities

The following table provides capacity guidelines for the Administration & Data Server VMs.

Virtual machine	Reporting user capacity	vDisks
Administration Server - AW	50	1x 40 GB
AW-HDS-DDS	200	1x 80 GB and 1x 500 GB 8
AW-HDS	200	1x 80 GB and 1x 500 GB 9
HDS-DDS	200	1x 80 GB and 1x 500 GB 10

⁸ The two vDisks can be in the same data store. You can customize the size of the DB vDisk at OVA deployment to meet your solution's needs with the DB Estimator tool.

⁹ The two vDisks can be in the same data store. You can customize the size of the DB vDisk at OVA deployment to meet your solution's needs with the DB Estimator tool.

- 10 The two vDisks can be in the same data store. You can customize the size of the DB vDisk at OVA deployment to meet your solution's needs with the DB Estimator tool.

CTI OS Server

Cisco requires that you collocate the CTI OS server component with the Agent PG. Standalone CTI OS servers are not supported.

See the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE for system requirements for the CTI OS server.

Silent Monitor Service for CTI OS

The silent monitor service is a single executable that you can deploy in two different ways:

- As a stand-alone server, which is called Silent Monitor Server
- Coresident with any CTI OS Client Toolkit application, which is called Silent Monitor Service for Unified CCE Toolkit

Silent Monitor Server

The Silent Monitor Server is a stand-alone server that provides silent monitor functionality for a set of mobile agents. When you deploy the silent monitor service as a stand-alone server, do not collocate the server with any other CTI OS or Unified CCE components. The following table gives the capacity for the standard Remote Silent Monitor VM.

Table 35: Virtual Machine Capacity

Virtual machine type	Capacity (concurrent sessions)
Remote Silent Monitor	80

Silent Monitor Service for Unified CCE Toolkit

You can also configure the silent monitor service to provide silent monitor functionality for a single Unified CCE agent. In this configuration, the silent monitor service runs on the same computer as the agent or supervisor's desktop. In a Citrix environment, the silent monitor service runs on the same computer as the agent or supervisor's Citrix client.

Cisco Unified Web and E-Mail Interaction Manager

See the *Hardware and System Software Specification for Cisco Unified Web and E-Mail Interaction Manager* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-web-interaction-manager/products-technical-reference-list.html> for system requirements, server configurations, capabilities, and limitations of the Cisco Unified Web and E-mail Interaction Manager.

Cisco Finesse Server

Cisco Finesse requires Unified Contact Center Enterprise and Unified Communications Manager. For more information about supported versions and system requirements for Cisco Finesse clients, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.

You can only install the Cisco Finesse server as a virtual machine running under VMware ESXi. For more information, see the *Virtualization for Cisco Finesse* DocWiki page at http://docwiki.cisco.com/wiki/Virtualization_for_Cisco_Finesse.

You can deploy Finesse on its own virtual machine or according to the coresidency policies outlined in the *Unified Communications Virtualization* DocWiki at http://docwiki.cisco.com/wiki/Unified_Communications_Virtualization.

Unified Contact Center Management Portal

This section helps you select servers for Cisco Unified Contact Center Management Portal (Unified CCMP) for Unified CCE environments.

Each of the deployment models described in this section assumes the possibility of an n-sided server configuration that replicates data between sites.

Map folder structure to organizational structure to facilitate resource management.

Table 36: Virtual Machine Capacities

VM	Capacity			
	Agents (concurrent/configured)	Configured CCMP users	Folders	Folder depth
Small CCMP single box	1500/1500	1500	200	5
Small CCMP with Config AW	1500/1500	1500	200	5
Large CCMP Web/App Svr	8000/48,000	8000	—	—
Large CCMP DB only	8000/48,000	8000	600	6

Network Connections

Consider these factors when you design the network connections for Unified CCMP:

- **LAN**—Connect Unified CCMP systems to Unified ICM/CCE and other servers via gigabit (1000BASE-T) connections.
- **WAN**—Allocate a dedicated link of at least 1.5-MB/s capacity to connect Unified CCMP systems to Unified ICM/CCE or for a distributed CCMP deployment across a WAN.

- **Load Balancing**—Distributed CCMP Systems can use a load balancer to distribute load across the sites. Use a dedicated load balancer, rather than using Windows built-in functionality. Any load balancing solution must support sticky connections to maintain the web session information between requests.

Software Requirements for Unified CCMP Servers

For information on the software requirements for a Unified CCMP server, see the Unified CCMP documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-management-portal/tsd-products-support-series-home.html>.



Note

For a single-server system, install the software prerequisites and Unified CCMP components for both the Web Application Server and the Database Server on the single server. A single-server system can only support the smallest deployments.

E.164 Dial Plan Considerations

Unified CCE supports E.164 dial plans and provides partial support for the ‘+’ prefix as follows:

- Agent extensions cannot include the ‘+’ character.
- Agent secondary lines cannot include the ‘+’ character if the agent peripheral has “All Lines” Agent Control enabled.
- Route DNs through a CTI Route Point or a VRU cannot include the ‘+’ prefix.
- Dialer-imported contact numbers and campaign prefixes cannot include the ‘+’ prefix.
- Agents can dial the ‘+’ prefix in combination with an E.164 number through Finesse or enabled customized desktops.
- Agents can dial the ‘+’ prefix in combination with an E.164 number through their phones.

For contact centers that advertise the agent extension outside of the contact center, these considerations apply:

- Use transformation patterns to add the ‘+’ prefix to the calling number on outgoing calls. You can use Calling Party Transformation CSS for phone configuration.
- To route incoming calls addressed to an E.164 number with the ‘+’ prefix, use called party transformations on the translation patterns to strip the ‘+’ prefix from the called number.
- The Attendant Console does not have visibility into the phone status.



Parent/Child

- Parent/Child Architecture, [page 329](#)
- Unified CCE High Availability with Unified ICM, [page 341](#)
- Traditional ACD Integration, [page 345](#)
- Traditional VRU Integration, [page 346](#)

Parent/Child Architecture

The Unified CCE Gateway PG allows Unified CCE to appear as a traditional ACD connected to the Unified ICM system. The Unified CCE Gateway PG provides the Unified ICM system with a PG that communicates with the CTI interface of the Unified CCE System PG.

In the parent/child model, you configure the child Unified CCE to function on its own. Unified CCE does not need a connection to the parent to route calls to agents. This independence provides local survivability for mission-critical contact centers during connection failures between the child and parent.

The child system can automatically send configuration objects to the parent Unified ICM for insertion into the Unified ICM configuration. This process eliminates the need to configure objects twice (on the local ACD and again on the Unified ICM). You can also turn off this functionality for situations where the customer does not want automatic configuration updates. For example, you do not want automatic updates from an outsourcer child system that also supports agents for another customer.

The Unified CCE Gateway PG can connect to a Unified CCE child that is using the Unified CCE System PG. If the child has multiple Unified CCE System PGs and peripherals, install and configure a separate Unified CCE Gateway PG in the parent system for each child PG. When deployed on a separate VM, a Unified CCE Gateway PG can manage multiple child Unified CCE peripherals.

In the Unified CCE child, you can deploy Unified IP IVR or Unified CVP for call treatment and queuing. If you deploy Unified CVP, configure another VRU PG. This model does not follow the single peripheral model used when Unified IP IVR is deployed. For this reason, information about calls queued at the child (and queue time of a call) is not available on the parent. Any computation involving queue time by the parent is inaccurate, which can lead to adverse impacts to routing on the parent.

Parent/Child Components

The following sections describe the components used in Unified ICM (parent) and Unified CCE (child) deployments.

Unified ICM (Parent) Data Center

The Unified ICM data center location contains the Unified ICM Central Controller. The data center has a redundant pair of Central Controllers. A Central Controller has Call Router and Logger servers. You can deploy the servers as individual Call Routers and Loggers and deploy the servers in two geographically distributed data centers for extra fault tolerance.

The Unified ICM Central Controllers control Peripheral Gateways (PGs) at the data center location. A redundant pair of VRU PGs controls Unified CVP across the architecture. You can insert more PGs at this layer to control TDM or legacy ACDs and VRUs. This approach can support a migration to Unified CCE or support outsource locations that use the TDM or legacy ACDs.

The Unified ICM parent does not support any directly controlled agents in this model. So, in parent/child deployments, Unified ICM does not support classic Unified CCE with a Unified Communications Manager PG installed on this Unified ICM parent. Control all agents externally to this Unified ICM parent system.

The Unified CVP VRU PG pair controls the Unified CVP Server. The CVP Server translates the VRU PG commands from Unified ICM into VXML and directs the VXML to the Voice Gateways (VGs) at the child sites. Calls from the data center location can come into the remote call centers under control of the Unified CVP at the parent location. The parent then has control over the entire network queue of calls across all sites. The parent holds the calls in queue on the VGs at the sites until an agent becomes available.

Unified CCE Call Center (Child) Site

The Unified CCE call center contains a Unified Communications Manager cluster for local IP-PBX functionality and call control for the IP phones. A local Unified IP IVR also provides local call queuing for the Unified CCE site. A redundant pair of Unified CCE Gateway PGs connects this site over the WAN to the Central Controller at the Unified ICM parent. You can deploy the Unified CCE Gateway PGs on separate VMs or coresident with the CCE System PG with the following caveats:

- If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are the same, then use different PG numbers for those PGs.
- If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are different, then you can use the same PG number for those PGs.
- Do not add other PGs (such as VRU PG or MR PG) to this VM.
- Follow the scalability limits of the coresident Unified CCE Gateway PG and Unified CCE System PG.

The Unified CCE Gateway PGs provide real-time event data and agent states to the parent from the Unified CCE child. The Unified CCE Gateway PGs also capture some, but not all, configuration data and send it to the parent Unified ICM configuration database.

You can replace the Unified IP IVR at the child site with a local Unified CVP instance. Unified CVP is not integrated as part of the System PG for the Agent Controller. The installation for Unified CCE with Unified CVP defines a separate VRU PG specifically for Unified CVP. Because Unified CVP is not part of the System PG, the Unified CCE Gateway PG does not report calls in queue or treatment to the parent Unified ICM. If

your parent routing requires queueing or treatment information, deploying with Unified CVP might not meet your needs.

A local Unified CCE child system provides ACD functionality. You can size the Unified CCE child system as a Rogger with a separate Unified CCE Agent PG server. The Rogger contains a Call Router and Logger. The set of redundant Agent PG servers contain the System PG for Unified Communications Manager and Unified IP IVR, CTI Server and CTI OS Server, and the optional VRU PG for Unified CVP.

In either configuration, you need a separate Administration & Data Server to host the configuration and scripting tools for the system, as well as the optional Historical Database Server, and the web-based Unified Intelligence Center reporting tool.

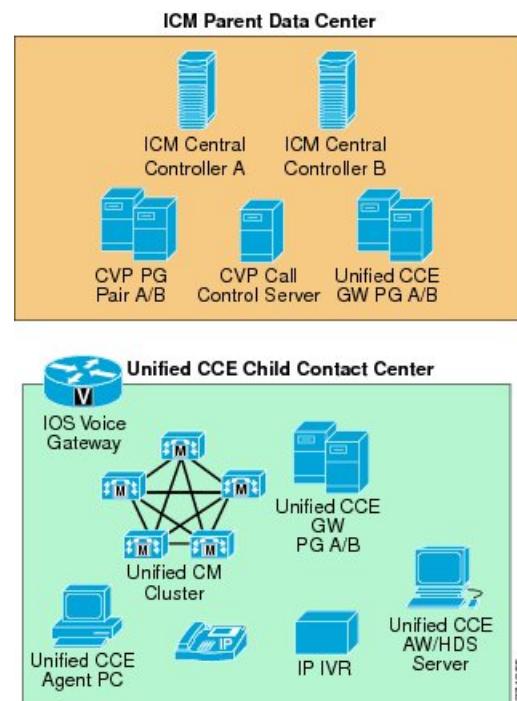
Related Topics

[Sizing Unified CCE Components and Servers, on page 215](#)

Unified CCE Gateway PGs at Data Center

You can deploy the Unified CCE Gateway PG at the Unified ICM data center as shown in the following figure. This deployment model allows you to manage and control the Unified CCE Gateway PGs centrally. When different companies own and manage the child and parent sites, that condition can force you to deploy the Unified CCE Gateway PG at the Unified ICM data center. For example, an Outsourcer/Service Bureau manages child site and connects to the Unified CCE Gateway PGs at the parent site.

Figure 101: Parent/Child Deployment with Unified CCE Gateway PGs at Data Center



There are several drawbacks with moving the Unified CCE Gateway PGs to the data center. One drawback is recovering reporting data after a network failure. If the network connection between the parent site and the Unified CCE System PGs at the child drops, all reporting at the parent site is lost for that period.

**Note**

You can deploy the Unified CCE Gateway PG locally to the Unified CCE System PG. Then, if the connection between the parent and child sites drops, the historical data in the parent site updates when the network connection comes back.

Another drawback with centralizing the Unified CCE Gateway PGs is higher network bandwidth requirements for the connections between the PGs.

Bandwidth for Unified CCE Gateway PG to Central Controller

No special considerations are necessary for the PG-to-CC connection over other TDM PGs.

If you do not use agent reporting, then leave that setting disabled to avoid sending unnecessary data over the link.

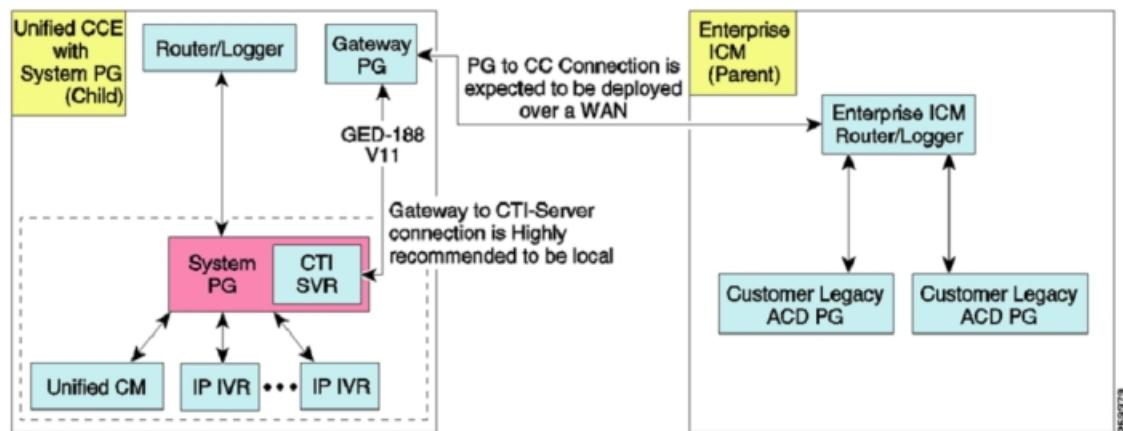
Related Topics

[Bandwidth and Latency Requirements, on page 245](#)

Bandwidth for Unified CCE Gateway PG to System PG

The following figure shows the connection between the parent PG/PIM and the child system PG.

Figure 102: Connection Between Gateway PG and System PG

**Note**

In general, you deploy the Gateway PG on the same VM with the System PG that it is monitoring. For an outsource model, you can deploy the Gateway PG remote from the System PG.

The following factors affect the amount of data coming over the link once it is initialized:

- Message sizes can vary depending on their content (such as the size of extensions, agent IDs, and call data). For example, a Route Request with no data is a small message. If all call variables and ECC variables are populated with large values, the data drastically affects the size of the message.

