# Cisco Unified Contact Center Enterprise Solution Reference Network Design (SRND)

Cisco Unified Contact Center Enterprise Release 8.x
September 24, 2012

**Americas Headquarters**
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
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.

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 used in this document are not intended to be actual addresses. Any examples, command display output, and figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses in illustrative content is unintentional and coincidental.

# CONTENTS

# Introduction

This document provides design considerations and guidelines for deploying Cisco Unified Contact Center Enterprise Release 8.0 and later releases in a Cisco Unified Communications System.

This document builds on ideas and concepts presented in the latest version of the *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*.

This document assumes that you are already familiar with basic contact center terms and concepts and with the information presented in the *Cisco Unified Communications SRND*. To review IP Telephony terms and concepts, see the documentation at the preceding link.

# New or Changed Information for This Release

Unless stated otherwise, the information in this document applies to Cisco Unified Contact Center Enterprise Release 8.0.

The references to Webview and Webview Server (WVS) have been completely removed from this document because Webview Reporting is *not* supported in Cisco Unified Contact Center Enterprise Release 8.5(1).

# Revision History

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

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

The following table lists the revision history for this document.

| Revision Date | Comments |
|---|---|
| September 18, 2012 | Updated Figure 4 and Figure 5 in Chapter 1, “Architecture Overview”. |
| August 10, 2012 | Added note that AW-HDS-DDS can only be enabled on primary AW to page 15. |
| August 9, 2012 | Added “Yes” to cell for Outbound Calls/Siebel in table 2 in chapter 4. |
| May 11, 2012 | Added content for Multiple PG/CTIOS or CTI PIM for Agent Expansion with respect to RSM. |
| April 24, 2012 | Clarified information about Windows server support in 8.5(3) on page 7. |
| March 5, 2012 | Added information to highlight that only T1 and E1 PRI interfaces to the PSTN are supported for SIP Outbound dialers. |
| February 29, 2012 | Added information about Cisco Media Blender sizing to Table 12. |
| January 25, 2012 | Updated to add information about Cisco Finesse. |
| December 2, 2011 | Editorial updates to help clarify titles in Table 13. |
| October 5, 2011 | Updated sizing information and deployment options to reflect the capacity for a maximum of 12,000 agents. |
| July 19, 2011 | Updated the Mobile Agent Over Broadband section in Chapter 2: Deployment Models.<br>Changed Max Skill Groups per PG from 1300 to 3000 in Table 13, Chapter 9 |
| June 20, 2011 | Updated content for Unified CCE Release 8.5(2). Updates include revisions for Mobile Agent, Agent Greeting and Whisper Announcement in Chapters 6, 8, and 9. |
| May 18, 2011 | Updated the sizing tools section to reflect that there is one tool only: Unified CST.<br>Updated “Scenario 2: Agent PG Side A Fails” in the “Peripheral Gateway Design Considerations” section of Chapter 3: Design Considerations for High Availability.<br>Updated the “Sizing Effects Due to Number of Skill Groups per Agent” table.<br>Added a note to the Additional Sizing Factors section to indicate that you can add a maximum of 50 agents per team. |

| Revision Date | Comments |
|---|---|
| April 25, 2011 | Updated link to sizing tools.<br>Corrected calculation on page 311.<br>Updated Cisco Unified Outbound Option Design Considerations section to describe the difference of multiple dialers for SIP and SCCP dialers.<br>Updated Contact Center Design Considerations.<br>Updated links to Hardware and System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5 (1).<br>Updated Best Practices and Recommendations for Cisco Agent Desktop Service Placement. |
| October 4, 2010 | Content was updated for Webview reporting tool removal in Unified CCE Release 8.5(1). Version of guide was modified from 8.01 to 8.x. Information about Agent Greeting and Whisper Announcement was added. |
| March 19, 2010 | Initial version of this document for Cisco Unified Contact Center Enterprise Release 8.0. |
| April 22, 2009 | Content was updated for Cisco Unified Communications System Release 7.1. |
| October 29, 2008 | Revised some of the sizing information for Cisco Unified Contact Center Enterprise components and servers, and added a chapter on Cisco Unified Contact Center Management Portal (Unified CCMP). |
| August 27, 2008 | Initial version of this document for Cisco Unified Contact Center Enterprise Release 7.5. |

# Obtaining Documentation and Submitting a Service Request

See the monthly *What's New in Cisco Product Documentation* for information about obtaining documentation, submitting a service request, and gathering additional information. This document also lists all new and revised Cisco technical documentation.

Subscribe to the *What's New in Cisco Product Documentation* as a Really Simple Syndication (RSS) feed and set content to be delivered directly to your desktop using a reader application. The RSS feeds are a free service and Cisco currently supports RSS version 2.0.

# Cisco Product Security Overview

This product contains cryptographic features and is subject to United States and local country laws governing import, export, transfer and use. Delivery of Cisco cryptographic products does not imply third-party authority to import, export, distribute, or use encryption. Importers, exporters, distributors and users are responsible for compliance with U.S. and local country laws. By using this product, you agree to comply with applicable laws and regulations. If you are unable to comply with U.S. and local laws, return this product immediately.

Read more information regarding U.S. export regulations.

CHAPTER 1

# 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. It combines software IP automatic call distribution (ACD) functionality with Cisco Unified Communications in a unified solution that enables companies to rapidly deploy an advanced, distributed contact center infrastructure.

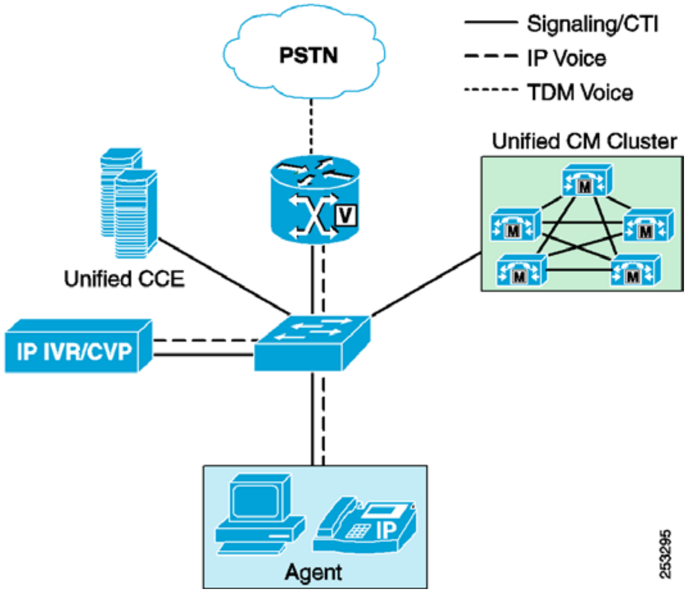
The reader of this document is expected to be familiar with the Unified CCE solution architecture and functionality as described in the *Installation and Configuration Guide for Cisco Unified Contact Center Enterprise & Hosted*. Make sure you become familiar with the concepts described in that manual 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 (Unified CM), Cisco Unified IP IVR, Cisco Unified Customer Voice Portal (Unified CVP), Cisco VoIP Gateways, and Cisco Unified IP Phones. Together these products provide Cisco Unified Communications and contact center solutions to achieve intelligent call routing, multichannel automatic call distribution (ACD) functionality, interactive voice response (IVR), network call queuing, and consolidated enterprise-wide reporting. Unified CCE can optionally integrate with Cisco Unified Intelligent Contact Manager to support networking 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 multi-site contact centers. It uses your existing Cisco IP network to lower administrative expenses and extend the boundaries of the contact center enterprise to include branch offices, home agents, and knowledge workers. Figure 1 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 CM
- Queuing and self-service: Cisco Unified IP Interactive Voice Response (Unified IP IVR) or Cisco Unified Customer Voice Portal (Unified CVP)
- 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 Agent Desktop, Cisco Toolkit Agent Desktop (CTI OS), Cisco Finesse Desktop, 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.

# Solution Components

## Cisco Unified Communications Manager

Cisco Unified Communications Manager (Unified CM) is a software application that controls the voice gateways and IP phones, thereby providing the foundation for a VoIP solution. Unified CM runs on Cisco Media Convergence Servers (MCS). The software running on a server is referred to as a Unified CM server. Multiple Unified CM servers can be grouped into a cluster to provide for scalability and fault tolerance. Unified CM communicates with the gateways using standard protocols such as H.323, Media Gateway Control Protocol (MGCP), and Session Initiation Protocol (SIP). Unified CM communicates with the IP phones using SIP or Skinny Call Control Protocol (SCCP). For details on Unified CM call processing capabilities and clustering options, see the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND)*.

Unified CM communicates with Unified CCE through the Java Telephony Application Programming Interface (JTAPI).  A single Unified CM subscriber server is capable of supporting hundreds of agents. In a fault-tolerant design, a Unified CM 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 guidelines defined in the *Sizing Cisco Unified Communications Manager Servers chapter*.

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 CM servers are needed within a Unified CM cluster, how many voice gateways are needed for each site and for the entire enterprise, how many servers and what types of servers are required for the Unified CCE software, and how many Unified IP IVR or Unified CVP servers are needed.

## 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

Prior to Release 8.0, Unified CCE supported monitoring of only a single line for all agent devices (Single Line Agent Mode).

Unified CCE Release 8.0 added support for monitoring multiple agent lines when Multi Line Agent Mode is enabled for the Peripheral.  This feature provides the following capabilities:

- Monitoring and reporting of calls on all lines on the phone. For additional details on reporting in a multiline environment, see the *Release Notes for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.0(1)*.

- Other than placing a call, all other call control on the non-ACD extensions is supported from multiline capable desktops, except for call initiation. Calls initiated from the hard phone can be controlled after initial call setup.
- Allows Unified CCE to support Join Across Line (JAL) and Direct Transfer Across Line (DTAL) features on the phone. These phone features are supported in all Cisco supported phone families.
- 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 not supported on ACD and non-ACD lines.
- 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.

For a list of supported agent phones, see the *Cisco Unified Contact Center Enterprise (Unified CCE) Software Compatibility Guide*.

The following three families of phones are in the Cisco portfolio:

**Cisco Unified IP Phones 7900 Series**

- The Unified CCE Agent Phone device supported prior to Release 8.0 of Unified CCE.
- JAL and DTAL are disabled by default.

**Cisco Unified IP Phones 6900 Series**

- Outbound campaign capability requires Cisco Unified Contact Center Enterprise 7.5(6) or later release.
- The 6900 Series phones do not support call waiting.
- JAL and DTAL features are enabled by default.
- The 6900 Series phones are supported in Single Line mode only when JAL and DTAL are disabled.
- Unified CM silent monitoring and recording and Remote Silent Monitoring (RSM) is supported with the 8.5(4) firmware load or higher.

**Cisco Unified IP Phones 8900 Series and 9900 Series**

- The 8900 Series and 9900 Series phones support cancel and swap features.
- JAL and DTAL features are on all of the time, so contact center agents can merge or transfer calls on the ACD line with calls on other extensions on the phone. Because of this, 8900 Series and 9900 Series phones are supported only when the Multi Line Agent Mode feature of Unified CCE is enabled.

Unified IP Phones 8900 Series and 9900 Series are not supported for Cisco Finesse Release 8.5(3).

Unified CCE Release 8.5 adds support for the Agent Greeting feature. This feature relies on the Unified CM phone Built-In-Bridge (BIB) functionality to play back the greeting to both the caller and the agent.

Agent Greeting requires:

1. The phones have the BiB feature.
2. The phones must be running the firmware version delivered with Unified CM 8.5(1) or greater.
3. The phones must be configured as BiB enabled in Unified CM.

**Note:** These requirements apply to local agents only.

For a list of agent phones and required firmware to support the Agent Greeting feature, please see the Cisco Unified Contact Center Enterprise (Unified CCE) Software Compatibility Guide.

In UCCE Version 8.5(1), Agent Greeting is not supported for mobile agent, or for parent/child deployments. Release 8.5(2) adds mobile agent and parent/child support for Agent Greeting.

**Note:** Agent Greeting for Parent/Child is only supported for a very specific Parent/Child configuration, where calls are queued at a CCX (IP-IVR) on the child system PG, and requires a dedicated CVP at the child on a dedicated VRU PG to provide the agent greetings. Agent Greeting in Parent/Child configurations must be approved by the Cisco Assessment to Quality (A2Q) process and requires Cisco Design Mentoring Service to assure that the deployment model is designed and sized correctly to support the Agent Greeting feature.

## Cisco Unified Customer Voice Portal

Unified Customer Voice Portal (Unified CVP) is a software application that runs on industry-standard servers such as Cisco Media Convergence Servers (MCS). It provides prompting, collecting, queuing, and call control services using standard web-based technologies. The Unified CVP architecture is distributed, fault tolerant, and highly scalable. With the Unified CVP system, voice is terminated on Cisco IOS gateways that interact with the Unified CVP application server using VoiceXML (speech) and H.323 or SIP (call control).

The Unified CVP software is tightly integrated with the Cisco Unified CCE software for application control. It 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 also invoke external VoiceXML applications to be executed by the Unified CVP VoiceXML Server, an Eclipse and J2EE-based scripting and web server environment. VoiceXML Server is well suited for sophisticated and high-volume IVR 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 Unified CVP solution components can be achieved by Cisco Content Services Switch (CSS) and Cisco IOS Gatekeepers or Cisco Unified Presence SIP Proxy Servers.

Unified CVP can support multiple grammars for prerecorded announcements in several languages. Unified CVP can optionally provide automatic speech recognition and text-to-speech capability. Unified 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 Unified 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; and when Unified CCE routes the call to an agent, the agent is able to send videos to a caller from the agent desktop application. For more information, see the latest version of the *Cisco Unified Customer Voice Portal (CVP) Solution Reference Network Design (SRND)*.

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 pre-recorded 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 CM 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 CM to transfer the call to the selected agent phone.

Unified IP IVR is a software application that runs on Cisco MCS Servers. You can deploy multiple Unified IP IVR servers with a single Unified CM cluster under control of Unified CCE.

Unified IP IVR has no physical telephony trunks or interfaces like a traditional IVR. The telephony trunks are terminated at the voice gateway. Unified CM 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 CM through the Java Telephony Application Programming Interface (JTAPI), and the Unified IP IVR communicates with Unified CCE through the Service Control Interface (SCI) with a VRU Peripheral Gateway or System Peripheral Gateway.

See the chapter Sizing Contact Center Resources for discussion about how to determine the number of IVR ports required. For deployments requiring complete fault tolerance, a minimum of two Unified IP-IVRs are required. See the chapter Design Considerations for High Availability, which provides details on Unified CCE fault tolerance.

Unified IP IVR supports the Whisper Announcement features starting in Release 8.5(1) SU1 with following requirements:

- You must have Unified CCE Release 8.5(2) or later.
- Only in Parent/Child deployment and IP IVR is deployed in Child with Unified CCE System PG.

## Unified Presence Server

Cisco Unified Presence Server is used by the Cisco Agent Desktop product in the solution to locate appropriate resources (agents or other employees of the enterprise) for managing the call.  See the chapter Unified Contact Center Enterprise Desktop for details on Cisco Agent Desktop.

## 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 emails 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. (See Unified CCE Software Components for a description of these components.)

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 Call Routers, data servers, and so on 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.

# Cisco Unified Contact Center Enterprise (Unified CCE)

Cisco Unified CCE is the software application that provides the contact center features, including agent state management, agent selection, call routing and queue control, IVR control, CTI Desktop screen pops, and contact center reporting. Unified Contact Center Enterprise (Unified CCE) runs on Cisco MCS servers or exact equivalents unless otherwise specified in the Sizing Unified CCE Components and Servers chapter and the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*. It relies on the Microsoft Windows Server 2003 operating system software (or Windows Server 2008 R2 starting in release 8.5(3)) and the Microsoft SQL Server 2005 database management system. Windows Server 2008 R2 is optional in release 8.5(3); Windows Server 2003 is still supported.

The supported servers can be single, dual, or quad Pentium CPU servers in single or multi-core variations with varying amounts of RAM. This variety of supported servers allows the Unified CCE software to scale and to be sized to meet the needs of the deployment requirements. (The Sizing Unified CCE Components and Servers chapter provides details on server sizing.)

Beginning with Release 8.5(1), Unified CCE supports the Whisper Announcement and Agent Greeting features.

## Unified CCE Software Components

This section 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 servers. 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 server being used (single, dual, or quad CPU). 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, Extended 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. |
| Logger | The database server that stores contact center configuration data and temporarily stores historical reporting data for |

| **Unified CCE Software Components** | **Description** |
|---|---|
| | distribution to the data servers. |
| CTI Object Server (CTI OS) | CTI interface for Agent Desktops. |
| Peripheral Gateway (PG) | Interfaces to various ‘peripheral’ devices, specifically to Unified CM, 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 CM PIM. |
| Unified CM Peripheral Interface Manager (PIM) | Part of a PG that interfaces to a Unified CM 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. |
| Network Interface Controller (NIC) | Interfaces to the public switched telephone network (PSTN), which enables Unified CCE to control how the PSTN routes a call. |
| 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 “lighter weight” 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 server, the server is referred to as a *Rogger*. When the CallRouter, Logger, and Peripheral Gateway modules run on the same server, the server is referred to as a *Progger*. In lab environments, the system Administration & Data Server can also be loaded onto the Progger to create a server known as a *Sprawler* configuration; however, this configuration is approved only for lab use and is not supported in customer production environments.

## Redundancy and Fault Tolerance

The CallRouter and Logger are deployed in a paired redundant fashion. This redundant configuration is also referred to as *duplex mode*, whereas a non-redundant configuration is said to be running in *simplex mode*. (Note: Simplex mode is *not* supported for production environments.) 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 servers. When both sides are running (normal operation), the configuration is in *duplex* mode. When one side is down, the configuration is said to be running in *simplex* mode. The two sides are for redundancy, not load-balancing. Either side is capable of running the full load of the solution. The A and B sides are both executing the same set of messages and, therefore, producing the same result. In this configuration, logically, there appears to be only one CallRouter. The CallRouters run in synchronized execution across the two servers, which mean both sides of the duplex servers process every call. In the event of a failure, the surviving CallRouter will pick up the call mid-stream and continue processing in real-time and without user intervention.

The Peripheral Gateway components run in hot-standby mode, meaning that only one of the Peripheral Gateways is actually active and controlling Unified CM or the IVR. When the active side fails, the surviving side automatically takes over processing of the application. During a failure, the surviving side is running in simplex mode and will continue to function this way until the redundant side is restored to service and the configuration automatically returns to duplex 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 may connect to either server. In the event of the failure of any one server, clients can automatically reconnect to the alternate server.

The Administration & Data Servers for configuration and real-time data are deployed in pairs for fault tolerance, with multiple pairs deployed for scalability. The data flows are described in the detailed section on Administration & Data Server and Administration Client. The Administration & Data Servers for historical data follow an n+1 architecture for redundancy and scalability with each having a Logger side (A or B) as its preferred and primary data source.

## Customer Instance and Unified CCH

The Cisco Unified Contact Center Hosted (Unified CCH) solution is largely the same as Unified CCE, but it supports multi-tenant or shared servers to manage multiple *customer instances*.

All Unified CCE systems are deployed as a single instance (using the same instance name and number in setup) across all the Unified CCE components.

## Peripheral Gateway (PG) and PIMs

For each Unified CM cluster in the Unified CCE environment, there is a Unified CM PIM on an Agent Peripheral Gateway. For scalability requirements, some deployments may require multiple PIMs for the same Unified CM cluster; they may be on the same PG and physical server or they may be separate.

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 Unified CM cluster.

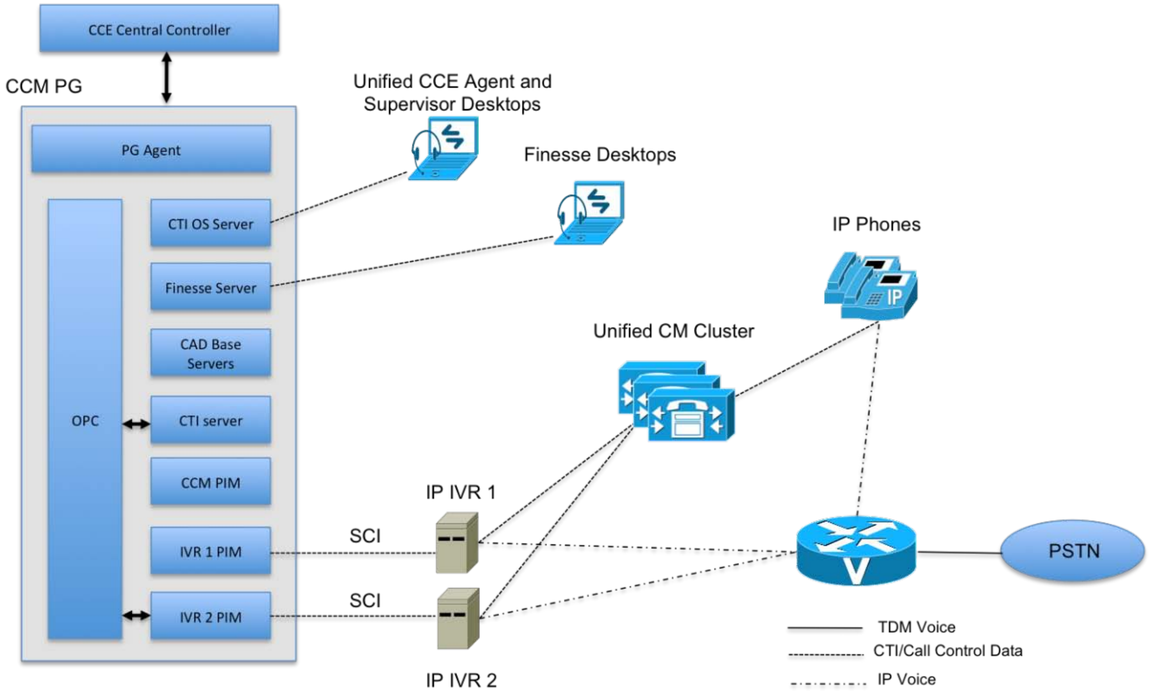
**Note** The CTI OS components on Side A and Side B are simultaneously active to load-balance desktop communication.

For each Unified IP IVR or CVP Call Server, there is one VRU PIM. VRU PIMs may be part of the Agent PG.

Often, the Unified CM PIM, the CTI Server, the CTI OS, and multiple VRU PIMs may run on the same server.

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. Figure 2 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 Unified CM 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.

## Network Interface Controller

The Network Interface Controller (NIC) is an optional component that interfaces to the public switched telephone network (PSTN). Intelligently routing a call before the call is delivered to any customer premise equipment is referred to as pre-routing. Only certain PSTNs have NICs supported by Unified CCE. For a detailed list of PSTN NICs and details on Unified CCE pre-routing, see the *Pre-installation Planning Guide for Cisco Unified ICM Enterprise & Hosted*.

## Unified CCE Agent Desktop Options

See the chapter Unified Contact Center Enterprise Desktop for detailed information about the options for Agent Desktops including CAD and CTI OS interfaces

Cisco offers the following interfaces for Unified CCE agents (see Figure 3):

- Cisco Agent Desktop

  Cisco Agent Desktop provides an out-of-the-box, feature-rich, desktop solution for Unified CCE. The desktop application can be deployed in various ways:

  - Windows application
  - Browser-based application
  - Cisco Unified IP Phone Agent, where there is no desktop application at all but just an XML application on the IP phone
  - 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.

- 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 only
  - an extensible desktop interface using standard OpenSocial gadgets
  - server features available to applications using documented REST APIs

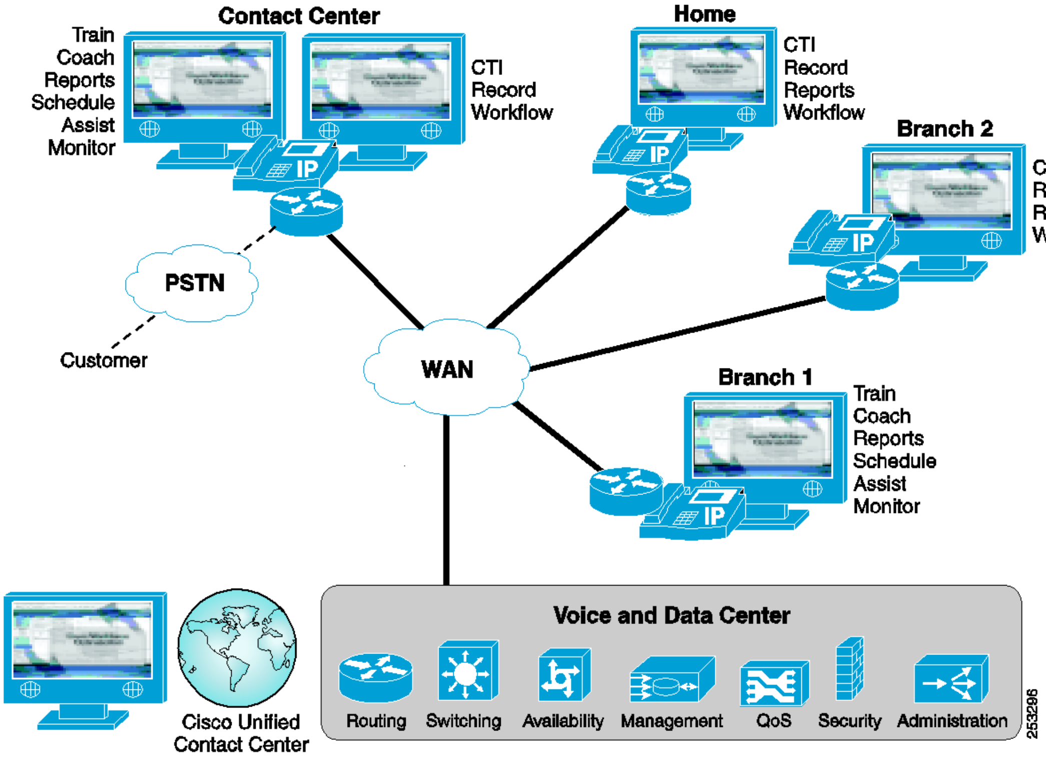
- CRM Connectors

  Cisco offers pre-built, certified CRM Connectors for CRM packages including SAP, Siebel (using CTI OS driver), Salesforce.com, Microsoft Dynamics CRM, and PeopleSoft. These integrated solutions enable call control from the CRM user interface (Answer, Drop, Hold, Un-Hold, Blind or Warm Transfers, and Conferences), outbound and consultative calls from the CRM desktop, and delivery and manipulation of Call Context Data (CTI screen pop).

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.

For more information about desktop selection and design considerations, see the chapter Unified Contact Center Enterprise Desktop.

**Figure 3** *Variety of Agent Interfaces for Unified CCE*



## Administration & 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 servers:

- 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 Deployment Models 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.

Each Administration & Data Server must be installed on a separate server 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 server as the CallRouter and Logger.

## 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 a 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 will handle a particular contact).

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

- Agent Re-skilling Web Tool (Unified CCE only)
- Configuration Management service (CMS) Node.
- Internet Script Editor Server—HTTPS (default protocol) connection for Script Editor clients

For details on the use of these and other configuration tools, see the *Administration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*.

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 (see Figure 4) 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.

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

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. They can be deployed in a multi-instance configuration for Hosted CCE.

**Figure 4** *Communication Between Unified CCE Central Controller and Administration & Data Server*

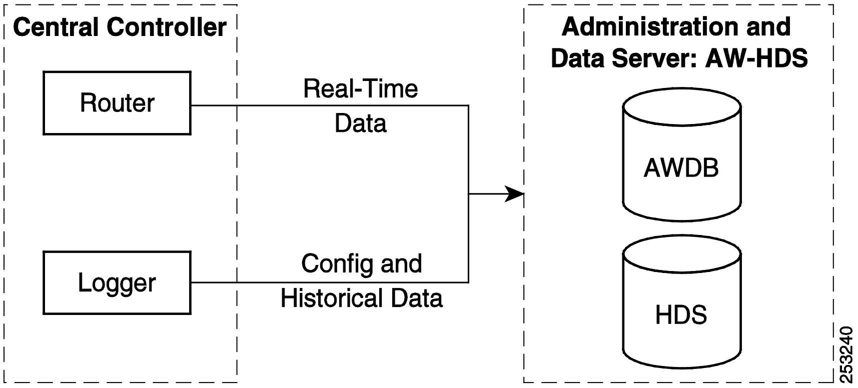
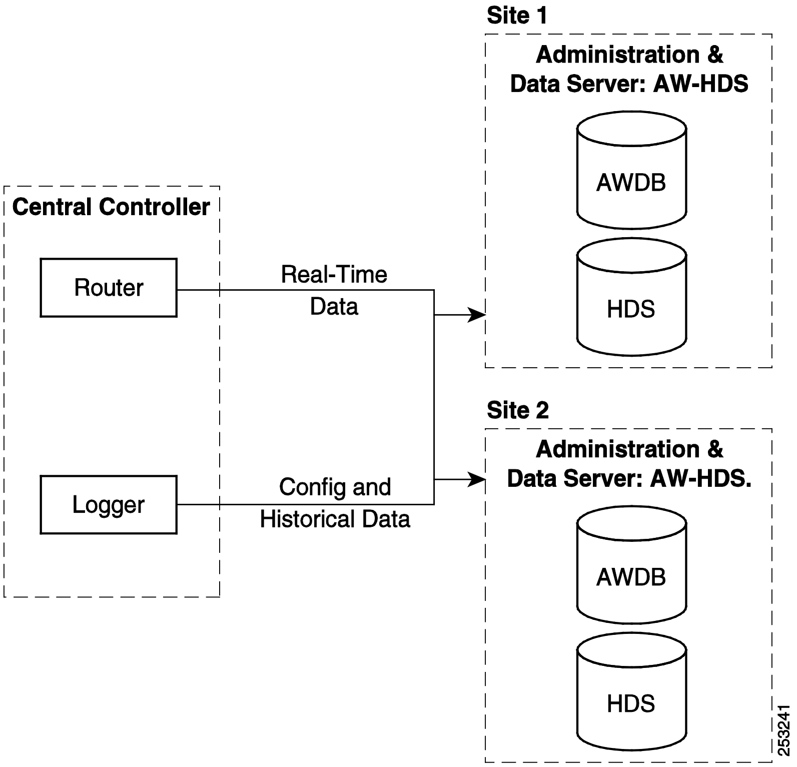


**Figure 5** *Communication Between Unified CCE Central Controller and Multiple Administration & Data Servers*



AWs, Configuration-Only Administration Servers, and Administration Clients may operate only as a single instance on a given server. 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.

These Data Servers are deployed with a primary AW as a single server serving all three roles (AW-HDS-DDS). AW-HDS-DDS can only be enabled on a primary AW. 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 Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*. (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 Suite Intelligence Center User Guide*. Unified Intelligence Center must be installed on a separate server; 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

See the chapter on Cisco Unified Contact Center Management Portal for more information about Cisco Unified Contact Center Management Portal.

## JTAPI Communications

In order for JTAPI communications to occur between Unified CM and external applications such as Unified CCE and Unified IP IVR, a JTAPI user ID and password must be configured within Unified CM. Upon startup of the Unified CM PIM or on startup of the Unified IP IVR, the JTAPI user ID and password are used to sign in to Unified CM. This sign-in process by the application (Unified CM PIM or Unified IP IVR) establishes the JTAPI communications between the Unified CM cluster and the application. A separate JTAPI user ID is required for each Unified IP IVR server. In a Unified CCE deployment with one Unified CM cluster and two Unified IP IVRs, three JTAPI user IDs are required: one JTAPI user ID for Unified CCE and two JTAPI user IDs for the two Unified IP IVRs. The best practice is one PG User for each PG Pair.

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

For example, suppose a deployment has a voice gateway homed to node 1 in a cluster, and node 2 executes the CTI Manager process that communicates to Unified CCE. When a new call arrives at this voice gateway and needs to be routed by Unified CCE, node 1 sends an intra-cluster message to node 2, which sends a route request to Unified CCE to determine how the call will be routed.

Each Unified IP IVR also communicates with only one CTI Manager (or node) within the cluster. The Unified CM PIM and the two Unified IP IVRs from the previous example could each communicate with different CTI Managers (nodes) or they could all communicate with the same CTI Manager (node). However, each communication uses a different user ID. The user ID is how the CTI Manager keeps track of the different applications.

When the Unified CM PIM is redundant, only one side is active and in communication with the Unified CM cluster. Side A of the Unified CM PIM communicates with the CTI Manager on one Unified CM node and side B of the Unified CM PIM communicates with the CTI Manager on another Unified CM node. The Unified IP IVR does not have a redundant side, but the Unified IP IVR does have the ability to fail over to another CTI Manager (node) within the cluster if its primary CTI Manager is out of service. (For more information about failover, see the Design Considerations for High Availability chapter.

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

- Routing control

  Routing control messages provide a way for Unified CM to request routing instructions from Unified CCE.

- Device and call monitoring

  Device monitoring messages provide a way for Unified CM 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 Unified CM 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 CM requests routing instructions from Unified CCE. For example, when Unified CM receives the routing response from Unified CCE, Unified CM attempts delivery of the call to the agent phone by instructing the phone to begin ringing. At that point, Unified CM notifies Unified CCE that the device (phone) has started ringing and that notification enables the agent's answer button on the desktop application. When the agent clicks the answer button, Unified CCE instructs Unified CM to make the device (phone) go off-hook and answer the call.

In order for the routing control communication to occur, Unified CM requires the configuration of a CTI Route Point. A CTI Route Point is associated with a specific JTAPI user ID, and this association enables Unified CM to know which application provides routing control for that CTI Route Point. Directory (Dialed) Numbers (DNs) are then associated with the CTI Route Point. A DN is associated to a CTI Route Point that is associated with Unified CCE JTAPI user ID. This enables Unified CM to generate a route request to Unified CCE when a new call to that DN arrives.

In order for the phones to be monitored and controlled, they also must be associated in Unified CM with a JTAPI user ID. Starting with Unified CM 8.0, when using Extension Mobility or Extension Mobility Cross Cluster, it is possible to associate an Extension Mobility device profile instead. In a Unified CCE environment, the IP phones or the corresponding Extension Mobility device profiles are associated with Unified CCE JTAPI user IDs. When an agent logs in from the desktop, the Unified CM PIM requests Unified CM to allow the PIM to begin monitoring and controlling that phone. Until the login has occurred, Unified CM does not allow Unified CCE to monitor or control that phone. If the device or the corresponding Extension Mobility device profile has not been associated with a Unified CCE JTAPI user ID, then the agent login request will fail.

Support for Extension Mobility Cross Cluster (EMCC) is provided in Release 8.0. The Unified CCE PIM phone registers to the local Unified CM after Extension Mobility login and it looks like an agent situated across a WAN. The Unified CCE peripheral is managing the agent devices based on the Extension Mobility profile rather than on a phone device in the Application User on Unified CM. For more details, see the *Cisco Unified Communications Solution Reference Network Design Guide (SRND)*.

Because EMCC is now supported, you can associate Extension Mobility devices by two methods; either by device or by user profile. The best practice is to associate the Extension Mobility profile to the CCE Application User on UC Manager.

Because the Unified IP IVR also communicates with Unified CM using the same JTAPI protocol, these same three types of communications also occur with the Unified IP IVR. Unlike Unified CCE, the Unified IP IVR provides both the application itself and the devices to be 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 IVR. Its ports are logical ports (independent software tasks or threads running on the Unified IP IVR application server) called CTI Ports. For each CTI Port on the Unified IP IVR, there needs to be a CTI Port device defined in Unified CM.

Unlike a traditional PBX or telephony switch, Unified CM does not select the Unified IP IVR port to which it will send the call. Instead, when a call needs to be made to a DN that is associated with a CTI Route Point that is associated with a Unified IP IVR JTAPI user, Unified CM asks the Unified IP IVR (through JTAPI routing control) which CTI Port (device) will handle the call. Assuming the Unified IP IVR has an available CTI Port, the Unified IP IVR will respond to the Unified CM routing control request with the Unified CM device identifier of the CTI Port that is going to handle that call.

The following scenarios are examples of device and call control performed by the Unified IP IVR.

When an available CTI Port is allocated to the call, a Unified IP IVR workflow is started within the Unified IP IVR. When the Unified IP IVR workflow executes the accept step, a JTAPI message is sent to Unified CM to answer the call on behalf of that CTI Port (device). When the Unified IP IVR workflow wants the call transferred or released, it again instructs Unified CM on what to do with that call.

When a caller releases the call while interacting with the Unified IP IVR, the voice gateway detects the caller release and notifies Unified CM by using H.323 or Media Gateway Control Protocol (MGCP), which then notifies the Unified IP IVR by using JTAPI. When DTMF tones are detected by the voice gateway, it notifies Unified CM via 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, the CTI Port devices on Unified CM must be associated with the appropriate Unified IP IVR JTAPI user ID. If you have two 150-port Unified IP IVRs, you would have 300 CTI ports. Half of the CTI ports would be associated with JTAPI user Unified IP IVR 1, and the other half of the CTI ports would be associated with JTAPI user Unified IP IVR 2.

While Unified CM can be configured to route calls to Unified IP IVRs on its own, routing of calls to the Unified IP IVRs in a Unified CCE environment will be done by Unified CCE (even if you have only one Unified IP IVR and all calls require an initial IVR treatment). Doing so will ensure 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.

## 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 (SRND) Guide for Unified Contact Center Enterprise, Hosted, and ICM*.

## Cisco Unified Outbound Option

Agents can handle both inbound and outbound contacts, which helps in optimizing contact center resources. The Cisco Unified 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. (See the Outbound Option for Cisco Unified Contact Center Enterprise and Hosted chapter for details.)

## 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. (For design guidance and considerations for implementing Cisco Unified Mobile Agent, see the Cisco Unified Mobile Agent chapter.)

## Unified SCCE 7.x

**Note** There is *no* 8.0 release for Unified SCCE; it remains at Release 7.5. Users can migrate their Unified SCCE 7.1, 7.2 or 7.5 systems directly to Unified CCE 8.0 using the Unified CCE installer.

Cisco Unified System Contact Center Enterprise 7.x (Unified SCCE 7.x) is a deployment model that simplifies installation and configuration by using three predefined configurations for Unified CCE. Unified SCCE 7.x uses a single installer to simplify installation and configuration and it provides web-based administration. Configuration of Unified SCCE 7.x is further simplified by removing Services, Translation Routes, Device Targets, Labels, and Sub Skill Groups.

Unified SCCE 7.x provides fault tolerance through the duplex operation on the Central Controller and Agent/IVR Controller. Unified SCCE 7.x can connect to a parent Unified ICM and the connection is made between the child Unified CCE System PG and the parent Gateway PG.

For further information about Unified SCCE 7.x, see the Cisco Unified Contact Center Enterprise 7.0, 7.1, and 7.2 Solution Reference Network Design (SRND).

## 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 & Hosted*.

### Solution Call Trace

Also integrated with the Unified CM Real-Time Monitoring Tool is the Analyze Call Path feature. This tool enables the administrator to search for failed calls (based on simple filter parameters) and then trace those calls through each component of a Unified Communications solution. The Analyze Call Path tool first identifies call records for failed calls (for example, dropped or abandoned calls) in a particular application. The administrator may then select a particular call and have the tool collect all related call records (and traces) from all other Unified Communications components such as Unified CVP or Unified CM. Analysis of the call record collection can then reveal the exact location and nature of the call failure.

Additionally, the Analyze Call Path tool allows the administrator to search for call records based on other criteria such as Originating Date and Time, Calling Number, or Called Number and to collect all associated call records for display and analysis.

For more information about Unified Communications Analyze Call Path capabilities in RTMT, see the *Analysis Manager User Guide*.

### 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 will alert 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.

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*.

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 & Hosted*.

# Combining IP Telephony and Unified CCE in the Same Unified CM Cluster

It is possible for a Unified CM 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 CM 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 *Cisco Unified Contact Center Enterprise Software Compatibility Guide*.

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 CM 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 CM is required to maintain to support those agents.

Because of these software and environmental limitations, it might sometimes be advantageous to separate the Unified CM clusters for IP Telephony extensions from the Unified CM clusters for Unified CCE extensions. It is important to consider the environment where Unified CCE is being deployed to determine whether a separate Unified CM 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 typically will not have voicemail or any 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 voicemail and other calling features.

The position of the line on the phone determines which line will be answered or used if the agent just picks up the handset. In a typical call center, the ACD line would be the first line on the phone to make it easier for the agent to answer inbound ACD calls and also to ensure that any calls the agent makes using the phone are tracked by the system as external calls for that agent. The agent's state will change based on this line. If the agent picks up the phone to place a call, the agent will be put into not-ready mode and the Unified CCE will 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 would not want to 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 will allow the users to pick up the phone and answer the first line as well as use this line for all calls they place by default. To answer an ACD call, they will have to select that line on the phone or use the agent desktop to answer that line appearance directly. Also, be aware that they will 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 will 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 could 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 CM instead of the agent DID in order 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 CM wildcard feature can be used, although Unified CCE will require distinct routing DNs (one for each agent).

# Agent Phones in Countries with Toll-Bypass Regulations

Some countries, such as India, have telecommunications regulations that 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 the organization, and a second one to access the local PSTN. To ensure adherence to the regulations in such countries, agents used to have only one line with access to customer calls and a different phone (for example, a soft phone) to access a VoIP line for contacting fellow teammates or experts located outside the contact center.

The Logical Partitioning feature in Cisco Unified CM provides the same capability through a telephony system to control calls and features on the basis of specific allowed or forbidden configurations. A common telephony system in a contact center environment can provide access to both the PSTN and VoIP networks, therefore configurations are required to provide controlled access and to avoid toll bypass.

The Logical Partitioning feature can be enabled and configured in Unified CM to prevent toll-bypass calls, thus allowing agents in a Unified CCE system to use the same phone for receiving customer calls and for making or receiving VoIP calls to and from other people within the organization. Although this 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 due to 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 re-queuing.

Call queuing in a Unified CCE deployment requires use of an IVR platform that supports the SCI interface to Unified CCE. The Unified IP IVR is one such platform. Cisco offers another IVR platform, Unified CVP, which can be used as a queuing point for Unified CCE deployments. The Deployment Models chapter provides considerations for deployments with Unified CVP. Traditional IVRs can also be used in Unified CCE deployments, and Deployment Models also provides considerations for deployments with traditional IVRs.

In a Unified CCE environment, an IVR is used to provide voice announcements and queuing treatment while waiting for an agent. The control over the type of queuing treatment for a call is provided by Unified CCE 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 IVR to play a particular queuing treatment.

While the IVR is playing the queuing treatment (announcements) to the caller, Unified CCE waits for an available agent with a particular skill (as defined within the routing script for that call). When an agent with the appropriate skill becomes available, Unified CCE reserves that agent and then instructs the IVR to transfer the voice path to that agent's phone.

# Transfers and Conferences in a Unified CCE Environment

Transfers and conferences are commonly used in contact centers, and they require special attention to ensure the proper system resources are available and configured correctly. (See Deployment Models)

When Call Recording is enabled in the DN configuration for an agent phone, the codec will not be renegotiated when establishing a conference. As a result, if two phones are connected using G.722 and a conference call is initiated, the codec will not be renegotiated to G.711 and a hardware conference bridge or transcoder is required.

CHAPTER 2

# Deployment Models

There are numerous ways that Unified Contact Center Enterprise (Unified CCE) can be deployed, but the deployments can generally be categorized into the following major types or models:

- Single Site
- Multisite Centralized Call Processing
- Multisite Distributed Call Processing
- Clustering over the WAN

Many variations or combinations of these deployment models are possible. The primary factors that cause variations within these models are as follows:

- Location of Unified CCE servers
- Location of voice gateways
- Choice of inter-exchange carrier (IXC) or local exchange carrier (LEC) trunks
- Pre-routing availability
- IVR queuing platform and location
- Transfers
- Traditional ACD, PBX, and IVR integration
- Sizing
- Redundancy

This chapter discusses the impact of these factors (except for sizing) on the selection of a design. With each deployment model, this chapter also lists considerations and risks that must be evaluated using a cost benefit analysis. Scenarios that best fit a particular deployment model are also noted.

In this chapter, section titles are prefaced by the type of factor discussed in the section. The factors are classified into the following categories:

- IPT—Cisco Unified Communications deployment factors (how Cisco Unified Communications Manager and the voice gateways are deployed)
- Unified CCE—Unified CCE deployment factors (such as what PG is used)
- IVR—IVR and queuing deployment factors (if Unified CVP or Unified IP IVR is used)

A combination of these deployment models is also possible. For example, a multisite deployment may have some sites that use centralized call processing (probably small sites) and some sites that use distributed call processing (probably larger sites). Examples of scenarios where combinations are likely are identified within each section.

Also in this chapter is a section on integration of traditional ACD and IVR systems into a Unified CCE deployment with considerations on hybrid PBX/ACD deployments. Sizing and redundancy are discussed in later chapters of this Unified CCE design guide. For more information about the network infrastructure required to support a Unified CCE solution, see the latest version of the *Cisco Network Infrastructure Quality of Service Design Guide*.

For more information about deployment models for Unified CCE and Cisco Unified Communications, see the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

# What’s New in This Chapter

The following topics are new in this chapter or have changed significantly from previous releases of this document.

- Whisper Announcement Feature Support.
- Agent Greeting Feature Support.

The new and changed information in this chapter is extensive; read the entire chapter.

# General Deployment Options

This section describes options that can apply to many of the specific deployment models listed in the rest of this document. It describes at a high level the tradeoffs that can be made when installing the Unified CCE software.

## Agent Peripheral Options

There are two types of Agent Peripherals that can be installed to handle Unified CCE agents. This section talks about those two types of peripherals and the strengths and weaknesses of each.

## Enterprise Unified CCE Peripheral

In Enterprise Unified CCE Peripheral deployments, the Unified CCE software treats the VRU and Unified CM as separate peripherals. This means that routing must be done once for each peripheral and Termination Call Detail records are created for each peripheral each time a call touches the peripheral. Translation routes must be used to send calls between the VRU and Unified CM.

The Unified CM PG and VRU PG can be deployed independently or the Unified CM and VRU can be deployed in a Generic PG with separate PIMs for Unified CM and VRU.

These deployments provide a large degree of flexibility in configuration. For example, this deployment is capable of using either a Unified CVP or a Unified IP IVR attached to the VRU peripheral and load balancing can be done between multiple IVRs.

These Unified CCE deployments may not be used where Unified CCE is a child to a Unified ICM; a Unified CCE System Peripheral deployment must be used for that solution. For more information, see the section on Parent/Child.

## Unified CCE System Peripheral

The Unified CCE System Peripheral combines the functionality of both the VRU peripherals (up to five Unified IP IVR peripherals) and a single Unified CM peripheral together into a single logical Unified ICM peripheral. The Unified CCE treats these Unified IP IVR and Unified CM 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.

Additionally, because the Unified CCE System PG is a single peripheral, Termination Call Detail (TCD) records and other reporting data include the information for the call during the entire time it is on the peripheral. Instead of getting up to three TCDs for each call (one for the original route, one for the IVR, and one for the agent handle time), only a single record is created with the Unified CCE System PG.

The Unified CCE System PG does not support Unified CVP, therefore all queuing and treatment in the Unified CCE System PG is done using Unified IP IVR. Note that a separate Unified CVP on its own PG can be used in conjunction with the Unified CCE System Peripheral.

## Unified CCE: Administration & Data Server

Beginning with Unified CCE 8.0(1), the Distributor AW (with or without HDS) is renamed as 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.

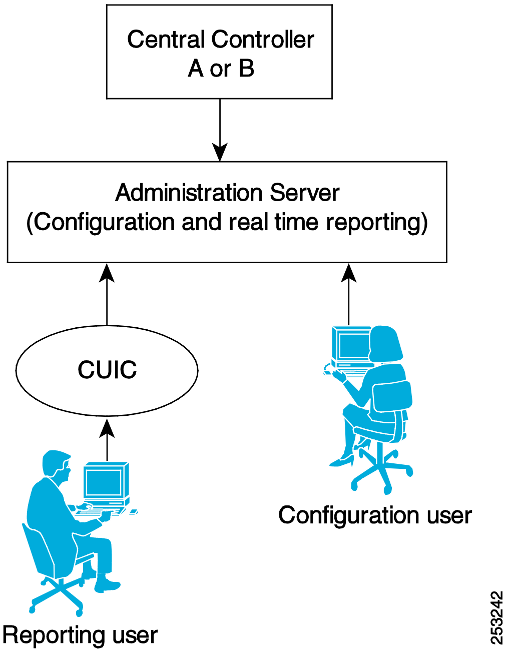
**Roles:**

The Administration & Data Servers are classified into 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). (See Figure 6) No historical reporting is supported.

**Figure 6** *Configuration and Real-Time Reporting*

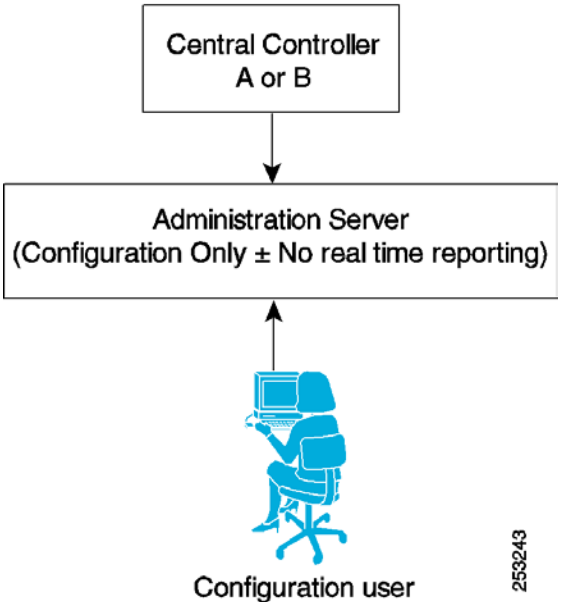


### Administration Server (Configuration-Only AW)

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. (See Figure 7)

This deployment role allows Contact Center Hosted using CCMP to configure a specific Unified CCE Customer Instance through the ConAPI interface. The load is low enough on such a lightweight Administration & Data Server that a single server is sufficient if deployed using VMware. For historical and real-time data, the customer can use a shared AW-HDS or AW-HDS-DDS.

**Figure 7** *Configuration-Only AW*

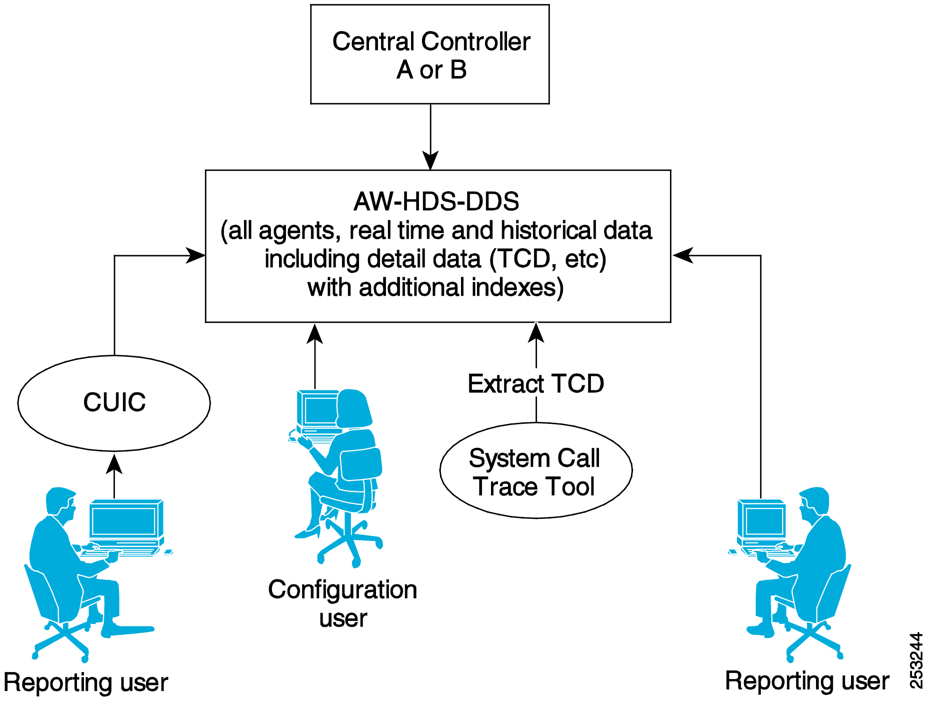


### Administration Server, Historical Data Server, and Detail Data Server (AW-HDS-DDS)

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 meet the requirements for the System Call Trace tool and to feed historical data. (See Figure 8)

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

**Figure 8** *Administration Server, Historical Data Server, and Detail Data Server (AW-HDS-DDS)*



### Administration Server and Historical Data Server (AW-HDS)

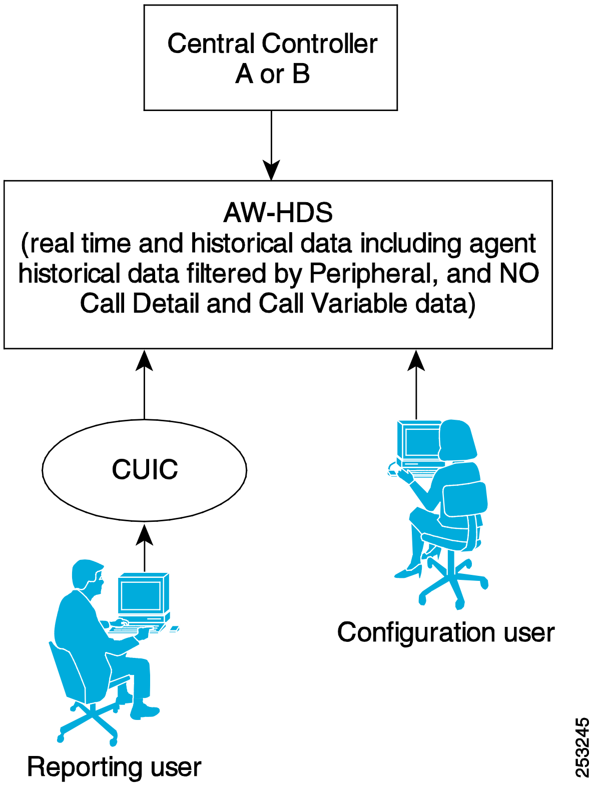
This Administration & Data Server deployment role provides the capability for configuration changes as well as both real-time and historical reporting. Real-time and historical reporting is supported using Cisco Unified Intelligence Center (Reporting client). (See Figure 9)

**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 procedure
- Data extraction for System Call Trace tool
- Feed to Cisco Unified Intelligence Suite (Unified IS) (Archiver)

**Figure 9** *Administration Server and Historical Data Server (AW-HDS)*


Central Controller
A or B
AW-HDS
(real time and historical data including agent historical data filtered by Peripheral, and NO Call Detail and Call Variable data)
CUIC
Configuration user
Reporting user
253245

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

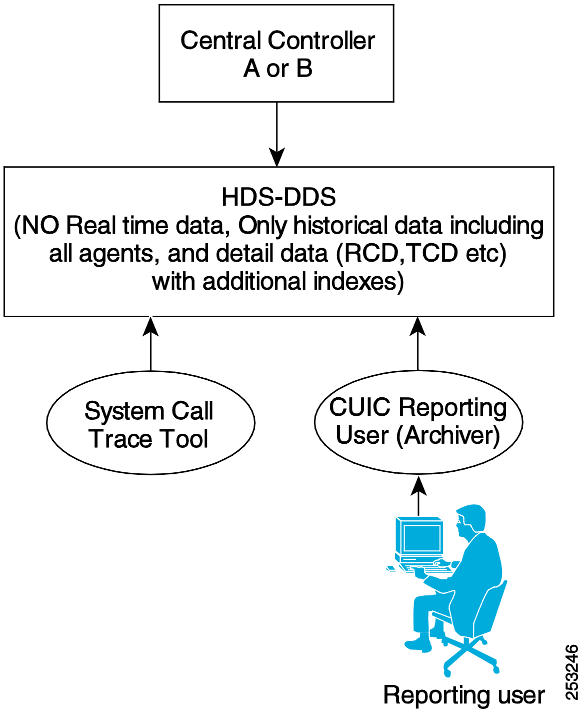
The HDS-DDS deployment model is used specifically for data extraction and for custom reports for call detail (TCD and RCD) only. (See Figure 10.)

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


Central Controller
A or B
HDS-DDS
(NO Real time data, Only historical data including all agents, and detail data (RCD,TCD etc) with additional indexes)
System Call
Trace Tool
CUIC Reporting
User (Archiver)
Reporting user
253246

## Deployment Options

The following deployment options are based on the varying capacity in terms of number of agents, call load, number of reporting users, and other operating conditions mentioned in the Sizing Unified CCE Components and Servers chapter.

### Option 1: Less than 4000 Agents

This deployment option is applicable to the following limitations:

- Sizing within the operating conditions listed in the Sizing Unified CCE Components and Servers chapter in the SRND and the Hardware and System Software Specification (Bill of Material) for Cisco Unified ICM/Contact Center Enterprise and Hosted specification
- A maximum of 4000 active agents
- A maximum of 400 reporting users

The number of HDSs is limited to 2 AW-HDS-DDSs per Logger side.

### Option 2: Greater than 4000 Agents

This deployment option is applicable to the following limitations:

- Sizing within the operating conditions listed in the Sizing Unified CCE Components and Servers chapter in the SRND and the Hardware and System Software Specification (Bill of Material) for Cisco Unified ICM/Contact Center Enterprise and Hosted specification
- More than 4000 active agents
- More than 400 reporting users

The number of Administration & Data Server roles is limited to three AW-HDSs and 1 HDS-DDS or AW-HDS-DDS per Logger side.

Cisco Unified Intelligence Center (Reporting client) is supported but not part of the out-of-the-box solution.

## Unified SCCE

There is no 8.0 release and above for Unified SCCE; it will remain at release 7.5. Users can migrate their Unified SCCE 7.1, 7.2, or 7.5 systems directly to Unified CCE 8.0 using the Unified CCE installer. Unified CCE 8.0 and above supports Parent/Child integrations to Unified SCCE 7.x.

## Parent/Child

The Unified CCE Gateway PG allows Unified CCE or Unified CCX to appear as a traditional ACD connected to the Unified ICM system. The Unified CCE Gateway PG does this by providing a PG to the Unified ICM system that communicates to the CTI interface of Unified CCE System PG or Unified CCX.

When the Unified CCE Gateway PG is used in a deployment, its relationship to Unified ICM is termed a *parent* and Unified CCE is called the *child*:

- Parent

  The Unified ICM system that serves as the network or enterprise routing point. The child looks like an ACD to the parent, which uses the appropriate Unified CCE Gateway PG (Enterprise or Express) to communicate to the CTI interface on the child Unified CCE. The parent can perform all functions that a Unified ICM can usually perform, including pre- and post-routing and end-to-end call tracking using translation routes.

- Child

  The Unified CCE System PG or Unified CCX system that is set up to function as an ACD. The child can receive calls that are translation-routed from the parent, but it is not aware of any other peripherals attached to the parent. The child can also post-route calls from the Unified CCE to the parent where the call can be handled like any other Unified ICM call. For example, the call could be translation-routed to any TDM or IP ACD controlled by the Unified ICM or queued in the Unified ICM network queue point with Unified CVP.

In the parent/child model, the child Unified CCE is configured to function completely on its own and does not need the connection to the parent to route calls to agents. This independence provides complete local survivability for mission-critical contact centers if the network between the child and parent goes down or if there is a problem with the parent or the Unified CCE Gateway PG connection.

Configuration objects entered into the child system can automatically be sent to the parent Unified ICM and inserted into the Unified ICM configuration, thus eliminating the need to configure objects twice (once in the local ACD and again to match the configuration in the Unified ICM itself for routing and reporting). This functionality can also be turned off for situations where the customer does not want automatic configuration updates, such as with an outsourcer using the Unified CCE child system where not all of the agents, skill groups, and call types on that child system apply to the customer's Unified ICM system.

The Unified CCE Gateway PG can connect to a Unified CCE child that is using the Unified CCE System PG or to Unified CCX. If the Unified CCE child has multiple Unified CCE System PGs and peripherals, a separate Unified CCE Gateway PG peripheral must be installed and configured for each one in the Unified ICM parent system. When deployed on a separate server, a Unified CCE Gateway PG can manage multiple child Unified CCE peripherals or multiple child Unified CCX systems (up to five child systems).

**Note** The Unified CCE Gateway PG does not support Unified CCE System PG and Unified CCX integration on the same CCE Gateway PG instance.

In the Unified CCE child, you can deploy IP IVR or Unified CVP for call treatment and queuing. If you deploy Unified CVP, you must configure an additional VRU PG. This model does not follow the single peripheral model used when 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, so any computation involving queue time (such as minimum expected delay (MED) and average answer wait time) are inaccurate.

**Special Note on Whisper Announcement and Agent Greeting for Parent/Child Systems**

Beginning with Unified CCE Release 8.5(2), the Whisper Announcement and Agent Greeting feature are supported on a child Unified CCE.

Agent Greeting for Parent/Child is only supported for a very specific Parent/Child configuration, where calls are queued at a CCX (IP-IVR) on the child system PG, and requires a dedicated CVP at the child on a dedicated VRU PG to provide the agent greetings. Agent Greeting in Parent/Child configurations must be approved by the Cisco Assessment to Quality (A2Q) process and requires Cisco's Design Mentoring Service to assure that the deployment model is designed and sized correctly to support the Agent Greeting feature.

To deploy the child Unified CCE, you must complete the following:

- Use 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. You must use a separate VRU PG for the CVP.

**Special Note on Network Consultative Transfer for Parent/Child Systems**

One restriction of the parent/child deployment is that calls terminating on child systems cannot be transferred by network consultative transfer (NCT) through the NIC or any other routing client on the parent. Although NCT works for TDM ACDs, and at first glance the parent/child deployment seems virtually identical in architecture, the parent/child deployment is not the same.

For a TDM PG, the CTI Server is connected to the ACD PG, which is part of the parent system. This is the equivalent of having a CTI-Server connected to the Gateway PG. To think of it another way, it is like using CTI directly to an ACD instead of the CTI Server, in which case network consultative transfer is not possible either. In parent/child deployments, CTI is connected to the child PG. Having CTI connected to the child PG does not provide the necessary network call ID and other information necessary to allow network consultative transfer.

**Note** Network blind transfer is still possible using any client (for example, Unified CVP or a NIC) on the parent system when a post route is initiated to the parent system from the child.

## SIP Support

Unified IP IVR is notified of caller entered digits (DTMF input) by way of JTAPI messages from Unified CM. Unified IP IVR and Unified CM do not support mechanisms to detect in-band DTMF digits. In deployments with SIP voice gateways or SIP phones that support only in-band DTMF (or are configured to use in-band DTMF In accordance with RFC 2833), Unified CM must invoke an MTP resource to convert the in-band DTMF signaling to out-of band signaling so that the Unified IP IVR can be notified of caller entered digits. Therefore, in environments that include SIP phones or gateways, it is necessary to provision sufficient MTP resources. Keep this in mind if the phones need to interact with Unified IP IVR. Likewise, CTI ports do not support in-band DTMF (RFC 2833). The Mobile Agent feature relies on CTI ports, so MTP resources are required when in-band DTMF (RFC 2833) is negotiated.

SIP trunking is supported using CVP deployments with Cisco IOS gateways (IOS GW) and Unified Border Elements (CUBE). For more information about deployment models for Unified CCE and Cisco Unified Communications, see the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

## Q.SIG Support

Cisco Unified CCE does not support using Q.SIG trunks with the Unified CM deployment.

## IPv6 Support

Unified CCE 8.0 supports interoperability with an IPv6-enabled Unified CM cluster. All the Unified CCE components run with IPv4 enabled including IP-IVR, Unified CVP, the CTI OS Agent Desktops, and Agent Phones. The Unified CCE Agent PG integration with the Unified CM CTI Manager uses IPv4.

Caller phones or voice gateways can run IPv4 or IPv6. If the caller's environment is IPv6 only, then Media Termination Point (MTP) resources are required for call treatment using IP-IVR and Unified CVP VXML Gateways.

Agent phones must run IPv4; IPv6 is not supported for agent phones. If the caller’s phone is IPv6, then an MTP must also be inserted between the resources.

Mobile Agent and Outbound Option endpoints (CTI ports and Dialer ports) must be configured as IPv4 devices.

## Service Advertisement Framework Call Control Discovery

The Service Advertisement Framework Call Control Discovery (SAF CCD) feature is a way to replace the need for gatekeepers and SIP proxies. The JTAPI impact has a new reason code for when calls are redirected from another cluster back to the local PSTN. For more details, see the *Cisco Unified Communications SRND*.

## Cisco Unified Mobile Agent

For deployments using Cisco Unified Mobile Agent, it is important to consider the location of the voice gateways that will be used to call agents because their location has design considerations for silent monitoring, call admission control, and other areas. For design guidance and considerations for implementing Cisco Unified Mobile Agent, see the chapter on Cisco Unified Mobile Agent.

## CTI-OS Multi-Server Support

Cisco Unified CCE supports multiple CTI OS servers connecting to a single CTI Server. Up to ten CTI OS servers are allowed per PG.

This deployment provides the following benefits:

- Simplification: multiple CTI OS Servers can be configured to use the same CTI Server.
- Many small sites may use a single PG with multiple PIMs rather than each requiring its own PG.
- Reduced box count because all of the PG processes (including the PIM and CTI OS Server processes) are running on the same box.
- Increased Scalability of a single Unified CCE PG because multiple PIMs under a single PG are used to connect to different Unified CM clusters.

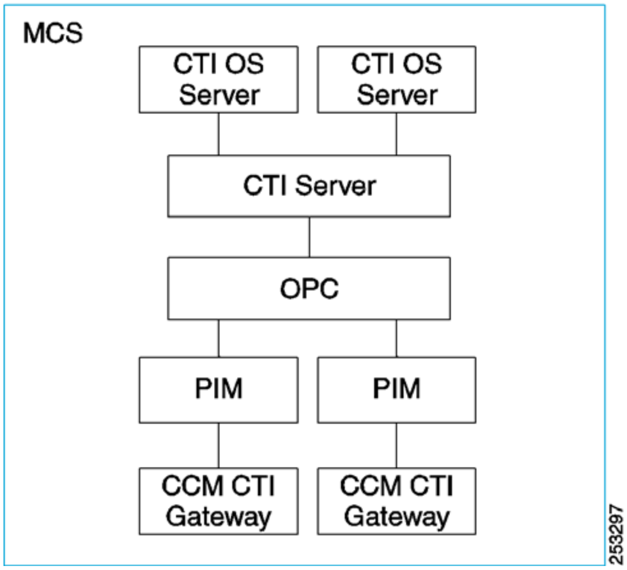
This deployment has the following requirements:

CTI OS Servers must reside on the same server as the rest of the PG processes.

- Each PG can be configured for only one peripheral type.
- Multi-instance deployments cannot add more than one CTI OS Server per instance.
- This deployment model may be used with Unified CCE peripherals.
- PGs are typically co-located with a peripheral. Allowing multiple peripherals per PG could result in some peripherals being situated remotely from the PG. This is not supported for some peripherals and remains unsupported in this case. For example, the Unified CM (Enterprise and System deployments) would not be aggregated in a single PG unless all the ACD and Unified CM peripherals were co-located on a local LAN with the PG. In general, the deployment rules associated with ACD integrations on a PG still apply. For those deployments supporting remote PGs, all network requirements (including bandwidth, latency, and availability) must be met.

When this deployment model is implemented, scaling is reduced to 75% of the scaling capacity for a single CTI OS. For example, if a given configuration supports 1000 agents with a single CTI OS server, it will support 750 agents using multiple CTI OS servers. This is due to the extra overhead of extra CTI OS processes and to the extra processing load incurred by the CTI Server due to the extra clients. The exception to this is when this feature is used for supporting over 2000 agents (the CTI Manager limit) on Unified CCE. (See Figure 11 for an example.) Note that this deployment is supported only when using the Unified CCE PG and it does not support a VRU under the same PG (no Generic type supported).

**Figure 11** ***Multiple CTI OS Servers***



## CAD Multi-Server Support

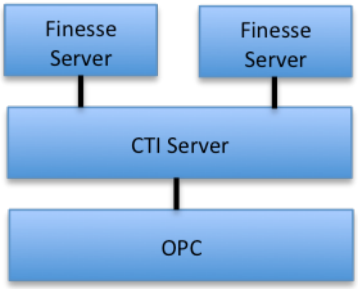
Multiple instances of CAD services are supported on a single PG. The CAD services must reside on separate servers, but they do not require a separate PG for each one. Up to 10 instances of CAD services are allowed per PG.

The benefit of this type of deployment is that the additional instances of CAD services required to achieve a specific agent number do not require a separate PG. While the separate instances of CAD services do have to be deployed on separate servers, the hardware requirements for CAD services are minimal. For information about those requirements, see the Cisco CAD Installation Guide/Cisco Unified Contact Center Enterprise and Hosted Release 8.0.

## Cisco Finesse Multi-Server Support

Two instances of Finesse are supported on a single PG, with each Finesse server capable of supporting the maximum of 2000 users supported by the CTI Server. Each Finesse server requires a single CTI connection.

**Figure 12** ***Multiple Finesse Servers***



Cisco Finesse must be deployed as a virtual machine in a VMware virtual environment. Finesse can be deployed on a stand-alone physical server or coresident with the Agent PG, as long as the Finesse server requirements are met.

If Finesse is deployed coresident with the Agent PG, only Finesse and the Agent PG can be deployed on the physical server and they both must be running in virtual machines.

Finesse can be deployed with CTI-OS with the following limitations:

- The maximum number of CTI all events connections supported by the CTI Server (7) is not exceeded.
- The total number of combined Finesse and CTI OS agents does not exceed the capacity of the common PG.

Finesse supports deployment only with the Enterprise Unified CCE PG as the Agent PG.

Finesse supports deployment where only a single PIM is configured on the Agent PG.

## Virtualization Support

For detailed information about Unified Communications applications running in virtual machines on UCS hardware, please see the Unified Communications Virtualization Doc Wiki page at: http://docwiki.cisco.com/wiki/Unified_Communications_Virtualization

## Whisper Announcement Support

Whisper Announcement is defined as the ability to play a pre-recorded announcement to an Agent right before the caller is connected. The announcement is played to the Agent only. The announcement is not heard by the Customer. Only one Whisper Announcement file will play per call.

Whisper Announcement is supported with Enterprise deployments of Unified CCE only. The feature operates with a type 10 Network VRU based on CVP with SIP as the CVP call control protocol. H323 is not supported.

Whisper time is not specifically categorized in Unified CCE reports. In agent and skill group reports, the period during which the announcement plays is reported as reserved agent state time. In the Termination Call Detail records, it is treated as ring time.

Whisper Announcement for Parent/Child is only supported for a very specific Parent/Child configuration, where calls are queued and whisper announcement are sourced by a CCX (IP-IVR) on the child system PG, and requires a dedicated CVP at the child on a dedicated VRU PG to provide agent greetings if so desired. Whisper Announcement in Parent/Child configurations must be approved by the Cisco Assessment to Quality (A2Q) process and requires Cisco Design Mentoring Service to assure that the deployment model is designed and sized correctly to support the Whisper Announcement feature.

## Agent Greeting Support

Agent Greeting is the capability for a contact center agent to record a greeting and then automatically play it to the caller and agent at the same time when a new call is received by the agent. The agent greeting playback is immediately followed by a caller and agent connection in a call. Only one Agent Greeting file will play per call.

Agent Greeting is supported with Enterprise deployments of Unified CCE only. The feature operates with a type 10 Network VRU based on CVP. SIP is the only CVP call control protocol supported (not H.323).

Agent Greetings are recorded using G.711 encoding. CVP supports a mixed codec environment in which the IVR uses G.711 encoding while customer/agent calls can use G.729. G.722 is not supported.

Greeting time is not specifically categorized in Unified CCE reports. In agent and skill group reports, the period during which the greeting plays is reported as talk time. Record time is counted as an internal call by the default skill group.

Agent Greeting is supported for in-house (that is, non-mobile) agents in Release 8.5(1). Release 8.5(2) adds support for mobile agents. It works with Unified CM supported IP phones with BIB. See the Unified CCE Compatibility Matrix for the list of supported phones.

The CVP VXML server is required for recording.

### Agent Greeting Deployment Models

For the agent greeting call leg, the Agent PIM sends a separate route request to the router when the agent answers the call. The PIM instructs the UCM to connect to the VRU to add media to the call and then the VRU plays the agent greeting based on the instruction from the router. This extra call leg adds extra load on the system and more capacity needs to be estimated when sizing the system. Special attention needs to be paid for this call leg when deploying systems with the agent greeting feature.

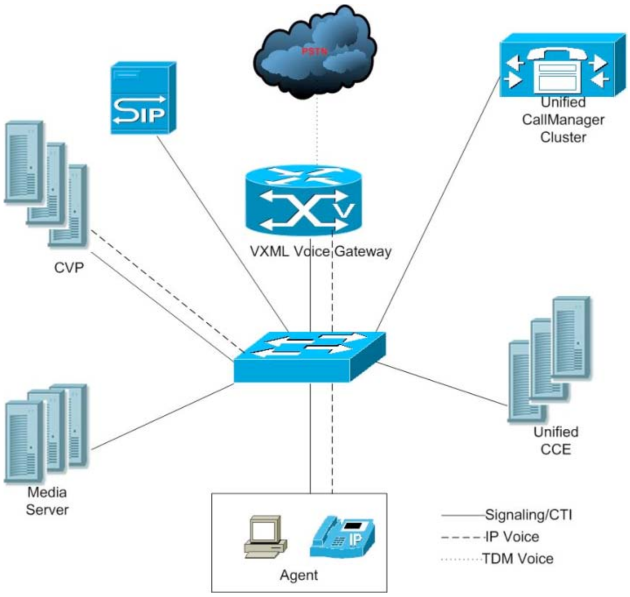
**One Network VRU is used for call treatment, call queuing, and playing agent greeting**

In this deployment model, one network VRU is configured in the UCCE and this VRU is used for playing prompts, call queuing, and playing or recording agent greeting. The SendToVRU node can be used in the routing script for the agent greeting call leg.

For small deployments where one CVP server is used for call queuing and treatment, a SIP trunk can be configured in UCM to directly send the call to the CVP to play the agent greeting.
For those deployments where more than one CVP servers are needed for call queuing and treatment, each CVP server is associated with a VRU PIM and all of those VRU PIMs are associated with the same network VRU.

- A proxy server can be configured between the UCM and CVP servers to balance the load among those CVP servers for the agent greeting call leg.

**Figure 13** ***Single-Site Deployment Model for Agent Greeting***



### Dedicated Network VRU for agent greeting

This is supported in Unified CCE Release 8.5(2) or later. For details, see the Agent Greeting and Whisper Announcement Feature Guide.

### Special requirements for multiple sites deployment

As the agent greeting is played to both the caller and agent, excessive network latency may cause the caller to hear silence before the greeting is played. The following requirements must be met when the system is deployed at multiple sites.

- The latency between the UCCE Central Controllers and remote PGs (both Agent PG and VRU PG) cannot exceed 50 ms one way (100 ms round trip).
- When multiple CVP servers are deployed for the call load, it is possible that the call is queued at one CVP and Agent Greeting is played by a different CVP at another site. For better performance, it is ideal to have the Agent Greeting played by the VXML voice gateway which is closest to the agent phone.

# IPT: Single Site

A single-site deployment refers to any scenario where all voice gateways, agents, desktops, phones, and call processing servers (Unified CM, Unified ICM, Unified CCE, and Unified IP IVR or Cisco Unified Customer Voice Portal (Unified CVP)) are located at the same site and have no WAN connectivity between any Unified CCE software modules. Figure 14 illustrates this type of deployment using the Unified CCE model.

**Figure 14** ***Single-Site Deployment***

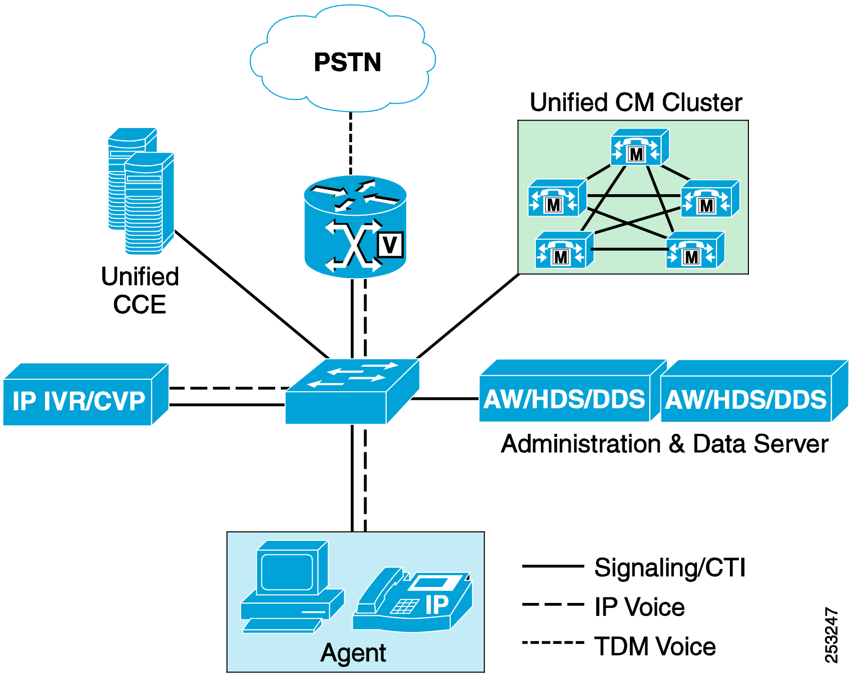


Figure 14 shows a Unified IP IVR, a Unified CM cluster, redundant Unified CCE servers, two Administration & Data Servers, and a direct connection to the PSTN from the voice gateways. The Unified CCE server in this scenario is running the following major software processes:

- Call Router
- Logger and Database Server
- Unified CCE System PG with Unified CM Peripheral Interface Manager (PIM) and Unified IP IVR PIM
- CTI Server
- CTI Object Server (CTI OS)
- Optionally, Cisco Agent Desktop (CAD) servers could be co-located on the Unified CCE servers as well.

Optionally, the Central Controller and Unified CCE System PG can be split onto separate servers. For information about when to install the Central Controller and PG on separate servers, see the chapter on Sizing Unified CCE Components and Servers.

Traditional Unified CCE must be deployed in a redundant fashion. Simplex deployments are supported only for lab or non-production deployments. For information about Unified CCE redundancy, see the Design Considerations for High Availability chapter.

The number of Unified CM nodes and the hardware model used is not specified along with the number of Unified IP IVRs. For information about determining the number and type of servers required, see the Sizing Unified CCE Components and Servers chapter.

Also not specified in this model is the specific data switching infrastructure required for the LAN, the type of voice gateways, or the number of voice gateways and trunks. Cisco campus design guides and Cisco Unified Communications design guides are available to assist in the design of these components. The chapter Sizing Contact Center Resources discusses how to determine the number of gateway ports.

Another variation of this model is to have the voice gateways connected to the line side of a PBX instead of the PSTN. Connection to multiple PSTNs and a PBX all from the same single-site deployment is also possible. For example, a deployment can have trunks from a local PSTN, a toll-free PSTN, and a traditional PBX/ACD. For more information, see Traditional ACD Integration and Traditional IVR Integration.

This deployment model also does not specify the type of signaling (ISDN, MF, R1, and so on) to be used between the PSTN and voice gateway, or the specific signaling (H.323, SIP or MGCP) to be used between the voice gateway and Unified CM.

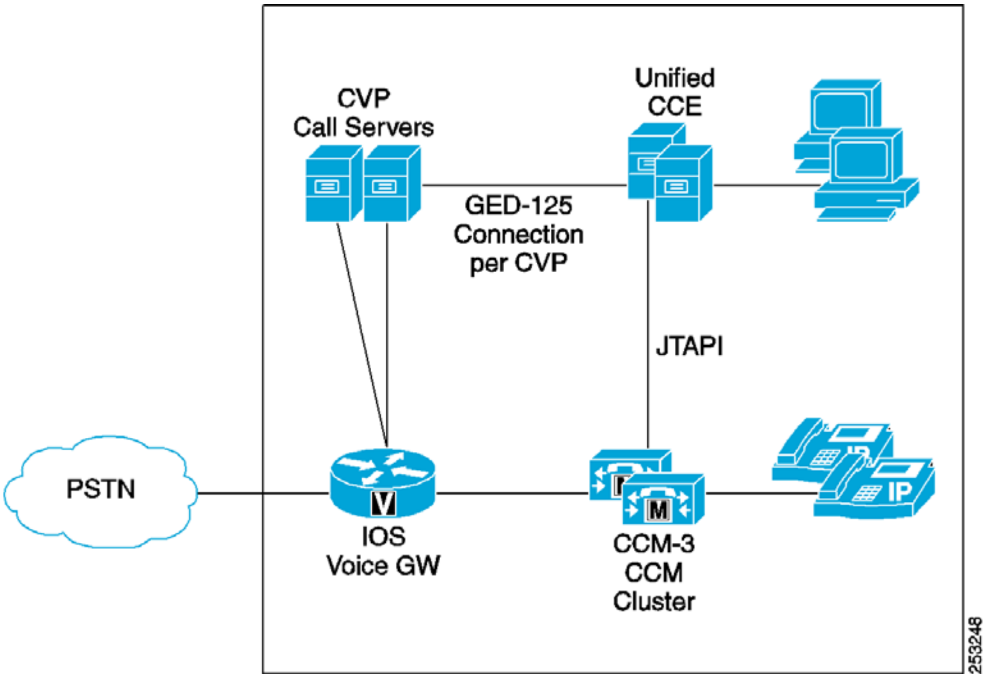
The amount of digital signal processor (DSP) resources required for placing calls on hold, consultative transfers, and conferencing is also not specified in this model. For information about sizing of these resources, see the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

The main advantage of the single-site deployment model is that there is no WAN connectivity required. Given that there is no WAN in this deployment model, there is generally no need to use G.729 or any other compressed Real-Time Transport Protocol (RTP) stream (so transcoding is not required).

## Unified CCE: Unified CCE System PG

In this deployment model, the agent PG that is deployed is a Unified CCE System PG. Only a single peripheral is needed to handle both the Unified CM and any Unified IP IVRs that may exist. This peripheral unifies the appearances of the multiple PIMs and also handles the load balancing of calls between multiple Unified IP IVRs. Alternatively, this model may be configured to use Unified CVP. When Unified CVP is used, its connectivity to Cisco Unified Presence handles load balancing by distributing the incoming calls among Unified CVP Call Servers. In this deployment, the VRU PIMs (up to 10) communicating with the Unified CVP Call Servers reside on their own PG and not under the Unified CCE System PG. Figure 15 shows a single-site deployment using Unified CVP instead of IP-IVR in a Unified CCE system. In this model, no longer do all calls reside under a single peripheral; Unified CVP is its own peripheral.

**Figure 15** ***Single-Site Deployment Using Unified CVP***



When using this configuration, the VRU PGs should be deployed in a duplex PG configuration with one to ten Unified CVP Call Servers (depending on the number of agents and call volume required).

## IVR: Treatment and Queuing with Unified IP IVR

In this deployment model, all initial and subsequent queuing is done on the Unified IP IVR. Up to five Unified IP IVRs can be deployed in this model (with the Unified CCE System PG). The Unified IP IVRs are placed behind Unified CM, using the Unified CM dial plan and call switching under control of Unified CCE. All calls come into a CTI Route Point on Unified CM, controlled by Unified CCE, and are then automatically translation-routed to the Unified IP IVR by the Unified CCE System PG. The Unified CCE handles load balancing between available Unified IP IVR ports. Configuring translation routes between the Unified IP IVR and Unified CM is not needed.

## IVR: Treatment and Queuing with Unified CVP

Unified CVP can be used to provide the call treatment and queuing in this model as well. Because Unified CVP is not part of the Unified CCE System PG peripheral, translation routes must be configured to transfer calls with call data between the peripherals.

In this deployment model, all initial and subsequent queuing is done using Unified CVP. A single server may be used with all Unified CVP processes co-located on that server. Multiple servers, on the other hand, allow scaling and redundancy. For more information about redundancy, see Design Considerations for High Availability.

For more information about Unified CVP, see the

*Cisco Unified Customer Voice Portal Solution Reference Network Design (SRND)*.

## Unified CCE: Enterprise Unified CCE PG

In these deployment models the Enterprise Unified CCE peripheral is used to handle interactions with Unified CM; a separately configured VRU peripheral is used to handle interactions with the Unified IP IVR or Unified CVP.

## IVR: Treatment and Queuing with Unified IP IVR

In this deployment model, all initial and subsequent queuing is done on the Unified IP IVR. If multiple Unified IP IVRs are deployed, use Unified CCE to load-balance calls across those Unified IP IVRs. Translation routes must be configured manually between the Unified CM peripheral and the Unified IP IVR peripherals. The translation routes are used to move calls and data between Unified CM and the Unified IP IVRs. Load balancing is done manually in the Translation Route To VRU node in the Unified CCE call routing script.

## IVR: Treatment and Queuing with Unified CVP

Unified CVP can be used to provide the call treatment and queuing in this model as well. Unified CVP has its own VRU PG, either loaded on the same server as the Unified CM PG or as part of a Generic PG combination.

In this deployment model, all initial and subsequent queuing is done using Unified CVP. A single server may be used with all Unified CVP processes co-located on that server. Multiple servers, on the other hand, allow scaling and redundancy. For more information about redundancy, see the Design Considerations for High Availability chapter.

For more information about Unified CVP, see the *Cisco Unified Customer Voice Portal Solution Reference Network Design (SRND)*.

## Unified CCE: Transfers

In this deployment model (as well as in the multisite centralized call processing model), both the transferring agent and target agent are on the same peripheral. This also implies that both the routing client and the peripheral target are the same peripheral. The transferring agent generates a transfer to a particular dialed number configured as a CTI Route Point in Unified CM (for example, looking for any specialist in the specialist skill group).

The Agent peripheral (either the Unified CCE System peripheral or the Enterprise Unified CCE peripheral) generates a route request to the Call Router. The Call Router matches the dialed number to a call type and activates the appropriate routing script. The routing script looks for an available specialist.

If a target agent (specialist) is available to receive the transferred call, the Call Router returns the appropriate label based on the Agent Target Rules (Dynamic) or a DeviceTarget Label (Static) to the requesting routing client (the Agent peripheral). In this scenario, the label is typically just the extension of the phone where the target agent is currently logged in. On receiving the route response (label), the Unified CM PIM initiates the transfer by sending a JTAPI transfer request to Unified CM.

At the same time that the label is returned to the routing client, pre-call data (which includes any call data that has been collected for this call) is delivered to the peripheral target. In this scenario, the routing client and peripheral target are the same Agent peripheral. This is because the transferring agent and the target agent are both associated with the same peripheral. In some of the more complex scenarios to be discussed in later sections, the routing client and peripheral target are not the same.

If a target agent is not available to receive the transferred call, the Call Routing script is typically configured to transfer the call to an IVR so that queue treatment can be provided. In this scenario, the logic in the Unified CCE System PG differs from the logic in the Unified CCE PG if the IP-IVR variant is used.

In both cases, the label is a dialed number that instructs Unified CM to transfer the call to an IVR. The translation-route or correlationID is not needed when using the Unified CCE System peripheral but is needed when deploying Unified CVP.

# IPT: Multisite with Centralized Call Processing

A multisite deployment with centralized call processing refers to any scenario where call processing servers (Unified CM, Unified CCE, and Unified IP IVR or Unified CVP) are located at the same site, while any combination of voice gateways, agents, desktops, and phones are either located centrally or remotely across a WAN link. Figure 16 illustrates this type of deployment.

There are two variations of this IPT model:

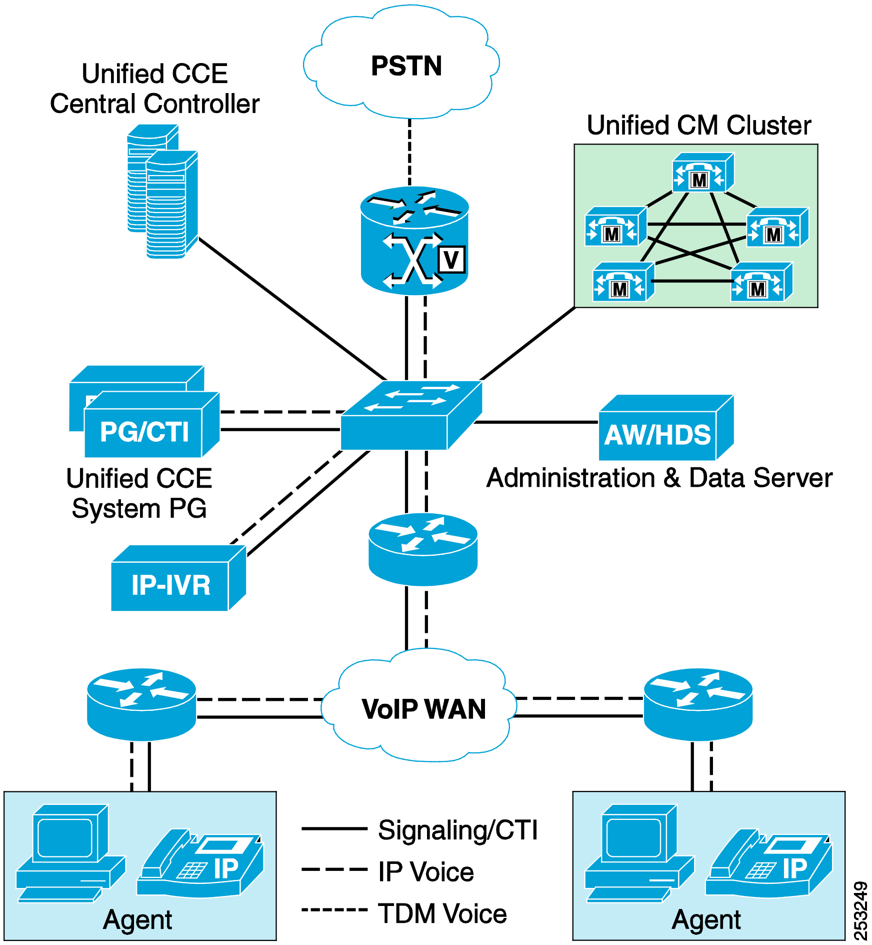
- IPT: Centralized Voice Gateways
- IPT: Distributed Voice Gateways

## IPT: Centralized Voice Gateways

If an enterprise has small remote sites or offices in a metropolitan area where it is not efficient to place call processing servers or voice gateways, then this model is most appropriate. As sites become larger or more geographically dispersed, use of distributed voice gateways might be a better option.

Figure 16 illustrates this model using a Unified CCE deployment. The illustration shows the deployment using IP-IVR, but it could also use Unified CVP instead of IP-IVR.

**Figure 16** ***Multisite Deployment with Centralized Call Processing and Centralized Voice Gateways***



### Advantages

- Only a small data switch and router, IP phones, and agent desktops are needed at remote sites where only a few agents exist. Only limited system and network management skills are required at remote sites.
- No PSTN trunks are required directly into these small remote sites and offices, except for local POTS lines for emergency services (911) in the event of a loss of the WAN link.
- PSTN trunks are used more efficiently because the trunks for small remote sites are aggregated.
- Unified CCE Queue Points (Unified IP IVR or Unified CVP) are used more efficiently because all Queue Points are aggregated.
- No VoIP WAN bandwidth is used while calls are queuing (initially or subsequently). Calls are extended over the WAN only when there is an agent available for the caller.

As with the single-site deployment model, all the same options exist when using Unified CCE configurations. For example, multisite deployments can run the Unified CCE software all on the same server or on multiple servers. The Unified CCE software can be deployed either with the Unified CCE System PG or the Unified CCE PG. The number of Unified CM and Unified IP IVR or Unified CVP servers is not specified by the deployment model, nor are the LAN/WAN infrastructure, voice gateways, or PSTN connectivity. For other variations, see IPT: Single Site.

**Best Practices**

- VoIP WAN connectivity is required for RTP traffic to agent phones at remote sites.
- RTP traffic to agent phones at remote sites might require compression to reduce VoIP WAN bandwidth usage. It may be desirable for calls within a site to be uncompressed, so transcoding might also be required depending on how the Cisco Unified Communications deployment is designed.
- Skinny Client Control Protocol (SCCP) or SIP call control traffic from IP phones to the Unified CM cluster flows over the WAN.
- CTI data to and from the Unified CCE Agent Desktop flows over the WAN. Adequate bandwidth and QoS provisioning are critical for these links.
- Because there are no voice gateways at the remote sites, customers might be required to dial a long-distance number to reach what would normally be a local PSTN phone call if voice gateways with trunks were present at the remote site. This situation could be mitigated if the business requirements are to dial toll- free numbers at the central site. An alternative is to offer customers a toll-free number to dial and have those calls all routed to the centralized voice gateway location. However, this requires the call center to incur toll-free charges that could be avoided if customers had a local PSTN number to dial.
- The lack of local voice gateways with local PSTN trunks can also impact access to 911 emergency services; this must be managed through the Unified CM dial plan. In most cases, local trunks are configured to dial out locally and for 911 emergency calls.
- Unified CM location-based call admission control failure will result in a routed call being disconnected. Therefore, it is important to provision adequate bandwidth to the remote sites. Also, an appropriately designed QoS WAN is critical.

**Note** For calls controlled by Unified CVP, call admission control failures can be recovered with the proper Unified CCE Scripting configuration with the Unified CCE Router Re-query feature enabled.

- Automated Alternate Routing (AAR) provides a mechanism to reroute calls through the PSTN or other network by using an alternate number when Unified CM blocks a call due to insufficient location bandwidth. For deployments with Unified CVP ingress call control, do not enable AAR. This allows Unified CVP Router Re-query to take control of the call in the event of a failure or timeout. For deployments using IP-IVR, enable AAR to route the call through the PSTN or other network component using an alternative number.

## IVR: Treatment and Queuing with Unified IP IVR

As in the single-site deployment, all call queuing is done on the Unified IP IVR at a single central site. While calls are queuing, no RTP traffic flows over the WAN. If re-queuing is required during a transfer or reroute on ring-no-answer, the RTP traffic flow during the queue treatment also does not flow over the WAN. This reduces the amount of WAN bandwidth required to the remote sites.

## IVR: Treatment and Queuing with Unified CVP

In this model, Unified CVP is used in the same way as Unified IP IVR.

## Unified CCE: Transfers

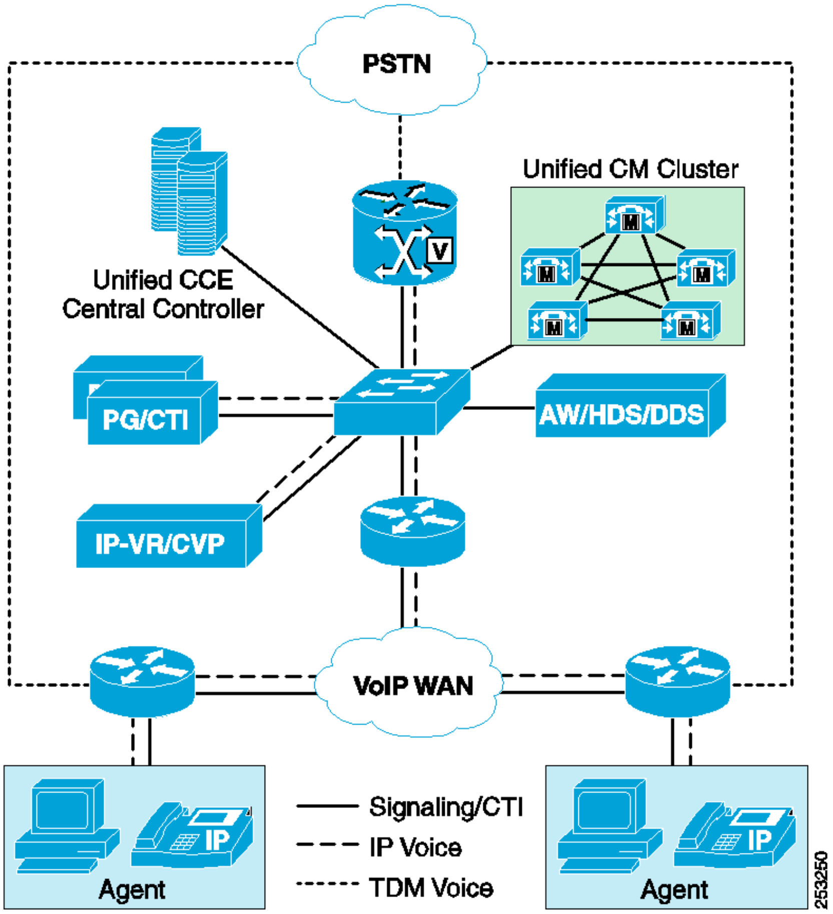
Transfers in this scenario are, from the point of view of the contact center, the same as in the single-site scenario. Therefore, the same call and message flows occurs as in the single-site model, whether the transferring agent is on the same LAN as the target or on a different LAN. The only differences are that QoS must be enabled and that appropriate LAN and WAN routing must be established. For details on provisioning your WAN with QoS, see the latest version of the *Cisco Network Infrastructure Quality of Service Design Guide*.

During consultative transfers where the agent (not the caller) is routed to a Unified IP IVR port for queuing treatment, transcoding is required because the Unified IP IVR can generate only G.711 media streams.

## IPT: Distributed Voice Gateways

A variation of the centralized call processing model can include multiple ingress voice gateway locations. This distributed voice gateway model might be appropriate for a company with many small sites, each requiring local PSTN trunks for incoming calls. This model provides local PSTN connectivity for local calling and access to local emergency services. Figure 17 illustrates this model.

**Figure 17** ***Multisite Deployment with Centralized Call Processing and Distributed Voice Gateways***



In this deployment model, shown with Unified IP IVR for queuing and treatment, it might be desirable to restrict calls arriving at a site to be handled by an agent within that site, but this is not required. By restricting calls to the site where it arrived, the following conditions apply:

- VoIP WAN bandwidth is reduced for calls going to agents from the ingress voice gateway.
- Calls will still cross the VoIP WAN during the time they are in a queue or are receiving treatment from the centralized Unified IP IVRs.
- Customer service levels for calls arriving into that site might suffer due to longer queue times and handling times.
- Longer queue times can occur because even though an agent at another site is available, the Unified CCE configuration may continue to queue for an agent at the local site.
- Longer handling times can occur because even though a more qualified agent exists at another site, the call may be routed to a local agent to reduce WAN bandwidth usage.

In order to restrict a call to the site at which it arrived in this deployment model, it is necessary to create separate skill groups for agents at each location. In order to route a call to any agent in a given skill group regardless of location, the location-specific skill groups can be combined using an enterprise skill group.

It is important for deployment teams to carefully assess the trade-offs between operational costs and customer satisfaction levels to establish the right balance on a customer-by-customer basis. For example, it may be desirable to route a specific high-profile customer to an agent at another site to reduce their queue time and allow the call to be handled by a more experienced representative, while another customer may be restricted to an agent within the site where the call arrived.

A Unified CCE deployment may actually use a combination of centralized and distributed voice gateways. The centralized voice gateways can be connected to one PSTN carrier providing toll-free services, while the distributed voice gateways can be connected to another PSTN carrier providing local phone services.

Inbound calls from the local PSTN could be both direct inward dial (DID) and contact center calls. It is important to understand the requirements for all inbound and outbound calling to determine the most efficient location for voice gateways. Important details include identifying who is calling, why they are calling, where they are calling from, and how they are calling.

### CVP Call Treatment and Call Routing

In multisite environments with distributed voice gateways, Unified CVP can be used to leverage the ingress voice gateways at the remote sites as part of the Unified CCE system. Unified CVP treats and queues calls locally in the ingress voice gateway rather than requiring the call to cross the VoIP WAN to a centralized queue platform. Unified CVP provides call treatment (VRU) using the VoiceXML Browser built into the Cisco IOS voice gateway. Only call signaling passes over the WAN to instruct the remote site voice gateway how to treat, queue, and transfer the call to an agent. In these models, pre-routing to the site might not be necessary because Unified ICM/CCE takes control of the call as soon as it arrives at the site. Basic carrier percent allocation can be used to allocate calls to the sites and failover (rollover) trunks can be used to address local failures as needed.