- Call scenarios can cause great variation in the number of messages per call that are sent over the line. A simple call scenario can send 21 messages over the line. More complex call scenarios involving queuing, hold retrieves, conferences, or transfers send even more messages over the line for each call.
- The more skill groups to which an agent belongs, the more messages are sent over the line. In a simple call scenario, each additional skill group adds two messages per call. These messages are approximately 110 bytes each, depending on field sizes.

Bandwidth Calculation for Basic Call Flow

A basic call flow (simple ACD call without other steps) with a single skill group typically generates 21 messages. Plan for a minimum required bandwidth of 2700 bytes/second.

In a basic call flow, there are four places where call variables and ECC data can be sent. If you use call data and ECC variables, they are sent four times during the call flow. Using many variables could easily cause the 2700 bytes/second of estimated bandwidth per call to more than double.

**Note**

Call variables used on the child PG are sent to the parent PG regardless of their use or the setting of the MAPVAR parameter. For example, assume that the child PG uses call variables 1 through 8 but the parent PG never uses those variables. If MAPVAR = EEEEEEEEEE, meaning Export all but Import nothing, the variables are sent to the PG where the filtering takes place. The bandwidth is still required. But, if the map setting is MAPVAR = IIIIIIII, Import all but Export nothing, then bandwidth is spared. Call variable data is not sent to the child PG on a ROUTE_SELECT response.

Basic Call Flow Example

Assume a call rate of 300 simple calls per minute (five calls per second). The agents are all in a single skill group with no passing of call variables or ECC data. The required bandwidth in this case is:

$$5 * 2700 \text{ bps} = 13,500 \text{ bps} = 108 \text{ kbps}$$

**Note**

A more complex call flow or a call flow involving call data could easily increase this bandwidth requirement.

Unified CCE System Peripheral

The Unified CCE System Peripheral acts as a single logical Unified ICM peripheral that combines the functionality of the VRU peripherals and a Unified Communications Manager peripheral. Unified CCE treats the Unified IP IVR and Unified Communications Manager peripherals as a single peripheral eliminating the need to translation-route calls to the Unified IP IVR for treatment and queuing. If multiple Unified IP IVRs are configured, the Unified CCE System peripheral automatically load-balances calls between the Unified IP IVRs that have available capacity.

As a single peripheral, Termination Call Detail (TCD) records and other reporting data include the information for the call during the entire time the call is on the peripheral. Instead of getting up to three TCDs for each call (one for the original route, one for the VRU, and one for the agent handle time), the Unified CCE System PG generates only a single record.

The Unified CCE System PG does not support Unified CVP. All queuing and treatment in the Unified CCE System PG uses Unified IP IVR. You can use a separate Unified CVP on its own PG with the Unified CCE System Peripheral.

Parent/Child Limitations

Precision Routing

Precision Routing is not supported in a parent/child deployment. Precision Routing does not support the Unified CCE System Peripheral.

Multichannel Routing

In parent/child configurations, there is no multichannel routing and integration through the parent Unified ICM. Media Routing PGs must connect to the child Unified CCE. A separate Cisco Interaction Manager or partition is required for each child.

Enterprise Unified CCE Peripheral deployments

You cannot use Enterprise Unified CCE (CallManager and VRU deployed as separate peripherals) deployments where Unified CCE is a child to a Unified ICM. Use a Unified CCE System Peripheral deployment for that solution.

Whisper Announcement

Whisper Announcement only supports a specific Parent/Child configuration. That configuration queues calls and sources whisper announcements with an Unified IP IVR on the child system PG. If you also use agent greetings, the configuration also requires a dedicated CVP at the child on a dedicated VRU PG to provide agent greetings. Cisco must approve any use of Whisper Announcement in Parent/Child configurations. Cisco must analyze and approve any such designs.

Whisper Announcement with Unified IP IVR in Parent/Child has no impact on agent sizing on the Child System PG, but incurs great impact on the Unified IP IVR.

Agent Greeting

Agent Greeting only supports a specific Parent/Child configuration. That configuration queues calls at an Unified IP IVR on the child system PG. The configuration requires a dedicated CVP at the child on a dedicated VRU PG to provide the agent greetings. Cisco must approve any use of Agent Greeting in Parent/Child configurations. Cisco must analyze and approve any such designs.

Your configuration must meet the following conditions:

- Use Unified IP IVR on the child for call treatment, queuing, and Whisper Announcement.
- Use CVP on the child only for Agent Greeting. Do not use CVP for call queuing. Use a separate VRU PG for the CVP.

Network Consultative Transfer

In a parent/child deployment, you cannot transfer calls terminating on child systems by network consultative transfer (NCT) through the routing clients on the parent. Although NCT works for TDM ACDs, and the parent/child deployment architecture is similar, the parent/child deployment does not work the same.

For a TDM PG, the CTI Server is connected to the ACD PG, which is part of the parent system. This arrangement is the same as a CTI-Server connected to the Gateway PG. In parent/child deployments, CTI connects to the child PG. Having CTI connected to the child PG does not provide the network call ID and other information that are needed for allow network consultative transfer.

**Note**

When a post route is initiated to the parent system from the child, network blind transfer is possible using any client (for example, Unified CVP) on the parent system.

Active Directory Deployments for Parent/Child

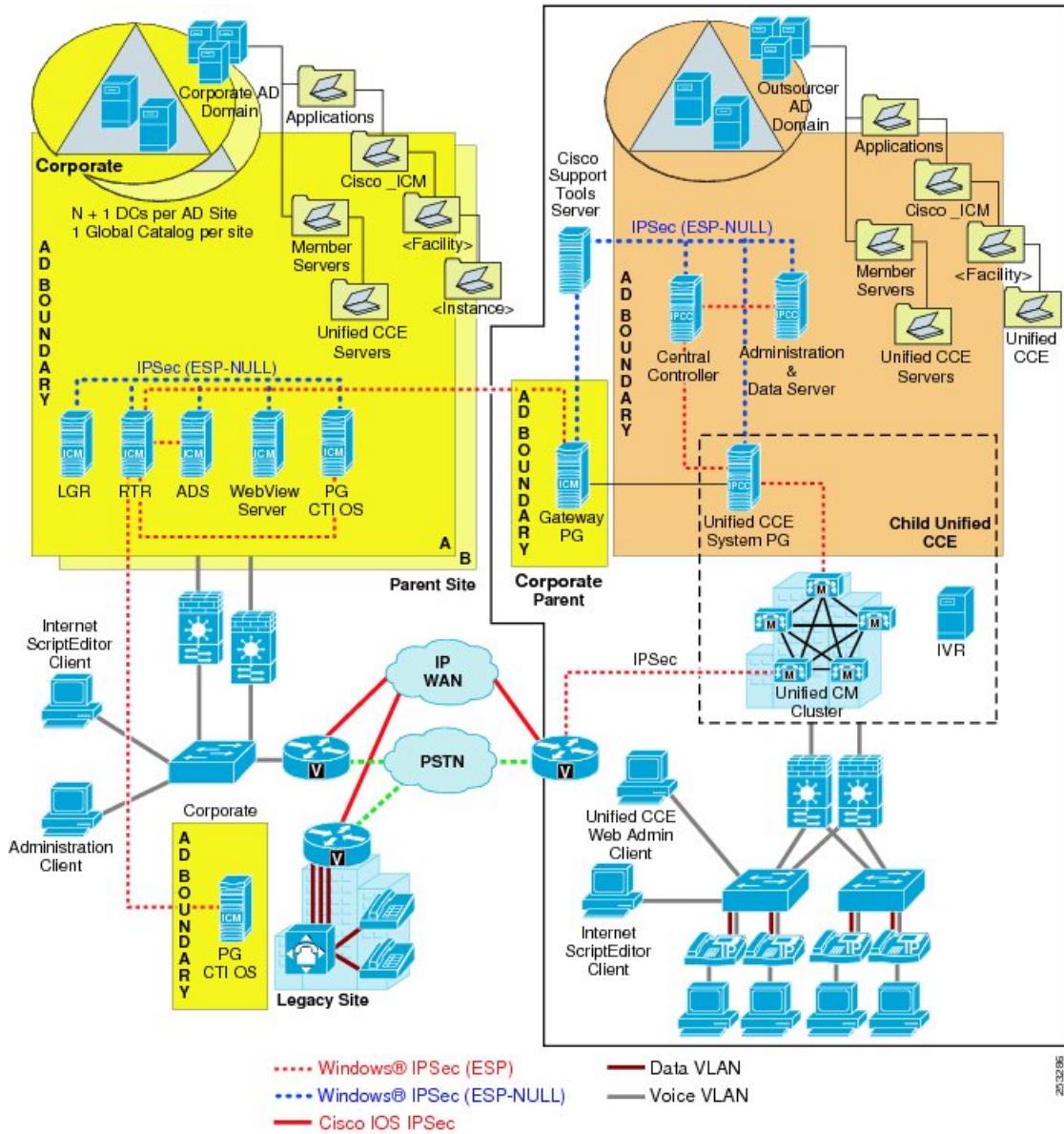
You can deploy parent/child systems on the same AD Domain or Forest, and in disparate AD environments. The common scenario for this deployment occurs when an outsourced contact center site houses the child Unified CCE system. In this case, the parent AD domain contains the Gateway PG that is a parent node. (Do *not* use Workgroup membership because of the administration limitations.) This deployment supports remote branch offices with PGs that are members of the central site domain which contains the Routers, Loggers, and Distributors.

The topology shown in the following figure represents the AD Boundaries for the two AD domains in this deployment. The figure also shows to which domain the application servers are joined. The parent AD Domain Boundary extends beyond the central data center. The parent AD Boundary includes the Unified ICM Central Controllers and accompanying servers with the ACD PG (at the legacy site) and Gateway PG at the child Unified CCE site. The child Unified CCE site and its AD Boundary have the Unified CCE servers as members.

Unified CCMP for Parent/Child Deployments

The AD Boundary can be part of an outsourcer corporate AD environment, or the AD Boundary can be a dedicated AD domain for Unified CCE.

Figure 103: Active Directory and Firewall Deployment



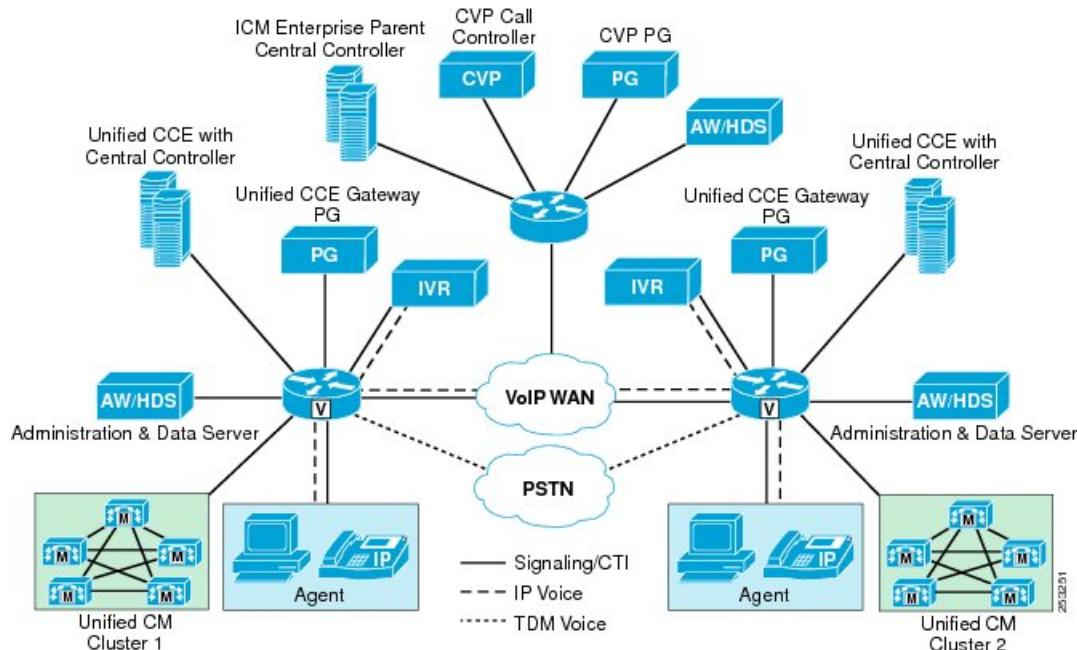
Unified CCMP for Parent/Child Deployments

In parent/child deployments, a single Unified CCMP instance connects to each of the child Unified CCE Administration & Data servers. Configure these servers as physically separate Primary Administration & Data Servers. Each child instance appears as a *tenant* within Unified CCMP. Resources added through Unified CCMP are linked to a tenant. The standard process replicates the added resource from the Unified CCE child to its parent.

Parent/Child Deployments Across Sites

The Parent/Child deployment provides local, distributed call processing with a local Unified Communications Manager and Unified CCE at each site (child). A centralized Unified ICM Enterprise parent controls the child site for enterprise-wide routing, reporting, and call control. This deployment is more tolerant of WAN outages, with each site continuing to operate during an outage. The following figure shows this deployment.

Figure 104: Multisite Deployment with Distributed Call Processing and Parent/Child



In this design, there is a parent Unified ICM Enterprise system deployed with Unified CVP and its own Administration & Data server. Each distributed child site is a complete Unified CCE deployment consisting of Central Controller on one or more VMs. A local Administration & Data Server for Unified CCE performs configuration, scripting, and reporting tasks for that specific site. A Unified CCE Gateway PG connects Unified CCE to the Unified ICM parent and it is part of the Peripheral Gateways deployed on the parent Unified ICM.

An optional deployment for the Unified CCE Gateway PG is to colocate it with the Unified CCE System PG, under the following guidelines:

- If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are the same, then use different PG numbers for the PGs.
- If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are different, then you can use the same PG number for the PGs.
- Do not add any other PGs (such as a VRU PG or MR PG) to this VM.

In this design, the local Unified CCE deployments act as their own local IP ACDs with no visibility to any of the other sites in the system. Agents at Site 1 cannot see any of the calls or reports from Site 2 in this

deployment. Only the Unified ICM Enterprise parent system has visibility to all activity at all sites connected to the Unified ICM Enterprise system.

The Unified CVP at the Unified ICM parent site controls the calls coming into the distributed sites. Unified CVP provides call queuing and treatment in the VXML Browser in the Voice Gateway (VG). Configure the Unified CVP on the parent to use Unified CVP Router Requery to take control of the call during a failure or answer timeout. The child Unified CCE cannot terminate the ingress call to a child Unified CVP or a child Unified IP IVR. The local Unified IP IVR servers only provide a local backup for the connection from these VGs to the parent Unified CVP Call Control server. The local Unified IP IVR also provides local queue treatment for calls that the local agents do not answer (RONA) rather than sending the call to the Unified CVP to be requeued.

The child Unified CCE deployments can also transfer calls across the system between the sites using Unified ICM post-routing by the Unified CCE Gateway PG. The Unified CCE Gateway PG allows the child Unified CCE to ask the Unified ICM to transfer a call to the best agent at another site or to queue it centrally for the next available agent.

Unlike traditional Unified CCE deployments with distributed Unified Communications Manager Peripheral Gateways, the parent/child deployment provides for complete local redundancy at the contact center site. The local Unified CCE takes over call processing for inbound calls from the Unified CVP gateways and provides local call queuing and treatment in the local Unified IP IVR. This configuration provides complete redundancy and 100% up-time for contact centers that cannot be down because of a WAN failure.