### Traditional Pre-routing

In a traditional deployment with Unified ICM (parent and child or hybrid) with multisite deployments and distributed voice gateways, the Unified ICM pre-routing capability can also be used to load-balance calls dynamically across the multiple sites. For a list of PSTN carriers that offer Unified ICM pre-routing services, see the *Pre-installation Planning Guide for Cisco ICM Enterprise & Hosted Editions*.

In multisite environments where the voice gateways have both local PSTN trunks and separate toll-free trunks delivering contact center calls, the Unified ICM pre-routing software can load-balance the toll-free contact center calls around the local contact center calls. For example, suppose you have a two-site deployment where Site 1 currently has all agents busy and there are many locally-originating calls in the queue. Site 2 has only a few calls in the queue or maybe even a few agents are currently available. In this scenario, you could have Unified ICM instruct the toll-free provider to route most or all of the toll-free calls to Site 2. This type of multisite load balancing provided by Unified ICM is dynamic and automatically adjusts as call volumes change at all sites.

Just as in the two previous deployment models, much variation exists in the number and types of Unified ICM/CCE, Unified CM, and Unified IP IVR or Unified CVP servers; LAN/WAN infrastructure, voice gateways, PSTN connectivity, and so forth.

### Advantages of Using Distributed Voice Gateways

- Only limited systems management skills are needed for the remote sites because most servers, equipment, and system configurations are managed from a centralized location.
- Unified CVP or Unified ICM/CCE pre-routing can be used to load-balance calls across sites including sites with local PSTN trunks in and toll-free PSTN trunks.
- No WAN RTP traffic is required for calls arriving at each remote site that are handled by agents at that remote site.
- Unified CVP provides call treatment and queuing at the remote site using the VoiceXML Browser in Cisco IOS on the voice gateway itself, thus eliminating the need to move the call over the VoIP WAN to a central queue and treatment point.

### Best Practices

- The Unified IP IVR or Unified CVP, Unified CM and PGs (for both Unified CM and IVR or Unified CVP) are co-located. In this model, the only Unified CCE communications that can be separated across a WAN are the following:
  - Unified Central Controller to PG
  - PG to Unified CCE Agent Desktops
  - Unified CM to voice gateways
  - Unified CM to phones
  - Unified CVP Call Control Server to remote voice gateway (call control)
- If calls are not going to be restricted to the site where calls arrive, or if calls will be made between sites, more RTP traffic will flow across the WAN. It is important to determine the maximum number of calls that will flow between sites or locations. Unified CM locations-based call admission control failure will result in a routed call being disconnected (rerouting within Unified CM is not currently supported). Therefore, it is important to provision adequate bandwidth to the remote sites and appropriately designed QoS for the WAN is critical. Calls that are treated by IP IVR at the central site must also be considered.

**Note** For calls controlled by Unified CVP, call admission control failures can be recovered with the proper Unified CCE Scripting configuration with the Unified CCE Router Re-query feature enabled.

- H.323, SIP, or MGCP signaling traffic between the voice gateways and the centralized Unified CM servers will flow over the WAN. Proper QoS implementation on the WAN is critical and signaling delays must be within tolerances listed in the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

## Unified CCE: Unified CCE System PG

Because the deployment of contact center components is essentially the same as in other multisite centralized call processing deployments, the same benefits and restrictions apply to Unified CCE deployed using the Unified CCE System PG.

Additionally, if Unified ICM/CCE pre-routing is used to interact with carriers and distribute calls to the voice gateways, translation routes for the NIC routing client to the Unified CCE System PG must be configured manually using the ConfigManager application on the Unified ICM/CCE Admin Workstation.

### Unified CCE & IVR: Treatment and Queuing with Unified IP IVR

WAN bandwidth must be provisioned to support all calls that will be treated and queued at the central site.

### IVR: Treatment and Queuing with Unified CVP

Unified CVP is not supported with a Unified CCE System PG. A separate VRU peripheral must be configured and deployed. This means that translation routes must be configured to transfer calls with call data between the peripherals. However, Unified CVP does provide the benefits of queuing and treatment for callers at the remote distributed ingress voice gateways in this model because the calls do not have to cross the VoIP WAN for treatment in the centralized Unified IP IVR.

Using Unified CVP for treatment and queuing allows you to reduce the amount of voice bearer traffic traveling across the WAN. Unified CVP queues and treats calls on the remote gateways, thus eliminating the need to terminate the voice bearer traffic at the central site. WAN bandwidth must still be provisioned for transfers and conferences that involve agents at other locations.

## Unified CCE: Unified CCE PG

Because the deployment of contact center components is essentially the same as in other multisite centralized call processing deployments, the same benefits and restrictions apply to Unified CCE deployed using the Unified CCE PG.

Additionally, if Unified ICM/CCE pre-routing is used to interact with carriers and to distribute calls to the voice gateways, translation routes must be configured for the NIC routing client using traditional Unified CCE with separate Unified CVP and Unified CM peripherals in the Unified ICM/CCE.

### IVR: Treatment and Queuing with Unified IP IVR

WAN bandwidth must be provisioned to support all calls that will be treated and queued at the central site.

### IVR: Treatment and Queuing with Unified CVP

Using Unified CVP for treatment and queuing allows you to reduce the amount of voice bearer traffic traveling across the WAN. Unified CVP queues and treats calls on the remote gateways, thus eliminating the need to terminate the voice bearer traffic at the central site. WAN bandwidth must still be provisioned for transfers and conferences that involve agents at other locations.

### Unified CCE: Transfers

Intra-site or inter-site transfers using the VoIP WAN to send the RTP stream from one site to another occur basically the same way as a single-site transfer or a transfer in a deployment with centralized voice gateways.

An alternative to using the VoIP WAN for routing calls between sites is to use a carrier-based PSTN transfer service. These services allow the Unified CCE voice gateways to out-pulse DTMF tones to instruct the PSTN to reroute (transfer) the call to another voice gateway location. Each site can be configured within the Unified ICM/CCE as a separate Agent Peripheral. The label then indicates whether a transfer is intra-site or inter-site using Take Back and Transfer (*8) or Transfer Connect. These transfer tones are played in-band over the voice path and must be played from a recorded file in Unified IP IVR or out-pulsed as digits from Unified CVP.

## IPT: Multisite with Distributed Call Processing

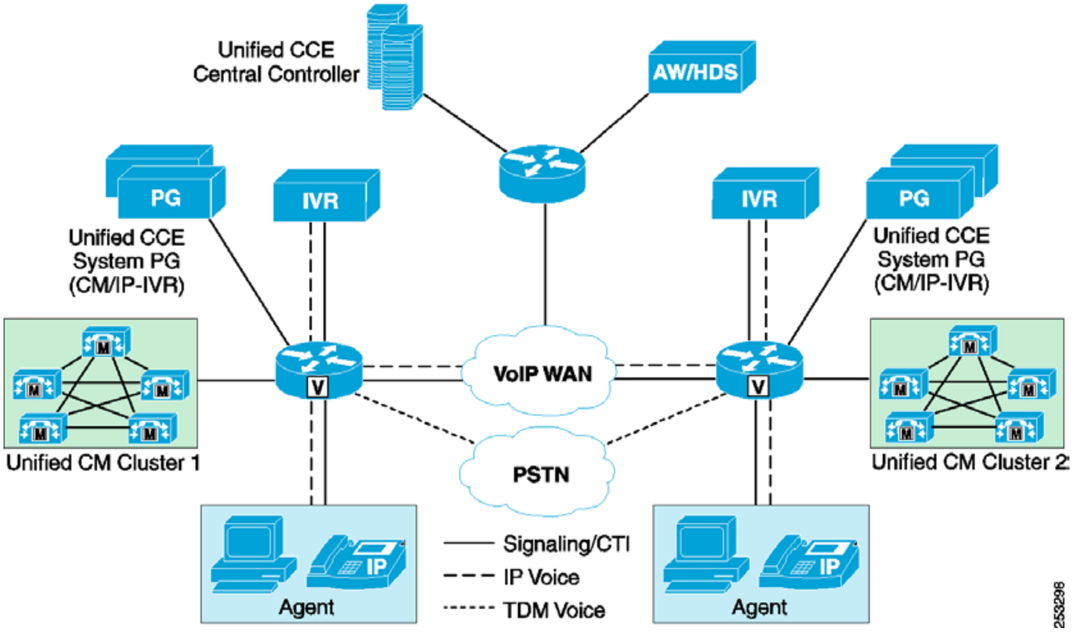
Enterprises with multiple medium to large sites separated by large distances tend to prefer a distributed call processing model. In this model, each site has its own Unified CM cluster, treatment and queue points, PGs, and CTI Server. However, as with the centralized call processing model, sites could be deployed with or without local voice gateways. Some deployments may also contain a combination of distributed voice gateways (possibly for locally dialed calls) and centralized voice gateways (possibly for toll-free calls) as well as centralized or distributed treatment and queue points.

Regardless of how many sites are being deployed in this model, there is still only one logical Unified CCE Central Controller. If the Unified CCE Central Controller is deployed with redundancy, sides A and B can be deployed side-by-side or geographically separated (remote redundancy). For details on remote redundancy, see the Unified ICM/CCE product documentation.

### Unified CCE: Distributed Voice Gateways with Treatment and Queuing Using Unified IP IVR

This deployment model is a good choice if the company has multiple medium to large sites. In this model, voice gateways with PSTN trunks terminate into each site. Just as in the centralized call processing model with distributed voice gateways, it might be desirable to limit the routing of calls to agents within the site where the call arrived (to reduce WAN bandwidth). An analysis of benefits from customer service levels versus WAN costs is required to determine whether limiting calls within a site is appropriate. Figure 18 illustrates this model using a traditional Unified CCE deployment with the Unified CCE System PG.

**Figure 18** ***Multisite Deployment with Distributed Call Processing and Distributed Voice Gateways with Unified IP IVR***



As with the previous models, many options are possible. The number and type of Unified CCE Servers, Unified CM servers, and Unified IP IVR servers can vary. The LAN and WAN infrastructure, voice gateways, PSTN trunks, redundancy, and so forth are also variable within this deployment model. Central processing and gateways may be added for self-service, toll-free calls and support for smaller sites. In addition, the use of a pre-routing PSTN Network Interface Controller (NIC) is also an option.

**Advantages**

- Scalability — Each independent site can scale up to the maximum number of supported agents per Unified CM cluster and there is no software limit to the number of sites that can be combined by the Unified CCE Central Controller to produce a single enterprise-wide contact center (provided that the total concurrent agent count is less than the maximum supported agent count in a Unified CCE System). For scalability and sizing information, see the chapter on Sizing Contact Center Resources.
- All or most VoIP traffic can be contained within the LAN of each site, if desired. The QoS WAN shown in Figure 18 would be required for voice calls to be transferred across sites. Use of a PSTN transfer service (for example, Take Back and Transfer or Transfer Connect) could eliminate 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/CCE pre-routing can be used to load-balance calls (based on agent or Unified IP IVR port availability) 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 Unified CCE Central Controller provides centralized management for configuration of routing for all calls within the enterprise.

- The Unified CCE Central Controller provides the capability to create a single enterprise-wide queue.
- The Unified CCE Central Controller provides consolidated reporting for all sites.

**Best Practices**

- The PG, Unified CM cluster, and Unified IP IVR must be co-located at the contact center site.
- The communication link from the Unified CCE Central Controller to the PG must be sized properly and provisioned for bandwidth and QoS. (For details, see the chapter Bandwidth Provisioning and QoS Considerations.)
- Gatekeeper-based or RSVP Agent-based call admission control could be 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 PG and the Unified CCE Central Controller is lost, then all contact center routing for calls at that site is also lost. Therefore, it is important to implement a fault-tolerant WAN. Even when a fault-tolerant WAN is implemented, it is important to identify contingency plans for call treatment and routing when communication is lost between the Unified CCE Central Controller and PG. For example, in the event of a lost Unified CCE Central Controller connection, the Unified CM CTI route points could send the calls to Unified IP IVR ports to provide basic announcement treatment or to invoke a PSTN transfer to another site. Another alternative is for the Unified CM cluster to route the call to another Unified CM cluster that has a PG with an active connection to the Unified CCE Central Controller. For more information about these options, see the Design Considerations for High Availability chapter.
- While two inter-cluster call legs for the same call will not cause unnecessary RTP streams, two separate call signaling control paths will remain intact between the two clusters (producing logical hair-pinning and reducing the number of inter-cluster trunks by two). Consider the percentage of inter-site transfers when sizing inter-cluster trunks capacities.
- Latency between Unified CCE Central Controllers and remote PGs cannot exceed 200 ms one way (400 ms round-trip).

## Treatment and Queuing

Initial call queuing is done on a Unified IP IVR co-located with the voice gateways, so no transcoding is required. When a call is transferred and subsequent queuing is required, perform the queuing on a Unified IP IVR at the site where the call is currently being processed. For example, if a call comes into Site 1 and gets routed to an agent at Site 2, but that agent needs to transfer the call to another agent whose location is unknown, queue the call to a Unified IP IVR at Site 2 to avoid generating another inter-cluster call. A second inter-cluster call would be made only if an agent at Site 1 was selected for the transfer. The RTP flow at this point would be directly from the voice gateway at Site 1 to the agent’s phone at Site 1. However, the two Unified CM clusters would still logically see two calls in progress between the two clusters.

## Transfers

Transfers within a site function just like a single-site transfer. Transfers between Unified CM clusters use either the VoIP WAN or a PSTN service.

If the VoIP WAN is used, sufficient inter-cluster trunks must be configured. An alternative to using the VoIP WAN for routing calls between sites is to use a PSTN transfer service. These services allow the

Unified CCE voice gateways to out-pulse DTMF tones to instruct the PSTN to reroute (transfer) the call to another voice gateway location. Another alternative is to have the Unified CM cluster at Site 1 make an outbound call back to the PSTN. The PSTN would then route the call to Site 2, but the call would use two voice gateway ports at Site 1 for the remainder of the call.

## Unified CCE: Unified CCE System PG

The Unified CCE System PG acts as a single peripheral that joins the Unified CM and Unified IP IVR peripherals of former versions to simplify installation, configuration, and routing. In this model, the PGs at the remote sites can be installed as Unified CCE System PGs to combine the Unified IP IVR and Unified CM peripherals under a single logical PG instance and peripheral.

This model is perhaps more typical of outsourcers that would set up a call center specifically for a single client and deploy it as a Unified CCE System PG to allow their client company to connect their Unified CCE Enterprise system to the outsourcer Unified CCE System PG with the Unified CCE Gateway PG, as they would any outsourced ACD.

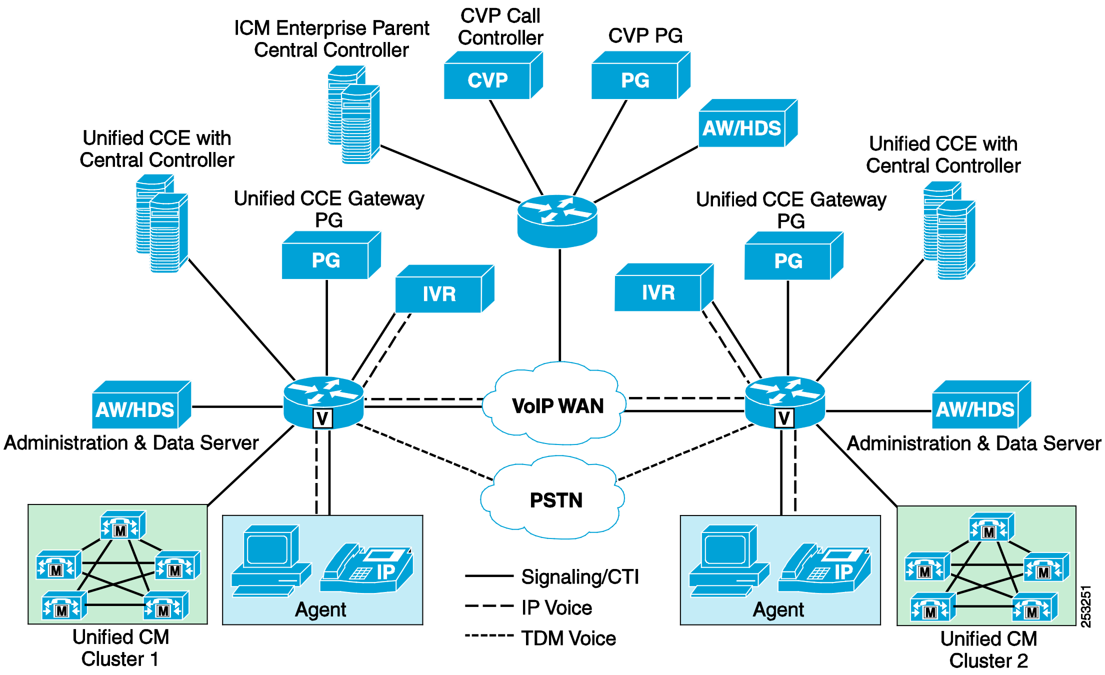
## Unified CCE: Unified CCE PG

This model, as designed with multiple remote locations, is more suited for the traditional Unified CCE design with multiple distributed peripheral gateways. The system could be deployed with the Generic PG or both Unified CM and Unified IP IVR PGs at the sites; however, the new Unified CCE System PG that combines both of these peripherals into a single peripheral for routing and reporting under the traditional model might be easier for new deployments of this solution. Existing customers upgrading to Unified CCE 8.0 may stay on their existing Generic PG or multi-PG model.

## Alternative: Parent/Child

The Unified ICM Enterprise (parent) and Unified CCE (child) model is an appropriate alternative deployment to provide local, distributed call processing with a local Unified CM and Unified CCE at each site (child), controlled under a centralized Unified ICM Enterprise parent for enterprise-wide routing, reporting, and call control. This model has the advantage of being more tolerant of WAN outages and each site is completely survivable. Figure 19 shows this same model deployed using the parent/child model.

**Figure 19** ***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. At each distributed site, there is a complete Unified CCE deployment consisting of Central Controller on one or more servers. There is also a local Administration & Data Server for Unified CCE to perform configuration, scripting, and reporting tasks for that specific site. There is a Unified CCE Gateway PG that 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 co-locate it with the Unified CCE System PG, while adhering to the following guidelines:

If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are the same, then the PG number for the Unified CCE Gateway PG and Unified CCE System PG must be different.

If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are different, then the PG number for the Unified CCE Gateway PG and the Unified CCE System PG may be the same.

No additional PGs (such as a VRU PG or MR PG) can be added to this Server.

The server co-resident Unified CCE Gateway PG and Unified CCE System PG is not supported with a Unified SCCE deployment.

For scalability limits of the co-resident Unified CCE Gateway PG and Unified CCE System PG, see the Sizing Unified CCE Components and Servers chapter for additional details.

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. Users at Site 1 cannot see any of the calls or reports from Site 2 in this model. 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 is used to control the calls coming into the distributed sites providing local call queuing and treatment in the VoiceXML Browser in the voice gateway. When configuring the Unified ICM parent CVP to use Unified CVP Router Re-query to take control of the call in the event of 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 are used only for a local backup if the connection from these voice gateways is lost to the parent Unified CVP Call Control server. The local Unified IP IVR also provides local queue treatment for calls that are not answered by the local agents (RONA) rather than sending the call back to the Unified CVP to be re-queued.

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 models with distributed Unified CM Peripheral Gateways, the parent/child model provides for complete local redundancy at the contact center site. The local Unified CCE will take over call processing for inbound calls from the Unified CVP gateways and provide local call queuing and treatment in the local Unified IP IVR. This is an excellent design for call center sites that require complete redundancy or 100% up-time and that cannot be down because of a WAN failure.

This design is a good approach for customers who have Unified ICM already installed with their TDM ACD platforms and who want either to add new sites with Unified CCE or to convert an existing site to Unified CCE. It allows the Unified ICM to continue performing enterprise-wide routing and reporting across all of the sites while inserting new Unified CCE technology on a site-by-site basis.

**Note** Unified CVP could be at both the parent and child. This is virtually identical to Unified CVP at the parent and IP-IVR at the child from a call flow perspective. One key difference is that information about queued calls at the child Unified CVP will not be available at the parent (through the Unified CCE Gateway PG), as is the case if IP-IVR is used. This means that minimum expected delay (MED) over services cannot be used.

### 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 will send 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 Unified CM cluster to scale up to the maximum number of supported agents per cluster. The Unified CCE systems are connected to the parent Unified ICM using the Unified CCE Gateway PG on the parent Unified ICM, which 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 shown in Figure 19 would be required for voice calls to be transferred across sites. Use of a PSTN transfer service (for example, Take Back and Transfer or Transfer Connect) could eliminate 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 be used to load-balance calls based on agent or Unified CVP session availability and to 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**

- Server count — The number of servers that are required to manage the parent/child model is usually higher due to the increased number of software components (additional Unified CCE Gateway PGs required if co-locating with Unified CCE System PG is not an option, additional Central Controller for each child, and so forth).

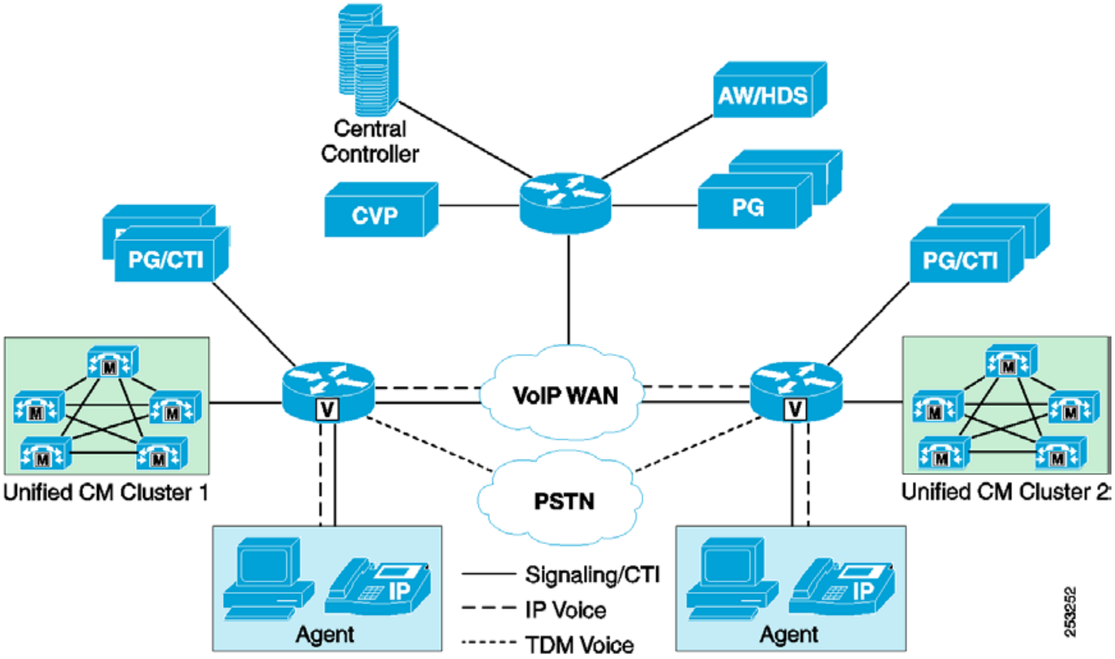
**Best Practices**

- Co-locate the Unified CCE Gateway PG, Unified CM 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. (For details, see the Bandwidth Provisioning and QoS Considerations chapter.)
- Gatekeeper-based or RSVP agent-based call admission control could be 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 would have survivability TCL scripts to redirect inbound calls to local Unified CM CTI route points and the local Unified IP IVR would be used to handle local queuing and treatment during the WAN outage. This is a major feature of the parent/child model to provide complete local survivability for the call center. For more information, see the Design Considerations for High Availability chapter.
- While two inter-cluster call legs for the same call will not cause unnecessary RTP streams, two separate call signaling control paths will remain intact between the two clusters (producing logical hair-pinning and reducing the number of inter-cluster trunks by two). Consider the percentage of inter-site transfers when sizing inter-cluster 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).

## IVR: Distributed Voice Gateways with Treatment and Queuing Using Unified CVP

This deployment model is the same as the previous model, except that Unified CVP is used instead of Unified IP IVR for call treatment and queuing. In this model, voice gateways with PSTN trunks terminate into each site. Just as in the centralized call processing model with distributed voice gateways, it might be desirable to limit the routing of calls to agents within the site where the call arrived (to reduce WAN bandwidth). Call treatment and queuing can also be achieved at the site where the call arrived, further reducing the WAN bandwidth needs. Figure 20 illustrates this model using a traditional Unified CCE deployment.

**Figure 20** *Multisite Deployment with Distributed Call Processing and Distributed Voice Gateways with Unified CVP*



As with the previous models, many options are possible. The number and type of Unified CCE Servers, Unified CM servers, and Unified CVP servers can vary. The LAN and WAN infrastructure, voice gateways, PSTN trunks, redundancy, and so forth are also variable within this deployment model. Central processing and gateways may be added for self-service, toll-free calls and support for smaller sites. The use of a pre-routing PSTN Network Interface Controller (NIC) is also an option.

### Advantages

- Unified CVP Servers can be located either centrally or remotely. Call treatment and queuing will still be distributed, executing on the local gateway, regardless of Unified CVP server location. Unified CVP is shown centrally located in Figure 20.
- For information about the number of agents supported per PG and across the entire system, see the Sizing Unified CCE Components and Servers chapter.
- All or most VoIP traffic can be contained within the LAN of each site, if desired. The QoS WAN would be required for voice calls to be transferred across sites. Use of a PSTN transfer service (for example, Takeback N Transfer) could eliminate 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 CCE pre-routing can be used to load-balance calls and route them 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 Unified CCE Central Controller provides centralized management for configuration of routing for all calls within the enterprise.

- The Unified CCE Central Controller provides the capability to create a single enterprise-wide queue.
- The Unified CCE Central Controller provides consolidated reporting for all sites.

**Best Practices**

- The Unified CM PG and Unified CM cluster must be co-located. The Unified CVP PG and Unified CVP servers must be co-located.
- The communication link from the Unified CCE Central Controller to PG must be properly sized and provisioned for bandwidth and QoS. Cisco provides a partner tool called the VRU Peripheral Gateway to Unified ICM Central Controller Bandwidth Calculator to assist in calculating the VRU PG-to-Unified ICM/CCE bandwidth requirement.
- If the communication link between the PG and the Unified CCE Central Controller is lost, then all contact center routing for calls at that site is lost. Therefore, it is important to implement a fault-tolerant WAN. Even when a fault-tolerant WAN is implemented, it is important to identify contingency plans for call treatment and routing when communication is lost between the Unified CCE Central Controller and PG.
- Latency between Unified CCE Central Controllers and remote PGs must not exceed 200 ms one way (400 ms round-trip)

## IVR: Treatment and Queuing

Unified CVP queues and treats calls on the remote gateways eliminating the need to terminate the voice bearer traffic at the central site. Unified CVP servers may be located at the central site or distributed to remote sites. WAN bandwidth must still be provisioned for transfers and conferences that involve agents at other locations.

Unlike Unified IP IVR, with Unified CVP the call legs are torn down and reconnected, avoiding signaling hairpins. With Unified IP IVR, two separate call signaling control paths will remain intact between the two clusters (producing logical hair-pinning and reducing the number of inter-cluster trunks by two).

## Transfers

Transfers within a site function just like a single-site transfer. Transfers between Unified CM clusters use either the VoIP WAN or a PSTN service.

If the VoIP WAN is used, sufficient inter-cluster trunks must be configured. An alternative to using the VoIP WAN for routing calls between sites is to use a PSTN transfer service. These services allow the Unified CCE voice gateways to out-pulse DTMF tones to instruct the PSTN to reroute (transfer) the call to another voice gateway location. Another alternative is to have the Unified CM cluster at Site 1 make an outbound call back to the PSTN. The PSTN would then route the call to Site 2, but the call would use two voice gateway ports at Site 1 for the remainder of the call.

## Unified CCE: Unified CCE System PG

The Unified CCE System PG is not a good fit for this model because it does not support Unified CVP for queuing and the IVR PIMs on the Unified CCE System PG would go unused.

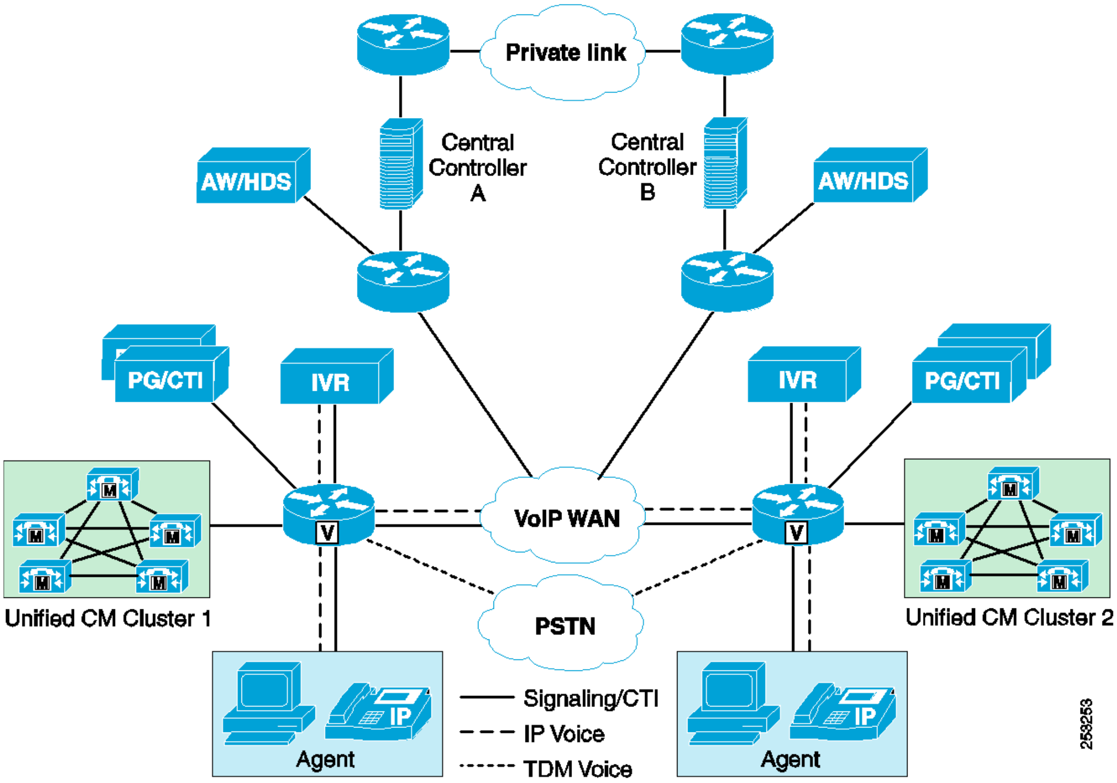
## Unified CCE: Unified CCE PG

The Unified CCE PG is the required PG for this deployment model.

## Unified CCE: Distributed Unified CCE Option with Distributed Call Processing Model

Figure 21 illustrates this deployment model.

**Figure 21** *Distributed Unified CCE Option Shown with Unified IP IVR*



### Advantages

The primary advantage of the distributed Unified CCE option is the redundancy gained from splitting the Unified CCE Central Controller between two redundant sites.

### Best Practices

- Unified CCE Central Controllers (Routers and Loggers) require a separate network path or link to carry the private communications between the two redundant sites. In a non-distributed Unified CCE model, the private traffic usually traverses an Ethernet crossover cable or LAN connected directly between the side A and side B Unified CCE Central Controller components. In the distributed Unified CCE model, the private communications between the A and B Unified CCE components travel across a dedicated link with at least as much bandwidth as a T1 line.
- Latency across the private separate link must not exceed 100 ms one way (200 ms round-trip), but 50 ms (100 ms round-trip) is preferred.

- Latency between Unified CCE Central Controllers and remote PGs must not exceed 200 ms one way (400 ms round-trip).
- The private link cannot traverse the same path as public traffic. The private link must have path diversity and must reside on a link that is completely path-independent from Unified CCE public traffic. This link is used as part of the system fault tolerant design. For more information, see the Design Considerations for High Availability chapter.
- The redundant centralized model is explored in the next section on IPT Clustering over the WAN.

# IPT: Clustering Over the WAN

As part of the centralization of call processing, many customers prefer to combine the redundancy of the distributed Unified CM call processing model with the simplicity of having a single Unified CM cluster for a single dial plan and voice system to administer. This combination of models provides for a single Unified CM cluster with its subscriber servers split across data center locations to provide a single cluster with multiple distributed call processing servers for a highly available and redundant design (known as clustering over the WAN).

Unified CM clustering over the WAN may also be used with Unified CCE for contact centers to allow full agent redundancy in the case of a data center (central site) outage. Implementation of clustering over the WAN for Unified CCE does have several strict requirements that differ from other models. Bandwidth between central sites for Unified CCE public and private traffic, Unified CM intra-cluster communication signaling (ICCS), and all other voice-related media and signaling must be properly provisioned with QoS enabled. The WAN between central sites must be highly available (HA) with redundant links and redundant routers.

### Advantages

- No single point of failure, including loss of an entire central site.
- Cisco Unified Mobile Agents (Remote Agent) require no reconfiguration to remain fully operational in case of site or link outage. When outages occur, agents and agent devices dynamically switch to the redundant site.
- Central administration for both Unified CCE and Unified CM.
- Reduction of servers for distributed deployment.

### Best Practices

- Deploy a minimum of three WAN links for systems that employ the clustered over the WAN model. Deploy at least two links for the high availability network that carries the Unified CCE public traffic (see Figure 22). Use a separate WAN link for the Unified CCE private traffic (see Figure 23). If QoS and bandwidth are configured correctly (see the guidelines in the Bandwidth Provisioning and QoS Considerations chapter), these WAN links can be converged with other corporate traffic as long as the private and public traffic are not carried over the same link. Carry the Unified CM ICCS traffic over the high availability network used by the Unified CCE public communications.

**Figure 22** ***High Availability  WAN Network for the Unified CCE Public Traffic***

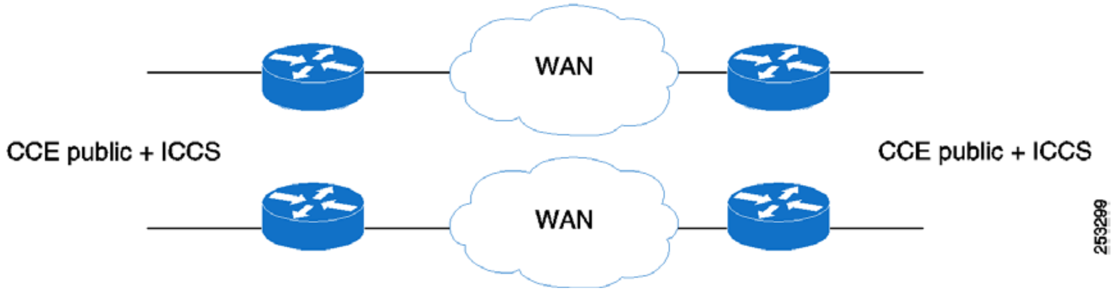
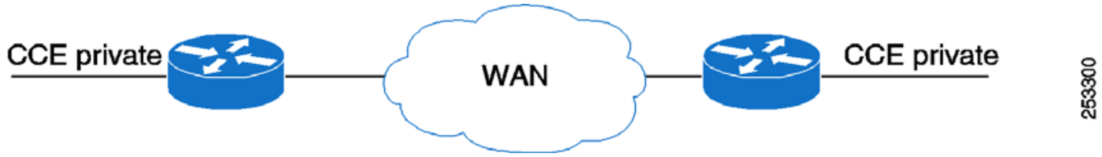
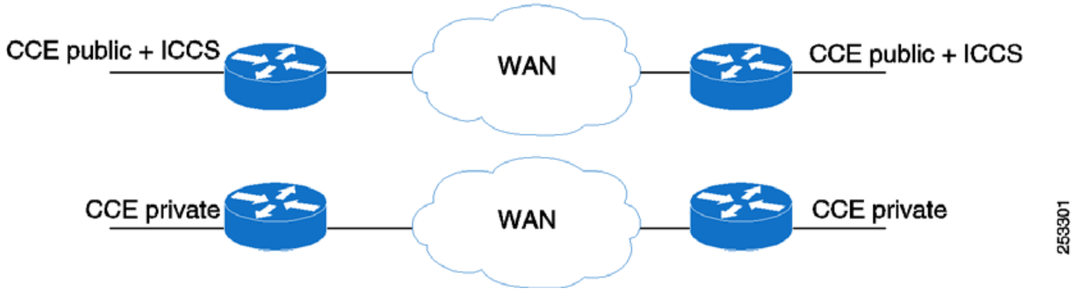


**Figure 23** ***Separate WAN link for Unified CCE Private Traffic***



- It is possible to deploy Unified CCE clustering over the WAN with two links if the following rules are applied:
  - During normal operations, the Unified CCE public and private traffic must be carried over separate links; they must not be carried over the same link.
  - Carry the Unified CM traffic over the Unified CCE public link in normal conditions (see Figure 24).
  - Two routers are required on each side of the WAN for redundancy; connect these to different WAN links.

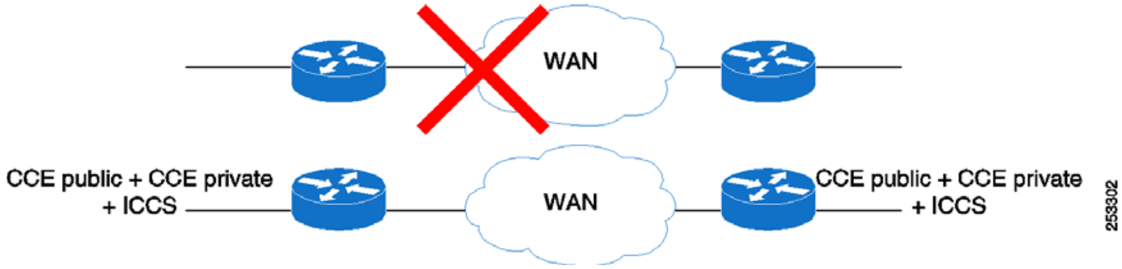
**Figure 24** ***Network Architecture Under Normal Operations***



- In case of network failure, configure the WAN link that carries the Unified CCE public traffic to fail-over to the other link that carries the Unified CCE private traffic (see Figure 25). Temporarily allow the Unified CM ICCS traffic to fail-over to the private link. This prevents situations where a CTI Manager that connects to the active Agent PG loses its WAN connection to the Unified CM node to which the agent phones are registered. Restore the link as fast as possible so that the public and private Unified CCE traffic are carried over separate links. If the redundant link that carries the Unified CCE private traffic also fails, Unified CCE instability and data loss can occur, including the corruption of one Logger database. Manual intervention could be required. This is why it is very important to actively monitor any network failure at all times.

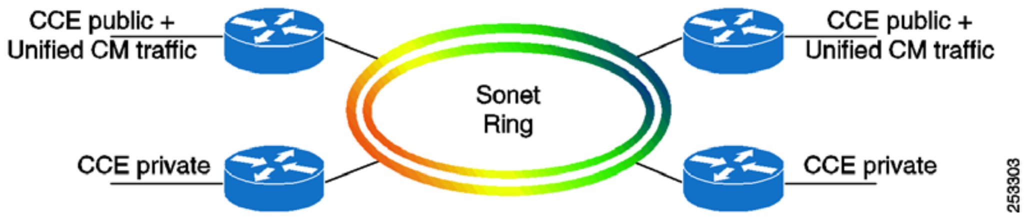
The links must also be sized correctly in order to accommodate this failover situation where the private link carries the entire WAN traffic, including the public and ICCS traffic. QoS and bandwidth must be configured according to the guidelines in the Bandwidth Provisioning and QoS Considerations chapter.

**Figure 25** ***Network Architecture After Failure of the Unified CCE Public Network***


WAN
CCE public + CCE private + ICCS
WAN
CCE public + CCE private + ICCS
253302

- It is also possible to allow the private link to fail-over to the public link. However, if the total failover latency takes more than 500 ms (five times the TCP keep alive interval of 100 ms), the Unified CCE system considers the private link to be down. If the public link is also down, Unified CCE instability and data loss can occur, including the corruption of one Logger database. Manual intervention could be required. The total failover latency typically includes the round-trip transmission latency, the routing protocol convergence delay, the HSRP convergence delay (if applicable), queuing and packetization delays, and any other delay that would be applicable. If the total failover latency is higher than 500 ms, or if you suspect possible recurrent network flapping, deploy three WAN links and keeping the private traffic separate from the public traffic at all times. Also, the links must be sized correctly in order to accommodate this failover situation where the public link carries the entire WAN traffic, including the private and ICCS traffic. Restore the link as fast as possible so that the public and private Unified CCE traffic are carried over separate links.
- If QoS and bandwidth are configured correctly (see the guidelines in the Bandwidth Provisioning and QoS Considerations chapter for more details), these WAN links can be converged with other corporate traffic.
- With a SONET fiber ring, which is highly resilient and has built-in redundancy, the public and private traffic can be carried 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. (See Figure 26.) The other rules described in this section also apply.

**Figure 26** ***Network Architecture Based on a SONET Ring Under Normal Operations***


CCE public + Unified CM traffic
CCE public + Unified CM traffic
Sonet Ring
CCE private
CCE private
253303

- The high availability (HA) WAN between the central sites must be fully redundant with no single point of failure. (For information regarding site-to-site redundancy options, see the available at WAN infrastructure and QoS design guides.) In case of partial failure of the high availability WAN, the redundant link must be capable of handling the full central-site load with all QoS parameters. For more information, see Bandwidth Requirements for Unified CCE Clustering Over the WAN.
- A high availability (HA) WAN using point-to-point technology is best implemented across two separate carriers, but this is not necessary when using a ring technology.
- Latency requirements across the high availability (HA) WAN must meet the current Cisco Unified Communications requirements for clustering over the WAN. Currently, a maximum latency of 40 ms one way (80 ms round-trip) is allowed with Unified CM 6.1 or later releases. With prior versions of Unified CM, the maximum latency is 20 ms one way. For full specifications, see the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.
- Unified CCE latency requirements can be met by conforming to Cisco Unified Communications requirements. However, the bandwidth requirements for Unified CM intra-cluster communications differ between Unified CCE and Cisco Unified Communications. For more information, see the section on Bandwidth Requirements for Unified CCE Clustering Over the WAN.
- Bandwidth requirements across the high availability (HA) WAN include bandwidth and QoS provisioning for:
  - Unified CM intra-cluster communication signaling (ICCS)
  - Communications between Unified CCE Central Controllers
  - Communications between Unified CCE Central Controller and PG
  - Communications between CTI Object Server (CTI OS) and CTI Server, if using CTI OS

  See Bandwidth Requirements for Unified CCE Clustering Over the WAN.
- Deploy separate dedicated links for Unified CCE private communications between Unified CCE Central Controllers Side A and Side B and between PGs Side A and Side B to ensure path diversity. Path diversity is required due to the architecture of Unified CCE. Without path diversity, the possibility of a dual (public communication and private communication) failure exists. If a dual failure occurs even for a moment, Unified CCE instability and data loss can occur, including the corruption of one Logger database. The separate links for Unified CCE private communications can be converged with other corporate traffic if QoS and bandwidth are configured correctly, but they cannot be converged with the Unified CCE public traffic.
- The separate private links may be either two links (one for Central Controller private traffic and one for Unified CM PG private traffic) or one converged link containing both Central Controller and PG private traffic. See Site-to-Site Unified CCE Private Communications Options for more information.
- Separate paths must exist from agent sites to each central site. Both paths must be capable of handling the full load of signaling, media and other traffic if one path fails. These paths may reside on the same physical link from the agent site with a WAN technology such as Frame Relay using multiple permanent virtual circuits (PVCs).
- The minimum cluster size using Unified IP IVR as the treatment and queuing platform is 5 nodes (publisher plus 4 subscribers). This minimum is required to allow Unified IP IVR at each site to have redundant connections locally to the cluster without traversing the WAN. JTAPI connectivity

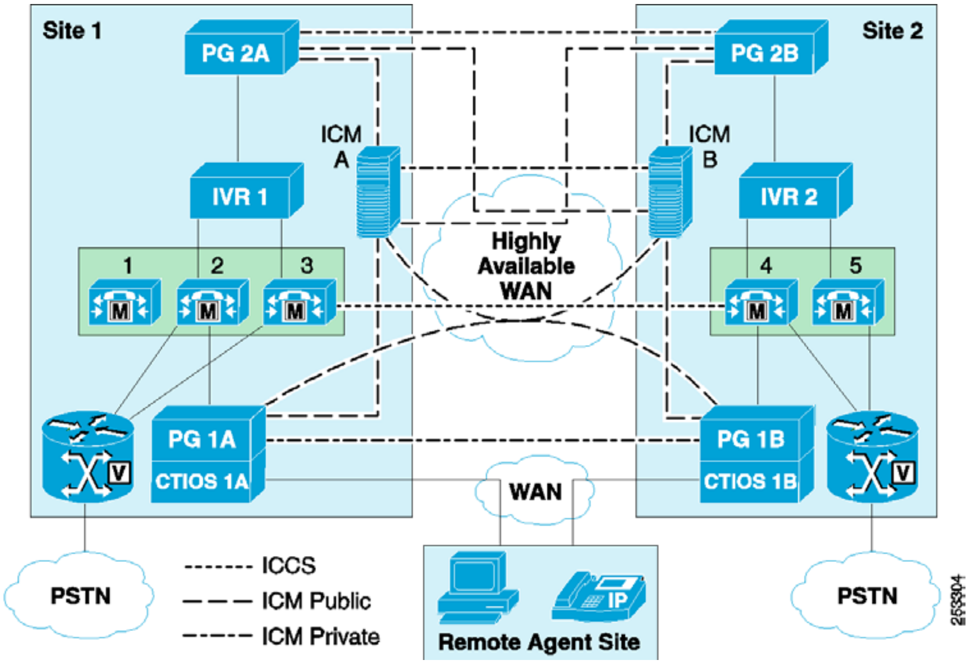
between Unified CM and Unified IP IVR is not supported across the WAN in this model. Local gateways also will need local redundant connections to Unified CM.

- The minimum cluster size using Unified CVP as the treatment and queuing platform is 3 nodes (publisher plus 2 subscribers). However, a deployment with 5 nodes is preferable, especially if there are phones (either contact center or non-contact center) local to the central sites, central gateways, or central media resources which would require local failover capabilities.
- In a deployment with clustering over the WAN, the VRU PG could connect to a local IP IVR or Unified CVP or to a redundant IP IVR or Unified CVP across the WAN. For information about bandwidth requirements, see the Bandwidth Provisioning and QoS Considerations chapter.
- In a deployment with clustering over the WAN, the Unified CM PG must be on the same LAN segment with the CTI Manager to which it is connected.
- Cisco Finesse does not support connection to the CTI Server over a WAN. All connections from Finesse to the CTI Server must be local to a given location.
- Cisco Finesse does not support side A and side B Finesse servers separated over a WAN connection. Both Finesse servers must be co-located in a single facility.

## Centralized Voice Gateways with Centralized Call Treatment and Queuing Using Unified IP IVR

In this model, the voice gateways are located in the central sites. Unified IP IVR is centrally located and used for treatment and queuing on each side. Figure 27 illustrates this model.

**Figure 27** *Centralized Voice Gateways with Centralized Call Treatment and Queuing Using Unified IP IVR*



### Advantages

- Component location and administration are centralized.

- Calls are treated and queued locally, eliminating the need for queuing across a WAN connection.

**Best Practices**

- WAN connections to agent sites must be provisioned with bandwidth for voice as well as control and CTI. See Bandwidth Requirements for Unified CCE Clustering Over the WAN for more information.
- A local voice gateway might be needed at remote sites for local out-calling and 911. For more information, see the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.
- Central site outages would include loss of half of the ingress gateways, assuming a balanced deployment. Gateways and IVRs must be scaled 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 in the case of a site or gateway loss. Pre-routing may be used to balance the load, but it will not be able to prevent calls from being routed to a failed central site. Pre-routing is best deployed only as a last resort.

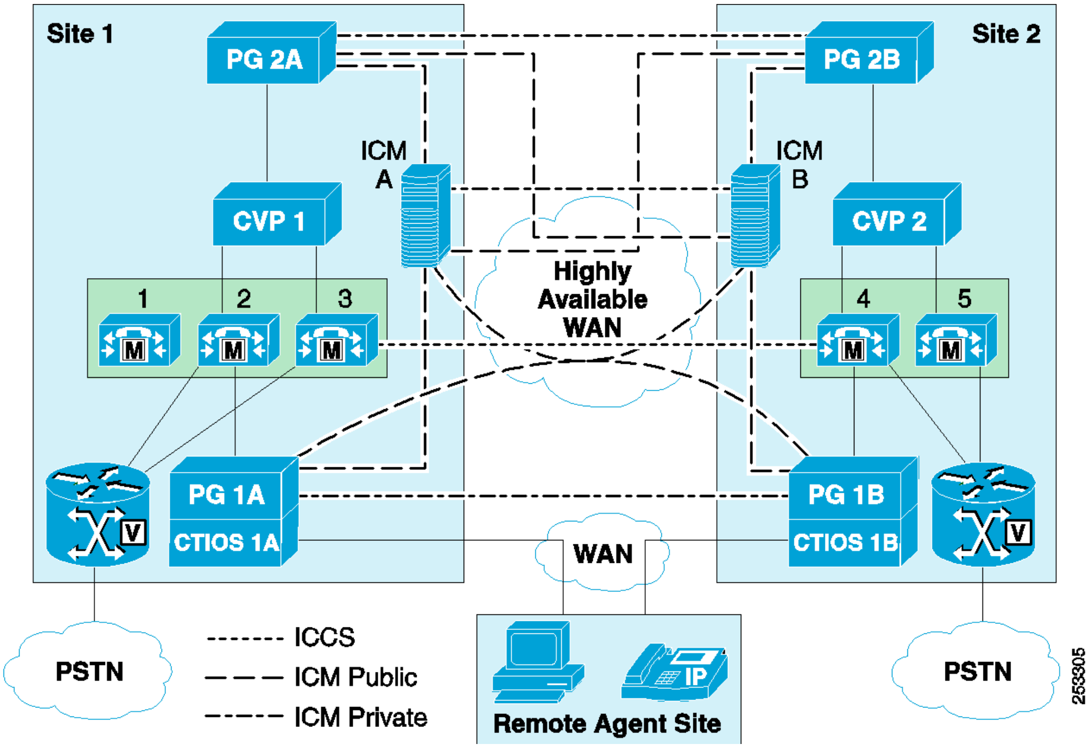
## Clustering Over the WAN with Unified CCE System PG

Clustering over the WAN with Unified CCE System PG is now supported. However, due to the fact that a single Unified CCE System peripheral is controlling all of the Unified IP IVRs and the Unified CM, the load-balancing of calls between Unified IP IVRs does not take into account which site the call came into; it simply distributes the calls to whichever Unified IP IVR is least loaded. This means that calls coming into Site A might be treated by a Unified IP IVR in Site B. Additionally, both the A-side and B-side Unified CCE System PG know about all of the Unified IP IVRs. PIM activation logic will determine if the A-side or the B-side PIM will connect to each of the Unified IP IVRs. This means that the PG at site A might connect to the Unified IP IVR at site B and traffic might not be sent optimally over the WAN. In this model, make sure the WAN is sized for proper operation given this fact. To avoid this bandwidth overhead, you can consider Clustering over the WAN deployments with Unified CVP in place of Unified IP-IVR.