For customers who have Unified ICM already installed with their TDM ACD platforms, this approach can be useful when they want to do the following:

- Add new sites with Unified CCE
- Convert an existing site to Unified CCE

The approach allows the Unified ICM to continue performing enterprise-wide routing and reporting across all the sites while inserting new Unified CCE technology on a site-by-site basis.



Note

Unified CVP can be at both the parent and child. The call flows are similar for Unified CVP at the parent and IP IVR at the child. One key difference is that information about queued calls at the child Unified CVP are not available at the parent (through the Unified CCE Gateway PG). This difference means that you cannot use routing elements like the minimum expected delay (MED) over services or CallsQNow in the parent.

Advantages

- Unified CVP provides a virtual network queue across all the distributed sites controlled by the parent Unified ICM. The parent Unified ICM has visibility into all the distributed sites and sends the call to the next available agent from the virtual queue.
- Each distributed site can scale up to the maximum number of supported agents on a single Unified CCE deployment. Multiple Unified CCE Central Controllers can be connected to a single cluster to scale up to the maximum number of supported agents per cluster. The Unified CCE Gateway PG on the parent connects the Unified CCE systems to the parent Unified ICM. The Unified CCE Gateway PG can scale up to the maximum number of supported agents per parent Unified ICM Enterprise system.
- All or most VoIP traffic can be contained within the LAN of each site, if desired. The QoS WAN is required for voice calls to be transferred across sites. Use of a PSTN transfer service (for example, Take

Back and Transfer or Transfer Connect) eliminates that need. If desired, a small portion of calls arriving at a particular site can be queued for agent resources at other sites to improve customer service levels.

- Unified ICM pre-routing can load-balance calls based on agent or Unified CVP session availability. Unified ICM can also route calls to the best site to reduce WAN usage for VoIP traffic.
- Failure at any one site has no impact on operations at another site.
- Each site can be sized according to the requirements for that site.
- The parent Unified ICM Central Controller provides centralized management for configuration of routing for all calls within the enterprise.
- The parent Unified ICM Central Controller provides the capability to create a single enterprise-wide queue.
- The parent Unified ICM Central Controller provides consolidated reporting for all sites.

Disadvantages

The parent/child deployment usually requires a higher number of VMs. The extra VMs are needed for the increased number of software components (additional Unified CCE Gateway PGs required if colocating with Unified CCE System PG is not an option, additional Central Controller for each child, and so forth).

Requirements

- Colocate the Unified CCE Gateway PG, Unified Communications Manager cluster, Unified IP IVR, and Unified CCE (if possible) at the contact center site.
- The communication link from the parent Unified ICM Central Controller to the Unified CCE Gateway PG must be sized properly and provisioned for bandwidth and QoS.
- Gatekeeper-based or RSVP agent-based call admission control is used to reroute calls between sites over the PSTN when WAN bandwidth is not available. It is best to ensure that adequate WAN bandwidth exists between sites for the maximum amount of calling that can occur.
- If the communication link between the Unified CCE Gateway PG and the parent Unified ICM Central Controller is lost, then all contact center routing for calls at that site is put under control of the local Unified CCE. Unified CVP-controlled ingress Voice Gateways have survivability TCL scripts to redirect inbound calls to local Unified Communications Manager CTI route points and the local Unified IP IVR are used to handle local queuing and treatment during the WAN outage. This feature of the parent/child deployment provides complete local survivability for the call center.
- While two intercluster call legs for the same call do not cause unnecessary RTP streams, two separate call signaling control paths remain intact between the two clusters (producing logical hair-pinning and reducing the number of intercluster trunks by two). Consider the percentage of intersite transfers when sizing intercluster trunks capacities.
- Latency between parent Unified ICM Central Controllers and remote Unified CCE Gateway PGs must not exceed 200 ms one way (400-ms round trip).

Geographically Redundant Child Data Centers (Using Unified IP IVR)

Globalization, security, and disaster recovery considerations are driving business to diversity locations across multiple regions. In addition, organizations want to distribute workloads between computers, share network

Geographically Redundant Child Data Centers (Using Unified IP IVR)

resources effectively, and increase the availability of critical applications. Geographically redundant data centers split critical applications across two data centers. Enterprises deploy geographically redundant data centers to minimize planned or unplanned downtime and share data across regions.

Geographically redundant data centers have a minimum of two load balancers, one in each data center. You can use two load balancers for each data center for local redundancy.

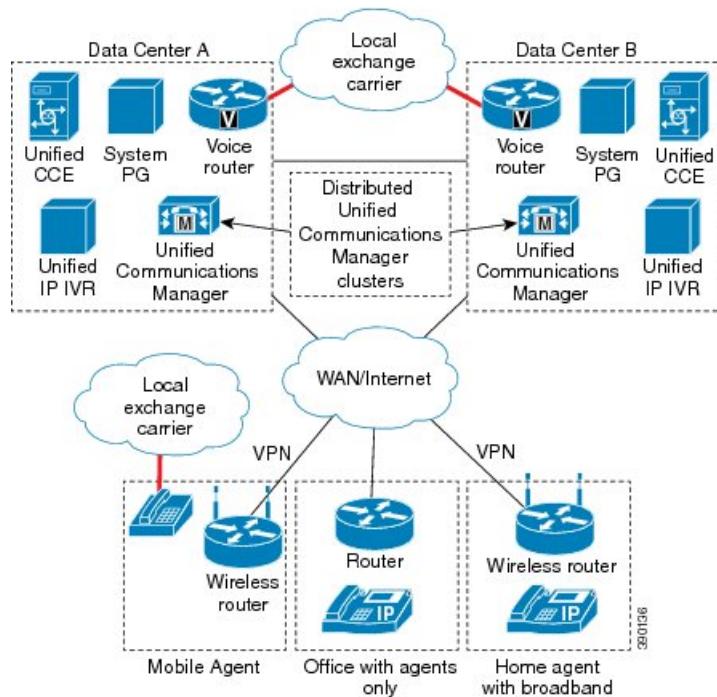
Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

Geographically Redundant Child Data Centers with CoW

The following figure shows geographically redundant Child data centers with clustering over the WAN (CoW) using Unified IP IVR.

Figure 105: Geographically Redundant Child Data Centers with Clustering over WAN (Using Unified IP IVR)



Geographically redundant data centers use clustering over the WAN, Unified Communications Manager clusters, and 1:1 redundancy for IP IVR, SIP proxy, voice gateways, and Cisco Unified Intelligence Center for example.

Latency requirements across the high-availability (HA) WAN must meet the current Cisco Unified Communications requirements for clustering over the WAN. Unified Communications Manager allows a maximum latency of 40 ms one way (80 ms for a round trip).

Certain fault tolerant networks can carry all your traffic on a single network, for example, Multiprotocol Label Switching (MPLS) or SONET. For such networks, keep the public and private traffic on separate routes within the network and respect standard latency and bandwidth.

Provision WAN connections to the agent sites with bandwidth for voice, control, and CTI. You can have a local voice gateway at remote sites for local and 911 calls. For more information, see Cisco Unified Communications and Collaboration Solutions Design Guidance at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html

In a balanced deployment, central site outages include loss of half of the ingress gateways. Scale gateways to handle the full load in both sites if one site fails.

Carrier call routing must be able to route calls to the alternate site during failures.

A single Unified CCE system peripheral controls all the Unified IP IVRs and the Unified Communications Manager. Unified IP IVR distributes the calls to the least loaded Unified IP IVR. A Unified IP IVR in data center B can treat calls coming into data center A. Both A- and B-sides of Unified CCE can identify all the Unified IP IVRs. PIM activation logic determines if the A- or B- side PIM connects to each of the Unified IP IVRs. This process means that the PG at data center A can connect to the Unified IP IVR at data center B. Make sure that you size the WAN appropriately.

To avoid the bandwidth overhead, consider using Unified CVP for clustering over the WAN deployments. Unified CVP allows higher scalability per cluster than Unified IP IVR.

Related Topics

[Geographically Redundant Data Centers with Clustering over WAN, on page 56](#)

Unified IP IVR-Based Child Data Centers with Distributed Unified Communications Manager

If you have remote offices with agents, gateways, and Unified Communications Manager clusters, the clusters at the data centers are typically independent.

The remote office has a WAN connection back to the data centers. Each cluster is independent, with its own agents and PG pairs. Each data center uses subscribers that are local to the data center because JTAPI is not supported over the WAN. For example, data center A cannot use the subscribers in data center B. The Unified CCE central controller, Unified Intelligence Center, load balancer, SIP proxy server, and Unified IP IVR are located in the data centers. TDM and VXML voice gateways are located at the remote office with local PSTN trunks.

Related Topics

[Bandwidth Provisioning and QoS Considerations, on page 237](#)

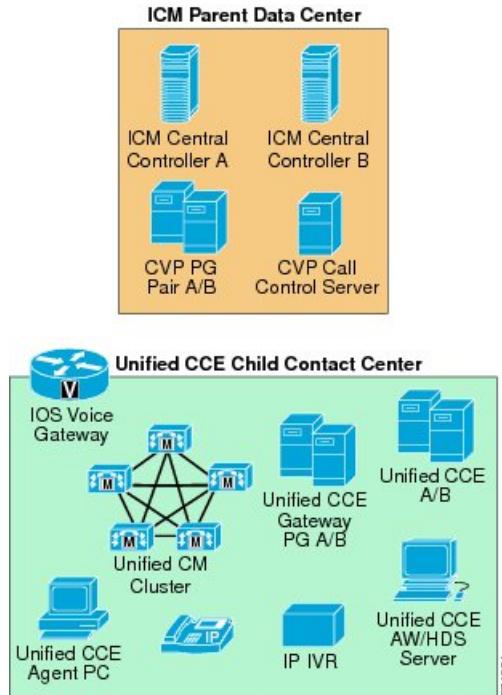
Unified CCE High Availability with Unified ICM

In a parent/child deployment, the Unified ICM acts as the parent controlling one or more Unified CCE child ACDs. The Unified ICM system is the network call routing engine for the contact centers. The network queuing uses the Unified CVP and Unified CCE Gateway PGs to connect child Unified CCE systems. The child Unified CCE systems are individual ACD systems fully functional with local call processing in case they lose their WAN connection to the parent system. This configuration provides a high level of redundancy

Parent/Child Call Flows

and availability to the Unified CCE solution. These qualities allow sites to remain functional as Unified CCE sites even if they are cut off from centralized call processing resources.

Figure 106: Parent/Child Deployment



Parent/Child Call Flows

The following sections describe the call flows between the parent and child.

Typical Inbound PSTN Call Flow

In a typical inbound call flow from the PSTN, the carrier network directs calls to the contact center sites using a predefined percent allocation or automatic routing method. These calls terminate in the Unified CVP VGs (Voice Gateways) at the contact center locations under control of Unified CVP on the Unified ICM parent .

The inbound call flow is as follows:

- 1 The call arrives on the Unified CVP VG at the Parent ICM Data Center.
- 2 The Unified CVP VG maps the call by dialed number to a particular Unified CVP Server at the parent site. The VG then sends a new call event to the Unified CVP Server.
- 3 The Unified CVP Server sends the new call event message to the Unified CVP VRU PG at the Unified ICM parent site.
- 4 The Unified CVP PG sends the new call message to the Unified ICM parent. The Unified ICM uses the inbound dialed number to qualify a routing script to determine the proper call treatment (messaging) or agent groups to consider for the call.
- 5 Unified ICM instructs Unified CVP to hold the call in the VG at the site. Unified CVP waits for an available agent while directing specific instructions to play .wav files for hold music to the caller.

- 6 When an agent becomes available, the Unified ICM instructs Unified CVP to transfer the call to the available agent by using a translation route. (The agent is not at the same physical site, but located across the WAN.) Any data collected about the call in the Unified ICM parent Unified CVP is transferred to the remote PG (either a TDM, legacy PG, or one of the Unified CCE Gateway PGs for Unified CCE).
- 7 The call arrives at the targeted site on a specific translation route DNIS that the Unified ICM parent selected. The PG at the child site is expecting a call to arrive on this DNIS to match up with any pre-call CTI data associated with the call. The local ACD or Unified CCE performs a post-route request to the PG (either TDM PG or Gateway PG depending on the target ACD) to request the CTI data as well as the final destination for the call (typically the lead number for the skill group of the available agent).
- 8 If the agent is no longer available, Unified CVP at the parent site uses the Router Requery function in the ICM Call Routing Script to select another target automatically.

Post-route Call Flow

You use post-routing for calls already at a peripheral ACD or VRU that you want to route to another agent or location intelligently. If an agent gets a call in the ACD or Unified CCE that must go to a different skill group or location, the agent can use the post-route functionality to reroute the call.

The post-route call flow is as follows:

- 1 The agent transfers the call to the local CTI route point for reroute treatment using the CTI agent desktop.
- 2 The reroute application or script makes a post-route request to the Unified ICM parent by using the local Unified CCE Gateway PG connection.
- 3 The Unified ICM parent maps the CTI route point from Unified CCE as the dialed number and uses that number to select a routing script. This script returns a label or routing instruction that can move the call to another site, into a different skill group on the same site, or to a Unified CVP node for queuing.
- 4 Unified CCE receives the post-route response from the Unified ICM parent system. Unified CCE then uses the returned routing label as a transfer number to send the call to the next destination.

Parent/Child Fault Tolerance

The parent/child model provides for fault tolerance to maintain a complete ACD with Unified CCE deployed at the site, with local IP-PBX and call treatment and queuing functionality.

Unified CCE Child Loses Connection to Unified ICM

A WAN failure between the child and the parent isolates the local Unified CCE system from the parent and the Unified CVP VG. Calls coming into the site no longer get treatment from the Unified CVP under control of the Unified ICM parent. Replicate the following functionality locally, depending on the configuration at the child site:

- For Unified CCE child configurations using local IP IVR resources for queue and treatment:
 - The local VG must have dial peer statements to pass control of the calls to the local Unified CM cluster if the parent Unified CVP Server cannot be reached. The local Unified CM cluster must have CTI route points mapped to the inbound DNIS or dialed numbers that the local VG presents if the parent Unified CVP Server is not reached.
 - Configure the local IP IVR with appropriate audio files and applications that the child system can call locally to provide basic call treatment.

- The child CCE Routing Script must handle queuing of calls for agents in local skill groups. The script must instruct the IP IVR to play the treatment in-queue while waiting for an agent.
- To allow the agents full access to customer data for routing and screen pops, provision locally any data lookup or external CTI access that parent system normally provides.
- Any post-routing transfer scripts fail during this outage, so configure Unified CCE to handle this outage or prevent the post-route scripts from being accessed.
- For Unified CCE child configurations using local Unified CVP resources for queue and treatment:
 - The local VG must have dial peer statements to pass control of the calls to the local Unified CVP Server at the child site. To process these calls locally at the child site, configure in the child Unified CCE the inbound DNIS or dialed numbers that the local VG presents to the child Unified CVP.
 - Configure the local VXML Gateways and Unified CVP Servers with appropriate .wav files and applications that the child system can call locally to provide basic call treatment.
 - Replicate self-service or Unified CVP Studio VXML applications that the parent Unified ICM normally provides. Use the Unified CVP Server (web application server) at the child site to generate the dynamic VXML for these applications.
 - The child Unified CCE Routing Script must handle queuing of calls for agents in local skill groups. The script must instruct the local Unified CVP at the child site to play the treatment in-queue while waiting for an agent.
 - To allow the agents full access to customer data for routing and screen pops, provision locally any data lookup or external CTI access that parent system normally provides.
 - Any post-routing transfer scripts fail during this outage, so configure Unified CCE to handle this outage or prevent the post-route scripts from being accessed.