## Centralized Voice Gateways with Centralized Call Treatment and Queuing Using Unified CVP

In this model, the voice gateways are VoiceXML gateways located in the central sites. Unified CVP is centrally located and used for treatment and queuing. Figure 28 illustrates this model.

**Figure 28** *Centralized Voice Gateways with Centralized Call Treatment and Queuing Using Unified CVP*



### Advantages

- Component location and administration are centralized.
- Calls are treated and queued locally, eliminating the need for queuing across a WAN connection.
- There is less load on Unified CM because Unified CVP is the primary routing point. This allows higher scalability per cluster compared to Unified IP IVR implementations. See the Sizing Unified CCE Components and Servers chapter for more information.

### Best Practices

- WAN connections to agent sites must be provisioned with bandwidth for voice as well as control and CTI. See Bandwidth Requirements for Unified CCE Clustering Over the WAN for more information.
- A local voice gateway might be needed at remote sites for local out-calling and 911.
- Cisco Finesse does not support connection to the CTI Server over a WAN. All connections from Finesse to the CTI Server must be local to a given location.
- Cisco Finesse does not support side A and side B Finesse servers separated over a WAN connection. Both Finesse servers must be co-located in a single facility.

## Centralized Voice Gateways with Centralized Call Treatment and Queuing Using Unified SCCE 7.x with Unified CVP

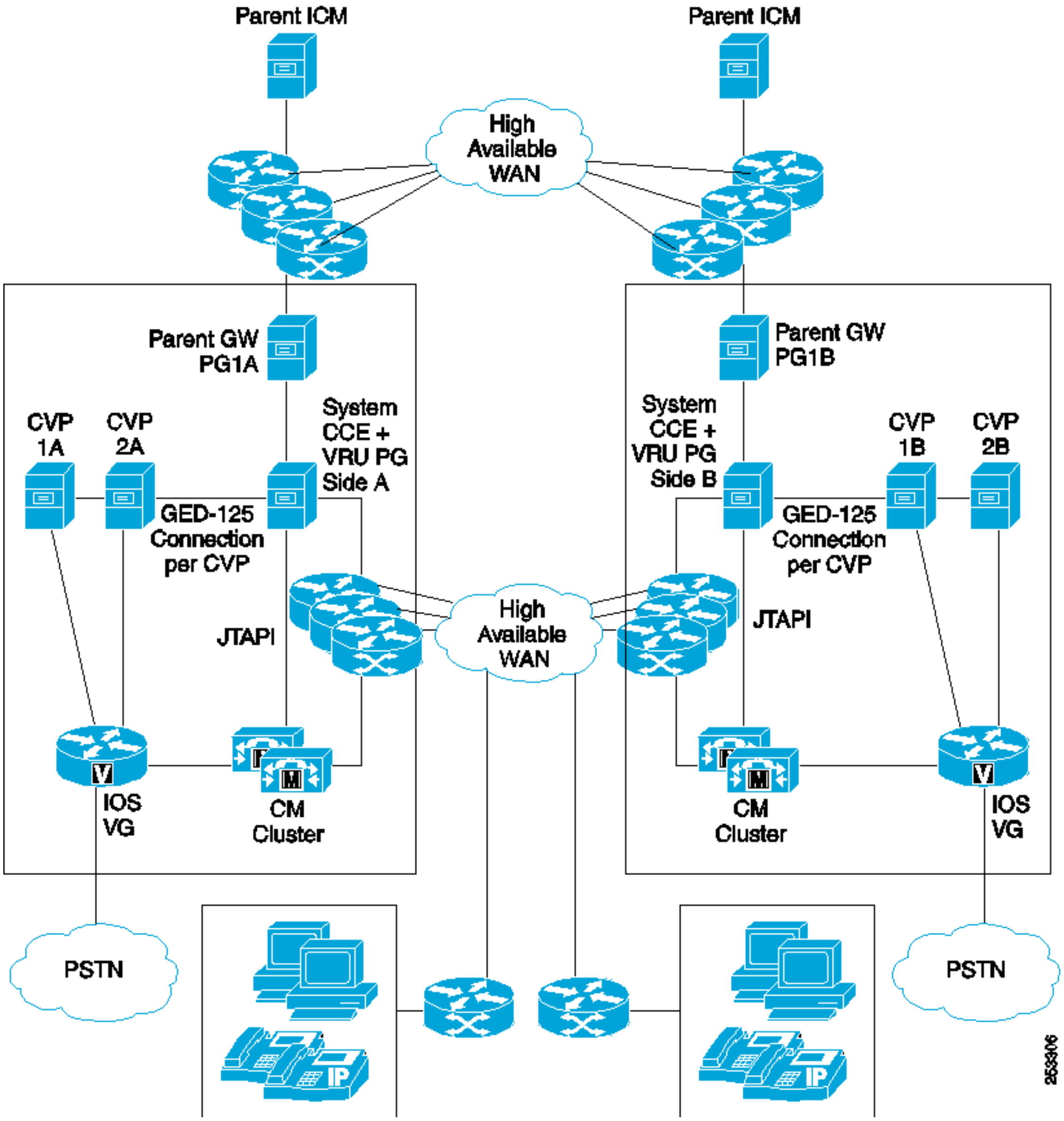
Load balancing of calls across Unified CVP Call Servers is managed by SIP and Cisco Unified SIP Proxy (CUSP). The load balancing does not take into account the site where the call came in, but calls are distributed based on simple load balancing rules define in Cisco Unified Presence (for example, alternate call distributions across configured Unified CVP Call Servers and preferential weighting of Call Servers).

Currently, if the system is designed to do so, Unified CVP can queue the call at the ingress gateway. This requires that Unified CVP be configured with **set transfer label** for H.323 or **Send To Originator** for SIP to match the Network VRU label. This will cause Unified CVP to send the call back to the ingress gateway for queuing when a label matching this Network VRU label is returned from Unified ICM/CCE. Currently Unified ICM/CCE is unaware of the location of the initial gateway; therefore it cannot make a label selection based on the original ingress location of the call.

## Considerations for Clustering Over the WAN

Figure 29 illustrates a deployment with clustering over the WAN.

**Figure 29** *Clustering Over the WAN*



The following guidelines and considerations apply to deployments with clustering over the WAN:

- The network deployment supports high availability, converged Visible, and Private Networks. The Unified ICM and Unified CCE Central Controller's Private traffic and Visible (Public) traffic are isolated and converge on different edge devices.
- WAN considerations for communications between the two Data Centers may include a Multiprotocol Label Switching (MPLS) backbone with VPN routing and forwarding table VRFs.
- Design the network to prevent any single points of failure. The visible network and private networks need to converge on separate switches and routers before connecting to the WAN.

- Isolation of the private network is not required. Central Controllers and Unified CCE System PGs can share a common private network path.
- Multiple private network paths can be provisioned. (Central Controllers and Unified CCE System PGs have separate private networks.)
- Bandwidth must be guaranteed across the WAN for the private network path traffic and visible (public) network traffic with appropriate traffic prioritization. For more information, see the Bandwidth Provisioning and QoS Considerations chapter.
- Currently there is no bandwidth calculator for the private network bandwidth between the gateway and system PG pairs because this has not been certified. For guidance, see the section on Bandwidth Provisioning.
- A side-to-side private network of a duplexed Central Controller and PGs has a maximum one-way latency of 100 ms, but 50 ms is preferred.

The underlying network infrastructure for LAN and WAN provisioning must meet all the above requirements. Key factors are isolation of visible and private paths as well as critical low-latency and bandwidth, especially on the private path. The isolated private networks for PGs and Central Controllers provide some degree of independence from each other's private link failures. The more path and route diversity provisioned, the greater the fault tolerance. For example, if the private network between the parent central controllers goes down, the child central controllers can still continue to function in duplex mode.

### MPLS Considerations

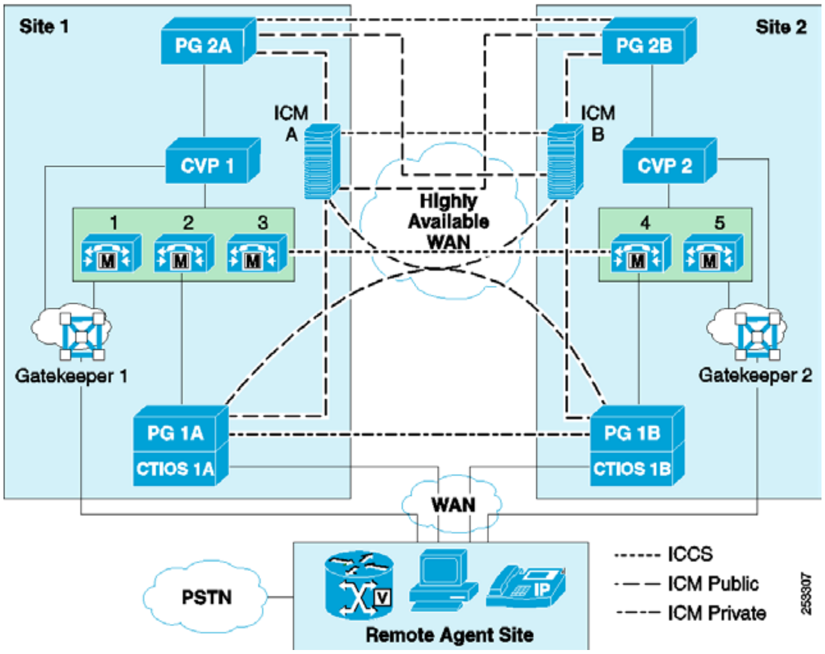
If an MPLS network can guarantee the route diversity, latency, and bandwidth; and if it is configured to support label switch paths that route through independent topologies and hardware elements to meet the above requirements, then the application will work as designed. It is important to ensure that the route autonomy is not compromised over time through adaptive change.

For additional information regarding best practices and high availability deployments, see the section on IPT: Clustering over the WAN.

## Distributed Voice Gateways with Distributed Call Treatment and Queuing Using Unified CVP

In this model, the voice gateways are VoiceXML gateways distributed to agent locations. Unified CVP is centrally located and used for treatment and queuing on the remote gateways. Figure 30 illustrates this model.

**Figure 30** *Distributed Voice Gateways with Distributed Call Treatment and Queuing Using Unified CVP*



### Advantages

- No or minimal voice RTP traffic across WAN links if ingress calls and gateways are provisioned to support primarily their local agents. Transfers and conferences to other sites would traverse the WAN.
- Calls are treated and queued at the agent site, eliminating the need for queuing across a WAN connection.
- Local calls incoming and outgoing, including 911, can share the local VoiceXML gateway.
- There is less load on Unified CM because Unified CVP is the primary routing point. This allows higher scalability per cluster compared to Unified IP IVR implementations. See Sizing Unified CCE Components and Servers for more information.

### Best Practices

- Distributed gateways require minimal additional remote maintenance and administration over centralized gateways.
- The media server for Unified CVP may be centrally located or located at the agent site. Media may also be run from gateway flash. Locating the media server at the agent site reduces bandwidth requirements but adds to the server count and maintenance costs.
- Cisco Finesse does not support connection to the CTI Server over a WAN. All connections from Finesse to the CTI Server must be local to a given location.
- Cisco Finesse does not support side A and side B Finesse servers separated over a WAN connection. Both Finesse servers must be co-located in a single facility.

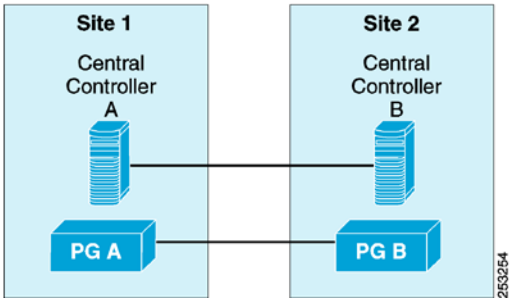
## Site-to-Site Unified CCE Private Communications Options

Unified CCE private communications must travel on a separate path from the public communications between Unified CCE components. There are two options for achieving this path separation: dual and single links.

## Unified CCE Central Controller Private and Unified CM PG Private Across Dual Links

Dual links, as shown in Figure 31, separate Unified CCE Central Controller Private traffic from VRU/CM PG Private traffic.

**Figure 31** ***Unified CCE Central Controller Private and Unified CM PG Private Across Dual Links***



### Advantages

- Failure of one link does not cause both the Unified CCE Central Controller and PG to enter simplex mode, thus reducing the possibility of an outage due to a double failure.
- The QoS configuration is limited to two classifications across each link, therefore links are simpler to configure and maintain.
- Resizing or alterations of the deployment model and call flow might affect only one link, thus reducing the QoS and sizing changes needed to ensure proper functionality.
- Unanticipated changes to the call flow or configuration (including misconfiguration) are less likely to cause issues across separate private links.

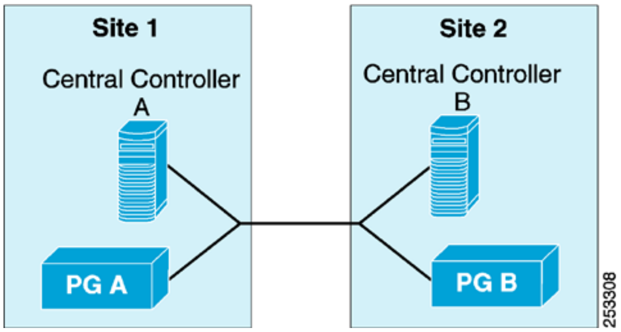
### Best Practices

- Deploy separate links across separate dedicated circuits. The links, however, do not have to be redundant and must not fail-over to each other.
- Link sizing and configuration must be examined before any major change to call load, call flow, or deployment configuration.
- Make the link a dedicated circuit; not tunneled across the high availability (HA) WAN. See Best Practices, at the beginning of the section about IPT: Clustering Over the WAN for more information about path diversity.

## Unified CCE Central Controller Private and Unified CM PG Private Across Single Link

A single link, as shown in Figure 32, carries both Unified CCE Central Controller Private traffic and VRU/CM PG Private traffic. Single-link implementations are more common and less costly than dual-link implementations.

**Figure 32** ***Unified CCE Central Controller Private and Unified CM PG Private Across a Single Link***



### Advantages

- Less costly than separate-link model.
- Fewer links to maintain, but is more complex.

### Best Practices

- The link does not have to be redundant. If a redundant link is used, however, latency on failover must not exceed 500 ms.
- Separate QoS classifications and reserved bandwidth are required for Central Controller high-priority and PG high-priority communications. For details, see the Bandwidth Provisioning and QoS Considerations chapter.
- Link sizing and configuration must be examined before any major change to call load, call flow, or deployment configuration. This is especially important in the single-link model.
- Make the link a dedicated circuit fully isolated from, and not tunneled across, the high availability (HA) WAN. See Best Practices, at the beginning of the section on IPT: Clustering Over the WAN for more information about path diversity.

## Failure Analysis of Unified CCE Clustering Over the WAN

This section describes the behavior of clustering over the WAN for Unified CCE during certain failure situations. The stability of the high availability (HA) WAN is extremely critical in this deployment model, and failure of the high availability WAN is considered outside the bounds of what would normally happen.

For illustrations of the deployment models described in this section, see the figures shown previously for IPT: Clustering Over the WAN.

## Entire Central Site Loss

Loss of the entire central site is defined as the loss of all communications with a central site, as if the site were switched off. This can result from natural disasters, power issues, major connectivity issues, and human error. If a central site retains some but not all connectivity, it is not considered a site loss but rather a partial connectivity loss. This scenario is covered in subsequent sections.

When an entire central site has completely lost Unified CCE clustering over the WAN, Mobile agents will fail-over properly to the redundant site. Failover times can range from 1 to 60 seconds for agents. Variations are due to agent count, phone registration location, and agent desktop server used.

When using distributed VoiceXML gateways and Unified CVP, the gateways must fail-over from one site to another if their primary site is lost. This failover takes approximately 30 seconds and calls coming into the remote gateways during those 30 seconds will be lost.

## Private Connection Between Site 1 and Site 2

If the private connection between Unified CCE Central Controller sides A and B fails, one Unified CCE Call Router will go out-of-service and the other Unified CCE Call Router will then be running in simplex mode until the link is reinstated. PGs with active connections to the Unified CCE Router that goes out-of-service realign message streams to the remaining active Unified CCE Router side. This situation will not cause any call loss or failure.

If the private connection between PG side A and PG side B fails, the disabled synchronizer (side B) initiates a test of its peer synchronizer by using the TOS procedure on the Public or Visible Network connection. If PG side B receives a TOS response stating that the A side synchronizer is enabled or active, then the B side immediately goes out of service, leaving the A side to run in simplex mode until the Private Network connection is restored. The PIM, OPC, and CTI SVR processes become active on PG side A, if not already in that state, and the CTI OS Server process still remains active on both sides as long as the PG side B server is healthy.

If the B side does not receive a message stating that the A side is enabled, then side B continues to run in simplex mode and the PIM, OPC, and CTI SVR processes become active on PG side B, if not already in that state. This condition (PG side B going Active) occurs only if the PG side A server is truly down or unreachable due to a double failure of visible and private network paths. The side A or Side B PG runs in simplex mode until the private link is reinstated.

There is no impact to the agents, calls in progress, or calls in queue because the agents stay connected to their already established CTI OS Server process connection. The system can continue to function normally; however, the PGs will be in simplex mode until the private network link is restored.

When using a combined private link, Unified CCE Central Controller and PG private connections will be lost if the link is lost. This will cause both components to switch to simplex mode as described above. This situation will not cause any call loss or failure.

## Connectivity to Central Site from Mobile Agent Site

If connectivity to one of the central sites is lost from a Mobile agent site, all phones and agent desktops will immediately switch to the second central site and begin processing calls. Failover typically takes between 1 and 60 seconds.

## High Availability WAN Failure

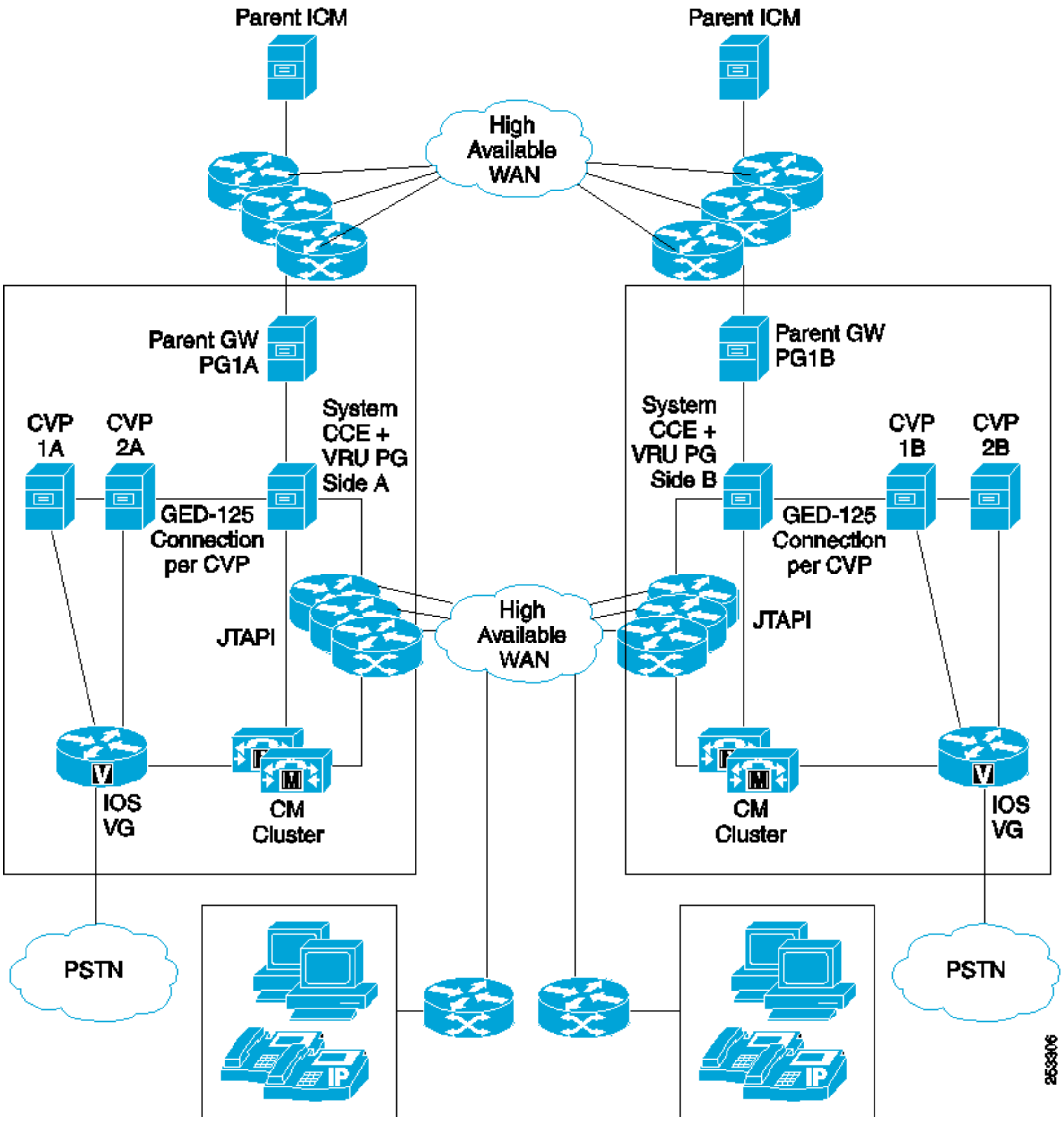
By definition, a high availability (HA) WAN does not fail under normal circumstances. If the HA WAN is dual-path and fully redundant, a failure of this type is highly unusual. This section discusses what happens in this unlikely scenario.

If the HA WAN is lost for any reason, the Unified CM cluster becomes split. The primary result from this occurrence is that Unified CCE loses contact with half of the agent phones. Unified CCE is in communication with only half of the cluster and cannot communicate with or see any phones registered on the other half. This causes Unified CCE to immediately log out all agents with phones that are no longer visible. These agents cannot log back in until the HA WAN is restored or their phones are forced to switch cluster sides.

## Split Unified CCE Gateway PGs

To enhance the distributed architecture of the Unified SCCE 7.x deployment, the support for geographically distributed Cisco Unified CCE Gateway PGs is needed. The Unified CCE Gateway PGs are deployed in the same location as the System PGs, adding maximum recovery capabilities in the event of a site failure. Figure 33 shows a distributed Unified SCCE 7.x deployment supporting two Remote Data Centers with Unified CCE Gateway PGs co-located (on separate servers) with each of the distributed Unified SCCE7.x systems. Note that the same would apply if the child servers were Unified CCEs utilizing the Unified CCE System PG.

**Figure 33** *Gateway PG Co-Located*



# Mobile Agent Over Broadband

An organization might want to deploy Unified CCE to support mobile agents (for example, at-home agents) using a Cisco Unified IP Phone over a broadband internet connection. This section outlines the mobile agent solution that can be deployed using a desktop broadband asymmetric digital subscriber line (ADSL) or Cable connection as the remote network. Another option is to use the Cisco Unified Mobile Agent solution (for details, see the Cisco Unified Mobile Agent chapter). Both Cisco Unified Mobile Agent and Mobile agent over Broadband can be supported concurrently using the same back-end infrastructure with the Cisco Virtual Office solution which is an underlying end-to-end secure infrastructure for remote teleworkers utilizing a converged VPN architecture.

The Cisco Voice and Video Enabled IPSec VPN (V3PN), ADSL, or Cable connection can use a Cisco 800 Series router as an edge router to the broadband network. The Cisco 800 Series router can provide the mobile agent with V3PN, Encryption, Network Address Translation (NAT), Firewall, Cisco IOS Intrusion Detection System (IDS), and QoS on the broadband network link to the Unified CCE campus. Mobile agent V3PN aggregation on the campus is provided through LAN to LAN VPN routers.

Use the Cisco 800 Series router with the following features for mobile agent over broadband:

- Quality of Service (QoS) with Low-Latency Queuing (LLQ) and Class-Based Weighted Fair Queuing (CBWFQ) support
- Managed Switch
- Power over Ethernet (optional)

The Cisco 830, 870, and 880 Series routers are examples of acceptable routers. Avoid using the Cisco 850 and 860 Series routers for this application because they have limited QoS feature support.

**Advantages**

- A mobile agent deployment results in money saved for a contact center enterprise, thereby increasing return on investment (ROI).
- Mobile agents can be deployed with standard Unified CCE agent desktop applications such as Cisco CTI OS, Cisco Agent Desktop, or customer relationship management (CRM) desktops.
- The Broadband Agent Desktop Always-on connection is a secure extension of the corporate LAN in the home office.
- Mobile agents have access to the same Unified CCE applications and most Unified CCE features in their home office as when they are working at the Unified CCE contact center and they can access those features in exactly the same way.
- The mobile-agent router provides high-quality voice using IP phones with simultaneous data to the agent desktop over existing broadband service.
- Unified CCE home agents and home family users can securely share broadband Cable and DSL connections with authentication of Unified CCE corporate users providing access to the VPN tunnel.
- The mobile-agent routers can be managed centrally by the enterprise using a highly scalable and flexible management product such as Cisco Unified Operations Manager.
- The mobile-agent-over-broadband solution is based on Cisco IOS VPN Routers for resiliency, high availability, and a building-block approach to high scalability that can support thousands of home agents.
- All traffic, including data and voice, is encrypted with the Triple Data Encryption Standard (3DES).
- The remote-agent router can be deployed as part of an existing Unified CM installation.
- Mobile agents can have the same extension type as campus agents.

### Best Practices

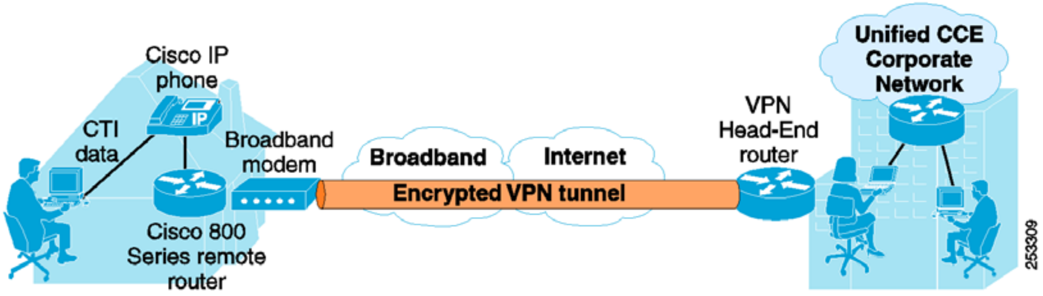
- Follow all applicable V3PN and Cisco Virtual Office design guidelines outlined in the documentation available at:
  - cisco.com/go/cvo
  - cisco.com/go/designzone
- Configure mobile agent IP phones to use G.729 with minimum bandwidth limits. Higher-quality voice can be achieved with the G.711 codec. The minimum bandwidth to support G.711 is 512 kbps upload speed.
- Implement fault and performance management tools such as NetFlow, Service Assurance Agent (SAA), and Internetwork Performance Monitor (IPM).
- Wireless access points are supported; however, their use is determined by the enterprise security polices for Mobile agents.
- Only one mobile agent per household is supported.
- Configure the conference bridge on a DSP hardware device. There is no loss of conference voice quality using a DSP conference bridge. This is the preferred solution even for pure Cisco Unified Communications deployments.
- The remote-agent-over-broadband solution is supported only with centralized Unified CCE and Unified CM clusters.
- There might be times when the ADSL or Cable link goes down. When the link is back up, you might have to reset your ADSL or Cable modem mobile agent router and IP phone. This task requires Mobile agent training.
- Only unicast Music on Hold (MoH) streams are supported.
- There must be a Domain Name System (DNS) entry for the mobile agent desktop, otherwise the agent will not be able to connect to a CTI server. DNS entries can be updated dynamically or entered as static updates.
- The mobile agent workstation and IP phone must be set up to use Dynamic Host Configuration Protocol (DHCP).
- The mobile agent workstation requires Windows XP Pro for the operating system. In addition, XP Remote Desktop Control must be installed.
- The Cisco Unified IP Phone requires a power supply if the remote-agent router does not have the Power over Ethernet option.
- Mobile agent broadband bandwidth requires a minimum of 256 kbps upload speed and 1.4 Mbps download speed for ADSL and 1 Mbps download for Cable. Before actual deployment, make sure that the bandwidth is correct. If you are deploying Cable, take into account peak usage times. If link speeds fall below the specified bandwidth, the home agent can encounter voice quality problems such as clipping.
- Mobile agent round-trip delay to the Unified CCE campus is not to exceed 180 ms for ADSL or 60 ms for Cable. Longer delay times can result in voice jitter, conference bridge problems, and delayed agent desktop screen pops.
- If the Music on Hold (MoH) server is not set up to stream using a G.729 codec, then a transcoder must be set up to enable outside callers to receive MoH.

- CTI OS Supervisor home and campus supervisors can silently monitor, barge in, and intercept (but not record) home agents. CTI OS home and campus supervisors can send and receive text messages, make an agent ready, and log out home agents.
- Connect the agent desktop to the RJ45 port on the back of the IP phone. Otherwise, CTI OS Supervisor will not be able to voice-monitor the agent phone.
- Only IP phones that are compatible with Cisco Unified CCE are supported. For compatibility information, see the following documentation:
  - *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*
  - Cisco Unified Contact Center Enterprise (Unified CCE) Software Compatibility Guide
  - Release Notes for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.0(1)
- You can find a test for the broadband line speed at Broadbandreports.com. From this website, you can execute a test that will benchmark the home agent's line speed (both upload and download) from a test server.
- Cisco Finesse does not support mobile agents. If you require support for mobile agents, you must use CTI OS.

## Mobile Agent with Unified IP Phones Deployed by using the Cisco Virtual Office Solution

In this model, the mobile agent IP phone and workstation are connected by using the VPN tunnel to the main Unified CCE campus. Customer calls routed to the mobile agent are handled in the same manner as campus agents, as shown in Figure 34.

**Figure 34** *Mobile Agent with IP Phones Deployed via the Cisco Virtual Office Solution*



### Advantages

- High-speed broadband enables cost-effective office applications.
- Site-to-site always-on VPN connection.
- Advanced security functions allow extension of the corporate LAN to the home office.
- Supports full range of converged desktop applications including CTI data and high-quality voice.

### Best Practices

- Minimum broadband speed supported is 256 kbps upload and 1.0 Mbps download for cable.
- Minimum broadband speed supported is 256 kbps upload and 1.4 Mbps download for ADSL.

- Agent workstation must have 500 MHz and 512 MB RAM or greater.
- IP phone must be configured to use G.729 on minimum broadband speeds.
- QoS is enabled only at the remote-agent router edge. Currently, service providers are not providing QoS.
- Enable security features on the remote-agent router.
- The Cisco 7200 VXR and Catalyst 6500 IPSec VPN Services Module (VPNSM) offer the best LAN-to-LAN performance for agents.
- The mobile agent home phone must be used for 911 calls.
- Use Redirect-on-no-answer (RONA) when a mobile agent is logged in and ready but is unavailable to pick up a call.

# Traditional ACD Integration

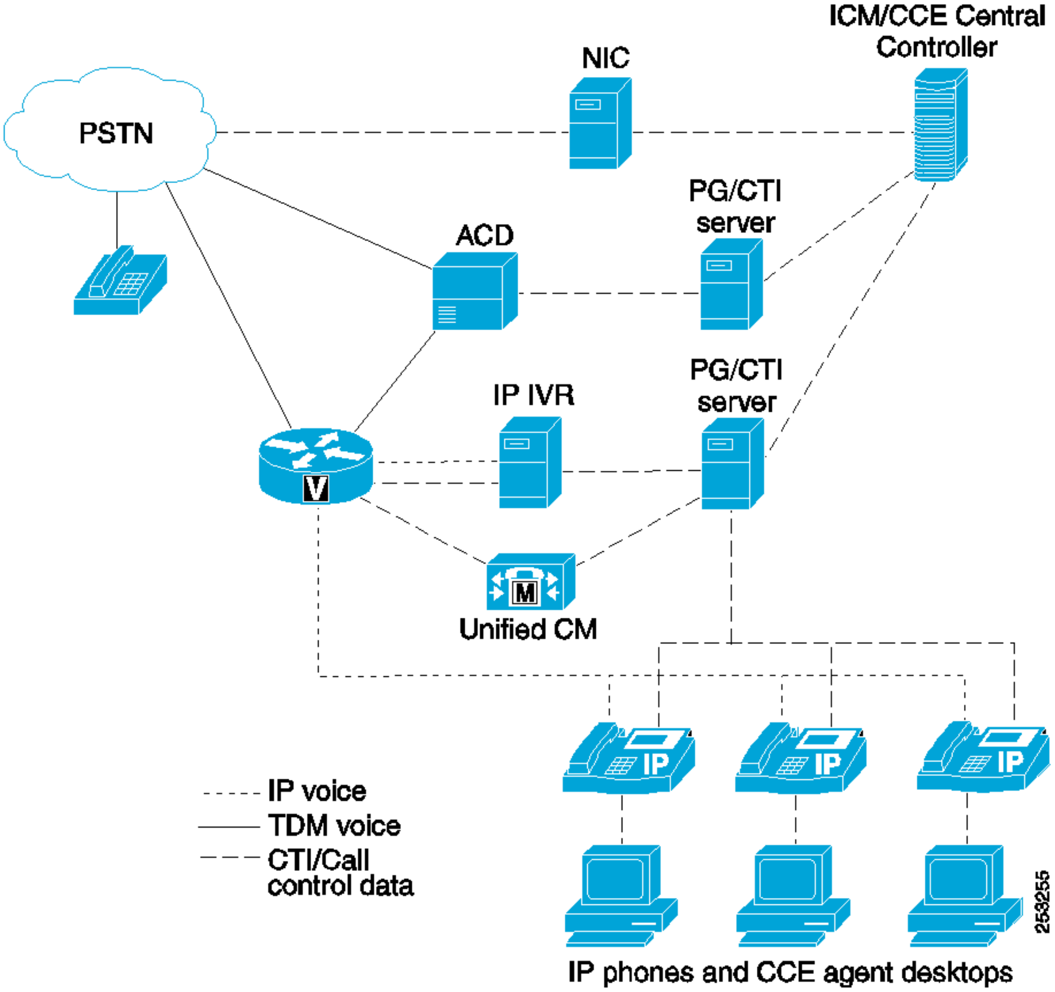
Enterprises that want to integrate traditional ACDs with their Unified CCE could use a parent/child deployment where the Unified ICM and Unified CCE each have a Central Controller or a hybrid deployment where Unified ICM and Unified CCE use a shared Central Controller. Several options exist within those categories depending on how the calls are routed within the deployment.

## Hybrid Deployment with PSTN Pre-routing

Enterprises that want to load-balance calls between a traditional ACD site and a Unified CCE site could add a pre-routing Network Interface Controller (NIC) as shown in Figure 35. This configuration requires that the Unified ICM have a NIC that supports the PSTN service provider. In this scenario, the PSTN queries the Unified ICM/CCE Central Controller (through the NIC) to determine which site is best and instructs the PSTN on where (which site) to deliver the call. Any call data provided by the PSTN to the Unified ICM/CCE is passed to the agent desktop (traditional ACD or Unified CCE).

In order to transfer calls between the two sites (ACD site and Unified CCE site), a PSTN transfer service could be used. Use of a PSTN transfer service avoids any double trunking of calls at either site. An alternative to using a PSTN transfer service is to deploy TDM voice circuits between the traditional ACD and Unified CCE voice gateways. In that environment, any transfer of a call back to the original site will result in double trunking between the two sites. Each additional transfer between sites will result in an additional TDM voice circuit being used.

**Figure 35** *Integrating a Traditional ACD with a Unified CCE Site Using a Hybrid Deployment and Pre-routing via PSTN*



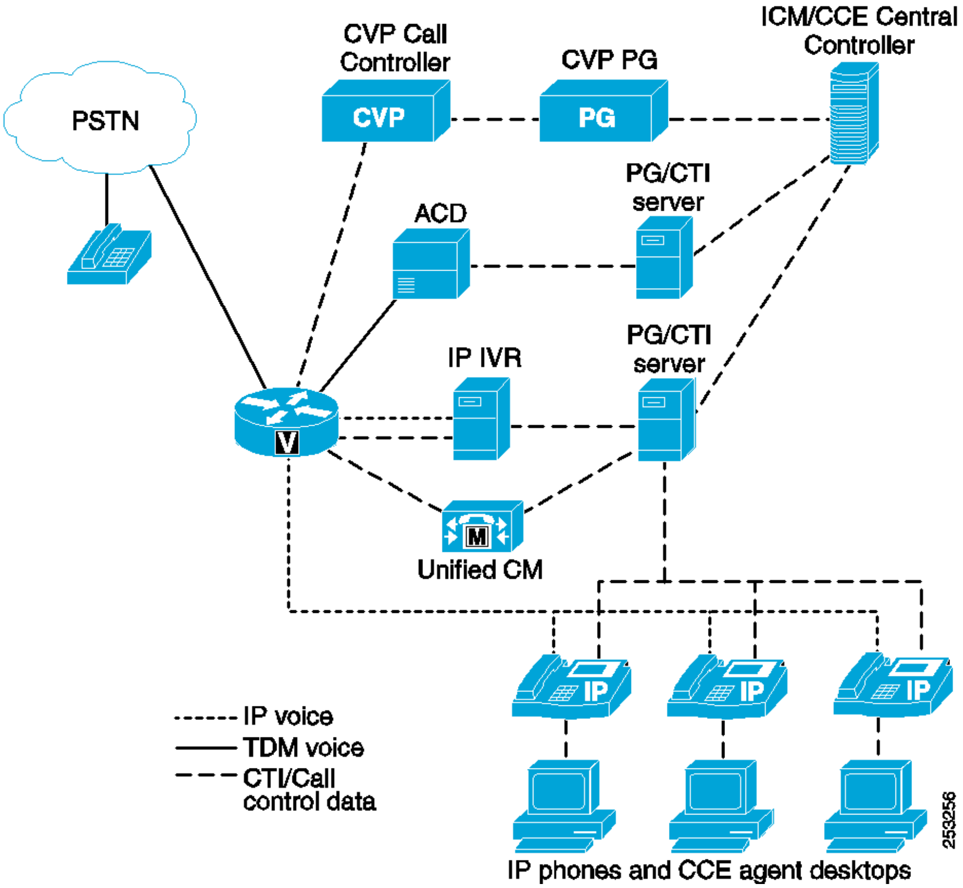
## Hybrid Deployment with Fixed PSTN Delivery

An alternative to pre-routing calls from the PSTN is to have the PSTN deliver calls to just one site or to split the calls across the two sites according to some set of static rules provisioned in the PSTN. When the call arrives at either site, either the traditional ACD or the Unified CM will generate a route request to the hybrid Unified ICM/CCE to determine which site is best for this call. If the call needs to be delivered to an agent at the opposite site from where the call was originally routed, then TDM circuits between sites will be required. The determination of where calls are routed (and if and when they are transferred between sites) will depend on the enterprise business environment, objectives, and cost components.

## Hybrid Deployment with Unified CVP

Alternatively, customers may 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, as shown in Figure 36.

**Figure 36** ***Integrating Unified CVP with a Traditional ACD and a Unified CCE Site Using a Hybrid Deployment and Unified CVP***

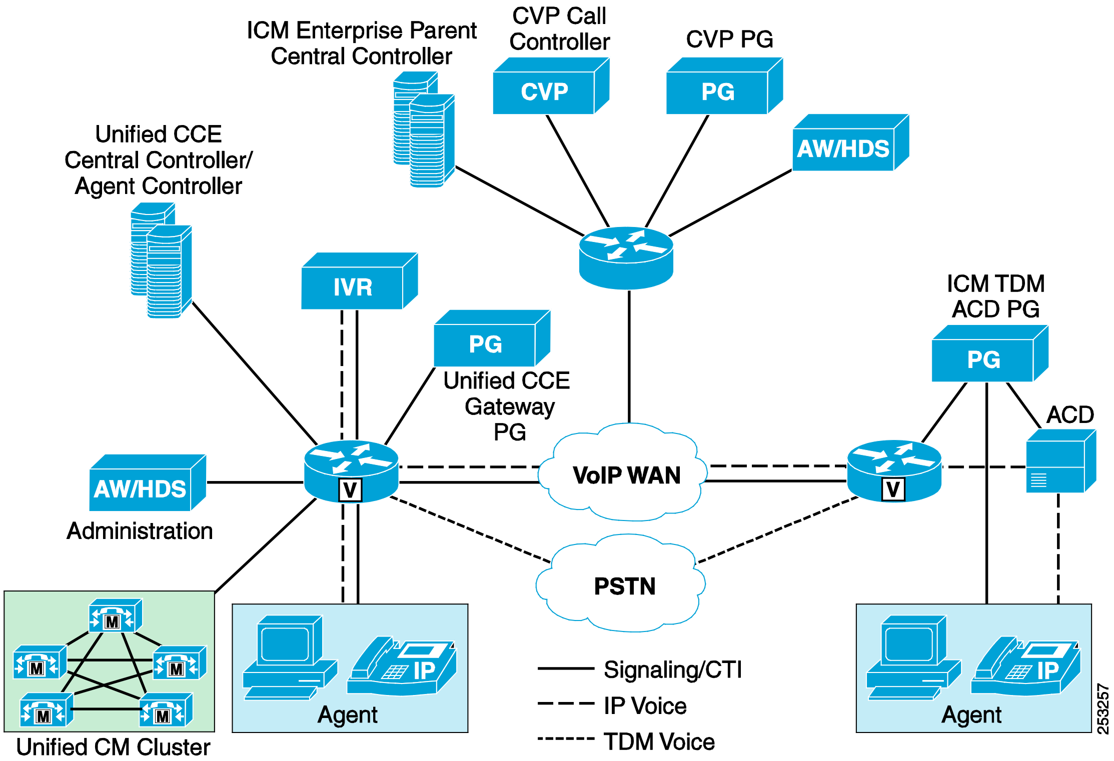


In this design, all calls first come to the voice gateway controlled by Unified CVP and they are then directed by the Unified ICM/CCE Call Router. Unified ICM/CCE uses the PG connections to the TDM ACD and Unified CCE PG to monitor for available agents. Calls are queued in Unified CVP until an agent becomes available in either environment. When a call needs to be transferred to the TDM ACD, it will hairpin in the voice gateway (meaning that it comes into the gateway 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. Most TDM ACDs are unable to accept inbound calls in IP from the voice gateway and require this physical T1 interface connection. Unified CCE agents receive their calls directly over the IP voice network.

## Parent/Child Deployment

The parent/child model is illustrated in Figure 37.

**Figure 37** ***Parent/Child Model for Integrating a Traditional ACD with a Unified CCE Site***



In this model, the Unified ICM Enterprise parent has PGs connected to a Unified CCE (using System PG) at one site (with a complete installation) and a second site with a TDM ACD that is using a Unified ICM TDM ACD PG. 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 will still be treated locally just as they would be at the TDM ACD site.

# Traditional IVR Integration

There are numerous ways that traditional IVRs can be integrated into a Unified CCE deployment. Determination of the best way will depend on many factors that are discussed in the following sections. The primary consideration is determining how to eliminate or reduce IVR double trunking when transferring the call from the IVR.

## Using PBX Transfer

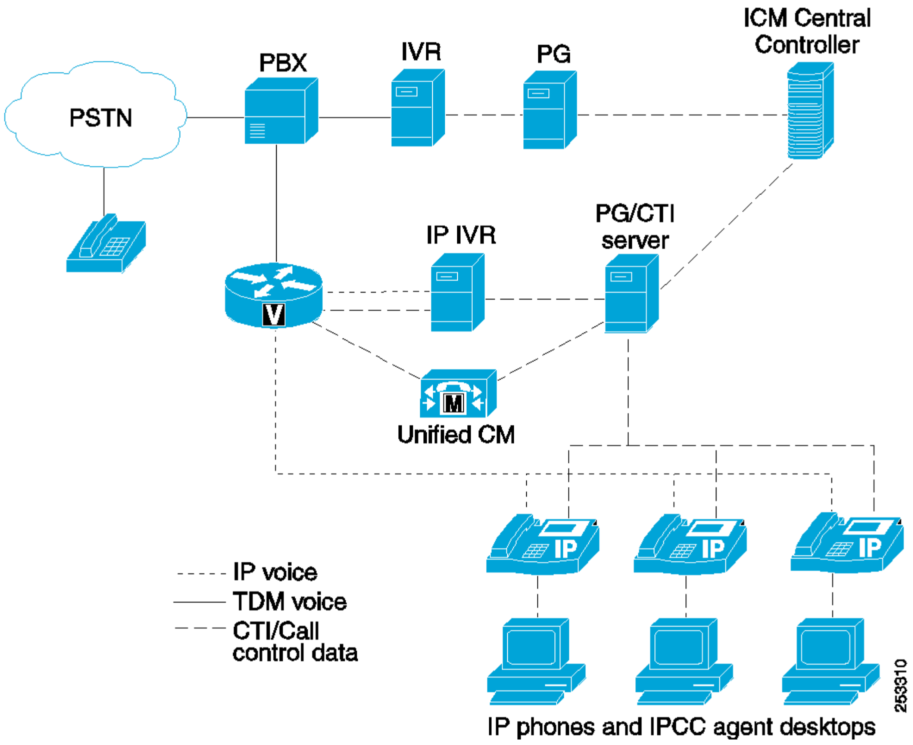
Many call centers have existing traditional IVR applications that they are not prepared to rewrite. In order to preserve these IVR applications and integrate them into a Unified CCE environment, the IVR must have an interface to Unified CCE. (See Figure 38)

There are two versions of the IVR interface to Unified CCE. One is simply a post-routing interface (Call Routing Interface or CRI), which just allows the IVR to send a post-route request with call data to Unified CCE. Unified CCE returns a route response instructing the IVR to transfer the call elsewhere. In this scenario, the traditional IVR invokes a PBX transfer to release its port and transfer the call into the Unified CCE environment. Any call data passed from the IVR is passed by Unified CCE to the agent desktop or Unified IP IVR.

The other IVR interface to Unified CCE is the Service Control Interface (SCI). The SCI enables the IVR to receive queuing instructions from Unified CCE. In the PBX model, the SCI is not required.

Even if the IVR has the SCI interface, it is still preferable to deploy Unified CVP or Unified IP IVR for all call queuing because this prevents any additional use of the traditional IVR 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 38** ***Traditional IVR 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 IVR, putting all of the IVR 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 IVR and it will have reporting only from the time the call arrived at the IVR and the IVR informed Unified CCE of the call.

When the caller opts out of the IVR application, the IVR sends a Post-Route to Unified CCE using the Call Routing Interface (CRI). Because this application does not require calls to queue in the IVR, the CRI is the preferred interface option. Unified CCE will look at the agent states across the system and select the agent to send the call to (by using agent phone number or device target) or translation-route the call to the Unified IP IVR for queuing.

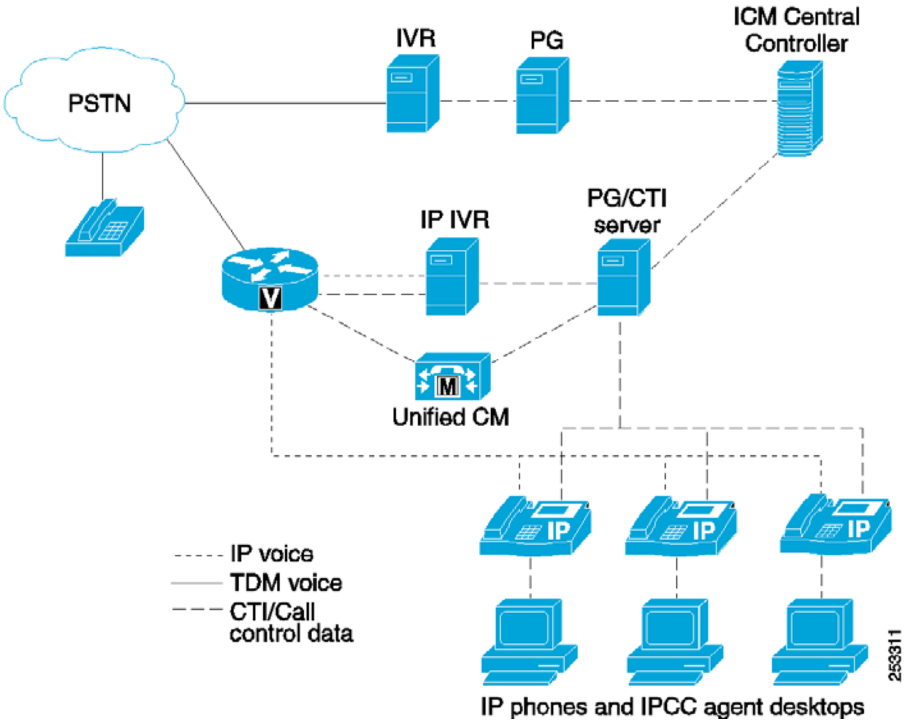
When the call is sent to an agent or into queue, it is hair-pinned in the PBX, coming in from the PSTN on a T1 trunk port and then going out to a voice gateway on a second T1 trunk port in the PBX. This connection is used for the life of the call.