Unified CCE Gateway PG Cannot Connect to Unified ICM

If the Unified CCE gateway PG fails or cannot communicate with the Unified ICM parent, the Unified ICM parent cannot detect the state of the agents at the child. But, in some cases, the Unified ICM parent Unified CVP can still control the inbound calls. In this case, the Unified ICM parent does not know if the remote Unified CCE gateway PG failed or if the actual Unified CCE ACD failed locally.

The Unified ICM at the parent location can automatically route around this site, considering it down until the PG comes back online and reports agent states again. Alternatively, the Unified ICM can also direct a percentage of calls as blind transfers to the site Unified CCE using the local inbound CTI route points on Unified CM. This method would present calls with no CTI data from Unified CVP, but it would allow the agents at the site to continue to get calls locally with their Unified CCE system.

If the local Unified CCE child system fails, the Unified CCE gateway PG cannot connect to it. The Unified ICM parent then considers all agents to be off-line and not available. If local cluster receives calls while the child Unified CCE system is down, the call-forward-on-failure processing takes over the call for the CTI route point. This method redirects the call to another site or an answering resource to play a message telling the caller there is a problem and to call again later.

Reporting and Configuration Impacts

If the Unified CCE child disconnects from the Unified ICM parent, the local ACD still collects reporting data and allows local administrators to change the child routing scripts and configuration. The Unified CCE gateway PG at the child site caches these objects and stores them to be sent when the Unified ICM parent is available. This functionality is available only if the Unified CCE gateway PG is colocated at the child Unified CCE site.

Other High Availability Considerations

You can install multichannel components, such as Cisco Unified Web and E-mail Interaction Manager and Cisco Outbound Option, only at the child Unified CCE level, not at the parent. They are treated as nodal implementations on a site-by-site basis.

Traditional ACD Integration

Enterprises that want to integrate traditional ACDs with their Unified CCE use a parent/child deployment. In these deployments, the Unified ICM and Unified CCE each have a Central Controller or a hybrid deployment where Unified Communications Manager PGs with TDM ACD PGs, Gateway PGs, or both use the same Central Controller. Several options exist within those categories depending on how the calls are routed within the deployment.

Hybrid Deployment with Fixed PSTN Delivery

As an alternative to pre-routing calls from the PSTN, the PSTN can deliver calls to just one site or to split the calls across the two sites according to a set of static rules provisioned in the PSTN. When the call arrives at either site, either the traditional ACD or the Unified Communications Manager generates a route request to the hybrid Unified ICM/CCE to determine which site is best for this call. If the call is for an agent at the opposite site from where the call was originally routed, you require TDM circuits between sites. The determination of where calls are routed (and if and when they are transferred between sites) depends on the enterprise business environment, objectives, and cost components.

Hybrid Deployment with Unified CVP

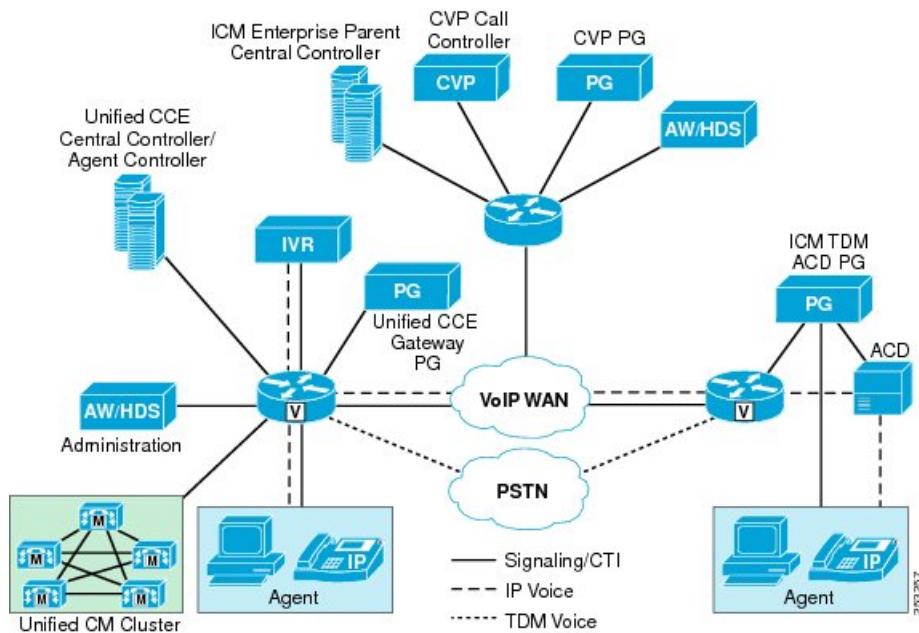
Customers can choose to front end all calls with Unified CVP to provide initial call treatment and queuing across both the TDM ACD and Unified CCE agents.

In this design, all calls first come to the Voice Gateway (VG) controlled by Unified CVP. The Unified ICM/CCE Call Router then directs the calls. Unified ICM/CCE uses the PG connections to the TDM ACD and Unified CCE PG to monitor for available agents. Calls are queued in Unified CVP until an agent becomes available in either environment. When a call transfers to the TDM ACD, the call comes into the VG on a T1 interface from the PSTN carrier network and goes out on a second physical T1 interface to appear as a trunk on the TDM ACD (known as "hairpinning the call"). Most TDM ACDs cannot accept inbound calls in IP from the VG and require this physical T1 interface connection. Unified CCE agents receive their calls directly over the IP voice network.

ACD Integration and Parent/Child Deployments

The following figure illustrates the parent/child.

Figure 107: Parent/Child Integration of Traditional ACD with Unified CCE



In this model, the Unified ICM Enterprise parent has PGs connected to a Unified CCE System PG at one site and a Unified ICM TDM ACD PG at a second site. In this model, Unified ICM still provides virtual enterprise-wide routing, call treatment, and queuing with the distributed Unified CVP Voice Gateways at the sites. Unified ICM also has full visibility to all the sites for agents and calls in progress. The difference in this model is that Unified CCE provides local survivability. If it loses connection to the Unified ICM parent, the calls are still treated locally just as they are at the TDM ACD site.

Traditional VRU Integration

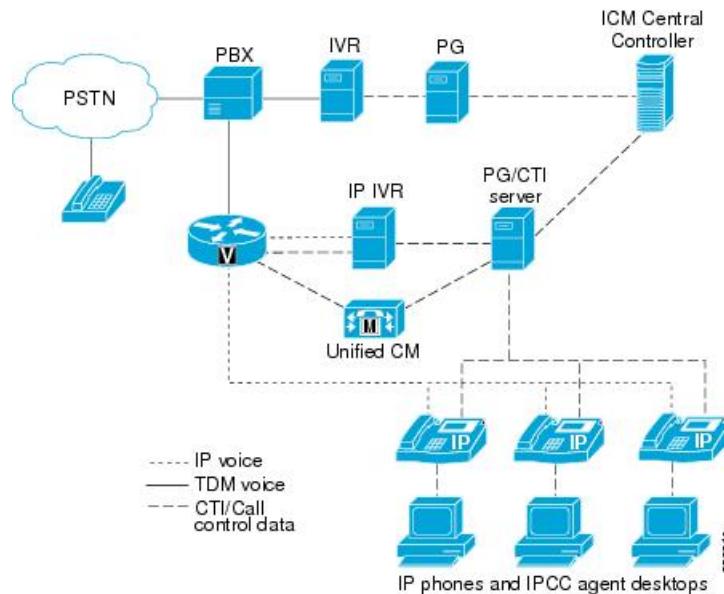
There are numerous ways to integrate traditional VRUs into a Unified CCE deployment. The following sections discuss the factors that can determine the best way. The primary consideration is determining how to eliminate or reduce VRU double trunking when transferring the call from the VRU.

Integration Through PBX Transfer

Many call centers have existing traditional VRU applications that they are not prepared to rewrite. To preserve these VRU applications and integrate them into a Unified CCE environment, the VRU must have an interface to Unified CCE. That VRU interface to Unified CCE is the Service Control Interface (SCI). The SCI enables the VRU to receive queuing instructions from Unified CCE. In the PBX model, the SCI is not required.

Even if the VRU has the SCI interface, deploy Unified CVP or Unified IP IVR for all call queuing. This method prevents any extra use of the traditional VRU ports. In addition, use of the Unified IP IVR for queuing provides a way to re-queue calls on subsequent transfers or RONA treatment.

Figure 108: Traditional VRU Integration Using PBX Transfer



In this design, calls come first to the PBX from the PSTN carrier network on a standard T1 trunk interface. The PBX typically uses a hunt group to transfer the call to the VRU, putting all VRU ports into the hunt group as agents in auto available mode. The PBX looks like the PSTN to Unified CCE because it does not have a PG connected to the PBX. Unified CCE cannot track the call from the original delivery to the VRU. Unified CCE only has call data from the time the call arrived at the VRU and the VRU informed Unified CCE of the call.

When the caller opts out of the VRU application, the VRU sends a Post-Route to Unified CCE. Unified CCE looks at the agent states across the system. Unified CCE then selects the agent to send the call to or translation-routes the call to the Unified IP IVR for queuing.

When the call is sent to an agent or into the queue, the call comes into the PBX from the PSTN on a T1 trunk port and then goes out to a VG on a second T1 trunk port in the PBX. This connection is used for the life of the call.

If you want to track the call from its entry at the PBX or if you want to capture the caller ANI or original dialed number, you can install a PG on the PBX. The PBX can request (through a Post-Route to Unified CCE) which VRU port to send the call to behind the PBX. The PBX cannot use a hunt group to deliver the call from the PBX to the VRU. Unified CCE requires direct DNIS termination to ensure that the translation route maintains the call data collected in the PBX and makes it available to the VRU.

Integration Through PSTN Transfer

This model is similar to the PBX Transfer model. In this model, the VRU invokes a PSTN transfer (instead of a PBX transfer) to release the traditional VRU port. Again, the Unified IP IVR does all queuing to avoid

Integration Through VRU Double Trunking

any additional occupancy of the traditional VRU ports and any double trunking in the VRU. Unified CCE passes any call data collected by the traditional VRU application to the agent desktop or Unified IP IVR.

In this model, the TDM VRU is set up as a farm of VRU platforms that have direct PSTN connections for inbound calls. The VRU has a PG connection to Unified CCE which tracks all calls in the system. When a caller opts out of the VRU treatment, the VRU sends a post-route request to Unified CCE. Unified CCE returns a label that directs the call either to an agent or to the Unified IP IVR for queuing.

The label that is returned to the TDM VRU instructs it to send an in-band transfer command using transfer tones (*8 with a destination label in the carrier network). The VRU out-pulses these tones to the service provider with tone generation or plays the tones by using a recorded file.

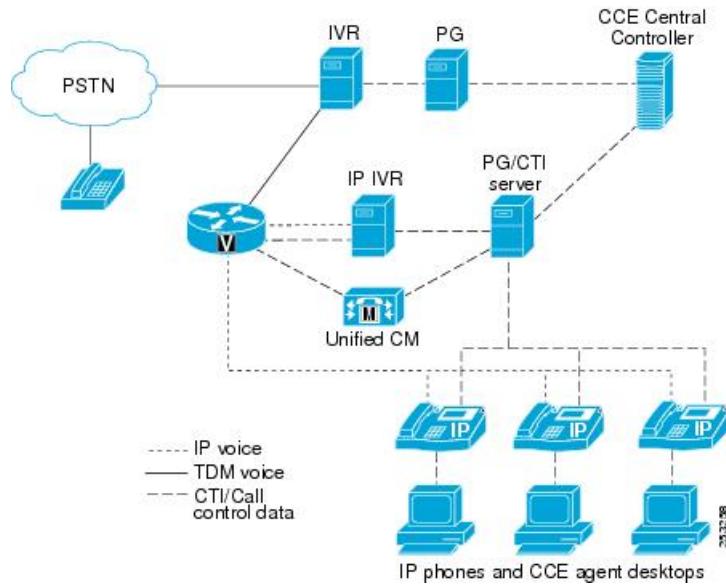
Integration Through VRU Double Trunking

Some traditional VRU applications have a high success rate where most callers are completely self-served in the traditional VRU and only a small percentage of callers are transferred to an agent. For those cases, you can double-trunk the calls in the traditional VRU for that small percentage of calls.

Unlike the previous model, if the traditional VRU has a Service Control Interface (SCI), then the initial call queuing is done on the traditional VRU. In this case, VRU double trunking does not use a second traditional VRU port to transfer the call to the Unified IP IVR. When the traditional VRU does the initial queuing, only one traditional VRU port is used for the call. However, any subsequent queuing because of transfers or RONA treatment must be done on the Unified IP IVR to avoid any double trunking.

If the traditional VRU does not have an SCI interface, then the VRU generates a post-route request to Unified CCE to determine where the call is transferred. All queuing in that scenario happens on the Unified IP IVR.

Figure 109: Traditional VRU Integration Using VRU Double Trunking



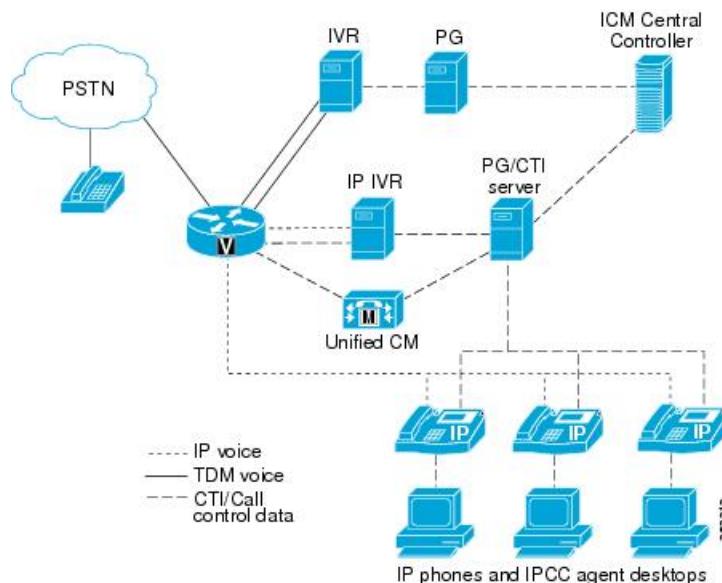
In this model, the TDM VRU is set up as a farm of VRU platforms that have direct PSTN connections for inbound calls. The VRU has a PG connection to Unified CCE which tracks all calls in the system. When a caller opts out of the VRU treatment, the VRU sends a post-route request to Unified CCE. Unified CCE returns a label that either directs the call to an agent or queues the call locally on the TDM VRU using the Service

Control Interface (SCI). The TDM VRU does the transfer to the agent by selecting a second port to hairpin the call to the Voice Gateway and to the Unified CCE agent. The hairpin connection takes up two ports for the time the call is at the agent.

Unified Communications Manager Transfer and VRU Double Trunking

Over time, you can migrate the traditional VRU applications to Unified CVP or Unified IP IVR. If a few traditional VRU applications still exist for specific scenarios, then the VRU can connect to a second Voice Gateway (VG). The Unified Communications Manager routes calls arriving at the VG from the PSTN. Unified Communications Manager can route specific DNs to the traditional VRU or let Unified CCE, Unified CVP, or Unified IP IVR determine when to transfer calls to the traditional VRU. To transfer calls in the traditional VRU to a Unified CCE agent, use a second VRU port, trunk, and VG port during the call. Ensure that transfer scenarios cannot create multiple loops or voice quality suffers.