Alternatively, if you want to track the call from its entry at the PBX or if you need 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 IVR 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 IVR. 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 IVR.

## Using PSTN Transfer

This model is very similar to the previous model, except that the IVR invokes a PSTN transfer (instead of a PBX transfer) so that the traditional IVR port can be released. (See Figure 39) Again, the Unified IP IVR is used for all queuing so that any additional occupancy of the traditional IVR ports is not required and so that any double trunking in the IVR is avoided. Any call data collected by the traditional IVR application is passed by Unified CCE to the agent desktop or Unified IP IVR.

**Figure 39** ***Traditional IVR Integration Using PSTN Transfer***



In this model, the TDM IVR is set up as a farm of IVR platforms that have direct PSTN connections for inbound calls. The IVR has a PG connection to Unified CCE which tracks all calls in the system. When a caller opts out of the IVR treatment, the IVR sends a post-route request to Unified CCE which 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 IVR instructs it to send an in-band transfer command using transfer tones (*8 with a destination label in the carrier network). The IVR out-pulses these tones to the service provider with tone generation or plays the tones by using a recorded file.

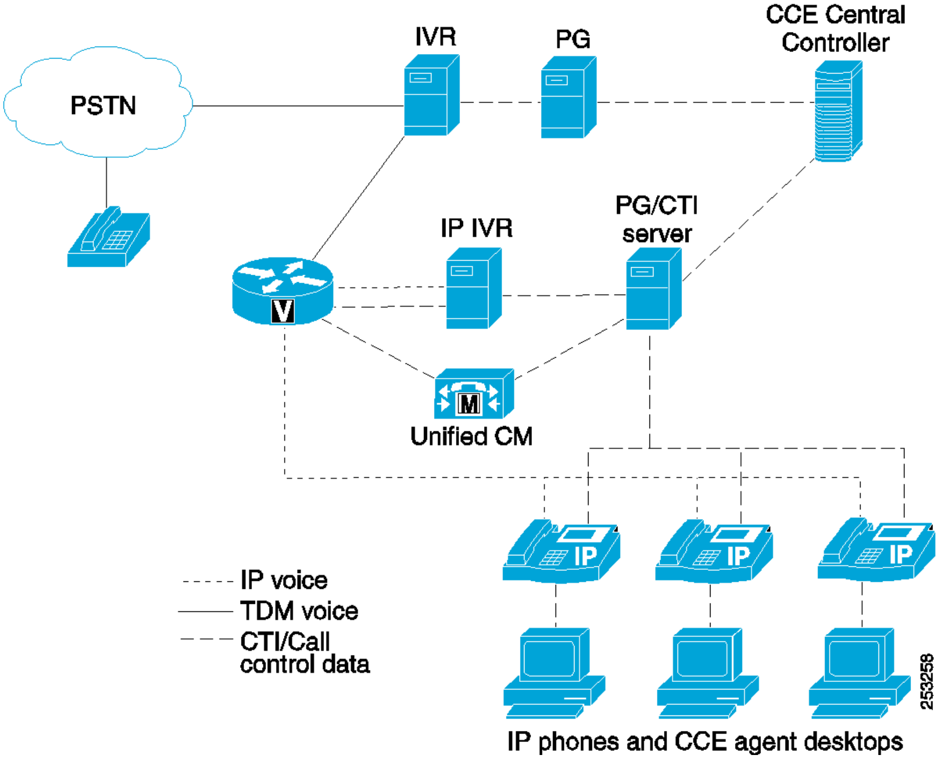
## Using IVR Double Trunking

If your traditional IVR application has a very high success rate where most callers are completely self-served in the traditional IVR and only a very small percentage of callers ever need to be transferred to an agent, then it might be acceptable to double-trunk the calls in the traditional IVR for that small percentage of calls. (See Figure 40)

Unlike the previous model, if the traditional IVR has a Service Control Interface (SCI), then the initial call queuing could be done on the traditional IVR. The reason this is beneficial is that in order to queue the call on the Unified IP IVR, a second traditional IVR port is used to transfer the call to the Unified IP IVR. By performing the initial queuing on the traditional IVR, only one traditional IVR port is used during the initial queuing of the call. However, any subsequent queuing as a result of transfers or RONA treatment must be done on the Unified IP IVR to avoid any double trunking.

If the traditional IVR does not have an SCI interface, then the IVR will just generate a post-route request to Unified CCE to determine where the call is transferred. All queuing in that scenario has to be done on the Unified IP IVR.

**Figure 40** ***Traditional IVR Integration Using IVR Double Trunking***

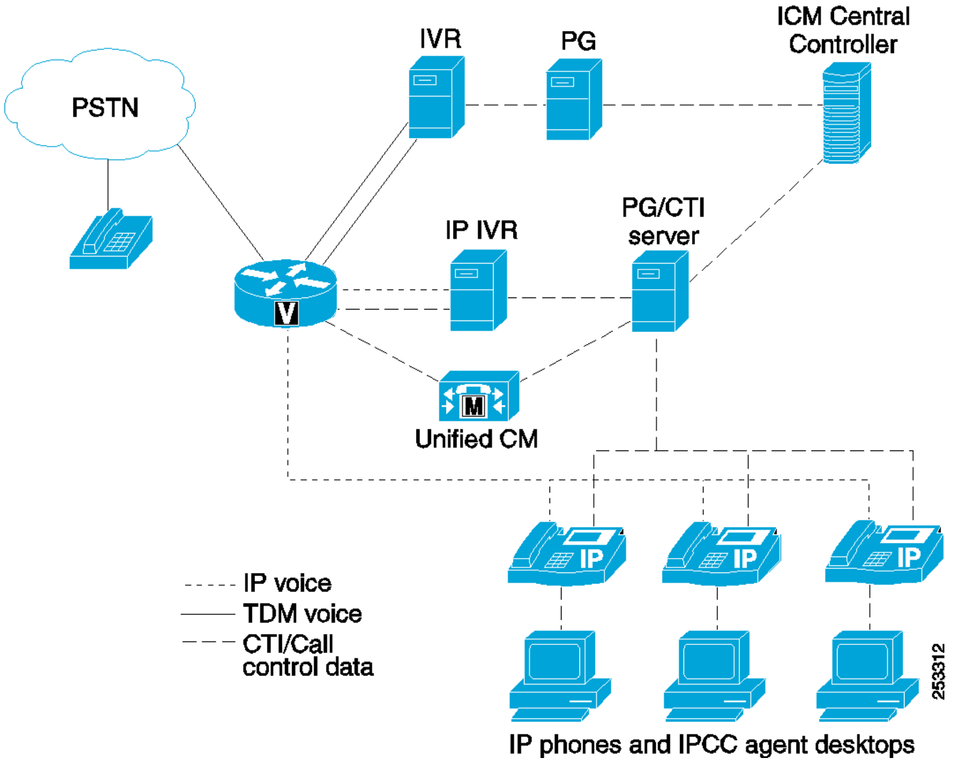


In this model, the TDM IVR is set up as a farm of IVR platforms that have direct PSTN connections for inbound calls. The IVR has a PG connection to Unified CCE which tracks all calls in the system. When a caller opts out of the IVR treatment, the IVR sends a post-route request to Unified CCE which returns a label that will either direct the call to an agent or queue the call locally on the TDM IVR using the Service Control Interface (SCI). The transfer to the agent is done by the TDM IVR selecting a second port to hairpin the call to the voice gateway and to the Unified CCE agent. This takes up two ports for the time the call is at the agent.

## Using Unified CM Transfer and IVR Double Trunking

Over time, it might become desirable to migrate the traditional IVR applications to Unified CVP or Unified IP IVR. However, if a small percentage of traditional IVR applications still exist for very specific scenarios, then the IVR could be connected to a second voice gateway (see Figure 41.) Calls arriving at the voice gateway from the PSTN are routed by Unified CM. Unified CM could route specific DNs to the traditional IVR or let Unified CCE, Unified CVP, or Unified IP IVR determine when to transfer calls to the traditional IVR. If calls in the traditional IVR need to be transferred to a Unified CCE agent, then a second IVR port, trunk, and voice gateway port are used for the duration of the call. Ensure that transfer scenarios do not allow multiple loops to be created because voice quality could suffer.

**Figure 41** ***Traditional IVR Integration Using Unified CM Transfer and IVR Double Trunking***



In this model, the TDM IVR is front-ended by either Unified CVP using the voice gateway or the Unified IP IVR and Unified CM with Unified CCE to determine the location to provide call treatment.

With Unified CVP, calls coming into the voice gateway 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 if the call needs to be sent to the TDM IVR for a specific self-service application or if Unified CVP has the application available for the caller. If the call was sent to the TDM IVR, the TDM IVR sends a route request to Unified CCE when the caller opts out. The reply is not sent back to the TDM IVR but back to Unified CVP as the original routing client. Unified CVP then takes the call leg away from the TDM IVR and transfers it to the Unified CCE agent over the VoIP network or holds it in a queue locally in the voice gateway.

With Unified CM, calls coming into the voice gateway hit a CTI route point for Unified CM to send a route request to Unified CCE to determine the appropriate call treatment device for the caller. If the CTI route point indicated an application that still is on the TDM IVR, Unified CCE instructs Unified CM to transfer the call to the TDM IVR by hairpinning the call using a second T1 port on the voice gateway to connect to the TDM IVR. Unified CCE could also instruct Unified CM to translation-route the call to the Unified IP IVR for call processing or prompting and then make a subsequent transfer to the TDM IVR for further processing. When the caller opts out of the TDM IVR, it sends a post-route request to Unified CCE, and Unified CCE returns a label to the TDM IVR. This label instructs the TDM IVR to transfer the call using a second T1 port on the IVR and to pass the call back to the voice gateway and over to the Unified CCE agent under the Unified CM dial plan.

In the model controlled by Unified CM, calls are initially received by the voice gateway and are hairpinned to the TDM IVR on a second T1 port. When the IVR sends the call back to the Unified CCE agent, it uses a second TDM IVR port and a third port on the voice gateway. All three ports would be tied up on the voice gateway as long as the agent is talking with the caller and both of the TDM IVR ports would be tied up for the duration of this call.

# Unified CCE/CCH: Integration with the Genesys Cisco T-Server

Unified CCE integration with the new Cisco T-Server from Genesys is designed to provide the following new capabilities:

- Allows integration of  the Genesys- Agent Desktop with Unified CCE/H deployments
- Allows Unified CCE to be the “site” ACD with Genesys Enterprise Routing
- Allows Unified CCE to be a backup ACD in the event of a Genesys disconnect, WAN failure, or other failure.

The deployment models listed in this document are supported for the integrated solution. See the *Cisco Unified Contact Center Enterprise (Unified CCE) Software Compatibility Guide*.

The following are noteworthy considerations from an overall deployment perspective:

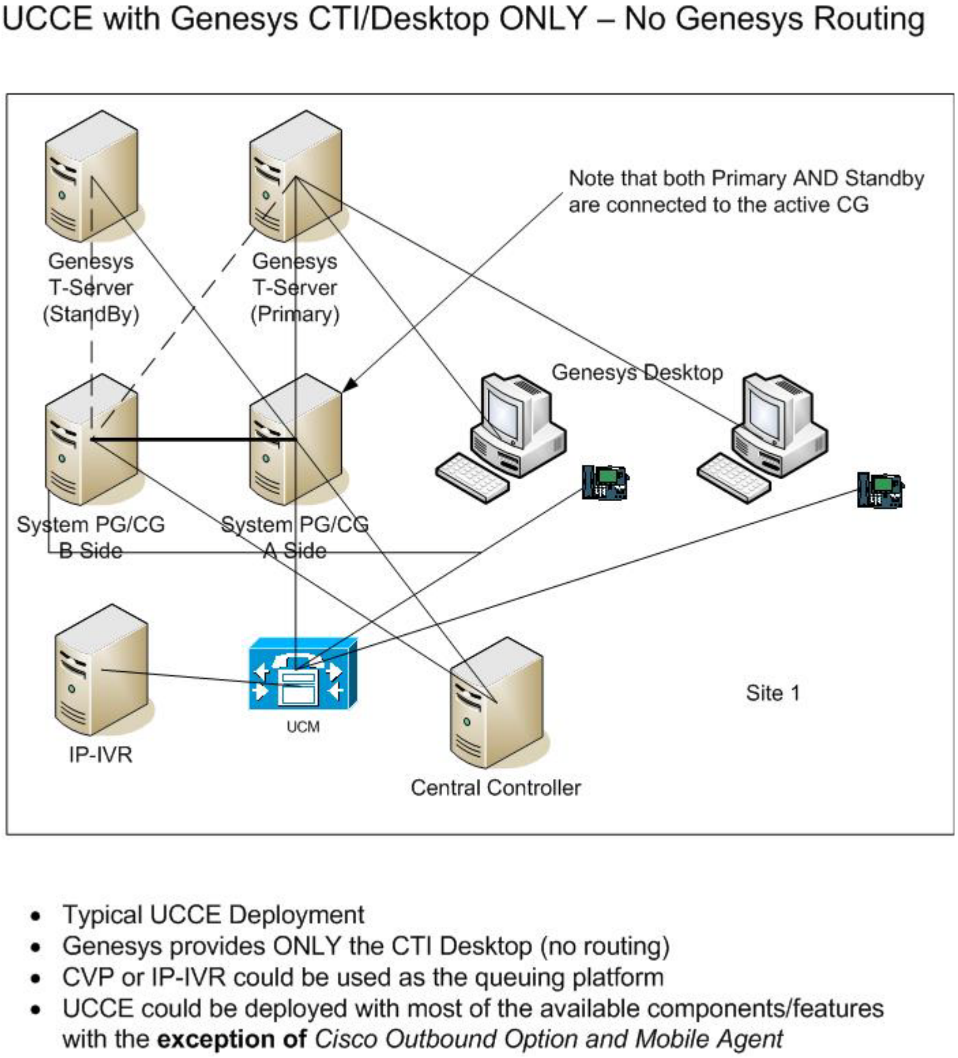
- In any integrated Unified CCE/Genesys deployment, enterprise level routing can be performed either by Cisco Unified CCE or Genesys Universal Routing Server (URS)—but not both
- If Genesys URS is the enterprise routing engine, Unified CCE can only support IP-IVR as the local queuing platform
- CVP as the queuing platform is only supported with Genesys in "CTI/Desktop only mode"
- If Unified CCE is the enterprise routing engine - the Genesys T-Server can only be used in a "CTI/Desktop only mode"
- Genesys Agent Desktop cannot be used for Unified CCE Mobile Agents (only Cisco CTI OS Desktop can be used)
- Cisco Outbound Option is not supported by any Unified CCE/Genesys integrated deployment
- Due to Unified CCE and Genesys having very different reporting architecture and terminology, the reports from Unified CCE and Genesys should not be used for correlation purposes
    - In case of Unified CCE routing (Genesys is CTI only), there is no impact to Unified CCE reporting data
    - In case of Genesys routing, Genesys enterprise reporting is used

- For deployments leveraging Genesys routing, Unified CCE must be configured to provide backup routing for the failure scenario where Unified CCE loses connectivity with the Genesys T-Server
- The Unified CCE configuration data is not synchronized with Genesys (both need to be configured separately)

**Note** Current deployments cannot mix Genesys and CTI-OS desktops. CTI OS may be installed for backup ACD functionality but cannot be operated concurrently.

- All standard Clustering across the WAN (COW) deployments are supported with the added caveat of when the T-Servers are split across the WAN, or in the case of a single T-Server that can link to CTI-Server across the WAN, the Genesys T-Server deployment guidelines should be followed.
- For all diagrams, Genesys Desktops can be exchanged with a CTI-OS desktop.

**Figure 42** ***Unified CCE with Genesys CTI/Desktop Only - No Genesys Routing***

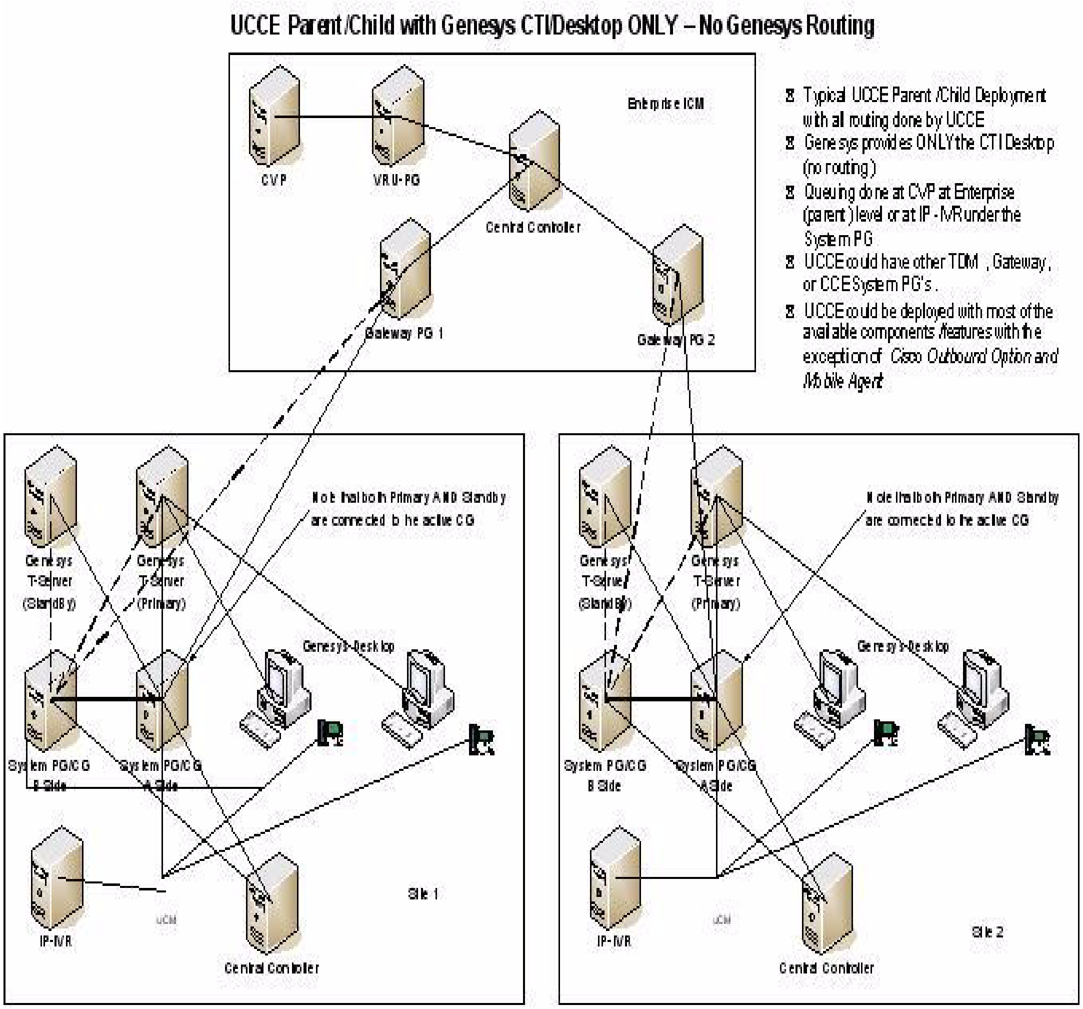


UCCE with Genesys CTI/Desktop ONLY – No Genesys Routing

Genesys T-Server (StandBy)
Genesys T-Server (Primary)
Note that both Primary AND Standby are connected to the active CG
Genesys Desktop
System PG/CG B Side
System PG/CG A Side
IP-IVR
UCM
Central Controller
Site 1

- Typical UCCE Deployment
- Genesys provides ONLY the CTI Desktop (no routing)
- CVP or IP-IVR could be used as the queuing platform
- UCCE could be deployed with most of the available components/features with the **exception of** *Cisco Outbound Option and Mobile Agent*

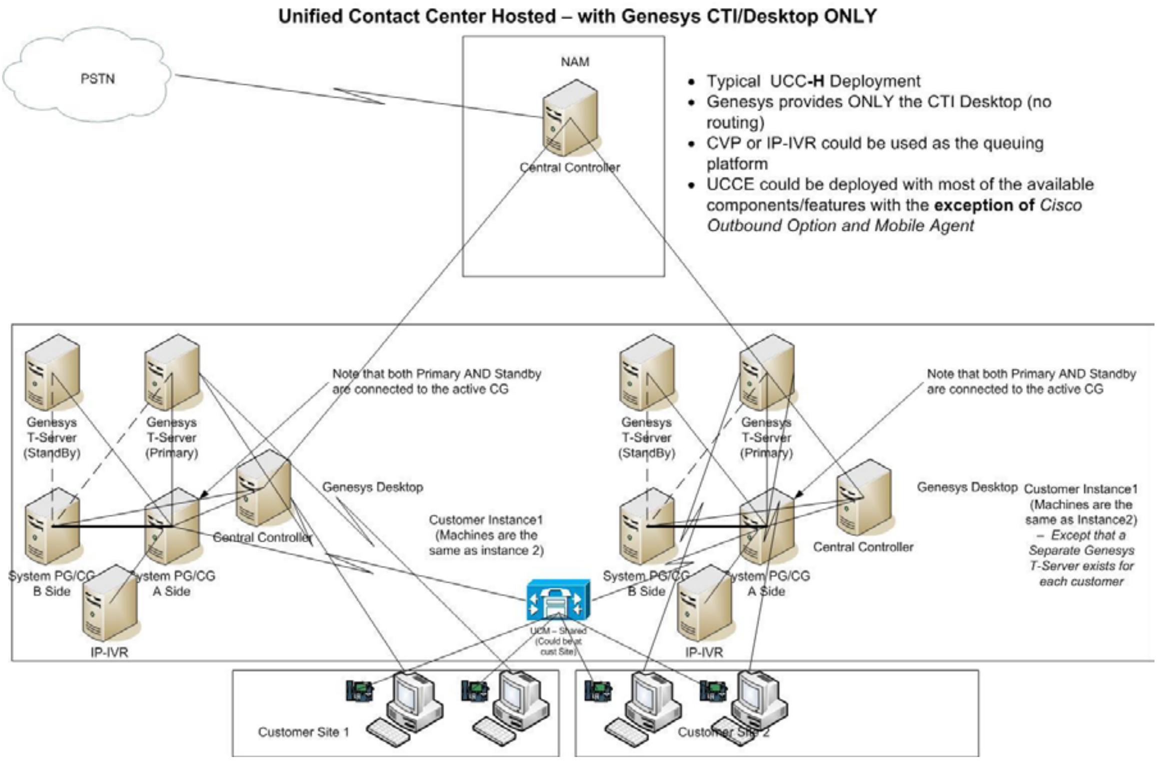
**Figure 43** ***Unified CCE Parent/Child with Genesys CTI/Desktop Only - No Genesys Routing***



UCCE Parent/Child with Genesys CTI/Desktop ONLY – No Genesys Routing

- Typical UCCE Parent /Child Deployment with all routing done by UCCE
- Genesys provides ONLY the CTI Desktop (no routing)
- Queuing done at CVP at Enterprise (parent) level or at IP - IVR under the System PG
- UCCE could have other TDM , Gateway, or CCE System PG's .
- UCCE could be deployed with most of the available components /features with the exception of *Cisco Outbound Option and Mobile Agent*

**Figure 44** ***Unified CCH with Genesys CTI/Desktop Only***

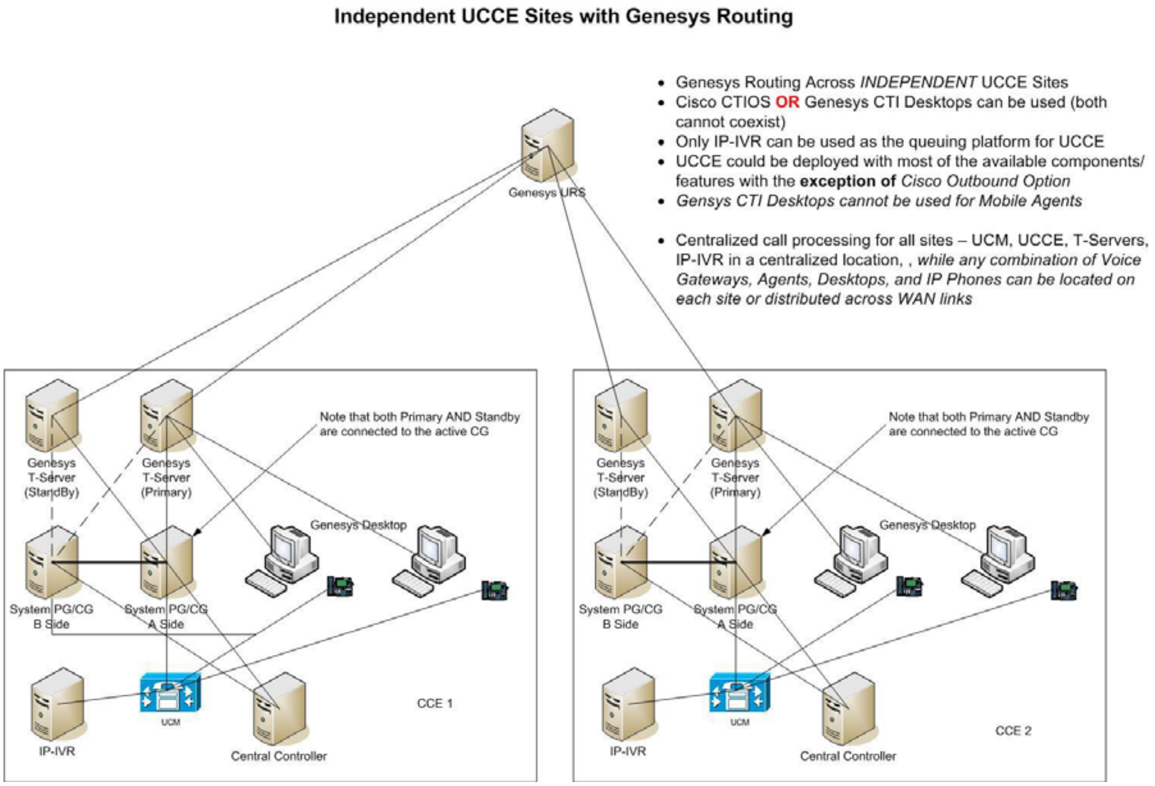


Unified Contact Center Hosted – with Genesys CTI/Desktop ONLY

- Typical UCC-**H** Deployment
- Genesys provides ONLY the CTI Desktop (no routing)
- CVP or IP-IVR could be used as the queuing platform
- UCCE could be deployed with most of the available components/features with the **exception of** *Cisco Outbound Option and Mobile Agent*

PSTN
NAM
Central Controller
Genesys T-Server (StandBy)
Genesys T-Server (Primary)
Note that both Primary AND Standby are connected to the active CG
Genesys Desktop
Central Controller
Customer Instance1 (Machines are the same as instance 2)
System PG/CG B Side
System PG/CG A Side
IP-IVR
UCM – Shared (Could be at cust Site)
Genesys T-Server (StandBy)
Genesys T-Server (Primary)
Note that both Primary AND Standby are connected to the active CG
Genesys Desktop
Customer Instance1 (Machines are the same as Instance2) – *Except that a Separate Genesys T-Server exists for each customer*
Central Controller
System PG/CG B Side
System PG/CG A Side
IP-IVR
Customer Site 1
Customer Site 2

**Figure 45** ***Independent Unified CCE Sites with Genesys Routing***



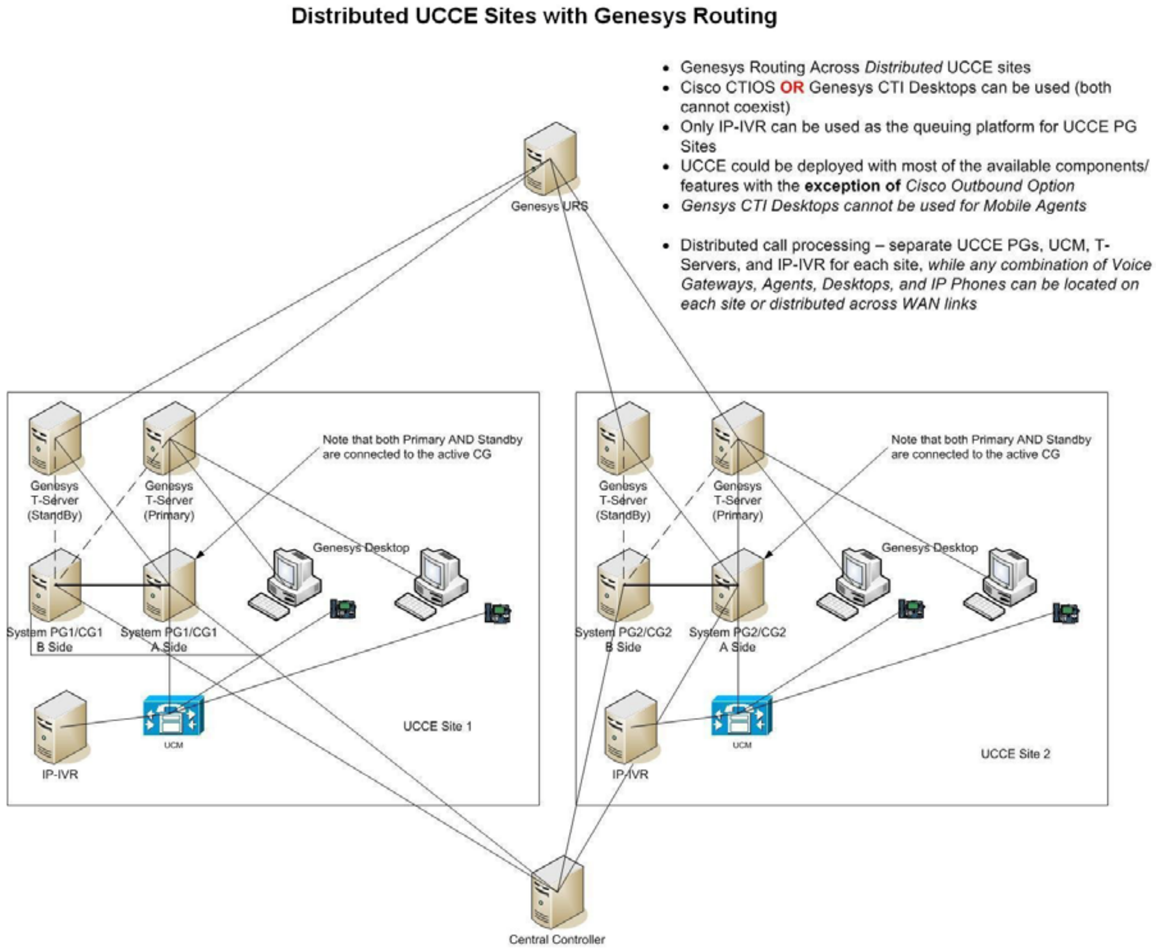
Independent UCCE Sites with Genesys Routing

- Genesys Routing Across *INDEPENDENT* UCCE Sites
- Cisco CTIOS **OR** Genesys CTI Desktops can be used (both cannot coexist)
- Only IP-IVR can be used as the queuing platform for UCCE
- UCCE could be deployed with most of the available components/ features with the **exception of** *Cisco Outbound Option*
- *Gensys CTI Desktops cannot be used for Mobile Agents*

- Centralized call processing for all sites – UCM, UCCE, T-Servers, IP-IVR in a centralized location, , *while any combination of Voice Gateways, Agents, Desktops, and IP Phones can be located on each site or distributed across WAN links*

Genesys URS
Genesys T-Server (StandBy)
Genesys T-Server (Primary)
Note that both Primary AND Standby are connected to the active CG
Genesys Desktop
System PG/CG B Side
System PG/CG A Side
IP-IVR
UCM
Central Controller
CCE 1
Genesys T-Server (StandBy)
Genesys T-Server (Primary)
Note that both Primary AND Standby are connected to the active CG
Genesys Desktop
System PG/CG B Side
System PG/CG A Side
IP-IVR
UCM
Central Controller
CCE 2

**Figure 46** ***Distributed Unified CCE Sites with Genesys Routing***



Distributed UCCE Sites with Genesys Routing

- Genesys Routing Across *Distributed* UCCE sites
- Cisco CTIOS **OR** Genesys CTI Desktops can be used (both cannot coexist)
- Only IP-IVR can be used as the queuing platform for UCCE PG Sites
- UCCE could be deployed with most of the available components/ features with the **exception of** *Cisco Outbound Option*
- *Gensys CTI Desktops cannot be used for Mobile Agents*
- Distributed call processing – separate UCCE PGs, UCM, T-Servers, and IP-IVR for each site, *while any combination of Voice Gateways, Agents, Desktops, and IP Phones can be located on each site or distributed across WAN links*

Note that both Primary AND Standby are connected to the active CG

Note that both Primary AND Standby are connected to the active CG

CHAPTER 3

# Design Considerations for High Availability

**Note** Many of the design considerations and illustrations throughout this chapter have been revised and updated. Review the entire chapter before designing a Unified CCE system.

## Designing for High Availability

Cisco Unified CCE is a distributed solution that uses numerous hardware and software components and it is important to design each system in a way that eliminates any single point of failure – or that at least addresses potential failures in a way that will impact the fewest resources in the contact center. The type and number of resources impacted will depend on how stringent your requirements are, the budget for fault tolerance, and which design characteristics you choose for the various Unified CCE components (including the network infrastructure). A good Unified CCE design will be tolerant of most failures (defined later in this section); but not all failures can be made transparent.

- Cisco Unified CCE is a solution designed for mission-critical contact centers. The successful design of any Unified CCE deployment requires a team with experience in data and voice internetworking, system administration, and Unified CCE application design and configuration.
- Simplex deployments are allowed for demo, laboratory, and non-production deployments. However, all production deployments *must* be deployed with redundancy for the core Unified CCE components (Call Routers, Loggers, PGs, and pre-routing gateways).

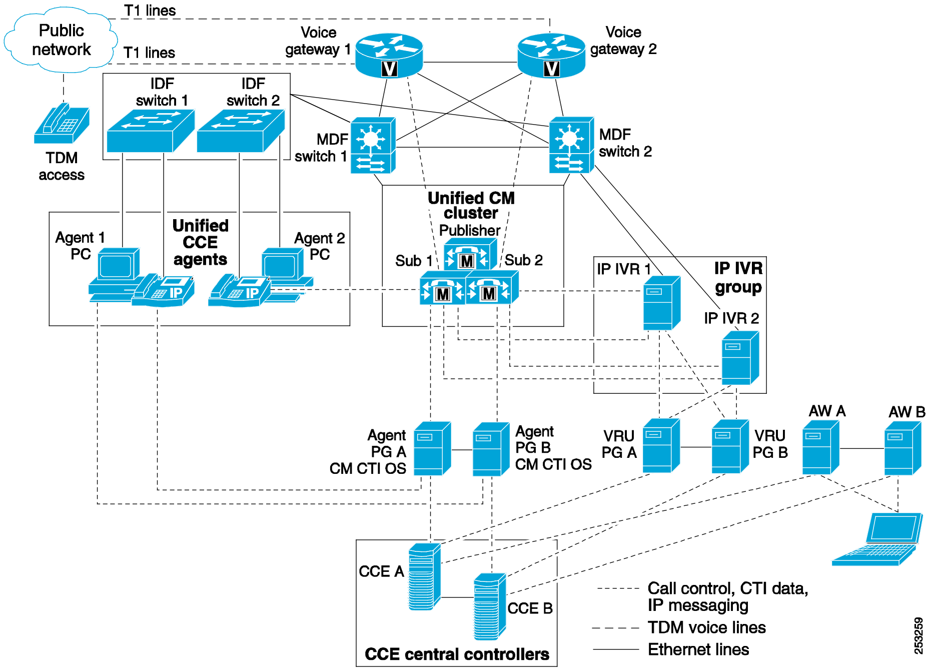
Before implementing Unified CCE, use careful preparation and design planning to avoid costly upgrades or maintenance later in the deployment cycle. Always design for the worst possible failure scenario with future scalability in mind for all Unified CCE sites.

In summary, plan ahead and follow all the design guidelines presented in this guide and in the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

For assistance in planning and designing your Unified CCE solution, consult your Cisco or certified Partner Systems Engineer (SE).

Figure 47 shows a high-level design for a fault-tolerant Unified CCE single-site deployment.

**Figure 47** *Unified CCE Single-Site Design for High Availability*



In Figure 47, each component in the Unified CCE solution is duplicated with a redundant or duplex component, with the exception of the intermediate distribution frame (IDF) switches for the Unified CCE agents and their phones. The IDF switches do not interconnect with each other but only with the main distribution frame (MDF) switches because it is better to distribute the agents among different IDF switches for load balancing and geographic separation (such as different building floors or different cities). If an IDF switch fails, route all calls to other available agents in a separate IDF switch or to a Unified IP IVR queue. Follow the design guidelines for a single-site deployment as documented in the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

If designed correctly for high availability and redundancy, a Unified CCE system can lose half of its core component systems or servers and still be operational. With this type of design, no matter what happens in the Unified CCE system, calls can still be handled in one of the following ways:

- Routed and answered by an available Unified CCE agent using an IP phone or desktop soft phone
- Sent to an available Unified IP IVR or Unified CVP port or session
- Answered by the Cisco Unified Communications Manager AutoAttendant or Hunt Group
- Prompted by a Unified IP IVR or Unified CVP announcement that the call center is currently experiencing technical difficulties and to call back later
- Rerouted to another site with available agents or resources to handle the call

The components in Figure 47 can be rearranged to form two connected Unified CCE sites, as illustrated in Figure 48.

**Figure 48** *Unified CCE Single-Site Redundancy*

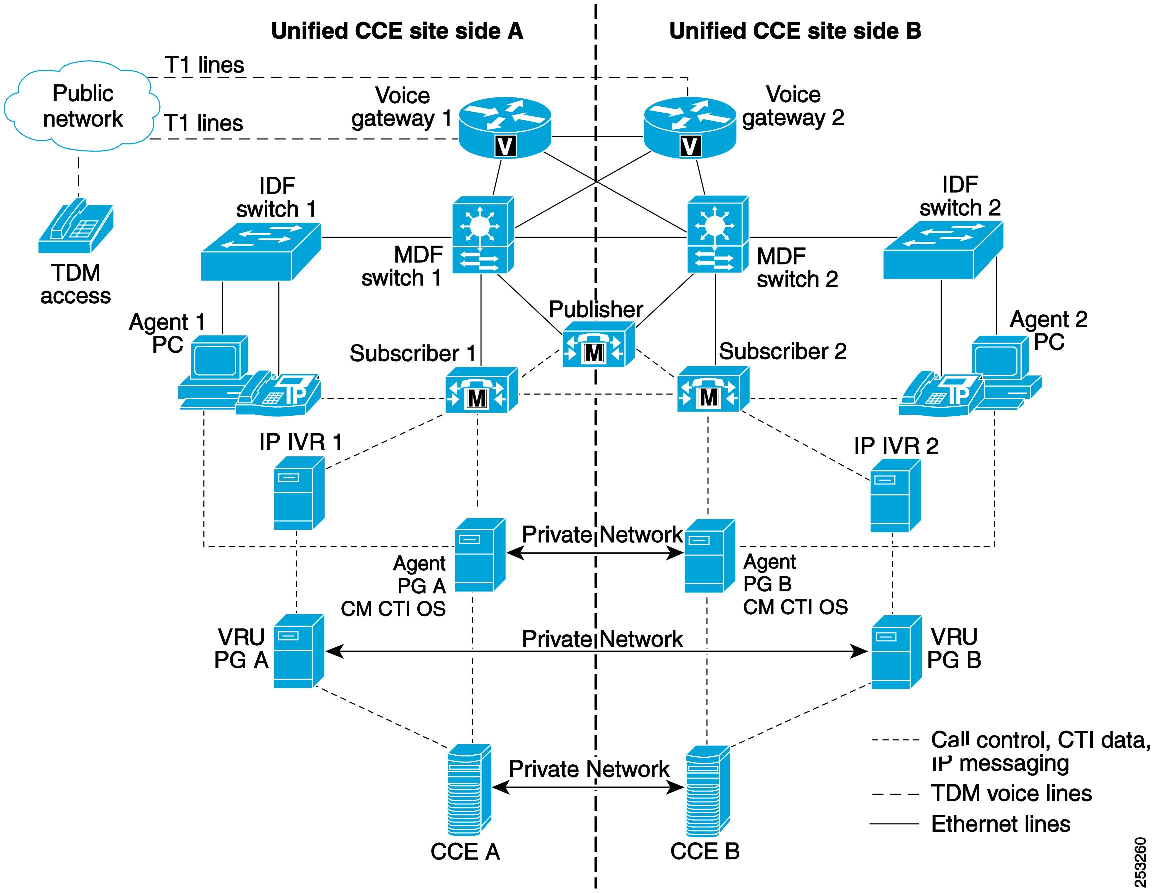


Figure 48 emphasizes the redundancy of the single site design in Figure 47. Side A and Side B are basically mirror images of each other. In fact, one of the main Unified CCE features to enhance high availability is its capability to add redundant or duplex components that are designed to automatically fail-over and recover without any manual intervention. Core system components with redundant components are interconnected to provide failure detection of the redundant system component with the use of TCP keep-alive messages generated every 100 ms over a separate Private Network path. The fault-tolerant design and failure detection and recovery method is described later in this chapter.

Other components in the solution use other types of redundancy strategies. For example, Cisco Unified Communications Manager (Unified CM) uses a cluster design that provides IP phones and devices with multiple Unified CM subscribers (servers) to register with when the primary server fails. The devices automatically reconnect to the primary server when it is restored.

The following sections use Figure 47 as the model design to discuss issues and features to consider when designing Unified CCE for high availability. These sections use a bottom-up model (from a network model perspective, starting with the physical layer first) that divides the design into segments that can be deployed in separate stages.

Use only duplex (redundant) Unified CM, Unified IP IVR or Unified CVP, and Unified CCE components for all Unified CCE deployments. This chapter assumes that the Unified CCE failover feature is a critical requirement for all deployments; therefore it presents only deployments that use a redundant configuration with each Unified CM cluster having at least one publisher and one subscriber. Additionally, where possible, deploy Unified CCE so that no devices, call processing, or CTI Manager Services are running on the Unified CM publisher.

# Data Network Design Considerations

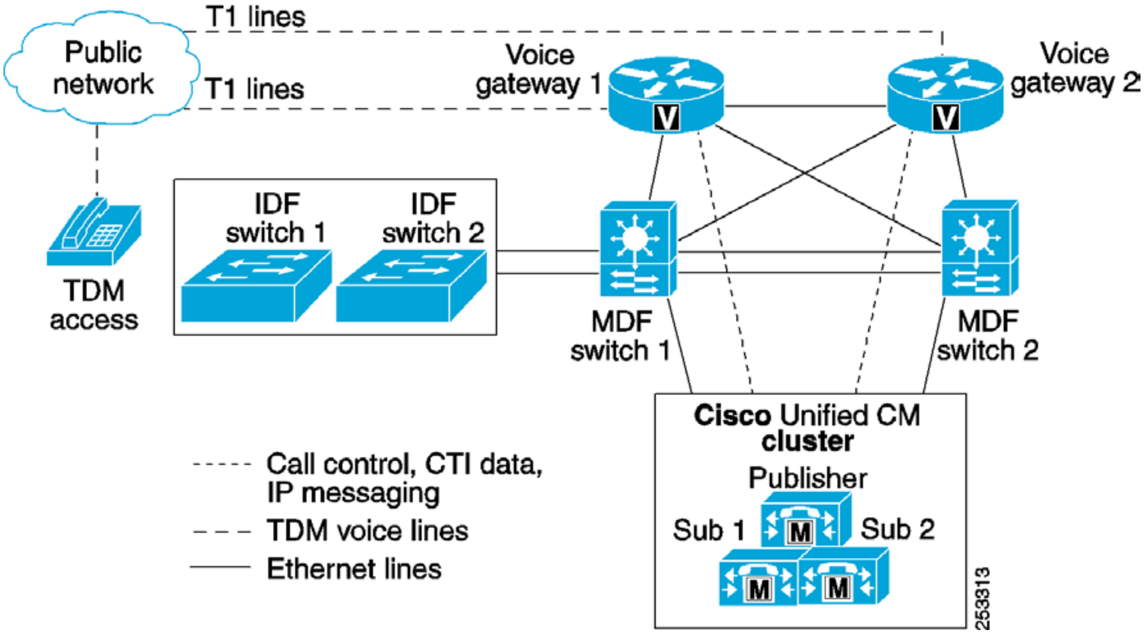
The Unified CCE design shown in Figure 49 illustrates the voice call path from the PSTN (public switched telephone network) at the ingress voice gateway to the call reaching a Unified CCE agent. The network infrastructure in the design supports the Unified CCE environment for data and voice traffic. The network, including the PSTN, is the foundation for the Unified CCE solution. If the network is poorly designed to handle failures, then everything in the contact center is prone to failure because all the servers and network devices depend on the network for highly available communications. The data and voice networks must be a primary part of your solution design and must be addressed in the early stages for all Unified CCE implementations.

Set the NIC card and Ethernet switch to 100 MB full duplex for 10/100 links, or set them to auto-negotiate for gigabit links for all the Unified CCE core component servers.

In addition, the choice of voice gateways for a deployment is critical because some protocols offer more call resiliency than others. This chapter provides high-level information about how to configure the voice gateways for high availability with the Unified CCE solution.

For more information about voice gateways and voice networks in general, see the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

**Figure 49** ***High Availability in a Network with Two Voice Gateways and One Unified CM Cluster***



The use of multiple voice gateways avoids the problem of a single gateway failure causing blockage of all inbound and outgoing calls. In a configuration with two voice gateways and one Unified CM cluster, register each gateway with a different primary Unified CM subscriber to spread the workload across the subscribers in the cluster. Configure each gateway to use another subscriber as a backup in case its primary fails. For details on setting up Unified CM for redundant service and redundancy groups related to call processing, see the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

With Cisco IOS voice gateways using H.323 or SIP, additional call processing is available by using TCL scripts and additional dial peers if the gateway is unable to reach its Unified CM for call control or call processing instructions. MGCP gateways do not have this built-in functionality and the trunks that are terminated in these gateways require backup routing or "roll-over service" from the PSTN carrier or service provider to reroute the trunk on failure or no-answer to another gateway or location.

As for sizing the gateway's trunk capacity, it is a good idea to account for failover of the gateways by building in enough excess capacity to handle the maximum busy hour call attempts (BHCA) if one or more voice gateways fail. During the design phase, first decide how many simultaneous voice gateway failures are possible and acceptable for the site. Based on this requirement, the number of voice gateways used, and the distribution of trunks across those voice gateways; you can determine the total number of trunks required for normal and disaster modes of operation. The more you distribute the trunks over multiple voice gateways, the fewer trunks you will need in a failure mode. However, using more voice gateways or carrier PSTN trunks will increase the cost of the solution, so compare the cost with the benefits of being able to service calls in a gateway failure. The form-factor of the gateway is also a consideration.

As an example, assume a contact center has a maximum BHCA that results in the need for four T1 lines and the company has a requirement for no call blockage in the event of a single component (voice gateway) failure. If two voice gateways are deployed, then provision each voice gateway with four T1 lines (a total of eight). If three voice gateways are deployed, then two T1 lines per voice gateway (a total of six) would be enough to achieve the same level of redundancy. If five voice gateways are deployed, then one T1 per voice gateway (a total of five) would be enough to achieve the same level of redundancy. Thus, you can reduce the number of T1 lines required by adding more voice gateways and spreading the risk over multiple physical devices.

The operational cost savings of fewer T1 lines might be greater than the one-time capital cost of the additional voice gateways. In addition to the recurring operational costs of the T1 lines, also factor in the carrier charges (like the typical one-time installation cost) of the T1 lines to ensure that your design accounts for the most cost-effective solution. Every installation has different availability requirements and cost metrics, but using multiple voice gateways is often more cost-effective. It is a worthwhile design practice to perform this cost comparison.

After you have determined the number of trunks needed, the PSTN service provider has to configure them so that calls can be terminated onto trunks connected to all of the voice gateways (or at least more than one voice gateway). From the PSTN perspective, if the trunks going to the multiple voice gateways are configured as a single large trunk group, then all calls will automatically be routed to the surviving voice gateways when one voice gateway fails. If all of the trunks are not grouped into a single trunk group within the PSTN, then you must ensure that PSTN rerouting or overflow routing to the other trunk groups is configured for all dialed numbers.

If a voice gateway with a digital interface (T1 or E1) fails, then the PSTN automatically stops sending calls to that voice gateway because carrier level signaling on the digital circuit has dropped. The loss of carrier level signaling on a digital circuit causes the PSTN to busy-out all trunks, thereby preventing the PSTN from routing new calls to the failed voice gateway. When the failed voice gateway comes back on-line and the circuits are back in operation, the PSTN automatically starts delivering calls to that voice gateway again.

With Cisco IOS voice gateways using H.323 or SIP, it is possible for the voice gateway itself to be operational but for its communication paths to the Unified CM servers to be severed (for example, a failed Ethernet connection). If this situation occurs, you can use the **busyout-monitor interface** command to monitor the Ethernet interfaces on a voice gateway. To place a voice port into a busyout monitor state, use the **busyout-monitor interface** *voice-port* configuration command. To remove the busyout-monitor state on the voice port, use the **no** form of this command.

As noted previously, these gateways also provide additional processing options if the call control interface is not available from Unified CM to reroute the calls to another site or dialed number or to play a locally stored .wav file to the caller and end the call.

With MGCP-controlled voice gateways, when the voice gateway interface to Unified CM fails, the gateway will look for secondary and tertiary Unified CM subscribers from the redundancy group. The MGCP gateway will automatically fail-over to the other subscribers in the group and periodically check the health of each, marking it as available once it comes back on-line. The gateway will then fail-back to the primary subscriber when all calls are idle or after 24 hours (whichever comes first).