Figure 110: Unified Communications Manager Transfer and VRU Double Trunking



In this model, Unified CVP can front end the TDM VRU using the VG to determine where to provide call treatment. Alternately, you can use Unified IP IVR and Unified Communications Manager with Unified CCE for this purpose.

With Unified CVP, calls coming into the VG immediately start a routing dialog with Unified CCE using the Service Control Interface (SCI). Based on the initial dialed number or prompting in Unified CVP, Unified CCE decides to send the call to the TDM VRU or to Unified CVP for a specific self-service application. If the call was sent to the TDM VRU, the TDM VRU sends a route request to Unified CCE when the caller opts out. The reply is not sent back to the TDM VRU but back to Unified CVP as the original routing client. Unified CVP then takes the call leg away from the TDM VRU. Unified CVP then transfers the call to the Unified CCE agent over the VoIP network or holds the call in a queue locally in the VG.

With Unified Communications Manager, calls in the VG use a subscriber CTI route point to send a route request to Unified CCE for the proper call treatment device. If the CTI route point indicates an application that still is on the TDM VRU, Unified CCE instructs the subscriber to transfer the call to the TDM VRU by hairpinning the call using a second T1 port on the VG. Unified CCE can also instruct the subscriber to

translation-route the call to the Unified IP IVR for call processing or prompting. Unified CCE can then make a subsequent transfer to the TDM VRU for further processing. When the caller opts out of the TDM VRU, it sends a post-route request to Unified CCE for a label to the TDM VRU. This label instructs the TDM VRU to transfer the call using a second T1 port on the VRU back to the VG. The VG transfers the call over to the Unified CCE agent under the Unified Communications Manager dial plan.

In the model controlled by Unified Communications Manager, the VG initially receives calls and sends them to the TDM VRU on a second T1 port. When the VRU returns the call to the Unified CCE agent, it uses a second TDM VRU port and a third port on the VG. All three VG ports are in use as long as the agent is talking with the caller. Both of the TDM VRU ports are used for the remainder of the call.



APPENDIX D

Cisco Agent Desktop

- [CAD Base Services, page 351](#)
- [Cisco Agent Desktop Solution, page 352](#)
- [CAD Silent Monitoring and Recording, page 357](#)
- [Cisco Agent Desktop Presence Integration, page 359](#)
- [Cisco Agent Desktop and NAT, page 361](#)
- [Support for IP Phones and IP Communicator, page 363](#)
- [Cisco Agent Desktop and Citrix, page 363](#)
- [Support for Mix of CAD and CTI OS Agents on the Same PG, page 364](#)
- [High Availability for Cisco Agent Desktop, page 364](#)
- [Cisco Agent Desktop Component Sizing, page 366](#)
- [Bandwidth Requirements for Cisco Agent Desktop, page 372](#)
- [Miscellaneous Deployment Considerations, page 379](#)

CAD Base Services

Cisco Agent Desktop (CAD) is a software suite that provides a feature-rich packaged solution. CAD consists of user applications and the CAD base services, which can run coresident on the Peripheral Gateway (PG) within a Unified CCE deployment and are required for CAD deployments only. The CAD base services provide redundancy and warm standby capabilities.

CAD Base Services

- Cisco Chat Service: Supports message passing and the text chat feature.
- Cisco Enterprise Service: Communicates with the Unified CCE components to provide call data to the user applications.
- Cisco Browser and IP Phone Agent Service: Provides services for CAD IPPA agent applications.
- Cisco Synchronization Service: Synchronizes the Unified CCE and CAD-specific configuration data.

- Cisco LDAP Monitor Service: Manages the storage and retrieval of CAD configuration data.
- Cisco Recording and Statistics Service: Manages the storage and retrieval of call recording, agent call, and agent state change data used in reports.
- Cisco Licensing and Resource Manager Service: Manages user licenses and controls fail-over behavior.
- Cisco Recording and Playback Service: Provides the call recording and playback feature.
- Cisco VoIP Monitor Service: Provides the voice streams for the call recording and Silent Monitoring features if server-based monitoring is used.

For more information about CAD, see the product documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/tsd-products-support-series-home.html>.

Cisco Agent Desktop Solution

The Cisco Agent Desktop (CAD) solution is a suite of packaged desktop applications and services. CAD offers a rich set of features for the contact center environment, including:

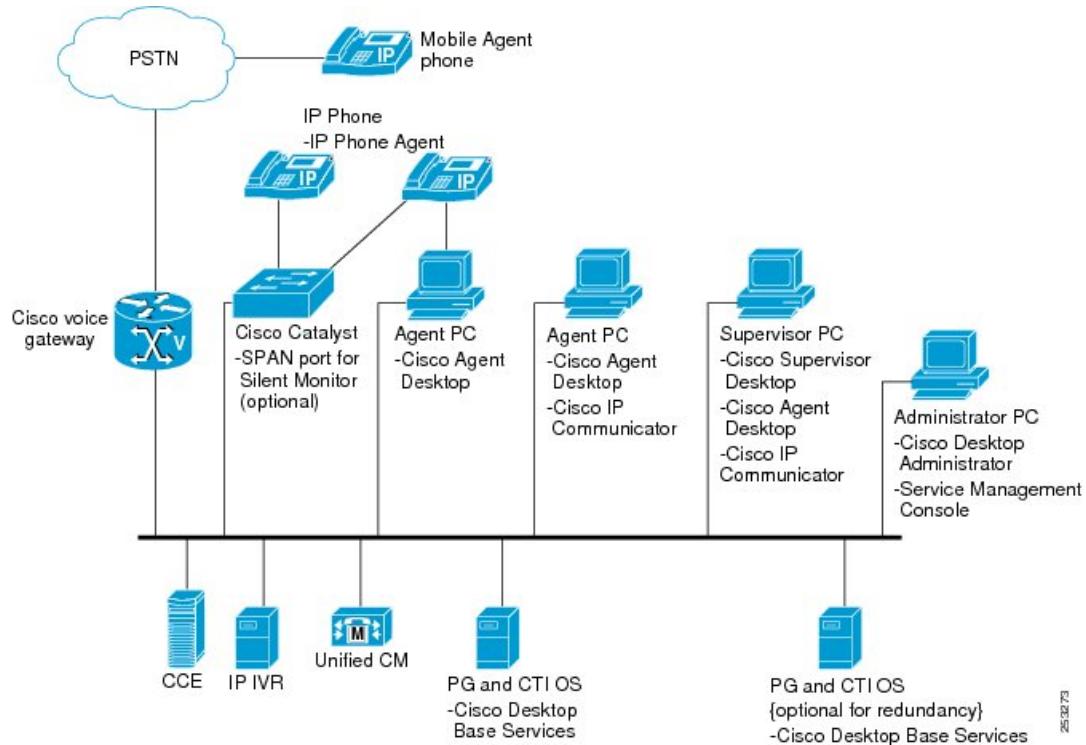
- Agent state and call control
Agent Desktop provides call control capabilities (call answer, hold, conference, and transfer) and ACD state control (ready/not ready, wrap up and so forth).
- Work Flow Automation
The Work Flow Automation feature allows an administrator to customize the agent environment and how the user applications interact with that environment. Work Flow Automation enables data processing actions to be scheduled based on telephony events (for example, popping data into a third-party application on the answer event and sending email on the dropped event). Work Flow Automation interfaces with applications written for Microsoft Windows browsers and terminal emulators. Some customizations can be as simple as using keystroke macros for screen pops.
- On-demand Recording
The supervisor (and, if enabled, the agent) can record a customer phone call for later review by a supervisor.
- Cisco IP Phone Agent service
With this XML service, agents using Cisco IP phones can log in and use their phone to perform most of the agent functions found in an agent desktop application.
- Collaboration
Supervisors can text-chat directly with agents or agent teams. Agents can text-chat with supervisors or other team members (if enabled). The supervisor can push web pages to agents and send team messages to agent desktops. This interactive collaboration enables the contact center to communicate better, increase productivity, improve customer responsiveness, and coach or train agents.
- Task Automation
Routine agent tasks, such as email, conferences to knowledge workers, launching other applications, and high-priority chat, can be configured as task buttons on the agent's toolbar to reduce call duration and improve customer responsiveness.
- Silent Monitoring
Supervisors can initiate a Silent Monitoring session with an agent on their team.

CAD User Applications

CAD user applications are for contact center agents, supervisors, and administrators and include the following:

- Cisco Agent Desktop: Windows-based agent application
- CAD IP Phone Agent (IPPA): IP phone service agent application
- Cisco Supervisor Desktop (CSD): Windows-based supervisor application
- Cisco Desktop Administrator (CDA): Web-based administrative application
- Cisco Desktop Work Flow Administrator: Windows-based work flow configuration tool

Figure 111: Cisco Agent Desktop System Configuration and Components



CAD Application Features

The following table compares some of the more important CAD features to assist users in selecting the appropriate agent application for their deployment.

Table 37: Comparison of Major CAD Features

Feature	CAD	CAD IPPA
Call Control	Yes	n/a**

Feature	CAD	CAD IPPA
VPN/Mobile Agent Support	Yes	Yes
Chat / Unified Presence Integration	Yes	No
Supports Cisco IP Communicator	Yes	No
Team Messages	Yes	No
Supports Mobile Agent	Yes	n/a
Real-time Queue and Agent Displays	Yes	Yes
Supports Cisco Outbound Dialer	Yes	No
Integrated Browser	Yes	n/a
Call event Work Flow Automation	Yes	No
Agent state Work Flow Automation	Yes	No
Supports thin client environment	Yes	n/a
Desktop Monitoring and Recording*	Yes	No
SPAN Monitoring and Recording*	Yes	Yes
Unified CM Monitoring and Recording*	Yes	Yes

*For more detailed information about supported monitoring and recording, refer to *Configuring and Troubleshooting VoIP Monitoring* at <http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/products-troubleshooting-guides-list.html>.

**Call control actions are performed by using the IP phone call control softkeys.

For more information about CAD agent applications, see the [appropriate user guide](#).

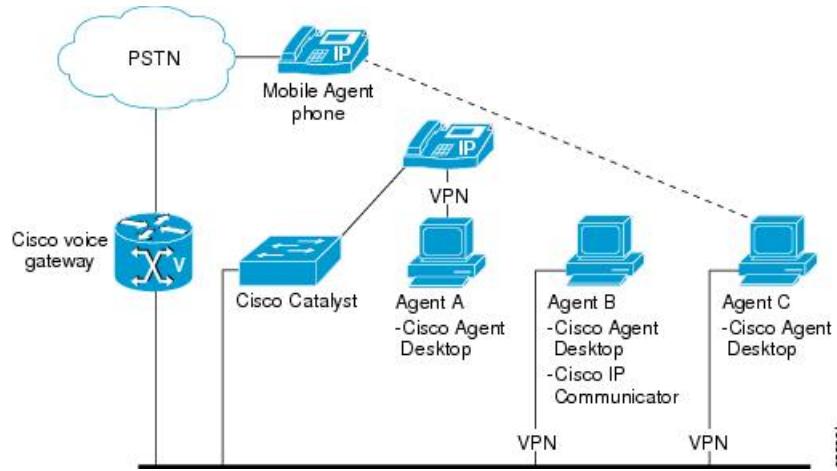
Cisco Agent Desktop

Cisco Agent Desktop is a Windows application that runs on the agent PC. It works with either a hardware IP phone or the Cisco IP Communicator soft phone. Agent Desktop interfaces with the CTI OS service for call control and agent state change events; for all other features, it communicates with the CAD services.

Agent Desktop supports Desktop, SPAN, and Unified CM Monitoring and Recording.

The figure below illustrates various ways agent desktops can be configured in a contact center.

Figure 112: CAD Agents and Components



- Agent A shows an agent who uses a hardware IP phone. The IP phone connects directly to the agent's PC through a network cable. This is the configuration required for desktop monitoring. CAD supports a VPN connection between the agent PC and the contact center network.
- Agent B shows an agent who uses Cisco IP Communicator. This configuration also supports a VPN connection to the contact center network. This is the most common configuration for mobile agents.
- Agent C shows Agent Desktop used with the Mobile Agent feature. Mobile agents are agents whose phones are not directly controlled by Unified CM. Agents might use their home phones or cell phones as their agent device. In this case, the agent provides a CTI port to associate with their remote phone when logging in. ACD calls for the logged-in agent are sent to the CTI port, which causes the call to appear at the mobile agent's phone device. There is a logical relationship (the dashed line) between the agent and the mobile phone. CAD supports a VPN connection between the agent and the contact center network in this configuration. Mobile agents can be monitored and recorded using SPAN monitoring.

For more information about Cisco Agent Desktop features and capabilities, see the *Cisco Agent Desktop User Guide* at <http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/products-user-guide-list.html>.

Cisco Agent Desktop IP Phone Agent

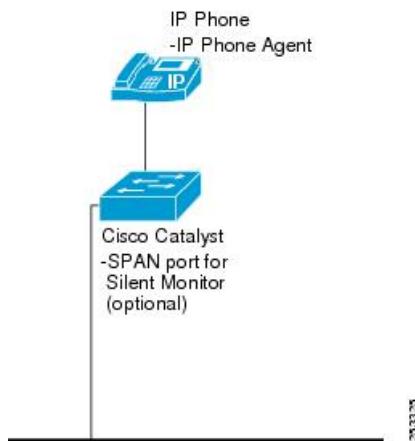
Cisco Agent Desktop IP Phone Agent (CAD IPPA) runs as an IP phone XML service. The agent is not required to have a PC. CAD IPPA includes all the basic features required by a contact center agent, as well as advanced features such as reason codes, wrap-up data, and On-demand Recording.

CAD IPPA agents can be monitored and recorded using server monitoring, and monitored using Unified CM monitoring.

For more information about CAD IPPA features and capabilities, see the CAD documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/products-user-guide-list.html>.

The following figure illustrates the components used by CAD IP Phone agents.

Figure 113: CAD IP Phone Agent Components



Cisco Supervisor Desktop

Cisco Supervisor Desktop provides a graphical view of the agent teams managed by the supervisor. An expandable navigation tree, similar to that in Windows Explorer, is used to navigate to and manage team resources.

Supervisors are able to view real-time information about the agents in a team as well as interact with those agents. The supervisor can:

- View and change an agent's state
- View contact information specific to the agent
- Silently monitor and/or record the agent's calls
- Barge-in or intercept an agent's call
- Chat with the agent using an instant message window
- Push a web page to the agent's desktop

When Supervisor Desktop is installed, an instance of Agent Desktop is installed as well. Agent Desktop is needed by the supervisor to take calls, barge in, intercept, and retrieve skill group statistics.

The Supervisor Work Flow module enables configurable actions to be triggered when specific events occur in the contact center. For example, a supervisor work flow can be set up so that whenever more than ten calls are in queue for a specified skill group, an audible alert sounds and the skill group name is highlighted in red on the supervisor's desktop. Another work flow sends an email to specified email addresses when certain events occur. The email contains information related to the condition that caused the event, as well as custom text.

Supervisors can use the Supervisor Record Viewer to review recordings and mark selected recordings for extended retention. The supervisor can also save recordings for permanent retention in a format that can be played by any media player.

For more information about Supervisor Desktop features and capabilities, see the *CTI OS Supervisor Desktop User Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* at <http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-user-guide-list.html>.

Cisco Desktop Administrator

Cisco Desktop Administrator enables an administrator to configure the CAD services and CAD client applications. Individual work flow groups containing agents and supervisors can be configured separately to provide specific functionality to particular groups of agents.

Desktop Administrator consists of two components:

- Cisco Desktop Work Flow Administrator, a Windows-based application
- Cisco Desktop Administrator, a web-based application

Cisco Desktop Work Flow Administrator is used to configure the following:

- Dial strings
- Phone books
- Reason codes
- Wrap-up data
- Record/monitor notification
- Work flow groups

Dial strings, phone books, reason codes, and wrap-up data can be configured on the global and work flow group level.

Work flows and user interfaces can be configured for specific agent types.

Cisco Desktop Administrator is used to configure the following:

- Enterprise data fields and layouts
- Silent Monitoring and recording
- Personnel and assigning users to work flow groups
- Cisco Unified Presence settings

For more information about Cisco Desktop Administrator features and capabilities, see the CAD documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/products-user-guide-list.html>.

Cisco Desktop Monitoring Console

The Cisco Desktop Monitoring Console is a Java application that monitors the status of the CAD services. It provides a convenient interface for an administrator to use to get real-time information about the CAD system.

CAD Silent Monitoring and Recording

This section describes Cisco Agent Desktop (CAD) Silent Monitoring.

**Note**

CAD recording is not suitable for use as a compliance recording solution. This functionality is best for on-demand recording or recording on a filtered list of calls only.

CAD-Based Monitoring

CAD-based monitoring consists of three types of monitoring:

- Desktop Monitoring
- Server Monitoring
- Mobile Agent Monitoring

Desktop Monitoring

Desktop Monitoring uses software running on the agent's desktop (Cisco Agent Desktop) to sniff the network traffic going to and from the agent's phone (hardware phone or software phone) for RTP packets. The monitoring software then sends the RTP packets to the appropriate software over the network for decoding. Desktop Monitoring relies on the ability for certain Cisco IP Phones to be daisy-chained with the agent's PC by using a network connection and for the phones to send all its network traffic along this connection to the software running on the PC. In this case, the packet-sniffing software can see the voice traffic flowing to and from the agent's phone. It will copy this traffic and send it to the supervisor that is monitoring the agent or to a recording service for the call to be stored and to be listened to at some later time. Desktop Monitoring is not a true service, at least from the perspective of the Service Control Manager. It is a Dynamic-Link Library (DLL), an executable module that is part of Cisco Agent Desktop.

Server Monitoring

Server monitoring uses one or more Cisco Desktop VoIP Monitor Services to sniff the network running over a Cisco Catalyst switch for voice streams. The Cisco Desktop VoIP Monitor Service looks for particular streams to and from phones being monitored or recorded. It then sends the voice packets to the supervisor desktop that is performing the monitoring or to a recording service for storage.

The Cisco Desktop VoIP Monitor Service uses the Switched Port Analyzer (SPAN) or Remote SPAN (RSPAN) monitoring feature of certain Cisco Catalyst switches to sniff the network. The switch uses the monitoring feature to copy the network traffic from one or more sources to a destination port. Sources can be ports and/or Virtual LANs (VLANs). RSPAN allows the source ports to reside on remote switches. The Cisco VoIP Monitor Service connects to the switch by using the destination port. This allows the Cisco VoIP Monitor Service to see the voice traffic going to and coming from IP phones.

Mobile Agent Monitoring

Cisco Agent Desktop has the ability to monitor and record mobile agents' RTP sessions by deploying a Cisco VoIP Monitor Service that can see traffic coming from Agent Voice Gateways (this also uses the SPAN feature).

For more information, see the Cisco Agent Desktop product documentation available on cisco.com.

Fault Tolerance for CAD-Based Monitoring and Recording

Desktop Monitoring

Desktop Monitoring is fault tolerant by design. If an agent's desktop fails, only that agent is unavailable for monitoring and recording.

Server Monitoring and Mobile Agent Monitoring

Server monitoring and mobile agent monitoring are not fault tolerant. If a Cisco Desktop VoIP Monitor Service fails, all agent phones and mobile agent Voice Gateways associated with that service is unavailable for monitoring and recording. No backup service can be specified. Monitoring and recording will continue to be available for devices associated with other Cisco Desktop VoIP Monitor Services.

Recording

Recording is fault tolerant. If a recording service fails in a high-availability deployment, the other recording service will assume all recording responsibilities.

Recording Playback

Playback of recordings is not fault tolerant. Recordings are tied to the recording service that captured the recording. If a recording service fails, all recordings associated with that service is unavailable until it is restored.

Load Balancing for CAD-Based Monitoring and Recording

Desktop Monitoring

Desktop Monitoring is load-balanced by design. Monitoring load is distributed between the agent desktops.

Server Monitoring and Mobile Agent Monitoring

Load balancing can be achieved when configuring SPAN ports for, and associating devices with, the Cisco Desktop VoIP Monitor Services. To achieve load balancing, have each VoIP Monitor Service monitor an equal number of agent phones.

Recording

Recording services are selected in round-robin fashion at runtime by the desktops. However, no attempt is made to ensure that the load is balanced between the recording services.

Cisco Agent Desktop Presence Integration

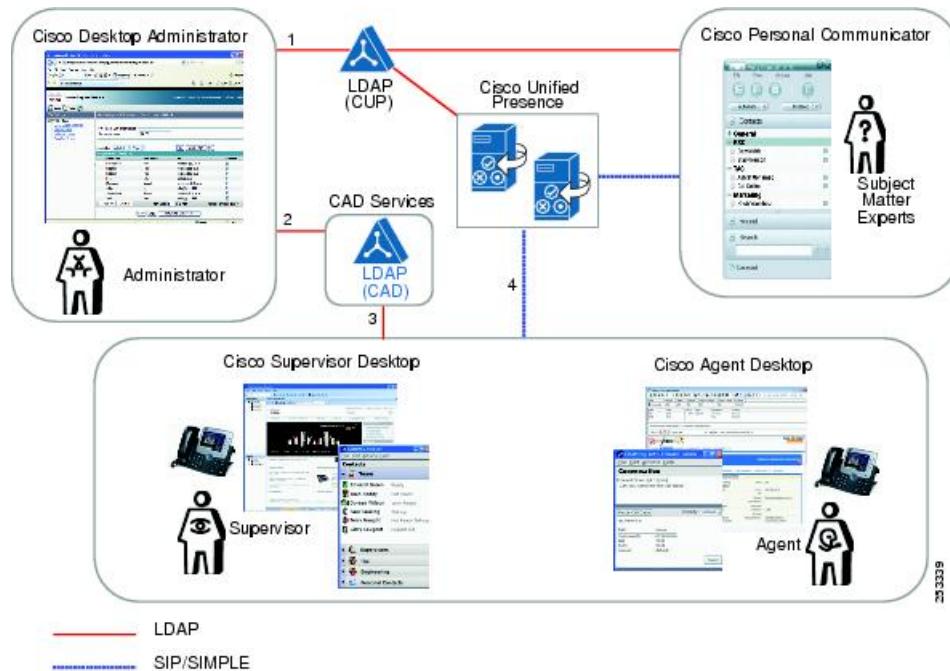
Cisco Agent Desktop agents and supervisors have long been able to communicate with each other by using the chat services built into the desktop applications. Now, for customers who have deployed Cisco Unified Presence in their environments, agents and supervisors can use these same desktop applications to see the

presence status of subject matter experts (SMEs) as well as other critical members of the enterprise and to initiate chat sessions with them. The subject matter experts use the familiar Cisco Unified Personal Communicator to initiate chat sessions with agents who are configured as Unified Presence users and to respond to chat requests from them. Subject matter experts can also use Microsoft Office Communicator if Cisco Unified Presence is configured to support federated users.

For example, suppose that a customer calls a Cisco Unified Contact Center that has integrated Cisco Unified Presence with CAD. The customer's call is routed to an available agent. If the agent requires assistance in addressing the caller's needs, the agent can launch the contact selection window from the Agent Desktop toolbar. The contact selection window will display the presence status of other agents, supervisors, and subject matter experts who are assigned to the agent's work flow group. The agent can then select a contact that is available and can initiate a chat session with the contact. If appropriate, the agent can also use the contact selection window to conference a contact into the call, or even transfer the customer's call to the contact.

The following figure and description explain how various components of CAD and Cisco Unified Presence interface with each other.

Figure 114: Interface Between CAD and Cisco Unified Presence



The figure above depicts the following sequence of events:

- 1 Cisco Desktop Administrator binds to the LDAP server for SME searches and information (name, telephone number, and so forth).
- 2 The Administrator places SMEs in logical groups called contact lists and then assigns them to specific work flow groups. In this way, administrators can segment contact lists and ensure that only those agents assigned to a specific work flow group have visibility to the appropriate contact list. This configuration is saved in the CAD LDAP directory so that each agent/supervisor does not have to access the Cisco Unified Presence LDAP server, which might have limitations on the number of connections and other parameters. Administrators can also control whether SMEs can see the agent's presence state.
- 3 CAD retrieves the contact list associated with the agent's workflow group.

- 4 CAD sends a SIP REGISTER message to register with Cisco Unified Presence, followed by individual SIP SUBSCRIBE messages for each user in its contact list. CAD also sends a SIP SUBSCRIBE message for "user-contacts" for contacts configured on Cisco Unified Presence. A SIP NOTIFY message is received whenever a contact in the contact list changes state. CAD does not allow agents to change their presence states; it only sends a single SIP PUBLISH message to Cisco Unified Presence when the agent logs in.

Call control is done via the existing CAD main window call controls using CTI.

All SIP traffic and presence information sent between CAD and Cisco Unified Presence is not encrypted and is done by using TCP or UDP.

Cisco Unified Presence can evenly assign the users registered with it across all nodes within the Cisco Unified Presence cluster. If a user attempts to connect to a node that is not assigned to him, CAD will connect to the Cisco Unified Presence server specified in redirect messages from the publisher.

Design Considerations

All communication between CAD agents and SMEs is through the Cisco Unified Presence Server and is not routed through any CAD servers. For deployment guidelines, see the *Cisco Collaboration System Solution Reference Network Designs* at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

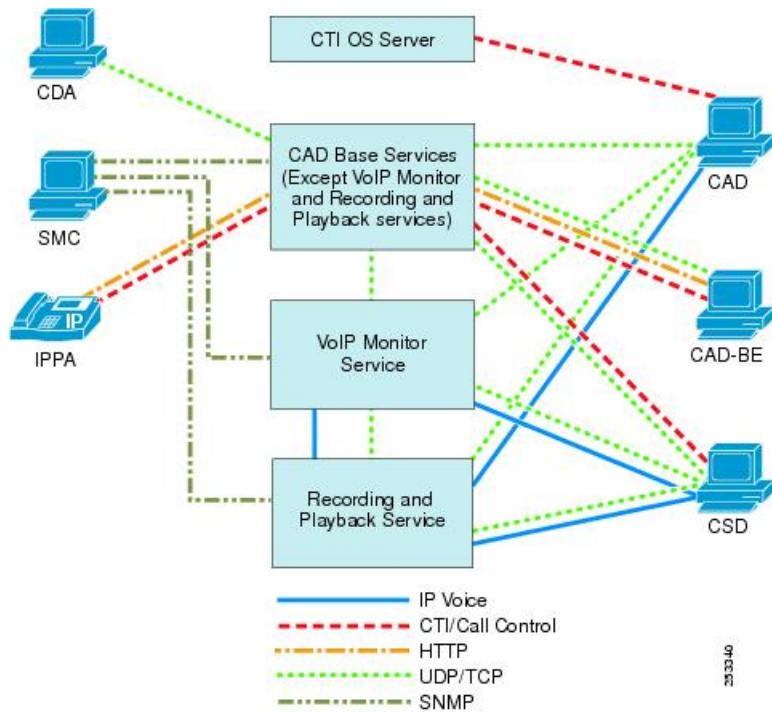
Cisco Agent Desktop and NAT

When the CAD desktop is deployed in a network environment where two or more disjointed networks are interconnected using NAT, the CAD Base Services must all be located on the same network. Network Address Translation (NAT) and Port Address Translation (PAT) are not supported between CAD Base Services servers. The CAD and Cisco Supervisor Desktop (CSD) applications support NAT and PAT but only over a VPN connection. Cisco Desktop Administrator (CDA) and Services Management Console (SMC) do not support NAT or PAT and must be installed on the same network as the CAD Base Services.

Firewalls are supported between the CAD services and desktop applications, and are supported between the desktop applications as long as the firewall allows the required type of traffic through and the appropriate ports are open. Internet Control Message Protocol (ICMP) must be allowed in the firewall for Unified CCE and Unified CM to communicate with CAD. ICMP is also needed for heartbeat time-out detection between CAD, the CTI Server (CTISVR), and Unified CM. The following figure shows the traffic types that are used between the CAD components.

For detailed port information, see the *Port Utilization Guide for Cisco Unified Intelligent Contact Management Enterprise & Hosted*.

Figure 115: Communication Between CAD Components



The figure above shows that IP voice streams are exchanged between the VoIP providers (CAD, the VoIP Monitor service, and the Recording and Playback Service) and the VoIP requestors (CSD and the Recording and Playback Service).

CTI and call control data (agent state, skill information, and call events) flow either from the CTI OS service (in the case of CAD) or from one or more of the CAD Base Services communicating directly with the CTI server (in the case of CSD and CAD IPPA agents).

Note that, in the case of the IP Phone Agent XML service, the CTI information exchanged applies only for agent state changes requested by the agent using the CAD IPPA application and for skill information displayed on the phone. Call control messages are still exchanged between the phone and Unified CM.

HTTP communication is performed between the SMC servlet and the SMC applet running on the CAD Base Services machine. HTTP is also the protocol used by the CAD IPPA service to communicate with the Browser and IP Phone Agent service.

The UDP/TCP traffic shown in the preceding figure represents the socket connections used to exchange messages between servers and clients, which includes the CORBA connections used by most of the clients to request services and information from the servers.

The SMC servlet that runs on the CAD Base Services machine uses SNMP to gather status information about all the CAD services that are part of an installation.

Support for IP Phones and IP Communicator

CAD support the use of Cisco IP hardware phones and the Cisco IP Communicator software phone.

Some CAD agent application features (CAD and CAD IPPA) require specific phone models. Some installations support either hardware phones or software phones but not both. For more information about the exact phone models and IP Communicator versions supported, see the [Cisco Agent Desktop documentation](#).

IP phones and silent monitoring

Silent monitoring of agents supports using either IP hardware phones or Cisco IP Communicator.

IP phones and Mobile Agent

The Mobile Agent feature does not require any specific type of phone. You can even use an analog phone with this feature.

IP Phones and Citrix or MTS

CAD supports both Cisco IP hardware phones and Cisco IP Communicator when using Citrix or MTS . In these environments, you must install Cisco IP Communicator on the agent desktop PC. You cannot deploy Cisco IP Communicator on the Citrix or MTS server.

Cisco Agent Desktop IP Phone Agent

The IP Phone XML service agent application supports only hardware IP phones because there is no desktop.