If no subscribers are available, the voice gateway automatically busies-out all its trunks. This action prevents new calls from being routed to this voice gateway from the PSTN. When the voice gateway interface to Unified CM homes to the backup subscriber, the trunks are automatically idled and the PSTN begins routing calls to this voice gateway again (assuming the PSTN has not permanently busied-out those trunks). The design practice is to spread the gateways across the Unified CM call processing servers in the cluster to limit the risk of losing all the gateway calls in a call center if the primary subscriber that has all the gateways registered to it fails.

Voice gateways that are used with the Cisco Unified Survivable Remote Site Telephony (SRST) option for Unified CM follow a similar failover process. If the gateway is cut off from the Unified CM that is controlling it, the gateway will fail-over into SRST mode, which drops all voice calls and resets the gateway into SRST mode. Phones re-home to the local SRST gateway for call control and calls will be processed locally and directed to local phones.

While running in SRST mode, it is assumed that the agents also have no CTI connection from their desktops. They will be seen as not ready within the Unified CCE routing application and no calls will be sent to these agents by Unified CCE. When the data connection is re-established to the gateway at the site, the Unified CM will take control of the gateway and phones again allowing the agents to be reconnected to the Unified CCE.

# Unified CM and CTI Manager Design Considerations

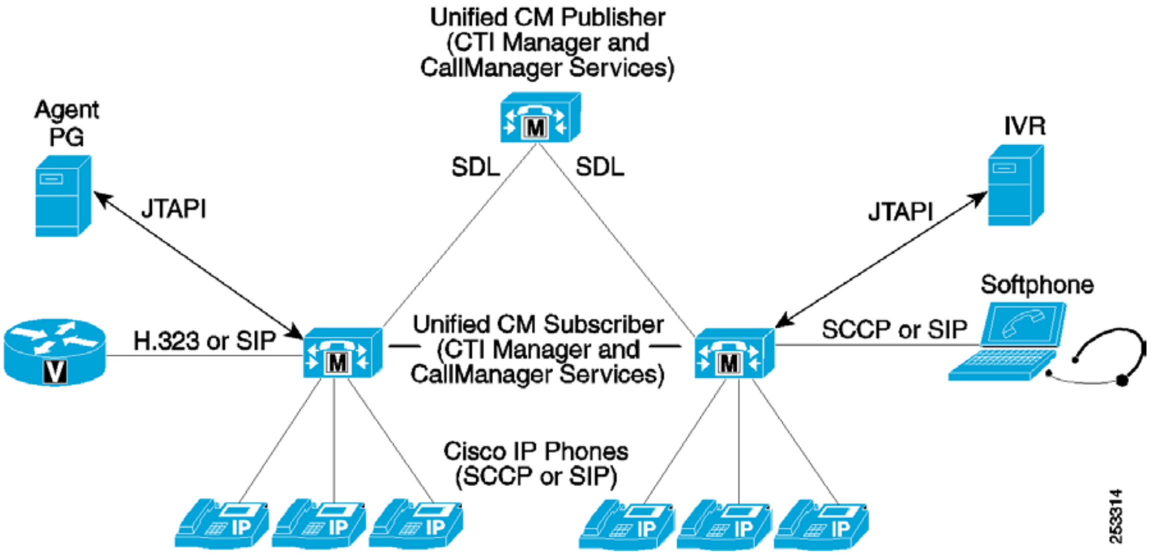
Cisco Unified CM uses CTI Manager, a service that acts as an application broker and abstracts the physical binding of the application to a particular Unified CM server, to handle all its CTI resources. (See the *Cisco Unified Communications Solution Reference Network Design (SRND)* guide for further details about the architecture of the CTI Manager.) The CTI Manager and CallManager are two separate services running on a Unified CM server. Some other services running on a Unified CM server include TFTP, Cisco Messaging Interface, and Real-time Information Server (RIS) data collector services.

The main function of the CTI Manager is to accept messages from external CTI applications and send them to the appropriate resource in the Unified CM cluster. The CTI Manager uses the Cisco JTAPI link to communicate with the applications. It acts like a JTAPI messaging router. The JTAPI client library in Cisco Unified CM connects to the CTI Manager instead of connecting directly to the CallManager service. In addition, there can be multiple CTI Manager services running on different Unified CM servers in the cluster that are aware of each other (by way of the CallManager service, which is explained later in this section). The CTI Manager uses the same Signal Distribution Layer (SDL) signaling mechanism that the Unified CM services in the cluster use to communicate with each other. However, the CTI Manager does not directly communicate with the other CTI Managers in its cluster. (This is also explained later in detail.)

The main function of the CallManager service is to register and monitor all the Cisco Unified Communications devices. It basically acts as a switch for all the Cisco Unified Communications resources and devices in the system while the CTI Manager service acts as a router for all the CTI application requests for the system devices. Some of the devices that can be controlled by JTAPI that register with the CallManager service include the IP phones, CTI ports, and CTI route points.

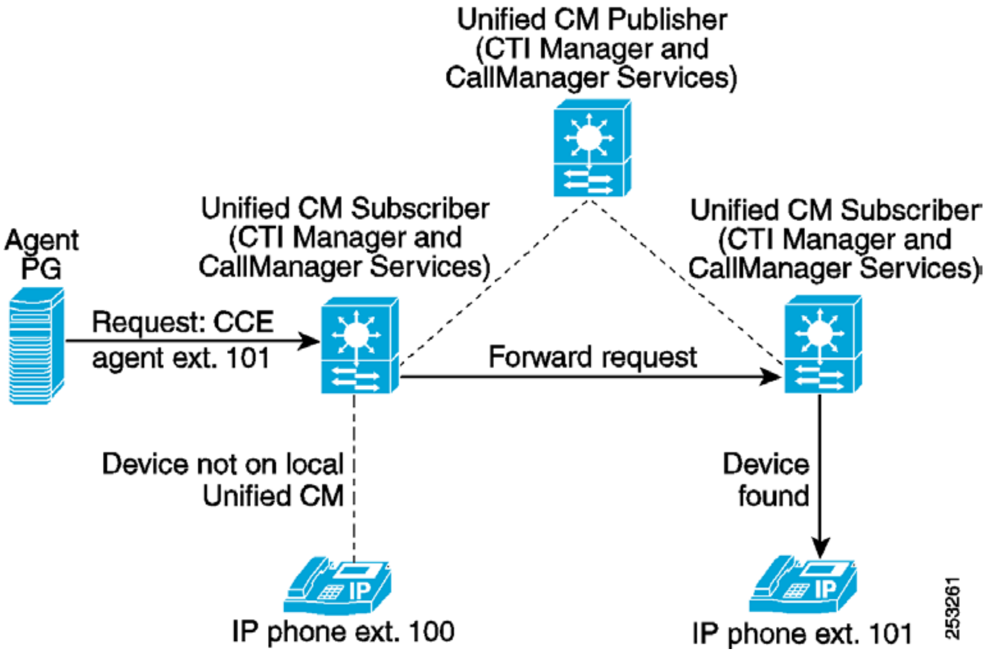
Figure 50 illustrates some of the functions of Unified CM and the CTI Manager.

**Figure 50** ***Functions of the CallManager and CTI Manager Services***



The servers in a Unified CM cluster communicate with each other using the Signal Distribution Layer (SDL) service. SDL signaling is used only by the CallManager service to talk to the other CallManager services to make sure everything is in sync within the Unified CM cluster. The CTI Managers in the cluster are completely independent and do not establish a direct connection with each other. CTI Managers route only the external CTI application requests to the appropriate devices serviced by the local CallManager service on this subscriber. If the device is not resident on its local Unified CM subscriber, then the CallManager service forwards the application request to the appropriate Unified CM in the cluster. Figure 51 shows the flow of a device request to another Unified CM in the cluster.

**Figure 51** ***CTI Manager Device Request to a Remote Unified CM***



Although it might be tempting to register all of the Unified CCE devices to a single subscriber in the cluster and point the Peripheral Gateway (PG) to that server, this configuration would put a high load on that subscriber. If the PG were to fail in this case, the duplex PG would connect to a different subscriber and all the CTI Manager messaging would have to be routed across the cluster to the original subscriber. It is important to distribute devices and CTI applications appropriately across all the call processing nodes in the Unified CM cluster to balance the CTI traffic and limit possible failover conditions.

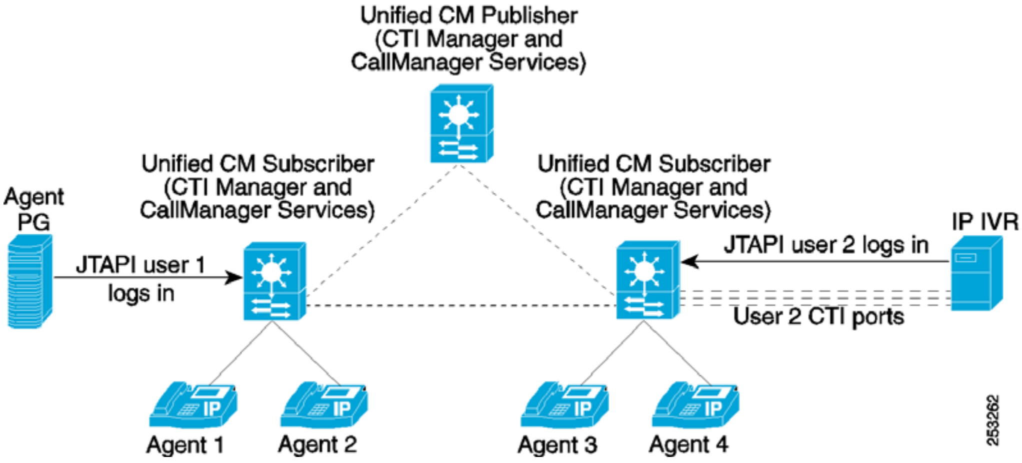
The external CTI applications use a CTI-enabled user account in Unified CM. They log into the CTI Manager service to establish a connection and assume control of the Unified CM devices associated to this specific CTI-enabled user account, typically referred to as the JTAPI user or PG user. In addition, given that the CTI Managers are independent from each other, any CTI application can connect to any CTI Manager in the cluster to perform its requests. However, because the CTI Managers are independent, one CTI Manager cannot pass the CTI application to another CTI Manager upon failure. If the first CTI Manager fails, the external CTI application must implement the failover mechanism to connect to another CTI Manager in the cluster.

For example, the Agent PG handles failover for the CTI Manager by using its duplex servers, sides A and B, each of which is pointed to a different subscriber in the cluster and by using the CTI Manager on those subscribers. It is important to note these connections from the PG are managed in hot standby mode which means that only one side of the PG is active at any given time and is connected to the CTI Manager on the subscriber.

The PG processes are designed to prevent both sides from trying to be active at the same time to reduce the impact of the CTI application on Unified CM. Additionally, both of the duplex PG servers (Side A and Side B) use the same CTI-enabled JTAPI or PG user to log into the CTI Manager applications. However, only one Unified CM PG side allows the JTAPI user to register and monitor the user devices to conserve system resources in the Unified CM cluster. The other side of the Unified CM PG stays in hot-standby mode waiting to connect, log in, register, and be activated upon failure of the active side.

Figure 52 shows two external CTI applications using the CTI Manager, the Agent PG, and the Unified IP IVR. The Unified CM PG logs into the CTI Manager using the JTAPI account User 1 while the Unified IP IVR uses account User 2. Each external application uses its own specific JTAPI user account and will have different devices registered and monitored by that user. For example, the Unified CM PG (User 1) monitors all four agent phones and the inbound CTI Route Points, while the Unified IP IVR (User 2) monitors its CTI Ports and the CTI Route Points used for its JTAPI Triggers. Although multiple applications could monitor the same devices, avoid this method because it can cause race conditions between the applications trying to take control of the same physical device.

**Figure 52** *CTI Application Device Registration*


Unified CM Publisher (CTI Manager and CallManager Services)
Unified CM Subscriber (CTI Manager and CallManager Services)
Unified CM Subscriber (CTI Manager and CallManager Services)
Agent PG
JTAPI user 1 logs in
IP IVR
JTAPI user 2 logs in
User 2 CTI ports
IP
Agent 1
Agent 2
Agent 3
Agent 4
253262

Unified CM CTI applications also add to the device weights on the subscribers, adding memory objects used to monitor registered devices. These monitors are registered on the subscriber that has the connection to the external application. It is a good design practice to distribute these applications to CTI Manager registrations across multiple subscribers to avoid overloading a single subscriber with all of the monitored object tracking.

Perform the design of Unified CM and CTI Manager as the second design stage, right after the network design stage. Perform deployment in this same order. The reason for this order is that the Cisco Unified Communications infrastructure must be in place to dial and receive calls using its devices before you can deploy any telephony applications.

Before moving to the next design stage, make sure that a PSTN phone can call an IP phone and that this same IP phone can dial out to a PSTN phone with all the call survivability capabilities considered for treating these calls. Also keep in mind that the Unified CM cluster design is paramount to the Unified CCE system, and any server failure in a cluster will take down two services (CTI Manager and CallManager), thereby adding an extra load to the remaining servers in the cluster.

## Configuring the Unified CCE Peripheral Gateway for CTI Manager Redundancy

To enable Unified CM support for CTI Manager failover in a duplex Unified CCE Peripheral Gateway model, perform the following steps:

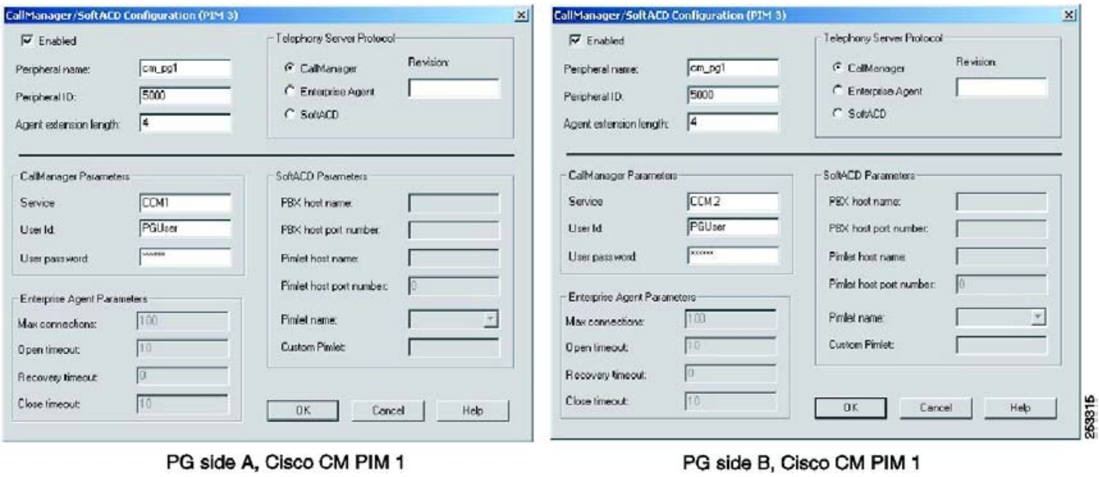
Step 1. Create a Unified CM redundancy group and add subscribers to the group. (Do not use Publishers and TFTP servers for call processing, device registration, or CTI Manager functions.)

Step 2. Designate two CTI Managers on different subscribers to be used for each side of the duplex Peripheral Gateway (PG), one for PG Side A and one for PG Side B.

Step 3. Assign one of the CTI Managers to be the JTAPI service of the Unified CM PG Side A. (See Figure 53.) Note that the setup panel on the left is for Side A of the Peripheral Gateway. It points to the CCM1 subscriber and uses the PGUser CTI-enabled user account on the Unified CM cluster.

Step 4. Assign the second CTI Manager to be the JTAPI service of the Unified CM PG Side B. (See Figure 53) Note that the setup panel on the right is for Side B of the Peripheral Gateway. It points to the CCM2 subscriber and uses the same PGUser CTI-enabled user account on the Unified CM cluster. Both sides of the duplex PG pair must use the same JTAPI user in order to monitor the same devices from either side of the PG pair.

**Figure 53** ***Assigning CTI Managers for PG Sides A and B***



CallManager/SoftACD Configuration (PIM 3)

☑ Enabled

Peripheral name: cm_pg1

Peripheral ID: 5000

Agent extension length: 4

Telephony Server Protocol: ● CallManager ○ Enterprise Agent ○ SoftACD

CallManager Parameters

Service: CCM1

User Id: PGUser

User password: ******

Enterprise Agent Parameters

Max connections: 100

Open timeout: 10

Recovery timeout: 0

Close timeout: 10

SoftACD Parameters

Pimlet host port number: 0

PG side A, Cisco CM PIM 1

CallManager/SoftACD Configuration (PIM 3)

☑ Enabled

Peripheral name: cm_pg1

Peripheral ID: 5000

Agent extension length: 4

Telephony Server Protocol: ● CallManager ○ Enterprise Agent ○ SoftACD

CallManager Parameters

Service: CCM 2

User Id: PGUser

User password: ******

Enterprise Agent Parameters

Max connections: 100

Open timeout: 10

Recovery timeout: 0

Close timeout: 10

SoftACD Parameters

Pimlet host port number: 0

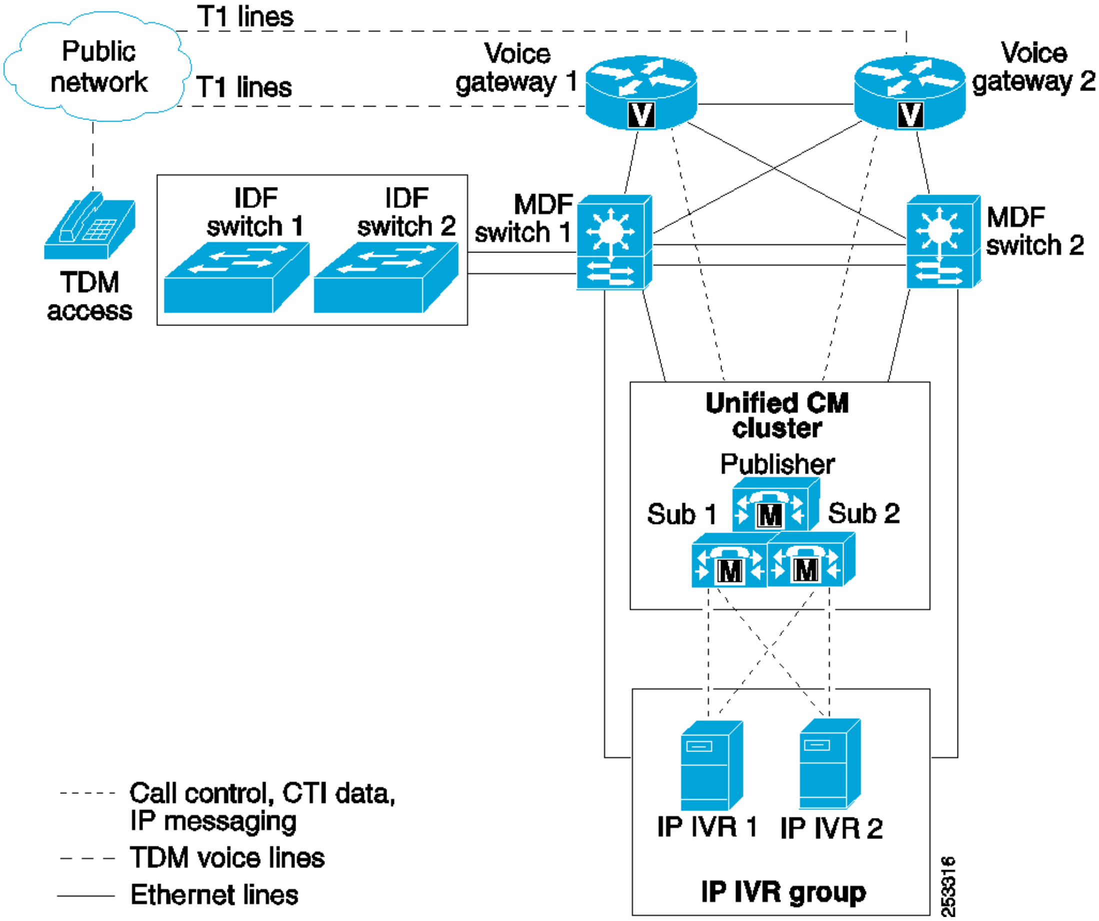
PG side B, Cisco CM PIM 1

# Unified IP IVR Design Considerations

The JTAPI subsystem in Unified IP IVR can establish connections with two CTI Managers on different subscribers in the Unified CM cluster. This feature enables Unified CCE designs to add Unified IP IVR redundancy at the CTI Manager level, such as the Unified CCE Peripheral Gateway connections. Additionally, deploy multiple, redundant IP-IVR servers in the design and allow the Unified CCE call routing script to load-balance calls automatically between the available IP-IVR resources.

Figure 54 shows two Unified IP IVR servers configured for redundancy within one Unified CM cluster. Configure the Unified IP IVR group so that each server is connected to a different CTI Manager service on different Unified CM subscribers in the cluster for high availability. Using the redundancy feature of the JTAPI subsystem in the Unified IP IVR server, you can implement redundancy by adding the IP addresses or host names of two Unified CMs from the cluster. Then, if one of the Unified CMs fails, the Unified IP IVR associated with that particular Unified CM will fail-over to the second Unified CM.

**Figure 54** *High Availability with Two Unified IP IVR Servers and One Unified CM Cluster*



## Unified IP IVR High Availability Using Unified CM

You can implement Unified IP IVR port high availability by using any of the following call-forward features in Unified CM:

- Forward Busy — forwards calls to another port or route point when Unified CM detects that the port is busy. This feature can be used to forward calls to another resource when a Unified IP IVR CTI port is busy due to a Unified IP IVR application problem, such as running out of available CTI ports.
- Forward No Answer — forwards calls to another port or route point when Unified CM detects that a port has not picked up a call within the timeout period set in Unified CM. This feature can be used to forward calls to another resource when a Unified IP IVR CTI port is not answering due to a Unified IP IVR application problem.
- Forward on Failure — forwards calls to another port or route point when Unified CM detects a port failure caused by an application error. This feature can be used to forward calls to another resource when a Unified IP IVR CTI port is busy due to a Unified CM application error.
- When using the call forwarding features to implement high availability of Unified IP IVR ports, avoid creating a loop in the event that all the Unified IP IVR servers are unavailable. Basically, do not establish a path back to the first CTI port that initiated the call forwarding.

## Unified IP IVR High Availability Using Unified CCE Call Flow Routing Scripts

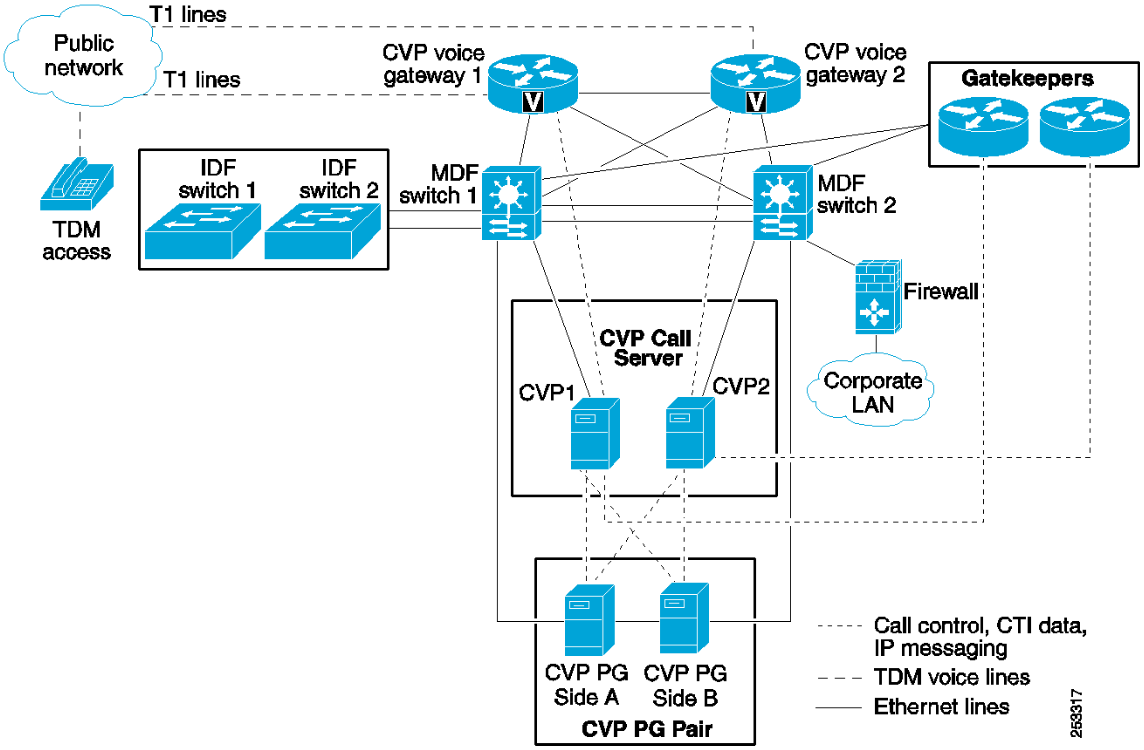
You can implement Unified IP IVR high availability through Unified CCE call flow routing scripts. You can prevent calls from queuing to an inactive Unified IP IVR by using the Unified CCE scripts to check the Unified IP IVR Peripheral Status before sending the calls to it. For example, you can program a Unified CCE script to check if the Unified IP IVR is active by using an IF node or by configuring a Translation Route to the Voice Response Unit (VRU) node (by using the **consider if** field) to select the Unified IP IVR with the most idle ports to distribute the calls evenly on a call-by-call basis. This method can be modified to load-balance ports across multiple Unified IP IVRs and it can address all of the Unified IP IVRs on the cluster in the same Translation Route or Send to VRU node.

- All calls at the Unified IP IVR are dropped if the Unified IP IVR server itself fails. It is important to distribute calls across multiple Unified IP IVR servers to minimize the impact of such a failure. In Unified IP IVR), there is a default script to handle cases where the Unified IP IVR loses the link to the IVR Peripheral Gateway so that the calls are not lost.

# Cisco Unified Customer Voice Portal (Unified CVP) Design Considerations

The Unified CVP can be deployed with Unified CCE as an alternative to Unified IP IVR for call treatment and queuing. Unified CVP is different from Unified IP IVR in that it does not rely on Unified CM for JTAPI call control. Unified CVP uses H.323 or SIP for call control and is used in front of Unified CM or other PBX systems as part of a hybrid Unified CCE or migration solution. (See Figure 55)

**Figure 55** ***High Availability with Two Unified CVP Call Control Servers Using H.323***



Unified CVP uses the following system components:

- Cisco Voice Gateway

  The Cisco Voice Gateway is typically used to terminate TDM PSTN trunks and calls to transform them into IP-based calls on an IP network. Unified CVP uses specific Cisco IOS voice gateways that support H.323 and SIP to enable more flexible call control models outside of the Unified CM MGCP control model. H.323 and SIP protocols enable Unified CVP to integrate with multiple IP and TDM architectures for Unified CCE. Voice gateways controlled by Unified CVP also provide additional functionality using the Cisco IOS built-in Voice Extensible Markup Language (VoiceXML) Browser to provide caller treatment and call queuing on the voice gateway without having to move the call to a physical device such as the IP-IVR or a third-party IVR platform. Unified CVP can also leverage the Media Resource Control Protocol (MRCP) interface of the Cisco IOS voice gateway to add automatic speech recognition (ASR) and text-to-speech (TTS) functions on the gateway under Unified CVP control.

- Unified CVP Call Server

  The Unified CVP Call Server provides call control signaling when calls are switched between the ingress gateway and another endpoint gateway or a Unified CCE agent. It also provides the interface to the Unified CCE VRU Peripheral Gateway and translates specific Unified CCE VRU commands into VoiceXML code that is rendered on the Unified CVP Voice Gateway. The Call Server can communicate with the gateways using H.323 or SIP as part of the solution.

- Unified CVP Media Server

  The Unified CVP caller treatment is provided either by using ASR/TTS functions through MRCP or with predefined .wav files stored on media servers. The media servers act as web servers and serve up the .wav files to the voice browsers as part of their VoiceXML processing. Media servers can be clustered using the Cisco Content Services Switch (CSS) products allowing multiple media servers to be pooled behind a single URL for access by all the voice browsers in the network.

- Unified CVP VXML Application Server

  Unified CVP provides a VoiceXML service creation environment using an Eclipse toolkit browser which is hosted in the Unified CVP Call Studio Application. The Unified CVP VXML server hosts the Unified CVP VoiceXML runtime environment where the dynamic VoiceXML applications are executed and Java and Web Services calls are processed for external systems and database access.

- H.323 Gatekeepers

  Gatekeepers are used with Unified CVP to register the voice browsers and associate them with specific dialed numbers. When a call comes into the network, the gateway will query the gatekeeper to find out where to send the call based on the dialed number. The gatekeeper is also aware of the state of the voice browsers and will load-balance calls across them to avoid sending calls to out-of-service voice browsers or ones that have no available sessions.

- SIP Proxy Servers

  SIP Proxy Servers are used with Unified CVP to select voice browsers and associate them with specific dialed numbers. When a call comes into the network, the gateway will query the SIP Proxy Server to find out where to send the call based on the dialed number.

Availability of Unified CVP can be increased by the following methods:

- Adding redundant Unified CVP Call Servers under control of the Unified CCE Peripheral Gateways allows calls to be balanced automatically across multiple Unified CVP Call Servers.

- Adding TCL scripts to the Unified CVP gateway to handle conditions where the gateway cannot contact the Unified CVP Call Server to direct the call correctly.
- Adding gatekeeper redundancy with HSRP or gatekeeper clustering in H.323.
- Adding a Cisco Content Server to load-balance .wav file requests across multiple Unified CVP Media Servers and VoiceXML URL access across multiple servers.
- Calls in Unified CVP are not dropped if the Unified CVP Call Server or Unified CVP PG fails because they can be redirected to another Unified CVP Call Server on another Unified CVP-controlled gateway as part of the fault-tolerant design using TCL scripts (which are provided with the Unified CVP images) in the voice gateway.

For more information about these options, review the Unified CVP product documentation.

# Cisco Multichannel Options with the Cisco Interaction Manager: E-Mail Interaction Manager and Web Interaction Manager

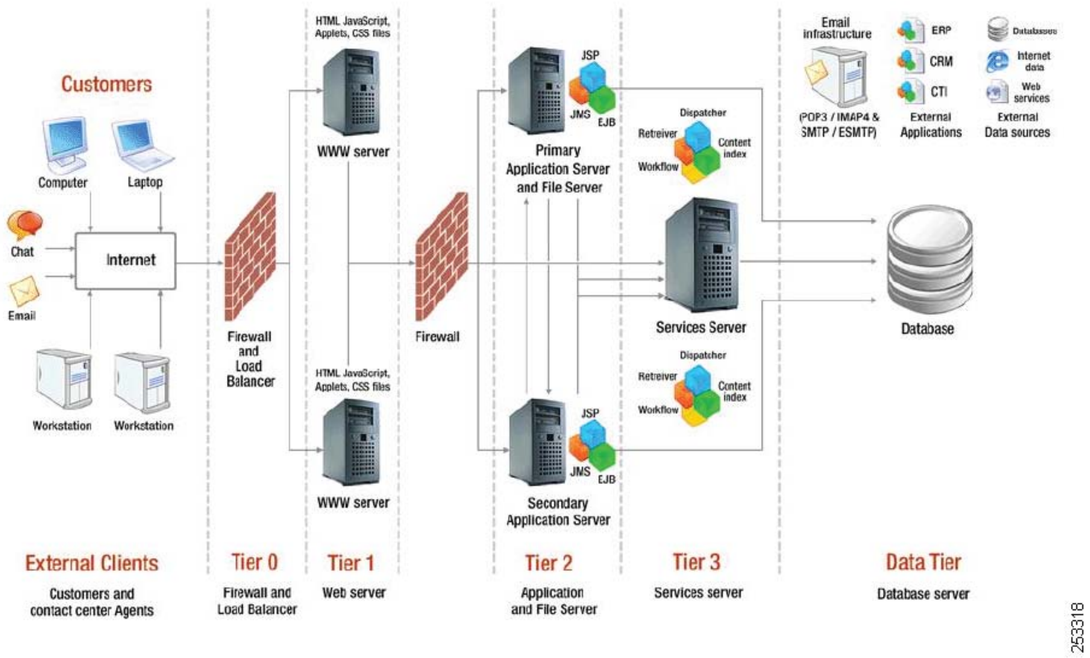
In 2007, Cisco introduced the replacement for the 5.*x* versions of the Multichannel products: Cisco E-Mail Manager (CEM) and Cisco Collaboration Server (CCS). These original products were two separate products that had their own integration methods and web interface for the agents and administrators. The new Cisco Interaction Manager (CIM) platform is a single application that provides both E-Mail and Web interaction management using a common set of web servers and pages for agents and administrators. The new offering is designed for integration with the Unified CCE platform to provide universal queuing of contacts to agents from different media channels.

For additional design information about the Interaction Manager platform, see the *Cisco Unified Web and E-Mail Interaction Manager Solution Reference Network Design (SRND) Guide for Unified Contact Center Enterprise, Hosted, and ICM* .

## Cisco Interaction Manager Architecture Overview

The Cisco Interaction Manager has several core components, as illustrated in Figure 56.

**Figure 56** ***Cisco Interaction Manager Architecture***



The architecture is defined by a multi-tiered model, with various components at each of the following levels of the design:

### External Clients

Cisco Interaction Manager is a 100% web-based product that agents and end-customers can access using a web browser from their desktops.

Agents can access the application using Microsoft Internet Explorer 6.0 or the embedded CAD browser and customers can access the chat customer console using specific versions of Microsoft IE, Mozilla, Firefox, or Netscape. Cisco Interaction Manager is not supported on agent desktops running in a Citrix terminal services environment.

### Tier 0: Firewall and Load Balancer

Agents and customers connect to the application from their respective browsers through a firewall, if so configured for the application.

A load balancer may also be used in case of a distributed installation of the application so that requests from agents and customers are routed to the least-loaded web servers.

### Tier 1: Web Server

The web server is used to serve static content to the browser. Cisco Interaction Manager is designed to be indifferent to the specific type of web server being used with the single requirement being that the application server vendor must provide a web server plug-in for the corresponding application server.

### Tier 2: Application and File Server

The application server is used as a web container (also known as the JSP or Servlet engine) and EJB Container. The core business logic resides in the Business Object Layer and stored procedures reside on the database server. The business logic residing in JAVA classes is deployed on the application server. The JSP or Servlets interact with the business objects through the business client layer and these in turn interact with the database to execute some business logic on data present in the database server.

Example: Outbound Task Creation

- A user logs in to the application and creates an outbound task.
- The JSP layer calls the Business Client layer which interacts with Business Objects residing in the same application server where JSPs or Servlets are deployed.
- The Business Objects execute queries and stored procedures residing on the database server.
- Activities are created and stored in database tables.
- The file server is used for storing all email and article attachment files, report templates, and all locale-specific strings used in the application.

### Tier 3: Services Server

Cisco Interaction Manager has processes that perform specific business functions such as fetching emails from a POP server, sending emails to an SMTP server, processing workflows, assigning chats to the agents, and so forth. All services run on the Services server and are managed by the Distributed Service Manager (DSM).

Cisco Interaction Manager facilitates the creation of multiple instances of services with work distributed among the various instances. For example, the service used to retrieve emails could be configured to have multiple instances to retrieve emails from different email addresses. This capability can be used to process increasing volumes of customer interactions coming into a contact center.

### Data Tier: Database Server

The data tier includes databases that are SQL-compliant, HTML/XML data-sources, and ultimately Web services that consume and produce SOAP messages. Business objects and data adapters use this layer to extract data from various third-party applications and data sources. This layer also deals with HTML and XML parsing using relevant J2EE-compliant packages to process data in other formats.

## Unified CCE Integration

As part of the system integration with Unified CCE, the services server consists of two additional services, the EAAS and the Listener Service, which interact with the Media Routing (MR) PG and Agent PG components of Unified CCE respectively through the Media Routing (MR) and Agent Resource Management (ARM) interfaces.

Additionally, the application server of Cisco Interaction Manager establishes a connection with the Unified CCE Administration & Data server to import relevant configuration data and to map the configuration to Cisco Interaction Manager objects in the Cisco Interaction Manager database. Note that Cisco Interaction Manager does not make use of the Configuration API (ConAPI) interface.

In parent/child configurations, there is no multichannel routing and integration through the parent Unified ICM. Media Routing PGs need to connect to the child Unified CCE. A separate Cisco Interaction Manager or partition is required for each child.

Likewise, in hosted Unified ICM/CCH environments, there is no multichannel routing through the Network Application Manager (NAM) layer and integration is at the individual Customer ICM (CICM) level only. The Media Routing (MR) PGs need to connect to the CICM.

## High Availability Considerations for Cisco Interaction Manager

The Cisco Interaction Manager offers high availability options using additional web and application servers and by using load balancing equipment to distribute agents and contact work more evenly across the platform. This also provides for failover in duplex or redundant models.

## Load Balancing Considerations

The web service component of a Cisco Interaction Manager deployment can be load balanced to serve a large number of agents accessing the application at the same time. The web (or Web and Application) servers can be configured behind the load balancer with a Virtual IP and an agent can access Cisco Interaction Manager through Virtual IP. Depending on the selected load balancing algorithm, the load balancer will send a request to one of the web and application servers behind it and send a response back to the agent. In this way, from a security perspective, the load balancer serves as a reverse proxy server too.

One of the most essential parameters for configuring a load balancer is to configure it to support sticky sessions with cookie-based persistence. After every scheduled maintenance task, before access is opened for users, verify that all web and application servers are available to share the load. In the absence of this, the first web and application server could be overloaded due to the sticky connection feature. With other configurable parameters, you can define a load-balancing algorithm to meet various objectives such as equal load balancing, isolation of the primary web and application server, or 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. If a problem is observed, the load balancer removes the given web and application server from the available pool of servers, preventing new web requests from being directed to the problematic server.

## Managing Failover

Cisco Interaction Manager supports clustered deployments. This ensures high availability and performance through transparent replication, load balancing, and failover. The following key methods are available for handling failure conditions within a Cisco Interaction Manager and Unified CCE integrated deployment:

- Implementing multiple Web and Application servers. If the primary server goes down, the load balancer can help handle the failure through routing requests to alternate servers. The load balancer detects application server failure and redirects requests to another application server, after which a new user session will be created and users will have to log in again to the Cisco Interaction Manager.
- Allowing servers to be dynamically added or removed from the online cluster to accommodate external changes in demand or internal changes in infrastructure.
- Allowing Cisco Interaction Manager services to fail-over with duplexed Unified CCE components (for example, MR PIM and Agent PIM of the MR PG and Agent PG, respectively) to eliminate downtime of the application in failure circumstances.

The single points of failure in Cisco Interaction Manager include the following.

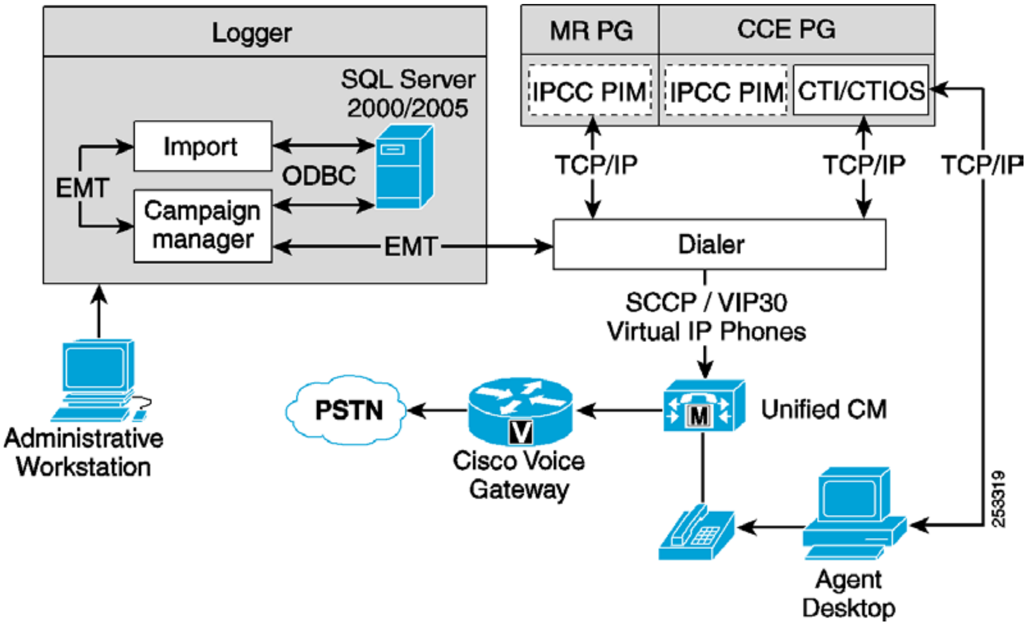
- The JMS server going down
- The Services server going down
- The Database server going down

# Cisco Unified Outbound Option Design Considerations

The Cisco Unified Outbound Option provides the ability for Unified CCE to place calls on behalf of agents to customers based on a predefined campaign. The major components of the Unified Outbound Option (shown in Figure 57) are:

- Outbound Option Campaign Manager—A software module that manages the dialing lists and rules associated with the calls to be placed. This software is loaded on the Logger Side A platform and is not redundant; it can be loaded and active only on the Logger A of the duplexed pair of Loggers in the Unified CCE system.
- Outbound Option Dialer—A software module that performs the dialing tasks on behalf of the Campaign Manager. In Unified CCE, the Outbound Option Dialer emulates a set of IP phones for Unified CM to make the outbound calls and it detects the called party and manages the interaction tasks with the CTI OS server to transfer the call to an agent. It also interfaces with the Media Routing Peripheral Gateway. Each Dialer has its own peripheral interface manager (PIM) on the Media Routing Peripheral Gateway.
- Media Routing Peripheral Gateway—A software component that is designed to accept route requests from non-inbound voice systems such as the Unified Outbound Option or the Multichannel products. In the Unified Outbound Option solution, each Dialer communicates with its own peripheral interface manager (PIM) on the Media Routing Peripheral Gateway.

**Figure 57** ***Unified CCE Unified Outbound Option***



The system can support multiple dialers across the enterprise, all of which are under control of the central Campaign Manager software.

For the new SIP Dialer introduced in Unified CCE Release 8.0, Dialers operate in a warm standby mode similar to the PG fault tolerance model. For more details on this, see the Outbound Option chapter.

For the pre-existing SCCP Dialers, although they do not function as a redundant or duplexed pair the way a Peripheral Gateway does, with a pair of dialers under control of the Campaign Manager, a failure of one of the dialers can be handled automatically and calls will continue to be placed and processed by the surviving dialer. Any calls that were already connected to agents would remain connected and would experience no impact from the failure.

In all deployments, the Dialers are coresident on the Unified CCE Peripheral Gateway for Unified CM.

Guidelines for high availability:

- Deploy the Media Routing Peripheral Gateways in duplex pairs.
- Deploy multiple Dialers with one on each side of the Duplex Unified CCE Peripheral Gateway and make use of them in the Campaign Manager to allow for automatic fault recovery to a second Dialer in the event of a failure. For the SCCP Dialer, there are two options with multiple Dialers: a second Dialer can be configured with the same number of ports (100% redundancy), or the ports can be split across the two Dialers since they operate independently and would both be active at the same time. In designs with a small number of Dialer ports, splitting them can impact the performance of the campaign.
- Deploy redundant voice gateways for outbound dialing to ensure that the dialers have enough trunks available to place calls in the event of a voice gateway failure. In some instances where outbound is the primary application, these gateways would be dedicated to outbound calling only.

# Peripheral Gateway Design Considerations

The Agent PG uses the Unified CM CTI Manager process to communicate with the Unified CM cluster with a single Peripheral Interface Manager (PIM) controlling agent phones and CTI route points anywhere in the cluster. The Peripheral Gateway PIM process registers with CTI Manager on one of the Unified CM servers in the cluster and the CTI Manager accepts all JTAPI requests from the PG for the cluster. If the phone, route point, or other device that is being controlled by the PG is not registered to that specific Unified CM server in the cluster, the CTI Manager forwards that request to the other Unified CM servers in the cluster using Unified CM SDL links. There is no need for a PG to connect to multiple Unified CM servers in a cluster.

## Multiple PIM Connections to a Single Unified CM Cluster

Although the Agent PG in this document is described as typically having only one PIM process that connects to the Unified CM cluster, the Agent PG can manage multiple PIM interfaces to the same Unified CM cluster. It can be used to create additional peripherals within Unified CCE for two purposes:

- Improving Failover Recovery for Customers with Large Numbers of CTI Route Points
- Scaling the Unified CCE PG Beyond 2,000 Agents per Server

## Improving Failover Recovery for Customers with Large Numbers of CTI Route Points

When a Unified CCE PG fails-over, the PIM connection that was previously controlling the Unified CM cluster is disconnected from its CTI Manager and the duplex or redundant side of the PG will attempt to connect it's PIM to the cluster using a different CTI Manager and Subscriber. This process requires the new PIM connection to register for all of the devices (phones, CTI Route Points, CTI Ports, and so forth) that are controlled by Unified CCE on the cluster. When the PIM makes these registration requests, all of them must be confirmed by Unified CM before the PIM can go into an active state and process calls.

To help recover more quickly, the Unified CCE PG can have a PIM created that is dedicated to the CTI Route Points for the customer, thus allowing this PIM to register for these devices at a rate of approximately five per second and allowing the PIM to activate and respond to calls hitting these CTI Route points faster than if the PIM had to wait for all of the route points, then all the agent phones, and all the CTI ports.

This dedicated CTI Route Point PIM could become active several minutes sooner and direct new inbound calls to queuing or treatment resources while waiting for the Agent PIM with the phones and CTI Ports to complete the registration process and become active.

This does not provide any additional scaling or other benefits for the design; the only purpose is to allow Unified CM to have the calls on the CTI Route Points serviced faster by this dedicated PIM. Use this only with customers who have more than 250 Route Points because anything less does not provide a reasonable improvement in recovery time. Additionally, associate only the CTI Route Points that would be serviced by Unified CCE with this PIM and provide it with its own dedicated CTI-Enabled JTAPI or PG user that is specific to the CTI Route Point PIM.

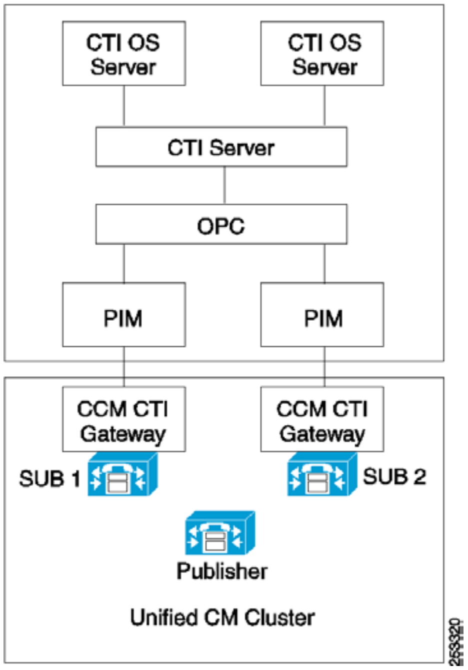
## Scaling the Unified CCE PG Beyond 2,000 Agents per Server

In Unified CCE, multiple PIMs in the same physical PG server may be used to connect either to the same Unified CM cluster or to a second Unified CM cluster. This design reduces the physical number of PG servers required in the Unified CCE design. This is different from the recovery strategy for multiple PIMs because both of these PIMs would be configured with up to 2,000 concurrent agents and their related CTI Route Points and CTI Ports as needed to support those agents. The additional PIM creates another Peripheral from the Unified CCE perspective, which might impact routing and reporting. Additionally, agent teams and supervisors cannot cross peripherals, so careful consideration must be given to which agent groups are allocated to each PIM and Peripheral in such a design.

In designs where Unified CCE is deployed with Unified CVP, the Cisco Unified Communications Sizing Tool might show that the Unified CM cluster can support more than 2,000 total agents; however, the CTI Manager and JTAPI interfaces are tested and supported with a maximum of only 2,000 agents. In order to allow for a design with a single Unified CM cluster with more than 2,000 agents, a second Agent PIM is configured to support the additional agents (up to a total of 4,000 agents per PG).

Figure 58 illustrates a single Unified CCE PG with two different PIMs pointing to the same Unified CM cluster.