Cisco Agent Desktop and Citrix

Cisco Unified CCE supports running Cisco Agent Desktop within a Citrix terminal services environment. When planning to use Citrix terminal services for CAD, take the following considerations into account:

- Cisco Supervisor Desktop (CSD) and Cisco Desktop Administrator (CDA) are not supported in a Citrix terminal services environment.
- Desktop Monitoring (for Silent Monitoring and Call Recording) is not supported with Citrix terminal services. SPAN Port Monitoring must be used instead.
- Macros work only if they involve applications that are running on the Citrix server.
- Macros do not work if they involve applications that are running on the client PC.
- Only one Citrix user name is supported per CAD application login.
- The login ID and extension that appear by default in the login dialog box when CAD is started, are those associated with the last login by any user.
- The Citrix web client is not supported.
- For more information on Unified CCE support for Citrix XenDesktop and XenApp, see the Compatibility Matrix at:

http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE

- For supported Cisco Agent Desktop per Citrix version, see the CAD documentation at <http://www.cisco.com/c/en/us/support/customer-collaboration/agent-desktop/tsd-products-support-series-home.html>.

For implementation details, refer to the CAD documentation.

Support for Mix of CAD and CTI OS Agents on the Same PG

Unified CCE deployments can support a mix of CAD and CTI OS agents on the same PG. If a mix is deployed, the sizing limitations of CAD apply. Note that Cisco Supervisor Desktop (CSD) can monitor only CAD agents, and the CTI OS supervisor application can monitor only CTI OS agents.

High Availability for Cisco Agent Desktop

This section looks at planning for high availability in Cisco Agent Desktop deployments.

Cisco Agent Desktop Failover Scenarios

Cisco Agent Desktop client applications are a client of CTI OS, which provides for automatic fail-over and redundancy for the Cisco Agent Desktop CTI connections. If the Unified Communications Manager Peripheral Gateway or CTI Server (CG) fails-over, the Cisco Agent Desktop client application displays a logged out state and automatically returns to a logged in state when an operational connection is established with the alternate Unified Communications Manager Peripheral Gateway or CTI Server (CG). Consequently, the scenarios outlined in the CTI OS Considerations section apply.

The Cisco Agent Desktop services (Enterprise, Chat, RASCAL, and so forth) can also be deployed redundantly to allow for fail-over of the core Cisco Agent Desktop components. The Cisco Agent Desktop client applications are aware of the redundant Cisco Agent Desktop services and automatically fail-over in the event of a Cisco Agent Desktop service process or hardware failure.

The following services are active on only one side at a time:

- Cisco Browser and IP Phone Agent Service
- Cisco Chat Service
- Cisco Enterprise Service
- Cisco Licensing and Resource Manager Service
- Cisco Recording and Statistics Service
- Cisco Sync Service

The following services are active on both sides at all times and are available to the Cisco Agent Desktop client applications as long as network connectivity is available:

- Cisco LDAP Monitor Service
- Cisco Recording & Playback Service
- Cisco VoIP Monitor Service

Active side Cisco License and Resource Manager service fails

In this scenario, the following events occur:

- The Cisco Agent Desktop services on Side A that are not always active go to an idle state.
- The Cisco Agent Desktop services on Side B (idle) activate.
- The Cisco Agent Desktop client applications recover to Side B.

Active side Cisco Agent Desktop service fails twice in 5 minutes

In this scenario, the following events occur:

- The Cisco Agent Desktop services on Side A that are not always active go to an idle state.
- The Cisco Agent Desktop services on Side B (idle) activate.
- The Cisco Agent Desktop client applications recover to Side B.

Active side Cisco Agent Desktop service down for 3 minutes

In this scenario, the following events occur:

- The Cisco Agent Desktop services on Side A that are not always active go to an idle state.
- The Cisco Agent Desktop services on Side B (idle) activate.
- The Cisco Agent Desktop client applications recover to Side B.

Network failure between active side and idle side Cisco Agent Desktop service

In this scenario, the following events occur:

- The Cisco Agent Desktop services on Side A remain active.
- The Cisco Agent Desktop services on Side B (idle) activate.
- The Cisco Agent Desktop client applications remain connected to Cisco Agent Desktop services on Side A.
- When network connectivity is restored between Sides A and B, the Cisco Licensing and Resource Manager Service renders inactive the non-preferred side. Recovery side preference is configurable in Post Install.

CAD IP Phone Agent

CAD IP Phone Agent communicates with CTI Server through the Cisco Browser and Unified IP Phone Agent service. When launching the desktop, the agent may use the URL for either side as long as the desired side is accessible. Once launched, desktop automatically connects to the active side. When launching CAD IP Phone Agent, the agent must select the active side from the services menu on the phone. If the idle side is selected, the user receives an error informing them that the side selected is idle and to try the other side.

Idle side Cisco Agent Desktop services become active after failure

In this scenario, the following events occur for a logged-in desktop agent:

- The desktop applet changes to a logged out state and the user is notified that the connection has been lost.
- The desktop applet automatically connects to services on Side B and logs-in the agent.

In this scenario, the following events occur for a logged-in CAD IP Phone agent:

- The phone agent is notified that the connection to the server has been lost.
- The phone agent manually selects Side B from their services list and logs in again.

Replacement of MSDE with SQL Database Server

As MSDE is no longer supported, at post install time the user has to choose a SQL database. Post install configures their system based on their selection. There is a value stored in LDAP that indicates which implementation is selected. After the initial configuration is completed (the implementation is selected and saved), the user cannot change implementations. For the database implementation, Unified CCE configures the FCRasSrv database, as it does now. Unified CCE continues to provide scripts for setup and teardown of database replication for HA. There are three tables in the database implementation: agent state data, call log data, and recording metadata. The data that is stored in the tables is identical.



Note

Starting release 9.0(3), all customers with new deployments of any version of Cisco Agent Desktop must use SQL Server as the data store, and not flat files. The rationale behind this change is that deployments with a fully replicated SQL Server database experience a more complete feature set, better performance, and stability.

Cisco Agent Desktop Component Sizing

Server capacities for the Cisco Agent Desktop CTI Option vary based on the total number of agents, whether or not Switched Port Analyzer (SPAN) Monitoring and Recording is used, and the number of simultaneous recordings.

This section presents sizing guidelines for the Cisco Agent Desktop Server components.

Related Topics

[Cisco Agent Desktop, on page 351](#)

Cisco Agent Desktop Operating Conditions

The sizing information presented in this chapter is based on the following operating conditions:

- Maximum of 30 busy hour call attempts (BHCA) per agent.
- Five skill groups or precision queues per agent.
- The total number of agents indicated in the following tables and figures consists of 90% agents and 10% supervisors. For example, if a table or figure indicates 100 agents, the assumption is that there are 90 agents and 10 supervisors.

- Supervisors do not handle calls.
- Total number of teams is equal to 10% of total number of agents.
- Team members consist of 90% agents and 10% supervisors.
- Call types consist of 85% straight calls, 10% consultative transfers, and 5% consultative conferences.
- The default refresh rate for skill group updates is 10 seconds.
- The default number of skill group statistics columns configured at the CTI OS server is 17 columns.
- Agent Statistics is turned ON.
- The default number of agent statistics columns configured at the CTI OS server is 6 columns.
- Average of five Voice Response Unit (VRU) scripts, running consecutively in the Unified CCE script, per VRU call.
- Five Expanded Call Context (ECC) scalars.
- Transport Layer Security (TLS) for CTI OS is turned OFF.
- No mobile agents.
- One all-events CTI server client.
- Outbound hit rate is averaged at 30%.

The following notes apply to all figures and tables in this topic:

- The number of agents indicates the number of logged-in agents
- Server types:
 - APG = Agent Peripheral Gateway
 - PGR = Lab deployment
 - RGR = Rogger

**Note**

The terms Rogger and Central Controller are used interchangeably throughout this chapter.

Figure 116: Minimum Servers Required for Unified CCE Deployments with Cisco Agent Desktop

Maximum Agent Count	210*	1,000*	2,000	6,000	8,000	12,000
Central Controller						
Peripheral Gateways Agent Services					(1) • • • (8)	(1) • • • (12)

* Deployment supported in Unified System CCE

3001-AB

The following notes apply to the figure above:

- Voice Response Unit (VRU), Administration & Data Server, and Unified Communications Manager components are not shown.
- The maximum values apply to all CAD clients (CAD and CAD IPPA).

Component	Notes
Voice Response Unit (VRU) PG	Use the number of ports instead of agent count. Average of 5 Run VRU Script Nodes per call.

Component	Notes
Agent PG with Outbound Voice (Includes Dialer and Media Routing PG)	<p>[Maximum inbound agent capacity] – 1.33 * [Number of SIP Dialer ports]</p> <p>To determine the maximum inbound agent capacity, see the Inbound Agent PG entry in this table. The capacity depends on your Unified CCE software release, hardware server class, and agent desktop type.</p> <p>The formula is an indicator of platform capacity. This is not an indicator of outbound resources in terms of how many agents can be kept busy by the number of dialer ports in the deployment. A quick but inexact estimate is that two ports are required for each outbound agent, but your outbound resources can vary depending on hit rate, abandon limit, and talk time for the campaigns in the deployment. Use the sizing tool to determine outbound resources required for your campaigns.</p> <p>Note The Cisco Media Blender is not supported when installed on a PG system.</p>
Cisco Unified Web and E-Mail Interaction Manager	<p>For the most current server specifications and sizing guidelines for Cisco Unified Web and E-Mail Interaction Manager, see the latest documentation at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-email-interaction-manager/products-implementation-design-guides-list.html.</p>
Cisco Unified Customer Voice Portal (CVP) Application Server And Voice Browser	<p>For the most current server specifications for Unified CVP, see the latest version of the <i>Hardware and System Software Specification for Cisco Unified CVP</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-technical-reference-list.html.</p>
Unified IP IVR Server	<p>For the most current Unified IP IVR server specifications, see the documentation available through valid Cisco Employee or Partner login.</p>
Cisco Unified Intelligence Center (Unified Intelligence Center)	<p>For the most current server specifications for Unified Intelligence Center, see the latest version of the <i>Hardware and System Software Specification (Bill of Materials)</i> at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-intelligence-center/products-implementation-design-guides-list.html.</p>

For further details, see the Virtualization for Unified CCE DocWiki at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE.

CTI OS for Cisco VXI

When you deploy VDI or VXI, the performance, bandwidth, and timing requirements for CTI-OS (as defined in this document) must still be met. Agents will observe delays or other negative side effects if the VDI or VXI is deployed in a way that doesn't give enough performance or bandwidth to the VDI clients.

The number of VDI clients that can run within the Citrix or VMWare server is dependent on a number of factors including the footprint of other applications that are run in the deployment. Verify proper sizing to ensure that CTI-OS desktops function properly (for more information, see the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE).

Cisco Agent Desktop Base Services

The Cisco Agent Desktop base services consist of a set of application servers that run as Microsoft Windows services. They include Chat Service, Directory Services, Enterprise Service, Unified IP Phone Agent Service, LDAP Monitor Service, Licensing and Resource Manager Service, Recording and Statistics Service, and Sync Service. In addition, there are application servers that may be placed on the same or separate VMs as the Base Servers. These additional applications include the VoIP Monitor Service and the Recording and Playback Service.

A set of Cisco Agent Desktop base services plus the additional application servers, single or redundant installation, correspond to a logical call center (LCC) and are associated with a PG pair.

Cisco Agent Desktop VoIP Monitor Service

The VoIP Monitor Service enables the Silent Monitoring and Recording features. For Desktop Monitoring, the VoIP Monitor Service has no impact on design guidance for Agent PG scalability. When using Switched Port Analyzer (SPAN) monitoring, the VoIP Monitor Service may be co-resident on the Agent PG for up to 100 agent phones. When SPAN Monitoring and Recording are required for more than 100 phones, the VoIP Monitor Service must be deployed on a dedicated VM. Each dedicated VoIP Monitor Service can support up to 1000 phones on a Gigabit NIC.

Cisco Agent Desktop Recording and Playback Service

The Recording and Playback Service stores the recorded conversations and makes them available to the Supervisor Log Viewer application.

A co-resident Recording and Playback Service can support up to 32 simultaneous recordings. A dedicated Recording and Playback Service (which is available in the Premium offering) can support up to 80 simultaneous recordings. The capacity of the Recording and Playback Service is *not* dependent on the codec that is used.

The following table summarizes the raw Recording and Playback Service capacity.

Table 38: Capacity of Recording and Playback Service

Recording and Playback Service Type	Maximum Simultaneous Recordings
Co-resident	32

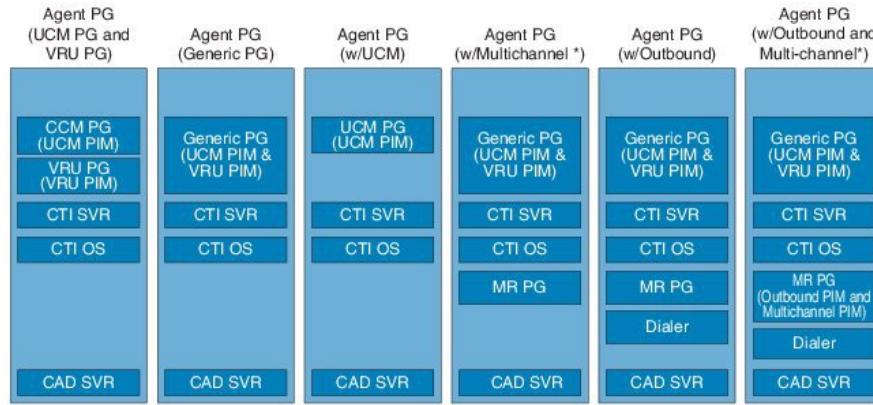
Recording and Playback Service Type	Maximum Simultaneous Recordings
Dedicated	80

Peripheral Gateway and Server Options for Cisco Agent Desktop

A Unified CCE Peripheral Gateway (PG) translates messages coming from the Unified Communications Manager servers, the Unified IP IVR, Unified CVP, or voice response units (VRUs) into common internally formatted messages that are then sent to and understood by Unified CCE. In the reverse, it also translates Unified CCE messages so that they can be sent to and understood by the peripheral devices.

The figures below illustrate various configuration options for the Agent PG with Cisco Agent Desktop.

Figure 117: Agent PG Configuration Options with Cisco Agent Desktop



* With CCE, the MR PG for outbound and multi-channel is co-resident with the Agent PG.
With System CCE, the MR PG for outbound (outbound controller) is co-resident with the Agent PG, but not the MR PG for multi-channel (multi-channel controller).

34378

The table below gives sizing guidelines for PGs and PIMs.

Table 39: PG and PIM Sizing Guidelines

Sizing variable	Guidelines based on Unified CCE Release 10
Maximum number of PGs per Unified CCE	150
Maximum number of PG types per VM	Up to two PG types are permitted per VM, but each VM must meet the maximum agent and VRU port limitations.
Maximum number of Unified Communications Manager PGs per VM	Only one Unified Communications Manager PG, Generic PG, or System PG is allowed per VM.
Maximum number of Unified Communications Manager PIMs per PG	1

Sizing variable	Guidelines based on Unified CCE Release 10
Can PGs be remote from Unified CCE?	Yes
Can PGs be remote from Unified Communications Manager?	No
Maximum number of VRUs controlled by one Unified Communications Manager	See the <i>Cisco Collaboration System Solution Reference Network Designs</i> at http://www.cisco.com/go/ucsrnd .
Maximum number of CTI servers per PG	1
Can PG be co-resident with Cisco Unified Communications Manager?	No

Bandwidth Requirements for Cisco Agent Desktop

This section presents some design considerations for provisioning network bandwidth, providing security and access to corporate data stores, and ensuring Quality of Service (QoS) for Unified CCE installations that include the Cisco Agent Desktop (CAD) product.

Silent Monitoring Bandwidth Usage

The Silent Monitoring feature of the Cisco Agent Desktop software, which includes listening to a live call, recording an agent call, and listening to a recorded call, has the largest bandwidth requirements for the Cisco Agent Desktop product. Properly configuring this feature is especially important for Unified Mobile Agents who are connected to the main site by a WAN connection.

To access the Silent Monitoring feature, a request is sent to a VoIP provider. The VoIP provider captures from the network, or reads from disk, the voice streams representing the call (two voice streams per call) and sends them back to the requestor. The requestor receives the streams and either decodes them for listening or stores them to disk. The bandwidth requirements detailed in this section are for the network links between the requestor and provider.

Silent Monitoring Requestors

There are two possible requestors in the Cisco Agent Desktop software:

- Cisco Supervisor Desktop
- Recording and Playback service

Cisco Supervisor Desktops send Silent Monitoring requests when the supervisor wants to listen to an agent's call in real-time or listen to a call that was recorded earlier. The Recording and Playback service send recording requests when a supervisor or agent wants to record a call. For listening to or recording a live call, the VoIP provider will capture the voice streams and send them to the requestor. On the supervisor's desktop, these streams are decoded and played through the supervisor's desktop sound card. For recording, the Recording and Playback service receives the voice streams and saves them to disk.

A Unified CCE installation may have one or two Recording services.

Silent Monitoring Providers

There are three possible VoIP providers in the Cisco Agent Desktop software:

- Cisco Agent Desktop
- VoIP Monitor service
- Recording & Playback service

The Cisco Agent Desktop application contains a module referred to as the Desktop Monitor service, which runs on the agent's desktop. The Desktop Monitor service processes Silent Monitoring requests only for the agent signed in to the Cisco Agent Desktop application on the desktop. The service captures voice packets sent to the phone or IP Communicator software phone associated with the signed-in agent. The phone must contain an extra network port that allows the phone to be connected to a network and also to an agent's computer. They also support the ability of hubs and switches to propagate network traffic through this additional port. This capability is what allows the Desktop Monitor service to see the phone conversations on the agent's phone. See the *Compatibility Matrix for Unified CCE* at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE for a list of supported phones.

By default, this service is active on all agent desktops when the application is started. After initial installation of the Cisco Agent Desktop servers, all agents are already configured to use the Desktop Monitor service for the Silent Monitoring feature.

A VoIP Monitor service is able to handle multiple requests for Silent Monitoring simultaneously. It captures packets directly from the switch through the switch's Switched Port Analyzer (SPAN) configuration. An installation may have up to five VoIP Monitor services on different machines. Off-board VoIP services may be installed at remote office locations. In some instances, this service may be required due to network complexity and capacity planning. Agents must be explicitly configured to use a VoIP Monitor service if this is the method desired for Silent Monitoring for that agent's device.

**Note**

CAD IP Phone Agents who do not have a desktop must be configured to use a VoIP Monitor service for the Silent Monitoring feature.

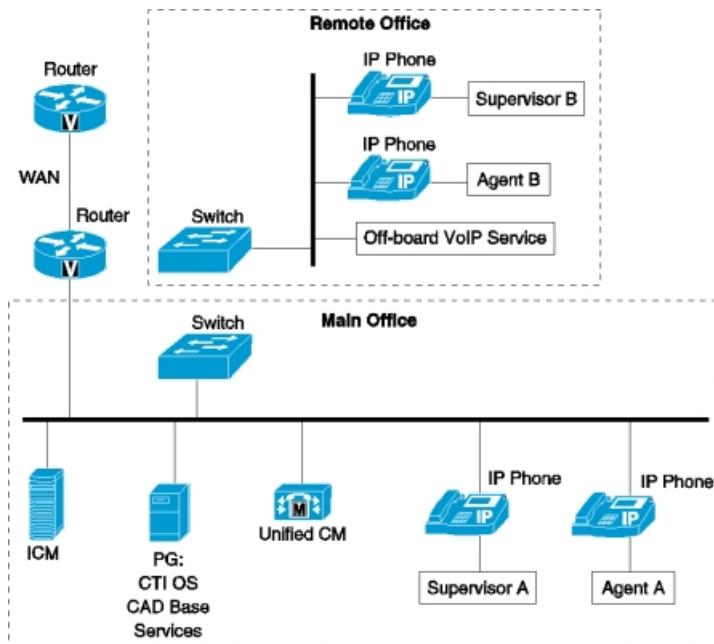
The Recording and Playback service may also provide the two streams representing a phone call when a supervisor plays back a recorded agent call. In this case, the streams have already been stored on disk from an earlier recording session. The Recording and Playback service reads the raw data files from the disk and sends the RTP streams over the network to the supervisor's desktop, where they are played through the sound card.

As this description indicates, the Recording and Playback service may be either the requestor (for recording a live call) or a provider (for playing back a recorded call).

A VoIP and Recording and Playback services are usually installed along with the Cisco Agent Desktop base services. Additional VoIP services and a second Recording and Playback service may be installed on other boxes.

The figure below shows a representative Unified CCE installation supporting a remote office over a WAN. Both the main office and the remote office have a VoIP Monitor service on-site.

Figure 118: VoIP Monitor Service at Main and Remote Sites



When you locate the requestors and providers, you can determine where the bandwidth is required for the Silent Monitoring feature. The following notes regarding bandwidth apply:

- Although an administrator can assign a specific VoIP service to an agent device, the Recording service that is used when calls are recorded is determined at the time the request is made. The same rule applies if two Recording services are installed to load-balance the installation. In some cases, the provider and requestor may be separated by a WAN and would require the bandwidth on the WAN. If a second Recording and Playback service is to be installed, install it on a server at the main office (on the LAN with the Cisco Agent Desktop base services).
- If the VoIP provider is a VoIP Monitor service, if the requestor is a Recording service, and if these services reside on the same machine, then there is no additional bandwidth used on the network to record the call.

Regardless of who is the requestor and VoIP provider, the bandwidth requirement between these two points is the bandwidth of the IP call being monitored and/or recorded. For purposes of calculating total bandwidth, you can think of each monitoring/recording session as being a new phone call. Therefore, to calculate bandwidth to support the Silent Monitoring feature, you can use the same calculations used to provision the network to handle call traffic, with the exception that the voice stream provided by the VoIP provider consists of two streams in the same direction. Whereas a normal IP phone call has one stream going to the phone and one stream coming from the phone, the VoIP provider has both streams coming from the provider. Keep this difference in mind when provisioning upload and download speeds for your WANs.

To determine the bandwidth requirements for these voice streams, see the *Cisco Collaboration System Solution Reference Network Designs* at <http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/tsd-products-support-series-home.html>.

Cisco Agent Desktop Applications Bandwidth Usage

The CAD desktop applications include:

- Cisco Agent Desktop
- Cisco Supervisor Desktop
- Cisco Desktop Administrator
- Cisco Desktop Monitoring Console

These applications also require a certain amount of bandwidth, although far less than the Silent Monitoring feature. In addition, the type of communication across the network is bursty. In general, bandwidth usage is low when the agents are not performing any actions. When features or actions are requested, the bandwidth increases for the time it takes to perform the action, which is usually less than one second, then the bandwidth usage drops to the steady-state level. From a provisioning standpoint, one must determine the probability of all the agents performing a particular action at the same time. It might be more helpful to characterize the call center and determine the maximum number of simultaneous actions (in the worst case) to determine instantaneous bandwidth requirements, and then determine what amount of delay is tolerable for a percentage of the requested actions.

For example, the raw bandwidth requirement for 1000 agents logging in simultaneously is about 6.4 kilobytes per second and the login time is about 9 seconds (with no network delay) for each agent. If the WAN link did not have this much bandwidth, logins would take longer as packets were queued before being sent and received. If this queuing delay caused the login attempts to take twice as long (18 seconds in this case), would this delay be acceptable? If not, provision more bandwidth.

Each of these applications communicates with the Cisco Agent Desktop base services running on VMs. In addition, the agent desktop application communicates with the CTI server through the CTI OS server for call control actions and state changes. The table below lists the types of messages for each application.

Table 40: Messaging Type by CAD Application

Application name	Message types
Cisco Agent Desktop	Login/logoff
	Agent state changes
	Call Control
	Call status information
	Desktop Monitoring and recording
	Chat messages
	Team performance messages
	Report generation
	Real-time data refresh

Application name	Message types
CTI OS Supervisor Desktop	Login/logout
	Agent state changes
	Call status updates
	Report Generation
	Silent Monitoring
	Call recording
	Call playback
	Chat messages
	Team performance messages
	Real-time data refresh
Cisco Desktop Administrator	Configuration information retrieval and storage
	Configuration data refresh
Cisco Desktop Monitoring Console	Service discovery
	SNMP Get messages

Cisco Agent Desktop Bandwidth Usage

Cisco Agent Desktop agents are able to sign in and sign off, change their agent state, handle calls, and send reporting information to the base servers. The bandwidth requirements for these activities are fairly small but can add up when many agents are considered.

The table below shows the average bandwidth requirements for various numbers of agents. This information is derived from bandwidth testing and extrapolation of bandwidth data. Because there are many variables that can affect bandwidth, a configuration that resulted in higher bandwidth usage was chosen to provide near worst-case scenarios. If the agent's WAN link meets or exceeds the bandwidth requirements shown in this table, Cisco Agent Desktop can run without delays in message passing.

The following configuration parameters affect bandwidth and apply to both the table below and [Cisco Agent Desktop Applications Bandwidth Usage, on page 375](#):

- Number of skills per agent: 10
- Number of agents per team: 20
- Number of teams: 50
- Number of agent state changes per agent per hour: 10 (Not including state changes due to handling calls)
- Calls per agent per hour: 60
- Team performance messages per team per hour: 8
- Chat messages sent or received per hour: 20
- Average chat message size (in bytes): 40

- Number of calls recorded per hour: 10

**Note**

The bandwidth requirements shown do not include the bandwidth of the RTP streams for the call, recording, or monitoring sessions, but include only the messaging needed to start and stop the sessions.

Table 41: Average Bandwidth Requirements for Cisco Agent Desktop

Number of agents	Average download bandwidth (kilobytes per second)	Average upload bandwidth (kilobytes per second)
1	0.02	0.003
100	1.7	0.1
200	3.4	0.3
300	5.0	0.4
500	8.4	0.7
600	10.0	0.8
700	11.7	1.0
800	13.4	1.1
900	15.1	1.3
1000	16.8	1.4

Cisco Supervisor Desktop Bandwidth Usage

A Cisco Supervisor Desktop receives events for all the agents of the team that the supervisor is logged into. This information includes state changes, call handling, login/logout, and so forth. The more agents, skills, and calls there are the more data is sent to supervisors. In addition, particular reports are automatically refreshed periodically to provide real-time data while the supervisor is viewing the report. Refreshing reports requires additional bandwidth.

The table below uses the same basic configuration parameters used to determine the bandwidth numbers in [Cisco Agent Desktop Applications Bandwidth Usage, on page 375](#). In addition, this table takes into account the fact that the Team Skill Statistics report is being viewed and refreshed.

Table 42: Average Bandwidth Requirements for Cisco Supervisor Desktop

Number of agents	Average download bandwidth (kilobytes per second)	Average upload bandwidth (kilobytes per second)
1	0.02	0.003

Number of agents	Average download bandwidth (kilobytes per second)	Average upload bandwidth (kilobytes per second)
100	1.3	0.1
200	2.5	0.3
300	3.7	0.4
400	5.0	0.5
500	6.2	0.6
600	7.5	0.8
700	8.7	0.9
800	10.0	1.0
900	11.2	1.1
1000	12.4	1.3

Cisco Desktop Administrator Bandwidth Usage

The bandwidth requirements for Cisco Desktop Administrator are very small and are seen only when an administrator is actively changing configurations. In general, the bandwidth used by Cisco Desktop Administrator is negligible from a provisioning standpoint.

Cisco Desktop Monitoring Console Bandwidth Usage

The bandwidth requirements for the Cisco Desktop Monitoring Console are very small and short-lived. In general, the bandwidth used by the Cisco Desktop Monitoring Console is negligible from a provisioning standpoint.

Cisco Agent Desktop Service Placement

In a Unified CCE installation using Cisco Agent Desktop, all CAD services except the VoIP Monitor service and the Recording and Playback service must coreside with the PG. You can install the VoIP Monitor Service and Recording and Playback Service on other VMs.

VoIP Monitor Server

A single VoIP Monitor Service can support up to 114 simultaneous Silent Monitoring sessions. More VoIP Monitor Services increase the SPAN-based monitoring capacity of the installation.

You can have a maximum of five VoIP Monitor servers in a CAD installation. Only one VoIP Monitor Service can exist on a single VM.

The main load on a VoIP Monitor Service is the amount of network traffic that is sent to the VoIP Monitor Service for the devices that are assigned to that VoIP service, not the number of simultaneous monitoring sessions. When Switched Port Analyzer (SPAN) is configured to send traffic from a device to a particular VoIP service, the VoIP services packet sniffer monitors network traffic even without active monitoring sessions. The amount of traffic monitored limits the number of devices that you can assign to a VoIP service.

If a VoIP Monitor Service coresides with the CAD base services on the PG, it supports the network traffic of up to 100 agents. You can dedicate a third virtual NIC for SPAN destination port in this environment, although it is not necessary. If more than 100 agents are configured to use a single VoIP Monitor Service, move that service to another VM. A single VoIP Monitor Service supports the network traffic of 1000 agent phones with a Gigabit NIC to connect to the switch.

**Note**

If the switch does not support ingress and egress traffic on the same switch port, then use a dedicated virtual NIC to support SPAN services.

Recording and Playback Server

You can have a maximum of two Recording and Playback Services in a CAD installation. As with the VoIP Monitor Service, only one of these services can exist on a single computer.

If the Recording and Playback Service coresides with CAD base services on the PG, it supports up to 32 simultaneous recording sessions. If you require more recording and playback sessions, move the Recording and Playback Service to another VM. The Recording and Playback Service can coexist with an off-board VoIP Monitor Service. An off-board Recording and Playback Service supports up to 80 simultaneous recordings.

The Recording and Playback Service converts copies of the RTP packets to RAW files and stores these files for play back using the Cisco Supervisor Desktop. Either the VoIP Monitor server (SPAN capture) or the Cisco Agent Desktop (Desktop capture) directs these RTP packets to the Recording and Playback server. So in a SPAN capture environment, a recording consumes a monitoring session and a recording and playback session.

A second Recording and Playback Service does not increase the recording capacity, but it does provide some load balancing and redundancy. When both Recording and Playback servers are active, the recording client alternates between the two servers and stores the recording files first on one server, then the other.

Miscellaneous Deployment Considerations

This section briefly describes the following additional deployment considerations.

Layer-3 Devices

Layer-3 network devices (routers and gateways) cannot exist between an agent telephone device (hardware or software phone) and the switch port used by the VoIP Monitor service that is configured to capture voice packets for Silent Monitoring and Recording. This restriction applies only if a VoIP Monitor Service is configured as the primary or back-up service for capturing voice streams. If desktop monitoring is configured as the primary method (with no secondary method), this information does not apply.

NDIS Compliance of NICs

The physical network interface cards (NICs) used by the VoIP Monitor services and on the agent's PC (when Desktop Monitoring is configured) must support promiscuous mode packet sniffing as stated. If the NIC card

or driver does not support this functionality through the NDIS interface, the Monitoring and Recording feature will not work.

Encrypted Voice Streams

If the voice streams are encrypted, the Silent Monitoring and Recording feature does not work correctly. Although the voice streams can still be captured, they will not be decoded correctly. The end result is that speech is unintelligible.