**Figure 58** *Two PIMs Configured to the Same Unified CM Cluster*



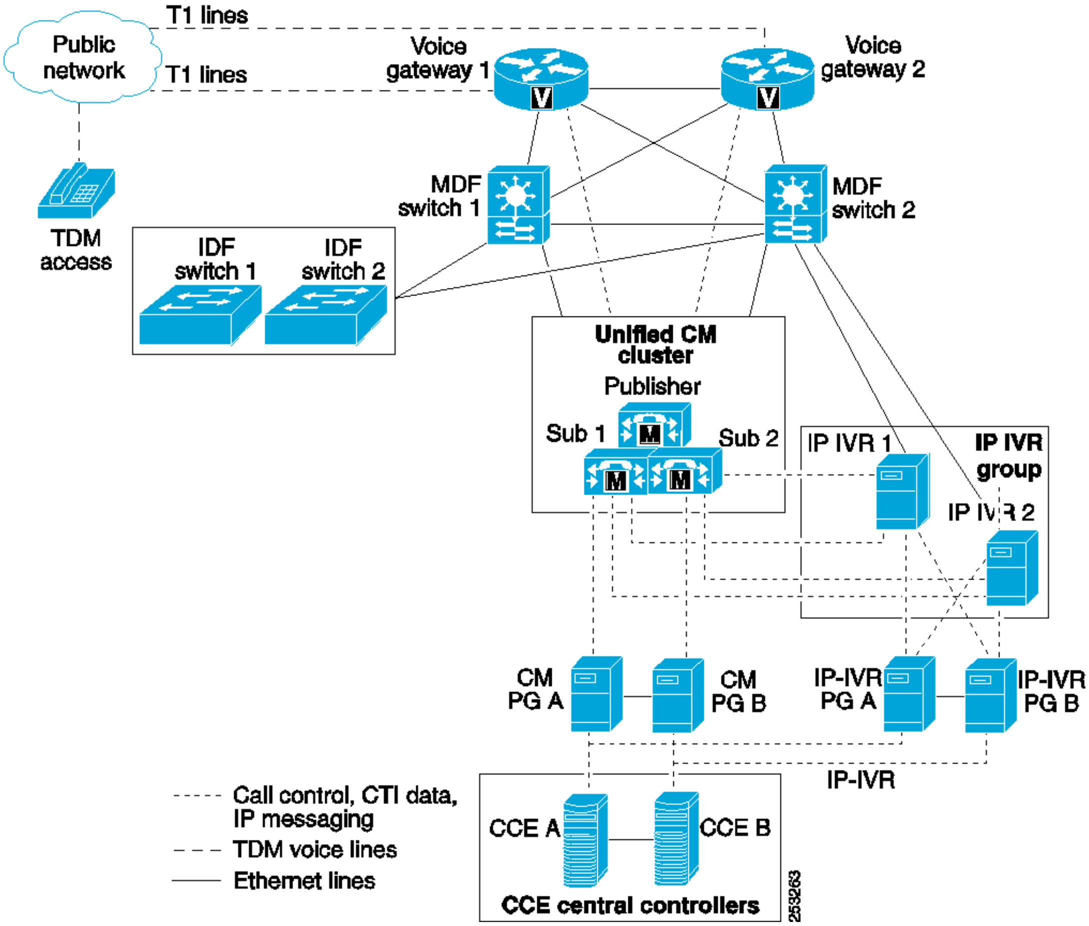
- Use the Cisco Unified Communications Sizing Tool (Unified CST) to size the Unified CM cluster properly for Unified CCE. This tool is only available to Cisco partners and employees with proper login authentication.

## Redundant or Duplex Unified CCE Peripheral Gateway Considerations

Unified CCE Agent PGs are deployed in a redundant or duplex configuration because the PG has only one connection to the Unified CM cluster using a single CTI Manager. If that CTI Manager were to fail, the PG would no longer be able to communicate with the Unified CM cluster. Adding a redundant or duplex PG allows Unified CCE to have a second pathway or connection to the Unified CM cluster using a second CTI Manager process on a different Unified CM server in the cluster.

The minimum requirement for Unified CCE high-availability support for CTI Manager and Unified IP IVR is a duplex (redundant) Agent PG environment with one Unified CM cluster containing at least two subscribers. Therefore, the minimum configuration for a Unified CM cluster in this case is one publisher and two subscribers. This minimum configuration ensures that if the primary subscriber fails, the devices re-home to the secondary subscriber and not to the publisher for the cluster. (See Figure 59) In smaller systems and labs, Cisco permits a single publisher and single subscriber. But if the subscriber fails, then all the devices will be active on the publisher. For specific details about the number of required Unified CM servers, see the chapter on Sizing Cisco Unified Communications Manager Servers.

**Figure 59** ***Unified CCE High Availability with One Unified CM Cluster***



To simplify the illustration in Figure 59, the Unified CCE Server or Unified CCE Central Controller is represented as a single server, but it is actually a set of servers sized according to the Unified CCE agent count and call volume. The Unified CCE Central Controllers include the following redundant or duplex servers:

- Call Router — The core of the CCE complex that provides intelligent call routing instructions based on real-time conditions that it maintains in memory across both the A-Side and B-Side Call Router processes.
- Logger and Database Server — The repository for all configuration and scripting information as well as historical data collected by the system. The Loggers are paired with Call Routers such that Call Router Side A will read and write data only to the Logger A and Call Router B will read and write only to the Logger B. Because both sides of the Call Router processes are synchronized, the data written to both Loggers is identical.

In specific deployment models, these two components can be installed on the same physical server which is then referred to as a Rogger, or combined Router and Logger. See the chapter on Sizing Unified CCE Components and Servers for more details on these specific configurations.

## Unified CM JTAPI and Peripheral Gateway Failure Detection

There is a heartbeat mechanism that is used to detect failures between the Unified CM JTAPI link and the Peripheral Gateway. However, unlike the Unified CCE heartbeat methods that use TCP keep-alive messages on the open socket ports, this method uses a specific heartbeat message in the JTAPI messaging protocol between the systems. By default, the heartbeat messages are sent every 30 seconds and the communications path is reset by the Unified CM or Peripheral Gateway after missing two consecutive heartbeat messages.

This failure detection can be enhanced by using the following procedure to change the heartbeat interval on the JTAPI Gateway client that runs on the Peripheral Gateway:

Step 1. From the Start Menu of the Peripheral Gateway, Select Programs -> Cisco JTAPI -> JTAPI Preferences.

Step 2. Set the Advanced -> Server Heartbeat Interval (sec) field to 5 seconds.

- Do not set this value lower than five seconds because it might impact system performance and trigger an inappropriate failover. This setting determines how often the heartbeats are generated. If it is set to five seconds, the system will fail-over this connection within ten seconds of a loss of network connection (because it must detect two consecutive missed heartbeats). The default of 30 seconds means that it takes up to one minute (60 seconds) to take action on a network connection failure.

Because this JTAPI connection between the Peripheral Gateway and Unified CM is only supported locally on the same LAN segment, there is no latency issue for this heartbeat value. However, if there are any additional network hops, firewalls, or other devices that cause delay between these two components, then set the heartbeat interval value accordingly to account for this delay.

## Unified CCE Redundancy Options

Duplex or redundant Unified CCE servers can be located at the same physical site or they can be geographically distributed. This applies specifically to the Central Controller (Call Router and Logger) and Peripheral Gateways.

Under normal operations, the Unified CCE Call Router and Logger and Database Server processes are interconnected through a Private Network connection that is isolated from the Visible or Public Network segment. Configure these servers with a second NIC card for the Private Network connection and isolate the Private connections from the rest of the Visible or Public Network in their own Cisco Catalyst switch if they are located at the same physical site.

If the Central Controllers are geographically separated (located at two different physical sites), under normal operations the same Private Network connections must continue to be isolated and connected between the two physical sites with a separate WAN connection. For normal operations, do not provision this Private Network connection on the same circuits or network gear as the Visible or Public Network WAN connection because that would create a single point of failure that could disable both WAN segments at the same time.

The Unified CCE Peripheral Gateway duplex pair of servers is also interconnected through a Private Network connection that is isolated from the Visible or Public Network segment under normal operations. If the two sides of the duplex pair (Side A and Side B) are both at the same physical site, the Private Network can be created by using an Ethernet Cross-Over Cable between the two servers to interconnect their Private Network NIC cards. If the two servers in the duplex pair are geographically distributed (located at two different physical sites), the Private Network connections must be connected with a separate WAN connection between the two physical sites. Do not provision this Private Network connection on the same circuits or network gear as the Visible or Public Network WAN connection because that would create a single point of failure that could disable both WAN segments at the same time.

For additional details on the Unified ICM network requirements for this connection, see the installation guides.

For additional details on the Unified CCE network requirements for clustered over the WAN, see the section on IPT: Clustering Over the WAN.

Within the Agent PG, two software processes manage the connectivity to the Unified CM cluster:

- JTAPI Gateway

  The JTAPI Gateway is installed on the PG by downloading it from the Unified CM cluster at the time of the PG installation. This ensures compatibility with the JTAPI and CTI Manager versions in the system. Note that when either the PG or Unified CM is upgraded, this JTAPI Gateway component must be removed and re-installed on the PG.

  The JTAPI Gateway is started by the PG automatically and runs as a node-managed process. The PG monitors this process and automatically restarts it if it fails for any reason. The JTAPI Gateway handles the low-level JTAPI socket connection protocol and messaging between the PIM and the Unified CM CTI Manager.

- Agent PG Peripheral Interface Manager (PIM)

  The PIM is also a node-managed process and is monitored for unexpected failures and automatically restarted. This process manages the higher-level interface between the Unified CCE and the JTAPI Gateway and Unified CM cluster, requesting specific objects to monitor and handling route requests from the Unified CM cluster.

In a duplex Agent PG environment, both JTAPI services from both Agent PG sides log into the CTI Manager upon initialization. Unified CM PG side A logs into the primary CTI Manager; PG side B logs into the secondary CTI Manager. Only the active side of the Unified CM PG register monitors for phones and CTI route points. The duplex Agent PG pair works in hot-standby mode with only the active PG side PIM communicating with the Unified CM cluster. The standby side logs into the secondary CTI Manager only to initialize the interface and make it available for a failover. The registration and initialization services of the Unified CM devices take a significant amount of time; therefore having the CTI Manager available significantly decreases the time for failover.

In duplex PG operation, the side that goes active is the PG side that is first able to connect to the Unified CCE Call Router Server and request configuration information. It is not determined based on the side-A or side-B designation of the PG device but depends only on the ability of the PG to connect to the Call Router. The Call Router ensures that only the PG side that has the best connection goes active.

The startup process of the PIM requires that all of the CTI route points be registered first, which is done at a rate of 5 route points per second. For systems with a lot of CTI route points (for example, 1000), this process can take as long as 3 minutes to complete before the system will allow any of the agents to log in. This time can be reduced by distributing the devices over multiple PIM interfaces to the Unified CM cluster, as noted above.

In the event that calls arrive at the CTI Route Points in Unified CM but the PIM is not yet fully operational, these calls fail unless these route points are configured with a recovery number in their "Call Forward on Unregistered" or "Call Forward on Failure" setting. These recovery numbers could be the Cisco Unity voicemail system for the Auto Attendant (or perhaps the company operator position) to ensure that the incoming calls are being answered.

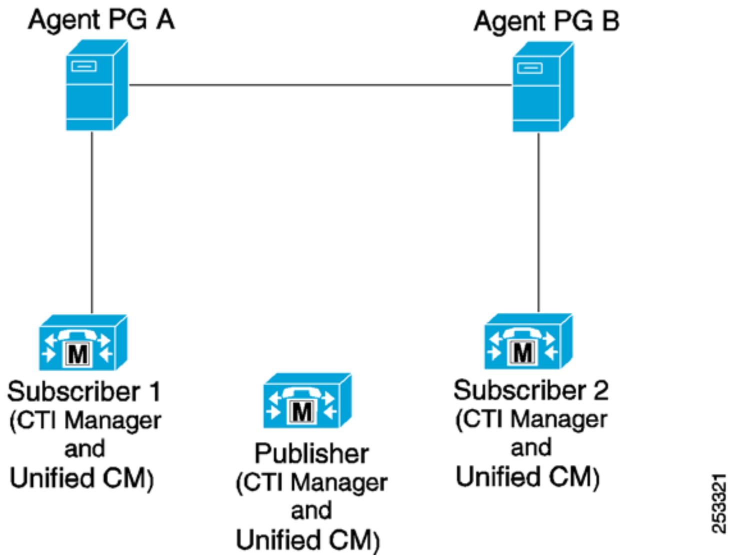
## Unified CM Failure Scenarios

A fully redundant Unified CCE system contains no single points of failure. However, there are scenarios where a combination of multiple failures can reduce Unified CCE system functionality and availability. Also, if a component of the Unified CCE solution does not itself support redundancy and failover, existing calls on that component are dropped. The following failure scenarios have the most impact on high availability and Unified CM Peripheral Interface Managers (PIMs) cannot activate if either of the following failure scenarios occurs (see Figure 60):

- Agent PG/PIM side A and the secondary CTI Manager that services the PG/PIM on side B both fail.
- Agent PG/PIM side B and the primary CTI Manager that services the PG/PIM on side A both fail.

In either of these cases, Unified CCE will not be able to communicate with the Unified CM cluster.

**Figure 60** ***Unified CM PGs Cannot Cross-Connect to Backup CTI Managers***


Agent PG A
Agent PG B
M
M
M
Subscriber 1 (CTI Manager and Unified CM)
Publisher (CTI Manager and Unified CM)
Subscriber 2 (CTI Manager and Unified CM)
253321

## Unified CCE Failover Scenarios

This section describes how redundancy works in the following failure scenarios:

- Scenario 1: Unified CM and CTI Manager Fail
- Scenario 2: Agent PG Side A Fails
- Scenario 3: The Unified CM Active Call Processing Subscriber Fails
- Scenario 4: The Unified CM CTI Manager Providing JTAPI Services to the Unified CCE PG Fails

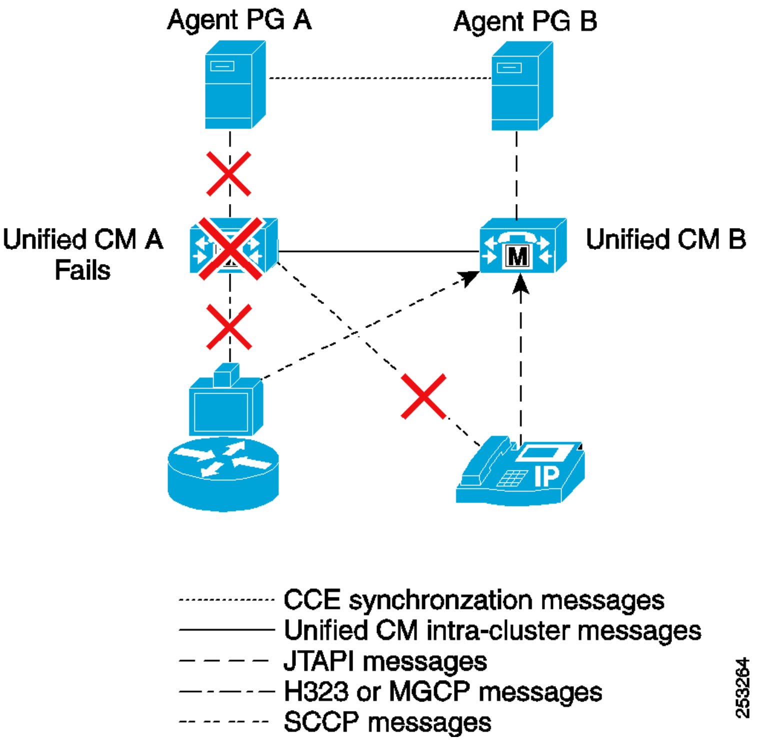
## Scenario 1: Unified CM and CTI Manager Fail

Figure 61 shows a complete system failure or loss of network connectivity on Cisco Unified CM subscriber A. The CTI Manager and Cisco CallManager services were initially both active on this same server and

Unified CM subscriber A is the primary CTI Manager in this case. The following conditions apply to this scenario:

- All phones and gateways are registered with Unified CM subscriber A as the primary server.
- All phones and gateways are configured to re-home to Unified CM subscriber B (that is, B is the backup server as part of the redundancy group in Unified CM).
- Unified CM subscribers A and B are each running a separate instance of CTI Manager within the same Unified CM cluster.
- When Unified CM subscriber A fails, all registered phones and gateways re-home to Unified CM subscriber B. Calls that are in progress with agent phones remain active, but the agents are not able to use phone services such as conference or transfer until they hang up the call and their phone re-registers with the backup subscriber. Although the call stays active, Unified CCE loses visibility to the call and writes a Termination Call Detail (TCD) record to the Unified CCE database for the call at the time of the failure. No additional call data (such as wrap-up codes) are written about the call after that point. Phones that are not active on a call re-home automatically.
- PG side A detects a failure and induces a failover to PG side B.
- Depending on the configuration of the Peripheral in Unified CCE, the CTI OS or CAD server keeps the agent logged in but "grays out" their desktop controls until the PG has completed its failover processing. The agents might not have to log in again but might have to manually make themselves "ready" or "available" to ensure that they are aware that call processing functionality has been restored.
- PG side B becomes active and registers all dialed numbers and phones, call processing continues.
- As noted above, when the PG fails-over, the Unified CCE Call Router writes a Termination Call Detail Record (TCD) in the Unified CCE database for any active calls. If the call is still active when the PG fails-over to the other side, a second TCD record is written for this call as if it were a "new" call in the system and not connected to the prior call that was recorded in the database.
- When Unified CM subscriber A recovers, all idle phones and gateways re-home to it. Active devices wait until they are idle before re-homing to the primary subscriber.
- PG side B remains active using the CTI Manager on Unified CM subscriber B.
- After recovery from the failure, the PG does *not* fail back to the A side of the duplex pair. All CTI messaging is handled using the CTI Manager on Unified CM subscriber B which communicates with Unified CM subscriber A to obtain phone state and call information.

**Figure 61** *Scenario—Unified CM and CTI Manager Fail*



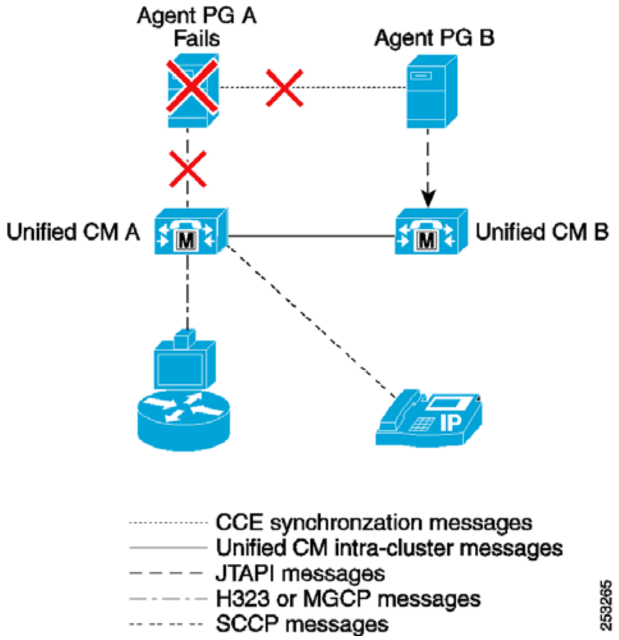
## Scenario 2: Agent PG Side A Fails

Figure 62 shows a failure on PG side A and a failover to PG side B. All CTI Manager and Unified CM services continue running normally. The following conditions apply to this scenario:

- All phones and gateways are registered with Unified CM subscriber A.
- All phones and gateways are configured to re-home to Unified CM subscriber B (that is, B is the backup server); however, they do not need to re-home as the primary subscriber continues to be functional.
- Unified CM subscribers A and B are each running a local instance of CTI Manager.
- When PG side A fails, PG side B becomes active.
- PG side B registers all dialed numbers and phones and call processing continues. Phones and gateways stay registered and operational with Unified CM subscriber A; they do not fail-over.
- Agents with calls in progress will stay in progress but with no third-party call control (conference, transfer, and so forth) available from their agent desktop soft phones. Agents that were not on calls may notice their CTI desktop disable their agent state or third-party call control buttons on the desktop during the failover to the B-Side PG. Once the failover is complete, the agent desktop buttons are restored; however the barge in and conference calls will not be rebuilt properly and calls will disappear from the desktop when either of the participants drops out of the call. Call Type indication of Transfer, Barge In, Intercept, Supervisor Assist, and Emergency Assist are not recovered in the agent desktop or in reporting.

- In most cases, after a PG failover, agents whose states are Available or Wrap Up are moved to Available. Alternatively, agents may receive a prompt to log in or to change their state from Not Ready to Available.
- When the PG fails-over, the Unified CCE Call Router writes a Termination Call Detail Record (TCD) in the Unified CCE database for any active calls. If the call is still active when the PG fails-over to the other side, a second TCD record is written for this call as if it were a "new" call in the system and not connected to the prior call that was recorded in the database.
- When PG side A recovers, PG side B remains active and uses the CTI Manager on Unified CM subscriber B. The PG does not fail-back to the A-Side, and call processing continues on the PG Side B.

**Figure 62** ***Scenario 2—Agent PG Side A Fails***



## Scenario 3: The Unified CM Active Call Processing Subscriber Fails

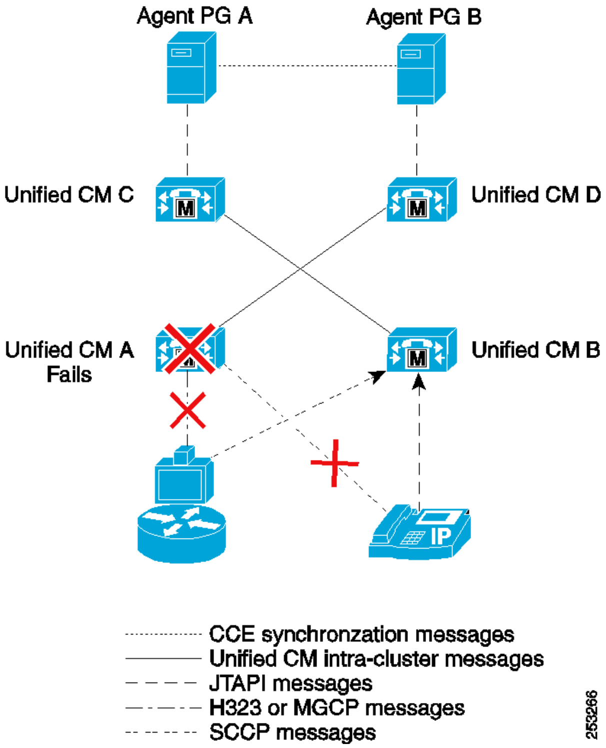
Figure 63 shows a failure on Unified CM active call processing subscriber A. In this model, the subscriber is actively processing calls and controlling devices but does not provide the CTI Manager connection to the Unified CCE PG. The CTI Manager services are running on all the Unified CM subscribers in the cluster, but only subscribers C and D are configured to communicate with the Unified CCE Peripheral Gateway.

The following conditions apply to this scenario:

- All phones and gateways are registered with Unified CM subscriber A.
- All phones and gateways are configured to re-home to Unified CM subscriber B (that is, B is the backup server).
- Unified CM subscribers C and D are each running a local instance of CTI Manager to provide JTAPI services for the Unified CCE PGs.

- If Unified CM subscriber A fails, phones and gateways re-home to the backup Unified CM subscriber B.
- PG side A remains connected and active with a CTI Manager connection on Unified CM subscriber C. It does not fail-over because the JTAPI-to-CTI Manager connection has not failed. However, it sees the phones and devices being unregistered from Unified CM subscriber A (where they were registered) and is notified of these devices being re-registered on Unified CM subscriber B automatically. During the time that the agent phones are not registered, the PG disables the agent CTI desktops to prevent the agents from attempting to use the system while their phones are not actively registered with a Unified CM subscriber. Also, they will be "logged out" by the system during this transition to avoid routing calls to them as well.
- Call processing continues for any devices not registered to Unified CM subscriber A. Call processing also continues for those devices on subscriber A when they are re-registered with their backup subscriber.
- Calls in progress on phones registered to Unified CM subscriber A continue; however, the agent desktop is disabled to prevent any conference, transfer, or other third-party call control during the failover. After the agent disconnects the active call, that agent’s phone re-registers with the backup subscriber, and the agent is logged out from the CT IOS desktop and will have to log in again.
- As noted above, when the Unified CM subscriber A fails, the calls in progress stay active; however, Unified CCE loses control and track of those calls because the phone has not re-homed (re-registered) with the backup subscriber in the cluster. In fact, the phone does not re-home until after the current call is completed. The Unified CCE Call Router writes a Termination Call Detail Record (TCD) in the Unified CCE database for calls that were active at the time of the subscriber failure with call statistics up to the time of the failure and loss of control. Any additional call information (statistics, call wrap-up data, and so forth) are not written to the Unified CCE database.
- When Unified CM subscriber A recovers, phones and gateways re-home to it. This re-homing can be set up on Unified CM to gracefully return groups of phones and devices over time or to require manual intervention during a maintenance window to minimize the impact to the call center. During this re-homing process, the CTI Manager service notifies the Unified CCE Peripheral Gateway of the phones being unregistered from the backup Unified CM subscriber B and re-registered with the original Unified CM subscriber A.
- Call processing continues normally after the phones and devices have returned to their original subscriber.

**Figure 63** ***Scenario—Only the Primary Unified CM Subscriber Fails***



## Scenario 4: The Unified CM CTI Manager Providing JTAPI Services to the Unified CCE PG Fails

Figure 64 shows a CTI Manager service failure on Unified CM subscriber C that is used to communicate with the Unified CCE PG. The CTI Manager services are running on all the Unified CM subscribers in the cluster, but only subscribers C and D are configured to connect to the Unified CCE PGs. During this failure, the PG detects the loss of the JTAPI connection and fails-over to the redundant PG side.

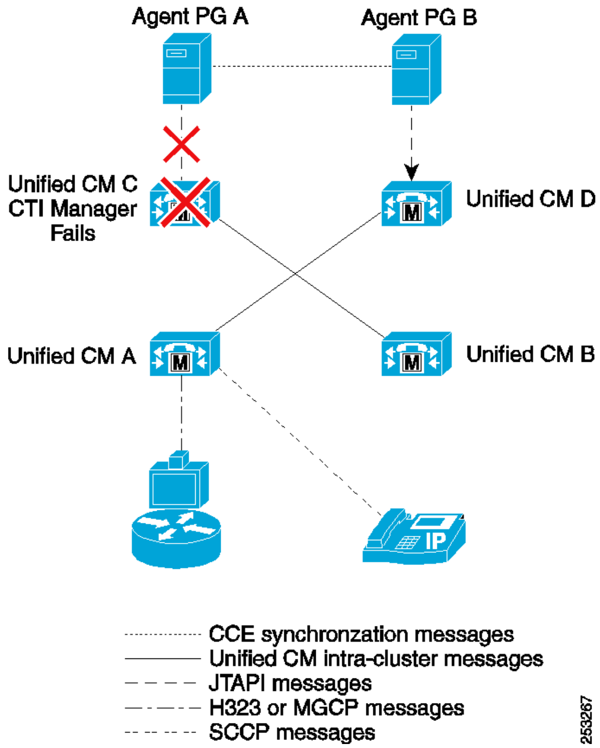
The following conditions apply to this scenario:

- All phones and gateways are registered with Unified CM subscriber A.
- All phones and gateways are configured to re-home to Unified CM subscriber B (that is, B is the backup server). In this case they do not re-home because subscriber A is still functional.
- Unified CM subscribers C and D are each running a local instance of CTI Manager and are designed to connect to the Unified CCE PGs.
- If the Unified CM CTI Manager service on subscriber C fails, the PG side A detects a failure of the CTI Manager service and induces a failover to PG side B.
- PG side B registers all dialed numbers and phones with the Unified CM CTI Manager service on subscriber D and call processing continues.
- Agents with calls in progress stay in progress but with no third-party call control (conference, transfer, and so forth) available from their agent desktop soft phones. After an agent disconnects from all calls, that agent’s desktop functionality is restored. Although the call stays active, Unified CCE loses visibility to the call and writes a Termination Call Detail (TCD) record to the Unified

CCE database for the call at the time of the failure. No additional call data such as wrap-up codes are written about the call after that point.

- When the Unified CM CTI Manager service on subscriber C recovers, PG side B continues to be active and uses the CTI Manager service on Unified CM subscriber D. The PG does not fail-back in this model.

**Figure 64** ***Scenario 4—Only the Unified CM CTI Manager Service Fails***



## Unified CCE Scenarios for Clustering over the WAN

Unified CCE can also be overlaid with the Unified CM design model for clustering over the WAN which allows for high availability of Unified CM resources across multiple locations and data centers. There are a number of specific design requirements for Unified CM to support this deployment model; Unified CCE adds its own specific requirements and new failover considerations to the model.

Specific testing has been performed to identify the design requirements and failover scenarios. The success of this design model relies on specific network configuration and setup and the network must be monitored and maintained. The component failure scenarios noted previously (see Unified CCE Failover Scenarios) are still valid in this model. Additional failure scenarios for this model include:

- Scenario 1: Unified CM and CTI Manager Fail
- Scenario 2: Agent PG Side A Fails
- Scenario 3: The Unified CM Active Call Processing Subscriber Fails
- Scenario 4: The Unified CM CTI Manager Providing JTAPI Services to the Unified CCE PG Fails

**Note** The terms *public network* and *visible network* are used interchangeably throughout this document.

## Scenario 1: Unified CCE Central Controller or Peripheral Gateway Private Network Failure

In clustering over the WAN with Unified CCE, provide a separate private network connection between the geographically distributed Central Controller (Call Router and Logger) and the split Peripheral Gateway pair to maintain state and synchronization between the sides of the system.

To understand this scenario fully, a brief review of the Unified CCE Fault Tolerant architecture is warranted. On each call router, there is a process known as the Message Delivery Service (MDS) which delivers messages to and from local processes such as router.exe and which handles synchronization of messages to *both* call routers. For example, if a route request comes from the carrier or any routing client to side A, MDS ensures that both call routers receive the request. MDS also handles the duplicate output messages.

The MDS process ensures that duplex Unified CCE sides are functioning in a synchronized execution, fault tolerance method. Both routers are executing everything in lockstep based on input the router receives from MDS. Because of this synchronized execution method, the MDS processes must always be in communication with each other over the private network. They use TCP keep-alive messages generated every 100 ms to ensure the health of the redundant mate or the other side. Missing five consecutive TCP keep-alive messages indicates to Unified CCE that the link or the remote partner system might have failed.

When running duplexed Unified CCE sides for all production system, one MDS will be the enabled synchronizer and will be in a *paired-enabled* state. Its partner will be the disabled synchronizer and is said to be *paired-disabled*. Whenever the sides are running synchronized, the side A MDS will be the enabled synchronizer in paired-enabled state. Its partner, side B, will be the disabled synchronizer in paired-disabled state. The enabled synchronizer sets the ordering of input messages to the router and also maintains the master clock for the Unified CCE system.

If the private network fails between the Unified CCE Central Controllers, the following conditions apply:

- The Call Routers detects the failure by missing five consecutive TCP keep-alive messages. The currently enabled side (side A in most cases) transitions to an isolated-enabled state and continues to function as long as it is in communication with at least half of the PGs configured in the system.
- The paired-disabled side (side B in most cases) transitions to an isolated-disabled state. This side will then check for device majority. If it is not communicating with either an Active or Idle DMP to more than half of the configured PGs in the system, it stops processing and stays disabled.
- If the B-Side has device majority (an Active or Idle connection to more than half the configured PGs), it transitions to a "Testing" state and sends "Test Other Side" (TOS) messages to each PG. This message is used to ask the PG if it can see the Call Router on the other side (in this case, Router A).
- As soon as any (even one) PG responds to the TOS message that the A-Side is still enabled, Router B remains in the Isolated-Disabled state and goes idle. Logger B also goes idle, as well as all the DMP connections to the PGs for Router B. All call processing continues on Side A without impact.
- If all of the PGs reply that Side A is down or not reachable, the B-Side Call Router re-initializes in simplex mode (isolated-enabled) and takes over all routing for Unified CCE.
- There is no impact to the agents, calls in progress, or calls in queue. The system can continue to function normally; however the Call Routers are in simplex mode until the private network link is restored.

### Additional Considerations

The Call Routers are "paired" with the Loggers and can read and write only to their own Logger for configuration and historical data over the Private Network locally. In the event that the failure is caused by the loss of a Private NIC card in the Call Router and that Call Router is the enabled side, it cannot write any historical data to the Logger nor can any configuration changes be made to the Logger database.

The Private NIC in the Call Router is also used in some cases to communicate with carrier-based Pre-Routing Network or SS7 interfaces. If the Private NIC fails, there is no access to these services.

If there is an even number of PGs specified in the Call Router Setup and only half of the PGs are available, then only Side A runs. For the B-Side to be operational during a private network failure, it must be able to communicate with more than half of the PGs in the system.

It is important to maintain the configuration so that "extra" PGs or PGs that are no longer on the network are removed from the Call Router Setup panels to avoid problems with determination of device majority for PGs that no longer exist.

If the private network fails between the Unified CM Peripheral Gateways, the following conditions apply:

- The Peripheral Gateway sides detect a failure if they miss five consecutive TCP keep-alive messages and they follow a process similar to the Call Routers of leveraging the MDS process when handling a private link failure. As with the Central Controllers, one MDS process is the enabled synchronizer and its redundant side is the disabled synchronizer. When running redundant PGs, the A side is always the enabled synchronizer.
- After detecting the failure, the disabled synchronizer (side B) initiates a test of its peer synchronizer by using the TOS procedure on the Public or Visible Network connection. If PG side B receives a TOS response stating that the A side synchronizer is enabled or active, then the B side immediately goes out of service, leaving the A side to run in simplex mode until the Private Network connection is restored. The PIM, OPC, and CTI SVR processes become active on PG side A, if not already in that state, and the CTI OS Server process still remains active on both sides as long as the PG side B server is healthy. If the B side does not receive a message stating that the A side is enabled, then side B continues to run in simplex mode and the PIM, OPC, and CTI SVR processes become active on PG side B if not already in that state. This condition occurs only if the PG side A server is truly down or unreachable due to a double failure of visible and private network paths.
- There is no impact to the agents, calls in progress, or calls in queue because the agents stay connected to their already established CTI OS Server process connection. The system can continue to function normally; however the PGs are in simplex mode until the private network link is restored.

If the two private network connections are combined into one link, the failures follow the same path; however, the system runs in simplex mode on both the Call Router and the Peripheral Gateway. If a second failure were to occur at that point, the system could lose some or all of the call routing and ACD functionality.

## Scenario 2: Visible Network Failure

The visible network in this design model is the network path between the data center locations where the main system components (Unified CM subscribers, Peripheral Gateways, Unified IP IVR/Unified CVP components, and so forth) are located. This network is used to carry all the voice traffic (RTP stream and call control signaling), Unified CCE CTI (call control signaling) traffic, as well as all typical data network traffic between the sites. In order to meet the requirements of Unified CM clustering over the WAN, this link must be highly available with very low latency and sufficient bandwidth. This link is critical to the

Unified CCE design because it is part of the fault-tolerant design of the system. It must also be highly resilient.

- The high availability (HA) WAN between the central sites must be fully redundant with no single point of failure. (For information regarding site-to-site redundancy options, see the WAN infrastructure and QoS design guides.) In case of partial failure of the high availability WAN, the redundant link must be capable of handling the full central-site load with all QoS parameters. For more information, see the section on Bandwidth Requirements for Unified CCE Clustering Over the WAN.
- An HA WAN using point-to-point technology is best implemented across two separate carriers, but this is not necessary when using a ring technology.

If the visible network fails between the data center locations, the following conditions apply:

- The Unified CM subscribers detect the failure and continue to function locally with no impact to local call processing and call control. However, any calls that were set up over this WAN link fail with the link.
- The Unified CCE Call Routers detect the failure because the normal flow of TCP keep-alives from the remote Peripheral Gateways stops. Likewise, the Peripheral Gateways detect this failure by the loss of TCP keep-alives from the remote Call Routers. The Peripheral Gateways automatically realign their data communications to the local Call Router and the local Call Router then uses the private network to pass data to the Call Router on the other side to continue call processing. This does not cause a failover of the Peripheral Gateway or the Call Router.
- Half the agents (or more) might be affected by this failure under the following circumstances:
  - If the agent desktop (Cisco Agent Desktop or CTI OS) is registered to the Peripheral Gateway on side A of the system but the physical phone is registered to side B of the Unified CM cluster.

    Under normal circumstances, phone events are passed from side B to side A over the visible network by using the CTI Manager Service to present these events to the side A Peripheral Gateway. The visible network failure does not force the IP phone to re-home to side A of the cluster and the phone remains operational on the isolated side B. The Peripheral Gateway is no longer able to see this phone and the agent is logged out of Unified CCE automatically because the system can no longer direct calls to the agent’s phone.
  - If the agent desktop (Cisco Agent Desktop or CTI OS) and IP phone are both registered to side A of the Peripheral Gateway and Unified CM, but the phone is reset and it re-registers to a side B of the Unified CM subscriber.

    If the IP phone re-homes or is manually reset and forced to register to side B of a Unified CM subscriber, the Unified CM subscriber on side A that is providing the CTI Manager service to the local Peripheral Gateway unregisters the phone and removes it from service. Because the visible network is down, the remote Unified CM subscriber at side B cannot send the phone registration event to the remote Peripheral Gateway. Unified CCE logs this agent out because it can no longer control the phone for the agent.
  - If the agent desktop (CTI Toolkit Agent Desktop or Cisco Agent Desktop) is registered to the CTI OS Server at the side-B site but the active Peripheral Gateway side is at the side-A site.

    Under normal operation, the CTI Toolkit Agent Desktop load-balances its connections to the CTI OS Server pair. At any given time, half the agent connections are on a CTI OS server that has to cross the visible network to connect to the active Peripheral Gateway CTI Server (CG). When the visible network fails, the CTI OS Server detects the loss of connection with the remote Peripheral Gateway CTI Server (CG) and disconnects the active agent desktop clients to force them to re-home to the redundant CTI OS Server at the remote site. The CTI Toolkit Agent Desktop is aware of the redundant CTI OS server and automatically uses this server. During this transition, the CTI Toolkit Agent Desktop is disabled and returns to an operational state as soon as it is connected to the redundant CTI OS server. (The agent may be logged out or put into not-read state depending on the /LOAD parameter defined for the Unified CM Peripheral Gateway in Unified CCE Configuration Manager.)

## Scenario 3: Visible and Private Networks Both Fail (Dual Failure)

Individually, the private and visible 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 is reduced to very limited functionality. This failure is considered catastrophic and can be avoided by careful WAN design with backup and resiliency built into the design.

If both the visible and private networks fail at the same time, the following conditions apply:

- The Unified CM subscribers detect the failure and continue to function locally with no impact to local call processing and call control. However, any calls that were set up and are sending the active voice path media over the visible WAN link fail with the link. When the call fails, the Unified CCE PG sees the call drop and writes a Termination Call Detail (TCD) record in the Unified CCE database for that call at the time it is dropped.
- The Call Routers and Peripheral Gateways detect the private network failure after missing five consecutive TCP keep-alive messages. These TCP keep-alive messages are generated every 100 ms, and the failure is detected within about 500 ms on this link.
- The Call Routers attempt to contact their Peripheral Gateways with the test-other-side message to determine if the failure was a network issue or if the remote Call Router had failed and was no longer able to send TCP keep-alive messages. The Call Routers determine which side will continue to be active (typically, this would be the A-Side of the system because it is the side with the most active Peripheral Gateway connections), and that side stays active in simplex mode while the remote Call Router and PGs are in isolated-disabled mode. The Call Routers send a message to the Peripheral Gateways to realign their data feeds to the active Call Router only.
- The Peripheral Gateways determine which side has the active Unified CM connection. However, it also considers the state of the Call Router and the Peripheral Gateway does not remain active if it is not able to connect to an active Call Router. Typically, this will force the A-Side PGs into active simplex enabled mode and the B-Side into isolated-disabled mode.
- The surviving Call Router and Peripheral Gateways detect the failure of the visible network by the loss of TCP keep-alives on the visible network. These keep-alives are sent every 400 ms so it can take up to two seconds before this failure is detected.
- The Call Router only sees the local Peripheral Gateways, which are those used to control local Unified IP IVRs or Unified CVP Call Servers and the local half of the Unified CM cluster. The remote Unified IP IVRs or Unified CVP Call Servers are off-line with no Unified CCE Call Control via the GED-125 IVR PG interface. The Unified CCE Call Routing Scripts automatically routes around these off-line devices using the peripheral-on-line status checks. Calls that were in progress in the off-line IP-IVRs either drop or use the local default script in the IP-IVR or the Call Forward on Error settings in Unified CM. Calls under Unified CVP control from the off-line Call Servers get treatment from the survivability TCL script in their ingress voice gateways. For calls that were in progress but are no longer visible to Unified CCE, a Termination Call Detail (TCD) record is written to the Unified CCE database for the call data up to the time of the failure. If the default or 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.

- Any new calls that come into the disabled side are not routed by Unified CCE but they can be redirected or handled using standard Unified CM redirect on failure for their CTI route points or the Unified CVP survivability TCL script in the ingress voice gateways.
- Agents are impacted as noted above if their IP phones are registered to the side of the Unified CM cluster opposite the location of their active Peripheral Gateway and CTI OS Server connection. Only agents that were active on the surviving side of the Peripheral Gateway with phones registered locally to that site are not impacted.

At this point, the Call Router and Unified CM Peripheral Gateway run in simplex mode and the system accepts new calls from only the surviving side for Unified CCE call treatment. The Unified IP IVR/Unified CVP functionality is also limited to the surviving side.

## Scenario 4: Unified CCE Agent Site WAN (Visible Network) Failure

The Unified CCE design model for clustering over the WAN assumes the Unified CCE agents are remotely located at multiple sites connected by the visible WAN. Each agent location requires WAN connectivity to both of the data center locations across the visible WAN where the Unified CM and Unified CCE components are located. These connections provide for redundancy as well as making use of basic SRST functionality in the event of a complete network failure, so that the remote site still has basic dial tone service to make emergency (911) calls.

If side A of the WAN at the Unified CCE Agent Site fails, the following conditions apply:

- Any IP phones that are homed to the side-A Unified CM subscribers automatically re-home to the side-B subscribers (provided the redundancy group is configured).
- Agent desktops that are connected to the CTI OS or Cisco Agent Desktop server at that site automatically realign to the redundant CTI OS server at the remote site. (Agent desktops are disabled during the realignment process.)

If both sides of the WAN at the Unified CCE Agent Site fail, the following conditions apply:

- The local voice gateway detects the failure of the communications path to the Unified CM cluster and goes into SRST mode to provide local dial-tone functionality. With Unified CVP, these gateways detect the loss of the Unified CVP Call Server and execute their local survivability TCL script to reroute the inbound calls. Active calls in Unified CVP locally are no longer be visible to Unified CCE, so a Termination Call Detail (TCD) record is written to the Unified CCE database at the time of the failure and tracking of the call stops at that point. The call executes the local survivability TCL script, which could redirect it using the PSTN to another Unified CCE site that remains active; however, the call then appears as a "new call" to Unified CCE and has no relationship with the original call information. If the call is retained locally and redirected by way of SRST to a local phone, Unified CCE does not have visibility to the call from that point forward.
- The agent desktop detects the loss of connectivity to the CTI OS Server (or Cisco Agent Desktop Server) and automatically logs the agent out of the system. While the IP phones are in SRST mode, they are not able to function as Unified CCE agents.

# Understanding Failure Recovery

This section analyzes the failover recovery of each individual part (products and subcomponents inside each product) of the Unified CCE solution.

## Unified CM Service

In larger deployments, it is possible that the Unified CM to which the agent phones are registered is not running the CTI Manager service that communicates with the Unified CM Peripheral Gateway for Unified CCE. When an active Unified CM (call processing) service fails, all the devices registered to it are reported "out of service" by the CTI Manager service locally and to any external client such as the Peripheral Gateway on a different subscriber CTI Manager service.

Unified CM call detail reporting (CDR) shows the call as terminated when the Unified CM failure occurred, although the call may have continued for several minutes after the failure because calls in progress stay in progress. IP phones of agents not on calls at the time of failure quickly register with the backup Unified CM subscriber. The IP phone of an agent on a call at the time of failure does not register with the backup Unified CM subscriber until after the agent completes the current call. If MGCP, H.323, or SIP gateways are used, then the calls in progress survive, but further call control functions (hold, retrieve, transfer, conference, and so on) are not possible.

Unified CCE also writes a call record to the Termination Call Detail (TCD) table because Unified CM has reported the call as terminated to the Unified CCE PG. If the call continues after the PG has failed-over, a second TCD record is written as a "new call" not related to the original call.

When the active Unified CM subscriber fails, the PG receives out-of-service events from Unified CM and logs out the agents. To continue receiving calls, the agents must wait for their phones to re-register with a backup Unified CM subscriber, then log back into their Unified CCE desktop application to have its functionality restored. On recovery of the primary Unified CM subscriber, the agent phones re-register to their original subscriber to return the cluster to the normal state with phones and devices properly balanced across multiple active subscribers.

In summary, the Unified CM call processing service is separate from the CTI Manager service which connects to the Unified CM PG through JTAPI. The Unified CM call processing service is responsible for registering the IP phones and its failure does not affect the Unified CM PGs. From a Cisco Unified CCE perspective, the PG does not go off-line because the Unified CM server running CTI Manager remains operational. Therefore, the PG does not need to fail-over.

## Unified IP IVR

When a CTI Manager service fails, the Unified IP IVR JTAPI subsystem shuts down and restarts by trying to connect to the secondary CTI Manager service on a backup Unified CM subscriber in the cluster. In addition, all voice calls at this Unified IP IVR are dropped. If there is an available secondary CTI Manager service on a backup subscriber, the Unified IP IVR logs into this CTI Manager service on that subscriber and re-registers all the CTI ports associated with the Unified IP IVR JTAPI user. After all the Unified CM devices are successfully registered with the Unified IP IVR JTAPI user, the server resumes its Voice Response Unit (VRU) functions and handles new calls. This action does not impact the Unified CVP because it does not depend on the Unified CM CTI Manager service for call control.

Unified IP IVR Release 3.5 provided for cold standby and Release 4.0 provides hot standby redundancy but this configuration is not supported for use with Unified CCE. These designs make use of a redundant server that is not used unless there is a failure of the primary Unified IP IVR server. However, during this failover processing, all calls that are in queue or treatment are dropped on the Unified IP IVR as part of the failover.

A more resilient design would be to deploy a second (or more) Unified IP IVR server(s) and have them all active; allowing Unified CCE to load-balance calls across them automatically. As shown in Figure 65, if one of the Unified IP IVR servers fails, only the calls on that server fail but the other active servers remain active and are able to accept new calls in the system.

## Unified CCE

Unified CCE is a collection of services and processes running on Unified CCE servers. The failover and recovery process for each of these services is unique and requires careful examination to understand the impact to other parts of the Unified CCE solution (including another Unified CCE service).

## Unified CM PG and CTI Manager Service

When the active CTI Manager service or PG software fails, the PG JTAPI Gateway or PIM detects an OUT_OF_SERVICE event and induces a failover to the redundant (duplex) PG. Because the redundant PG is logged into the backup Unified CM subscriber CTI Manager service already, it registers the IP phones and configured dialed numbers or CTI route points automatically. This initialization service takes place at a rate of about 5 devices per second. The agent desktops show them as being logged out or not ready and a message displays stating that their routing client or peripheral (Unified CM) has gone off-line. (This warning can be turned on or off depending on the administrator's preference.) All agents and supervisors lose their desktop third-party call control functionality until the failure recovery is complete. The agents and supervisors can recognize this event because call control action buttons on the desktop gray out and they cannot do anything with the desktop. Any existing calls remain active without any impact to the caller.

In the event that calls arrive at the CTI Route Points in Unified CM during a PG failover and the PIM is not yet fully operational, these calls fail unless these route points are configured with a recovery number in their "Call Forward on Unregistered" or "Call Forward on Failure" setting. These recovery numbers could be the Cisco Unity voicemail system for the Auto Attendant (or perhaps the company operator position) to ensure the incoming calls are getting answered.

**Note** Do not push any buttons during desktop failover because these keystrokes can be buffered and sent to the CTI server when it completes its failover and restores the agent states.

When an active PG fails over to the idle side, calls still in progress are recovered by querying Unified CM as part of the activation sequence. There is one Termination Call Detail record providing information about the call after the PG transition when the call terminates. Peripheral call variables and ECC variables are maintained on the agent desk top. Call Type indication of Transfer, Barge In, Intercept, Supervisor Assist, and Emergency Assist are not recovered in the desktop or in reporting after the fail over. In most cases, after a PG failover, agents whose states are Available or Wrap Up are moved to Available. Alternatively, agents may receive a prompt to log in or to change their state from Not Ready to Available. Agents can release, transfer, or conference calls from their agent desktop after activation completes. During conference tear down, a call appearance from the desk top of an active call but agent state is not affected. Calls that end while the PG is down end after a dead call time out after two hours.

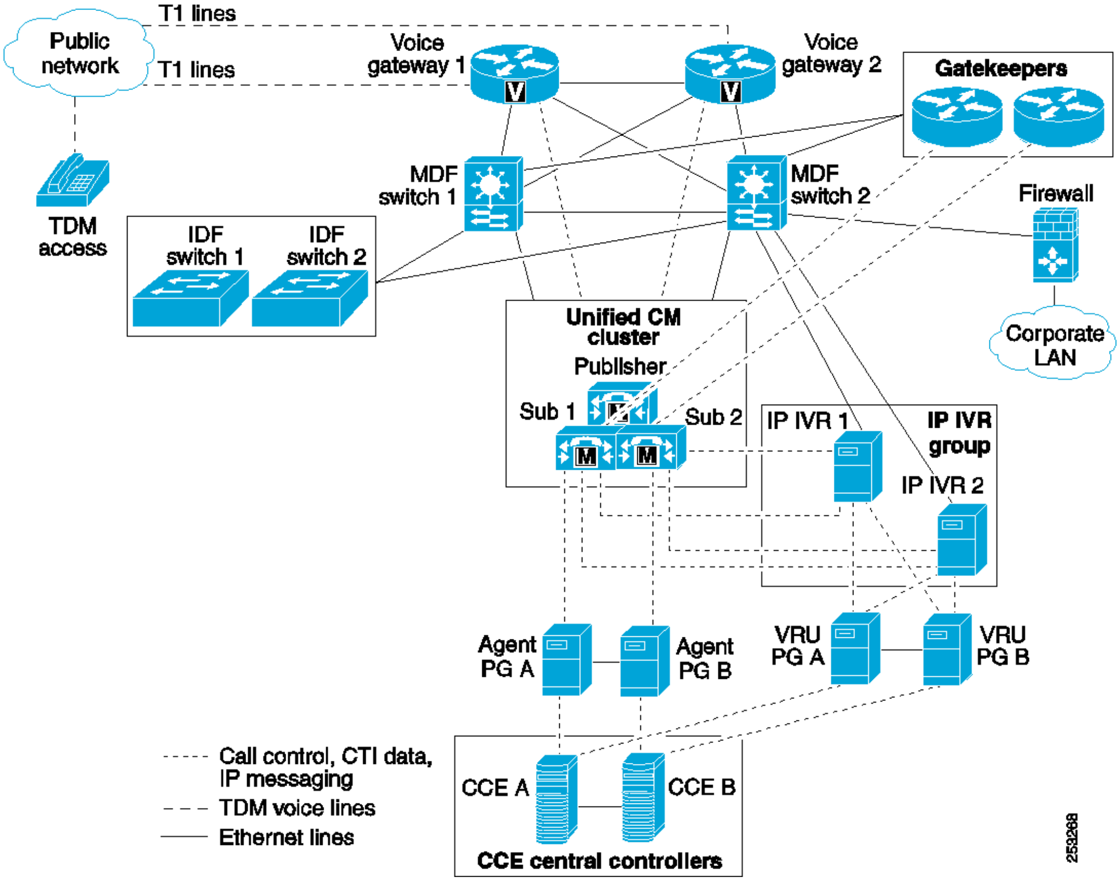
**Note** Call and agent state information might not be complete at the end of a failover if there are call status and agent state changes during the failover window.

## Unified CCE Voice Response Unit PG

When a Voice Response Unit (VRU) PG fails, all the calls currently in queue or treatment on that Unified IP IVR are dropped unless there is a default script application defined or the CTI Ports have a recovery number defined in Unified CM for their "Call Forward on Failure" setting. Calls in progress or queued in Unified CVP are not dropped and are redirected to a secondary Unified CVP or number in the H.323 or SIP dial plan, if available by the Survivability TCL script in the voice gateway.

The redundant (duplex) VRU PG side connects to the Unified IP IVR or Unified CVP and begins processing new calls upon failover. On recovery of the failed VRU PG side, the currently running VRU PG continues to operate as the active VRU PG. Therefore, having redundant VRU PGs adds significant value because it allows a Unified IP IVR or Unified CVP to continue to function as an active queue point or to provide call treatment. Without VRU PG redundancy, a VRU PG failure would block use of that IP IVR even though the IP IVR is working properly. (See Figure 65)

**Figure 65** ***Redundant Unified CCE VRU PGs with Two IP IVR Servers***



## Unified CCE Call Router and Logger

The Unified CCE Central Controllers or Unified CCE Servers are shown in these diagrams as a single set of redundant servers. However, depending on the size of the implementation, they could be deployed with multiple servers to host the following key software processes:

- Unified CCE Call Router

The Unified CCE Call Router is the brain of the system and it maintains a constant memory image of the state of all the agents, calls, and events in the system. It performs the call routing in the system; executing the user-created Unified CCE Routing Scripts and populating the real-time reporting feeds for the Administration & Data Server. The Call Router software runs in synchronized execution with both of the redundant servers running the same memory image of the current state across the system. They keep this information updated by passing the state events between the servers on the private LAN connection.

- Unified CCE Logger and Database Server

The Unified CCE Logger and Database Server maintain the system database for the configuration (agent IDs, skill groups, call types, and so forth) and scripting (call flow scripts) as well as the historical data from call processing. The Loggers receive data from their local Call Router process to store in the system database. Because the Call Routers are synchronized, the Logger data is also synchronized. In the event that the two Logger databases are out of synchronization, they can be resynchronized manually by using the Unified ICMDBA application over the private LAN. The Logger also provides a replication of its historical data to the customer Administration & Data Server over the visible network.

In the event that one of the Unified CCE Call Routers fails, the surviving server detects the failure after missing five consecutive TCP keep-alive messages on the private LAN. The Call Routers generate these TCP keep-alive messages every 100 ms, so it takes up to 500 ms to detect this failure. On detection of the failure, the surviving Call Router contacts the Peripheral Gateways in the system to verify the type of failure that occurred. The loss of TCP keep-alive messages on the private network could be caused by either of the following conditions:

- Private network outage — It is possible for the private LAN switch or WAN to be down but for both of the Unified CCE Call Routers to still be fully operational. In this case, the Peripheral Gateways still see both of the Unified CCE Call Routers even though they cannot see each other over the private network to provide synchronization data. If the disabled synchronizer (Call Router B) can communicate with a majority of the PGs, it then sends a Test Other Side (TOS) message to the PGs sequentially to determine if the Call Router on the other side (Side A) is enabled. If Call Router B receives a message that side A is in fact enabled, then Call Router A runs in simplex until the private network is restored. If all the PGs reply to the TOS message and indicate that side A is down, then side B re-initializes in simplex mode.
- Call Router hardware failure — It is possible for the Call Router on the other side to have a physical hardware failure and be completely out of service. In this case, the Peripheral Gateways report that they can no longer see the Call Router on the other side and the surviving Call Router takes over the active processing role in simplex mode. This failure is detected by the Call Routers from the loss of heartbeat keep-alives on the Private Network.

During Call Router failover processing, any Route Requests sent to the Call Router from a Carrier Network Interface Controller (NIC) or Peripheral Gateway are queued until the surviving Call Router is in active simplex mode. Any calls in progress in the IVR or at an agent are not impacted.

If one of the Unified CCE Logger and Database Servers were to fail, there is no immediate impact except that the local Call Router is no longer be able to store data from call processing. The redundant Logger continues to accept data from its local Call Router. When the Logger server is restored, the Logger contacts the redundant Logger to determine how long it had been off-line. If the Logger was off-line for less than 12 hours, it automatically requests all the transactions it missed from the redundant Logger while it was off-line. The Loggers maintain a recovery key that tracks the date and time of each entry recorded in the database and these keys are used to restore data to the failed Logger over the private network.

If the Logger was off-line for more than 12 hours, the system does not automatically resynchronize the databases. In this case, resynchronization is done manually using the Unified ICMDBA application.

Manual resynchronization allows the system administrator to decide when to perform this data transfer on the private network, perhaps scheduling it during a maintenance window when there is little call processing activity in the system.

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

There is no impact to call processing during a Logger failure; however, the historical data on the Administration & Data Server that is replicated from that Logger stops until the Logger is restored.

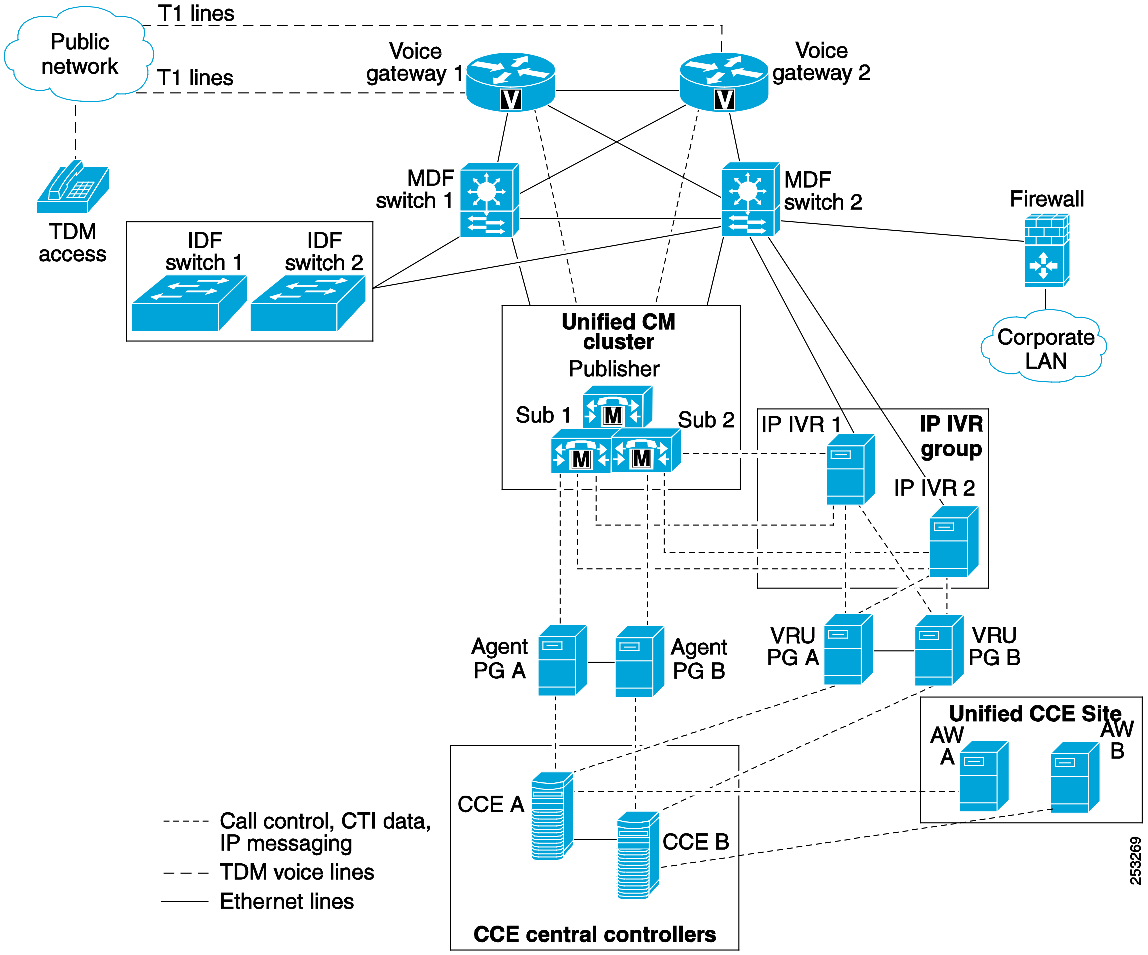
Additionally, if the Unified Outbound Option is used, the Campaign Manager software is loaded on Logger A only. If that platform is out of service, any outbound calling stops until the Logger is restored to operational status.

## Administration & Data Server

The Administration & Data Server provides the user interface to the system for making configuration and scripting changes. It can also host the web-based reporting tool and Internet Script Editor.

These servers do not support redundant or duplex operation as the other Unified CCE system components do. However, you can deploy multiple Administration & Data Servers to provide redundancy for Unified CCE. (See Figure 66)

**Figure 66** ***Redundant Unified CCE Administration & Data Servers***



Administration & Data Server Real-Time Distributors are clients of the Unified CCE Call Router real-time feed that provides real-time information about the entire Unified CCE across the enterprise. Real-Time Distributors at the same site can be set up as part of an Admin Site that includes a designated primary real-time distributor and one or more secondary real-time distributors. Another option is to add Administration Clients that do not have their own local SQL databases and are homed to a Real-Time Distributor locally for their SQL database and real-time feed.

The Admin Site reduces the number of real-time feed clients the Unified CCE Call Router has to service at a particular site. For remote sites, this is important because it can reduce the required bandwidth to support remote Administration & Data Servers across a WAN connection.

When using an Admin Site, the primary Administration & Data Server is the one that will register with the Unified CCE Call Router for the real-time feed and the other Administration & Data Servers within that Admin Site register with the primary Administration & Data Server for the real-time feed. If the primary real-time distributor is down, the secondary real-time distributors will register with the Unified CCE Call Router for the real-time feed. Administration Clients that cannot register with the primary or secondary Administration & Data Server will not be able to perform any tasks until the distributors are restored.

Alternatively, each Administration & Data Server could be deployed in its own Admin Site regardless of the physical site of the device. This deployment creates more overhead for the Unified CCE Call Router to maintain multiple real-time feed clients; however, it prevents a failure of the primary Administration & Data Server from taking down the secondary Administration & Data Server at the site.

Additionally, if the Administration & Data Server is used to host the ConAPI interface for the Cisco Unified Contact Center Management Portal (Unified CCMP), any configuration changes made to the Unified CCE or Unified CCMP systems are not passed over the ConAPI interface until it is restored.

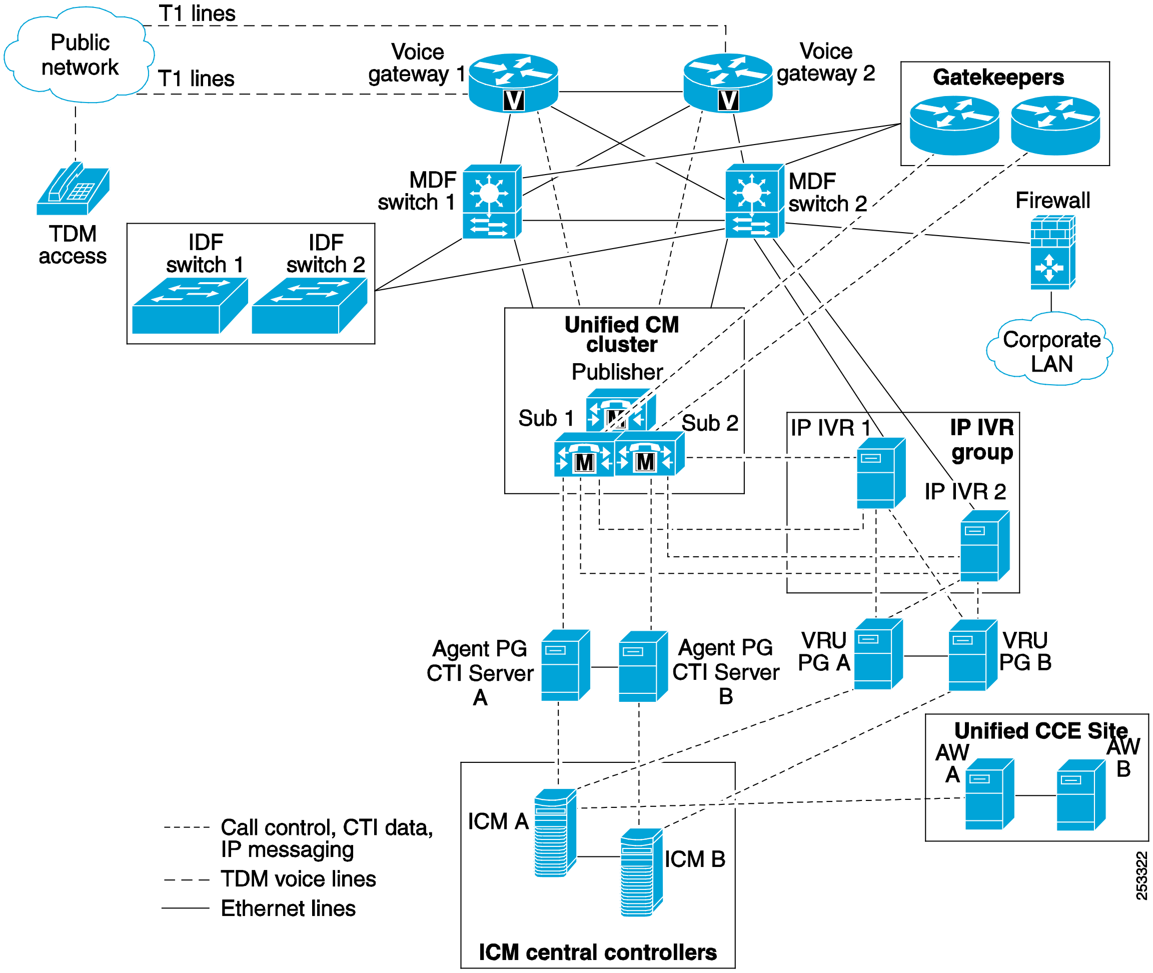
## CTI Server

The CTI Server monitors the data traffic of the Unified CM PIM on the Agent PG for specific CTI messages (such as call ringing or off-hook events) and makes those messages available to CTI clients such as the CTI OS Server or Cisco Agent Desktop Enterprise Server. It also processes third-party call control messages (such as make call or answer call) from the CTI clients and sends those messages by using the PIM interface of the PG to Unified CM to process the event on behalf of the agent desktop.

CTI Server is redundant and co-resident on the Agent PG servers. (See Figure 67) It does not, however, maintain agent state in the event of a failure. On failure of the CTI Server, the redundant CTI server becomes active and begins processing call events. Both CTI OS and Finesse Servers are clients of the CTI Server and are designed to monitor both CTI Servers in a duplex environment and maintain the agent state during failover processing. CTI OS agents see their desktop buttons dim during the failover to prevent them from attempting to perform tasks while the CTI Server is down. The buttons are restored as soon as the redundant CTI Server is restored and the agent does not have to log on again to the desktop application. Most call context is maintained, but ANI and DNIS are lost in this instance where only the CTI Server component is impacted.

Finesse servers return an Out of Service status to clients during the failover, preventing the clients from initiating actions. The Finesse Desktop user interface retains its last state until the redundant CTI Server is restored, at which time the Finesse server updates each client with the current CTI state.

**Figure 67** ***Redundant CTI Servers Coresident on Agent PG***

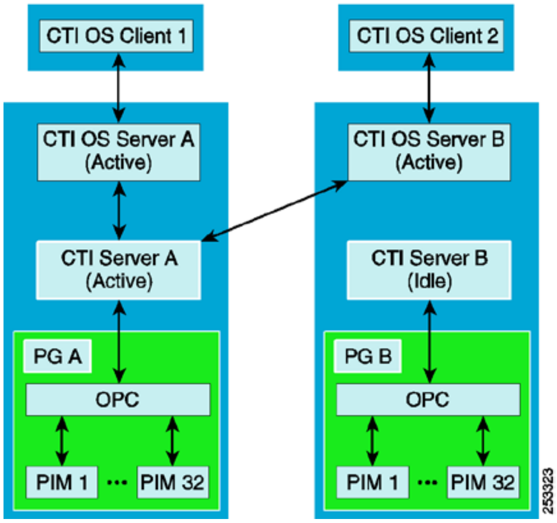


# CTI OS Considerations

CTI OS Server is a software component that runs co-resident on the Unified CM Peripheral Gateway. CTI OS Server software is designed to be fault-tolerant and is typically deployed on redundant physical servers; however, unlike the PG processes that run in hot-standby mode, both of the CTI OS Server processes run in active mode all the time. The CTI OS Server processes are managed by Node Manager, which monitors each process running as part of the CTI OS service and which automatically restarts abnormally terminated processes.

CTI OS handles failover of related components as described in the following scenarios (see Figure 68).

**Figure 68** ***Redundant CTI OS Server Processes***



### Scenario 1: CTI Server Side A (Active) Fails

In this scenario, CTI Server side A is co-resident on PG side A, and the following events occur:

- CTI Server side B detects the failure of side A and becomes active.
- Node Manager restarts CTI Server side A and becomes idle.
- Both CTI OS Server sides A and B drop all CTI OS client and agent connections and restart after losing the connection to CTI Server A. At startup, CTI OS Server sides A and B stay in CONNECTING state until they connect to CTI Server side B, and then they go into CONFIGURING state where they download agent and call states and configuration information. CTI OS Client connections are not accepted by CTI OS Server A and B during CONNECTING and CONFIGURING states. When CTI OS Server synchronizes with CTI Server, the state becomes ACTIVE and it is now ready to accept CTI OS Client connections.
- Both CTI OS Clients 1 and 2 lose connections to CTI OS Servers and they each randomly select one CTI OS Server to connect to. CTI OS Client 1 can be connected to either CTI OS Server A or B, and the same is true for CTI OS Client 2. During this transition, the buttons of the CTI Toolkit Agent Desktop will be disabled and will return to operational state as soon as it is connected to a CTI OS.

### Scenario 2: CTI Server B (Idle) Fails

In this scenario, CTI Server side B is co-resident on PG side B but is not the active side. The following events occur:

- CTI Server side A stays active.
- Node Manager restarts CTI Server side B and stays idle.
- Neither CTI OS Clients nor CTI OS Servers are affected by this failure.

### Scenario 3: CTI OS Server A Fails

In this scenario, CTI OS Server side A processes are co-resident on PG/CTI Server side A. The following events occur:

- CTI OS Client 1 detects the loss of network connection and automatically connects to CTI OS server B. During this transition, the buttons of the CTI Toolkit Agent Desktop are disabled and return to the operational state as soon as it is connected to CTI OS server B.
- CTI OS Client 2 stays connected to CTI OS Server B.
- NodeManager restarts CTI OS Server A.

**Scenario 4: CTI OS Server B Fails**

In this scenario, CTI OS Server side A processes are co-resident on PG/CTI Server side B. The following events occur:

- CTI OS Client 2 detects the loss of network connection and automatically connects to CTI OS server A. During this transition, the buttons of the CTI Toolkit Agent Desktop are disabled and return to the operational state as soon as it is connected to CTI OS server A.
- CTI OS Client 1 stays connected to CTI OS Server A.
- NodeManager restarts CTI OS Server B.

**Scenario 5: CTI OS Client 1 Fails**

In this scenario, the following events occur:

- The agent manually restarts CTI OS Client 1 application.
- CTI OS Client 1 randomly selects one CTI OS Server to connect to. (CTI OS Client 1 can be connected to either CTI OS Server A or B.)
- Once connected, the agent logs in and CTI OS Client 1 recovers its state by getting agent and call states through the CTI OS Server to which it is connected.

**Scenario 6: CTI OS Client 2 Fails**

In this scenario, the following events occur:

- The agent manually restarts CTI OS Client 2 application.
- CTI OS Client 2 randomly selects one CTI OS Server to connect to. (CTI OS Client 2 can be connected to either CTI OS Server A or B.)
- Once connected, the agent logs in and CTI OS Client 2 recovers its state by getting agent and call states through the CTI OS Server to which it is connected.

**Scenario 7 - Network Failure Between CTI OS Client 1 and CTI OS Server A**

In this scenario, the following events occur:

- CTI OS Server A drops the connection of CTI OS Client 1
- CTI OS Client 1 detects the loss of network connection and automatically connects to CTI OS server B. During this transition, the buttons of the CTI Toolkit Agent Desktop are disabled and return to the operational state as soon as it is connected to CTI OS server B.

**Scenario 8: Network Failure Between CTI OS Client 1 and CTI OS Server B**

CTI OS Client 1 is not affected by this failure because it is connected to CTI OS Server A.

### Scenario 9: Network Failure Between CTI OS Client 2 and CTI OS Server A

CTI OS Client 2 is not affected by this failure because it is connected to CTI OS Server B.

### Scenario 10: Network Failure Between CTI OS Client 2 and CTI OS Server B

In this scenario, the following events occur:

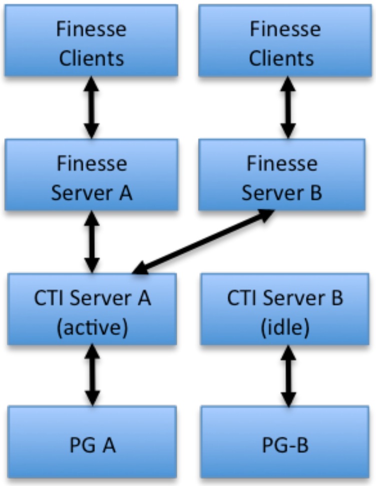
- CTI OS Server B drops the connection of CTI OS Client 2.
- CTI OS Client 2 detects the loss of network connection and automatically connects to CTI OS server A. During this transition, the buttons of the CTI Toolkit Agent Desktop are disabled and return to the operational state as soon as it is connected to CTI OS server A.

# Cisco Finesse Considerations

Cisco Finesse is a software component that runs exclusively in a dedicated virtual machine, either stand-alone or coresident on the Unified CM Peripheral Gateway server (if virtualized). The Cisco Finesse software is designed to be fault-tolerant and is deployed on redundant physical servers. Both Finesse servers are active. Cisco Finesse runs on the Cisco VOS platform, where the local servm process restarts any failed Finesse processes.

Finesse handles failover of related components as described in the following scenarios (see Figure 69).

**Figure 69** ***Redundant Finesse Processes***



### Scenario 1: CTI Server Side A (Active) Fails

In this scenario, CTI Server side A is coresident on PG side A, and the following events occur:

- CTI Server side B detects the failure of side A and becomes active.
- Node Manager restarts CTI Server side A and becomes idle.
- Both Finesse Server side A and side B go into OUT OF SERVICE state and do not allow any further client actions or sign-in attempts. The Finesse servers attempt to reconnect to CTI Server A and, on a reconnect failure, attempt to connect to CTI Server B. After the connection to CTI Server B is achieved, both Finesse servers attempt to rebuild their internal agent state through CTI

Server B. After the internal state is rebuilt, the Finesse servers transition to IN SERVICE and allow new sign-in attempts, as well as any CTI actions. Each connected Finesse client also receives an XMPP update with the refreshed agent state as received from CTI Server B, including the updated call state as received from CTI Server B

#### Scenario 2: CTI Server B (Idle) Fails

In this scenario, CTI Server side B is coresident on PG side B but is not the active side. The following events occur:

- CTI Server side A stays active.
- Node Manager restarts CTI Server side B and stays idle.
- Finesse clients and Finesse servers are not affected by this failure.

#### Scenario 3: Finesse Server A Fails

In this scenario, Finesse server A fails. The following events occur:

- The Finesse Desktop application detects the loss of network connection to the Finesse server and automatically connects to Finesse server B. Failover is handled through an HTML redirect to the sign-in page of the Finesse Desktop application on Finesse server B. Agents are prompted to enter their sign-in credentials.
- After the agents sign in to Finesse server B, their desktops are updated to reflect their current state.
- All Finesse clients remain on Finesse server B until they sign out of their current session.
- Third- party applications that use the Finesse REST API must perform the failover within their application logic to move to Finesse server B. Finesse provides a REST API call that contains the host addresses of Finesse servers A and B.
- If the cause of the failure on Finesse server A was the Cisco Tomcat process stopping for any reason, the local servm process attempts to restart Cisco Tomcat.

#### Scenario 4: Finesse Server B Fails

In this scenario, all Finesse clients failover to Finesse server A with same set of server state transitions and client updates as in Scenario 3.

#### Scenario 5: Finesse Client Fails

Finesse runs in a web browser on the client desktop. If the browser process fails for any reason, the following occurs:

- After a 10-second timeout, the Finesse server automatically instructs the CTI Server to sign out the user.
- If the user is currently active on a call, the automatic sign-out occurs after the call is complete.
- The user may restart the browser and sign in again to the Finesse server. The desktop is then updated to reflect the current state of the user.

#### Scenario 6 - Network Failure Between Finesse Client and Finesse Server A

In this scenario, the following events occur:

- Finesse server A automatically signs the user out of the CTI Server unless the user is active on a call.

- The Finesse client detects the loss of connectivity and automatically redirects the browser to Finesse server B.
- The user is prompted for sign-in credentials.
- The Finesse Desktop application loads from server B and the current state of the user is updated (including agent state and call state).

**Scenario 7: Network Failure Between Finesse Client and Finesse Server B**

This scenario mirrors scenario 6. The Finesse client automatically redirects the user to Finesse Server A to sign in again.

# Cisco Agent Desktop Considerations

## Cisco Agent Desktop

Cisco Agent Desktop client applications are a client of CTI OS, which provides for automatic failover and redundancy for the Cisco Agent Desktop CTI connections. If the Unified CM 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 CM 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 failover of the core Cisco Agent Desktop components. The Cisco Agent Desktop client applications are aware of the redundant Cisco Agent Desktop services and automatically failover 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 CAD client applications as long as network connectivity is available:

- Cisco LDAP Monitor Service
- Cisco Recording & Playback Service
- Cisco VoIP Monitor Service

**Scenario 1: Cisco Agent Desktop License and Resource Manager Service Side A (Active) Fails**

In this scenario, the following events occur:

- The CAD services on side A that are not always active go to an idle state.
- The CAD services on side B (idle) activate.
- The CAD client applications recover to side B.

#### Scenario 2: A Cisco Agent Desktop Service on Side A (Active) Fails Twice Within Five Minutes

In this scenario, the following events occur:

- The CAD services on side A that are not always active go to an idle state.
- The CAD services on side B (idle) activate.
- The CAD client applications recover to side B.

#### Scenario 3: A Cisco Agent Desktop Service on Side A (Active) Fails and Remains Down for Three Minutes

In this scenario, the following events occur:

- The CAD services on side A that are not always active go to an idle state.
- The CAD services on side B (idle) activate.
- The CAD client applications recover to side B.

#### Scenario 4: Network Failure Between CAD Services on Side A (Active) and CAD Services on Side B (Idle)

In this scenario, the following events occur:

- The CAD services on side A remain active.
- The CAD services on side B (idle) activate.
- The CAD client applications remain connected to CAD 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.

## Cisco Agent Desktop Browser Edition and IP Phone Agent

Cisco Agent Desktop Browser Edition and IP Phone Agent communicate with CTI Server through the Cisco Browser and IP Phone Agent service. When launching CAD-BE, the agent may use the URL for either side as long as the desired side is accessible. Once launched, CAD-BE automatically connects to the active side. When launching IPPA, 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.

#### Scenario 1: Cisco Agent Desktop Services on Side A (Active) Fail and Side B (Idle) Becomes Active

In this scenario, the following events occur for a logged-in CAD-BE agent:

- The CAD-BE applet changes to a logged out state and the user is notified that the connection has been lost.
- The CAD-BE applet automatically connects to services on side B and logs-in the agent.

In this scenario, the following events occur for a logged-in IPPA agent:

- The IPPA agent is notified that the connection to the server has been lost.
- The IPPA agent manually selects side B from their services list and logs in again.

### Replacement of MSDE with Flat Files

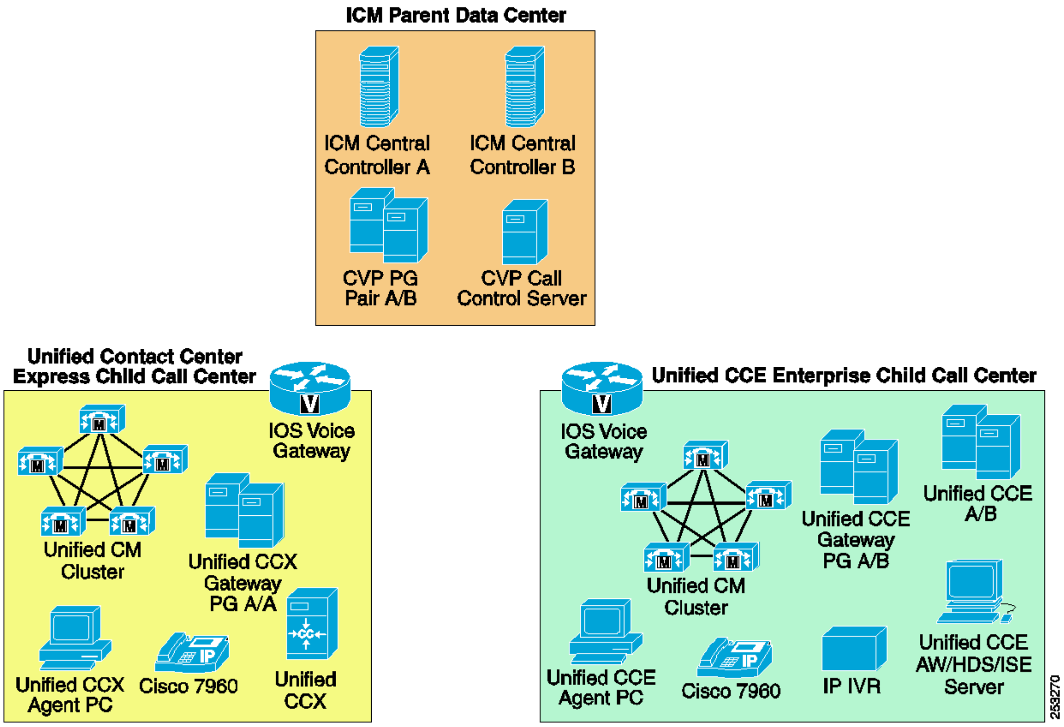
As MSDE is no longer supported, at post install time the user can choose flat files or a full SQL database. Post install configures their system based on their selection. There is a value stored in LDAP that indicates which implementation is selected. After the initial configuration is completed (the implementation is selected and saved), the user cannot change implementations. For the database implementation, Unified CCE configures the FCRasSvr database, as it does now. Unified CCE continues to provide scripts for setup and teardown of database replication for HA. In the database implementation, there are three tables: agent state data, call log data, and recording metadata. In the flat-file implementation, each of these tables is represented by a set of text files. Regardless of which implementation is used in a particular installation, the data stored is identical.

There are currently no minimum or maximum file sizes set.

## Design Considerations for Unified CCE Deployment with Unified ICM Enterprise

The parent/child deployment is where the Unified ICM acts as the parent controlling one or more Unified CCE child IP ACDs. (See Figure 70) In this model, the Unified ICM Enterprise system is designed to be the network call routing engine for the contact centers, with network queuing using the Unified CVP and Unified CCE Gateway Peripheral Gateways to connect child Unified CCE systems (either Unified CCE with system PG or Unified CCX). The child Unified CCE systems are individual IP-ACD systems fully functional with local call processing in case they lose their WAN connection to the parent Unified ICM system. This configuration provides a high level of redundancy and availability to the Unified CCE solution to allow sites to remain functional as Unified CCE sites even if they are cut off from centralized call processing resources.

**Figure 70** *Parent/Child Deployment Model*



## Parent/Child Components

The following sections describe the components used in Unified ICM Enterprise (parent) and Unified CCE System (child) deployments.

## The Unified ICM Enterprise (Parent) Data Center

The Unified ICM parent data center location contains the Unified ICM Central Controller. In Figure 70, it is shown as a redundant (duplex) pair of Central Controllers which represents Call Router and Logger servers. These servers can be deployed as individual Call Routers and Loggers and they can be deployed in two different data centers to be geographically distributed for additional fault tolerance.

The Unified ICM Central controllers control Peripheral Gateways at the data center location. In Figure 70, there is only a redundant (duplex) pair of IVR PGs used to control Unified CVP across the architecture. Additional PGs can be inserted at this layer to control TDM or legacy ACDs and IVRs, perhaps to support a migration to Unified CCE or to support out-source locations that still use the TDM or legacy ACDs. The Unified ICM parent at this level can also support standard pre-routing with inter-exchange carriers (IXCs) such as AT&T, MCI, and others allowing Unified ICM to select the best target for the call while it is still in the carrier network.

The Unified ICM parent is not designed to support any directly controlled agents in this model, which means that it does not support classic Unified CCE with a Unified CM Peripheral Gateway installed on this Unified ICM parent. All agents must be controlled externally to this Unified ICM parent system.

The Unified CVP or IVR PG pair controls the Customer Voice Portal Call Server which translates the IVR PG commands from Unified ICM into VoiceXML and directs the VoiceXML to the voice gateways at the remote contact center sites. This allows calls from the data center location to 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 and holds the calls in queue on the voice gateways at the sites until an agent becomes available.

## The Unified Contact Center Express (CCX) Call Center (Child) Site

The Unified Contact Center Express (CCX) Call Center location contains a local Unified CM cluster that provides local IP-PBX functionality and call control for the IP phones and local Unified CVP voice gateway. There is also a local Unified CCX Server that provides IP-ACD functionality for the site. Prior to Unified CCX Server Release 8.0, the Unified CCX Server had the Unified CCE Gateway PG installed on it, which reduces the number of servers required to support this contact center site.

Unified CCX 8.0(1) is deployed on the Unified Communications Operating System platform requiring the Unified CCE Gateway PG to be installed on a separate (Windows) server. The deployment model changes for new and existing customers require the Unified CCE Gateway PG and the CCX ACMI Manager to be installed on separate (Windows) servers. In either of these deployments, the Unified CCE Gateway PG connects to the Unified ICM Call Router (Rogger) at the Unified ICM parent data center location over the WAN and provides real-time event data and agent states to the parent from the Unified CCX. The Unified CCE Gateway PG also captures configuration data (skill groups, CSQs, services, applications, and so forth) and sends it to the parent Unified ICM configuration database as well.

Additional Unified CCX servers may be used and included in this site to provide redundant Unified CCX Servers, historical reporting database services, recording and monitoring servers, and ASR/TTS servers. High availability deployments of Unified CCX Release 8.0 or above require the deployment of two "SideA" Unified CCE Gateway PGs on separate (Windows) servers. The Unified CCX Servers are configured with the IP Addresses of the two "SideA" Unified CCE Gateway PGs.

## The Unified CCE Call Center (Child) Site

The Unified CCE Call Center location contains a local Unified CM cluster that provides local IP-PBX functionality and call control for the IP phones and local Unified CVP voice gateway. There is also a local Unified IP IVR to provide local call queuing for the Unified CCE site. There is a redundant pair of Unified CCE Gateway PGs that are used to connect this site to the Unified ICM parent Central Controller at the Unified ICM parent data center location over the WAN. The Unified CCE Gateway PGs can be deployed on separate servers or co-resident 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 the PG number for the Unified CCE Gateway PG and Unified CCE System PG must be different.

If the Unified CCE Gateway PG and Unified CCE System PG Instance Numbers are different, then the PG number for Unified CCE Gateway PG and Unified CCE System PG may be the same.

No additional PGs (such as VRU PG or MR PG) can be added to this Server.

For scalability limits of the co-resident Unified CCE Gateway PG and Unified CCE System PG, refer to Sizing Unified CCE Components and Servers for additional details.

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 configuration data (skill groups, services, call types, and so forth) and send it to the parent Unified ICM configuration database as well.

The IP-IVR at the Child site can be replaced with a local Unified CVP instance. Unified CVP is not integrated as part of the Agent Controller's System PG; there is a separate IVR PG defined specifically for Unified CVP as part of the installation for System CCE with Unified CVP. Because Unified CVP is not part of the System PG, calls in queue or treatment in Unified CVP are not reported to the Parent ICM through the Unified CCE Gateway PG.

A local Unified CCE child system is used to provide IP-ACD functionality and it can be sized depending on the type of deployment required:

- Progger configuration

  Single (or duplex) server that contains the Unified CCE components: Call Router and Logger, System PG for Unified CM and IP IVR, CTI Server and CTI OS Server, and optionally the VRU PG for Unified CVP.

- Rogger configuration with separate Unified CCE Agent Controller (System PG and optional Unified CVP controller and CTI/CTI OS Server)

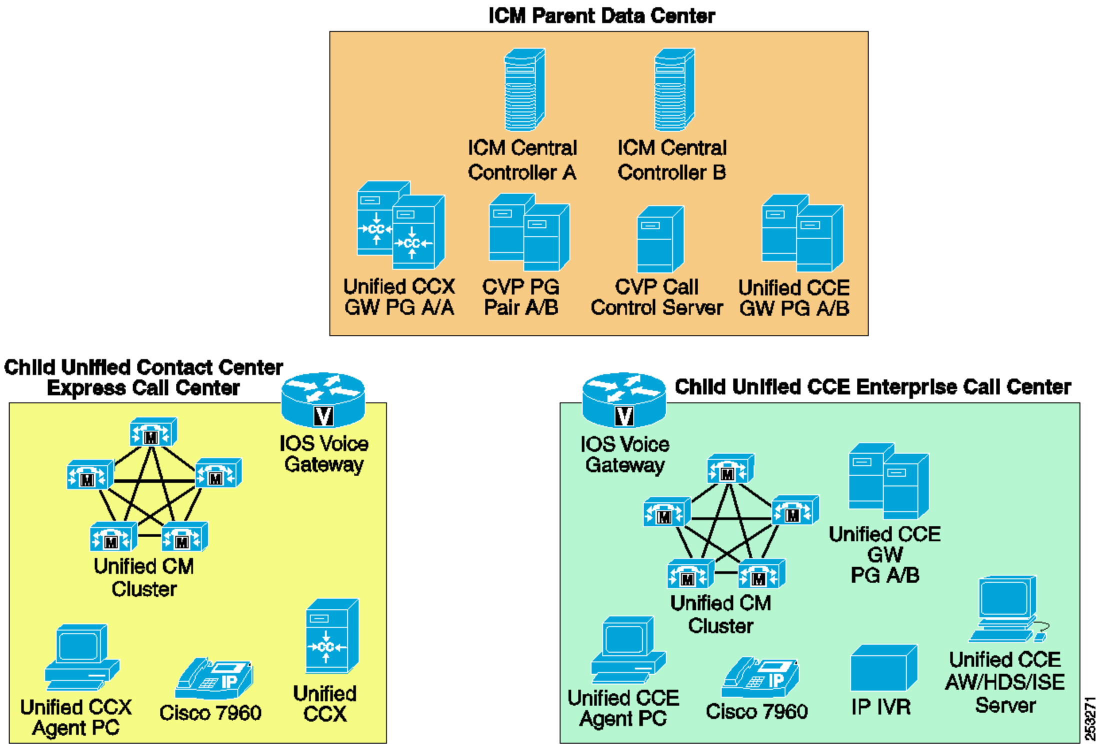
  The Rogger configuration contains the Unified CCE components: Call Router and Logger as a single set of duplex Central Controllers, and a separate Agent Controller set of duplex servers that contain the System PG for Unified CM and IP IVR, CTI Server and CTI OS Server, and the optional VRU PG for Unified CVP.

For more details about the capacity of these configurations, refer to Sizing Unified CCE Components and Servers.

In either configuration, a separate Administration & Data Server is required to host the configuration and scripting tools for the system as well as an optional Historical Database Server role and Web based Unified Intelligence Center reporting tool.

## Unified ICM Enterprise (Parent) with Unified CCE Gateway PGs at Data Center

**Figure 71** *Parent/Child Deployment Model with Unified CCE Gateway PGs at Data Center*



The Unified CCE Gateway PG may be deployed at the ICM Parent Data Center as illustrated in Figure 71. Some advantages with this deployment model include centrally managing and controlling Unified CCE Gateway PGs as well as configuring Unified CCE Gateway PGs with multiple PIMs to help reduce the server TCO requirements. Another condition forcing the Unified CCE Gateway PG to be deployed at the ICM Parent Data Center is where the ownership and management of the Unified CCE and Parent Data Center are different. For example the Unified CCE is managed by an Outsourcer/Service Bureau and the Parent manages the Unified CCE Gateway PG.

There are several drawbacks with moving the Unified CCE Gateway PGs to the Data Center. One is specific to recovering reporting data in the event of a network failure. If the network connection between the Parent Data Center and the Child Unified CCE System PGs drops, all reporting at the parent is lost for that period.

**Note** If the Unified CCE Gateway PG is deployed locally to the Unified CCE System PG and the connection between the Unified CCE Gateway PG and the Parent Data center drops, the parent historical data is updated when the network connection is restored.

A second drawback with centralizing the Unified CCE Gateway PGs is that the network bandwidth requirements for the connections between the Parent CCE Gateway PG and the Child CCE System PG are significantly higher. See the "Bandwidth Requirements for Unified CCE Gateway to System PG" section in the Bandwidth Provisioning and QoS Considerations chapter for additional details.

## Parent/Child Call Flows

The following sections describe the call flows for the parent and child.

## Typical Inbound PSTN Call Flow

In a typical inbound call flow from the PSTN, calls would be directed by the carrier network to the contact center sites using some predefined percent allocation or automatic routing method. These calls are terminated in the Unified CVP voice gateways at the call center locations under control of the Unified ICM parent Unified CVP. The inbound call flow is as follows:

1. The call arrives on the Unified CVP voice gateway at the Unified CCE call center location.
2. The Unified CVP voice gateway maps the call by dialed number to a particular Unified CVP Call Server at the Unified ICM parent site and sends a new call event to the Unified CVP Call Server.
3. The Unified CVP Call Server sends the new call event message to the Unified CVP or IVR PG at the Unified ICM parent site.
4. The Unified CVP PG sends the new call message to the Unified ICM parent, which 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 voice gateway at the site and wait for an available agent while directing specific instructions to play .wav files for hold music to the caller in the gateway.
6. When an agent becomes available, the Unified ICM instructs Unified CVP to transfer the call to the site with the available agent by using a translation route. (The agent might not be at the same physical site but across the WAN.) Any data collected about the call in the Unified ICM parent Unified CVP will be transferred to the remote system's PG (either a TDM, legacy PG, or one of the Unified CCE Gateway PGs for Unified CCX or Unified CCE).
7. When the call arrives at the targeted site, it arrives on a specific translation route DNIS that was selected by the Unified ICM parent. The PG at the 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 local PG 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 for the call (walked away or unplugged), Unified CVP at the Parent site uses the Router Re-query function in the ICM Call Routing Script to select another target for the call automatically.

## Post-Route Call Flow

Post-routing is used when a call is already at a peripheral ACD or IVR and needs to be routed intelligently to another agent or location. If an agent gets a call in the ACD or Unified CCE that needs to be sent to a different skill group or location, the agent can make use of 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 will return a label or routing instruction that can move the call to another site, or to the same site but into a different skill group, or to a Unified CVP node for queuing.
4. The Unified CCE receives the post-route response from the Unified ICM parent system and 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 IP-ACD with either Unified CCX or Unified CCE deployed at the site, with local IP-PBX and call treatment and queuing functionality.

## Unified CCE Child Loses WAN Connection to Unified ICM Parent

If the WAN between the Unified CCE child site and the Unified ICM parent fails, the local Unified CCE system will be isolated from the parent as well as the Unified CVP voice gateway. Calls coming into the site will no longer get treatment from the Unified CVP under control of the Unified ICM parent, so the following functionality must be replicated locally, depending on the Child configuration.

- For Unified CCE Child configurations using local IP IVR resources for queue and treatment:
  - The local voice gateway must have dial peer statements to pass control of the calls to the local Unified CM cluster if the Parent Unified CVP Call Server cannot be reached. Also, the local Unified CM cluster must have CTI route points mapped to the inbound DNIS or dialed numbers that the local voice gateway will present if the Parent Unified CVP Call Server is not reached.
  - The local IP IVR must be configured with appropriate .wav files and applications that can be called by the Unified CCE Child system locally to provide basic call treatment such as playing a welcome greeting or other message.
  - The Child CCE Routing Script must handle queuing of calls for agents in local skill groups, instructing the IP IVR to play treatment in-queue while waiting for an agent.
  - Any data lookup or external CTI access that is normally provided by the Parent Unified CVP or the Parent Unified ICM must be provisioned locally to allow the agents to have full access to customer data for routing and screen pops.
  - Any post-routing transfer scripts will fail during this outage, so Unified CCE must be configured 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 with Unified CCE 7.5*(x)*:
  - The local voice gateway must have dial peer statements to pass control of the calls to the local Unified CVP Call Server at the Child site. Also, the inbound DNIS or dialed numbers that the local voice gateway will present to the Child Unified CVP must be configured in the Child Unified CCE to process these calls locally at the Child.
  - The local VXML Gateways and Unified CVP Call Servers must be configured with appropriate .wav files and applications that can be called by the Unified CCE Child system locally to provide basic call treatment such as playing a welcome greeting or other messages.

  - Self-service or Unified CVP Studio VXML applications normally provided by the Parent Unified ICM must be replicated using Unified CVP VXML 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, instructing the local Unified CVP at the Child site to play treatment in-queue while waiting for an agent.
  - Any data lookup or external CTI access that is normally provided by the Parent Unified CVP or the Parent Unified ICM must be provisioned locally to allow the agents to have full access to customer data for call routing and screen pops.
  - Any post-routing transfer scripts will fail during this outage, so Unified CCE must be configured to handle this outage or prevent the post-route scripts from being accessed.

## Unified Contact Center Express Child Loses WAN Connection to Unified ICM Parent

If the WAN between the Unified Contact Center Express (CCX) child site and the Unified ICM parent fails, the local Unified CCX system will be isolated from the parent as well as the Unified CVP voice gateway. Calls coming into the site will no longer get treatment from the Unified CVP under control of the Unified ICM parent, so the following functionality must be replicated locally:

- The local voice gateway must have dial peer statements to pass control of the calls to the local Unified CM cluster if the Parent Unified CVP Call Server cannot be reached.
- Unified CCX JTAPI applications have to be mapped to these CTI route points to provide any typical inbound call treatment, such as playing a welcome greeting or other message.
- The application has to provide for call queuing and treatment in queue while waiting for a local Contact Service Queue (CSQ) agent.
- Any data lookup or external CTI access that is normally provided by the Parent Unified CVP or the Parent Unified ICM must be provisioned locally to allow the agents to have full access to customer data for call routing and screen pops.
- Any post-routing applications or transfer scripts will fail during this outage, so the Unified CCX must be configured to handle this outage or prevent the post-route applications from being accessed.

A similar failure would occur if the local Unified CVP ingress voice gateways controlled by the Parent Unified CVP Call Server could not see the Unified ICM Parent CVP Call Servers. The local Unified CVP gateways would be configured to fail-over to the local Unified CM (or Child Unified CVP) to route calls to the Unified CCX agents as described above. Likewise, if the entire Unified ICM parent were to fail, the local voice gateways controlled by the Parent Unified CVP at the sites would no longer have call control from the Unified ICM parent, and calls would forward to the local sites for processing.

## Unified CCE Gateway PG Fails or Cannot Communicate with Unified ICM Parent

If the Unified CCE gateway PG fails or cannot communicate with the Unified ICM parent, the local agents are no longer seen as available to the Unified ICM parent, but the inbound calls to the site may still be under control of the Unified ICM parent Unified CVP. In this case, the Unified ICM parent will not know if the remote Unified CCE gateway PG has failed or if the actual Unified CCE IP-ACD has 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 or Unified CCX 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/CCX system.

If the local Unified CCE or Unified CCX child system were to fail, the Unified CCE gateway PG would not be able to connect to it, and the Unified ICM parent would then consider all of the agents to be off-line and not available. If calls were sent to the local Unified CM while the child Unified CCE or Unified CCX system was down, the call-forward-on-failure processing would take over the call for the CTI route point. This method would redirect the call to another site or an answering resource to play a message telling the caller there was an error and to call again later.

## Parent/Child Reporting and Configuration Impacts

During any time that the Unified CCE child is disconnected from the Unified ICM parent, the local IP-ACD is still collecting reporting data and allows local users to make changes to the child routing scripts and configuration. The Unified CCE gateway PG at the child site will cache these objects and store them in memory (and eventually to disk) to be sent later to the Unified ICM parent when it is available. This functionality is available only if the Unified CCE gateway PG is co-located at the child Unified CCE site.

## Other Considerations for the Parent/Child Model

Multichannel components such as EIM/WIM and Unified Outbound Option may be installed only at the child Unified CCE level, not at the parent. They are treated as nodal implementations on a site-by-site basis.

# Other Considerations for High Availability

A Unified CCE failover can affect other parts of the solution. Although Unified CCE may stay up and running, some data could be lost during its failover, or other products that depend on Unified CCE to function properly might not be able to handle a Unified CCE failover. This section examines what happens to other critical areas in the Unified CCE solution during and after failover.

### Reporting

The Unified CCE reporting feature uses real-time, five-minute and reporting interval (15 or 30 minute) data to build its reporting database. Therefore, at the end of each five-minute and reporting interval (15 or 30 minute), each Peripheral Gateway will gather the data it has kept locally and send it to the Call Routers. The Call Routers process the data and send it to their local Logger for historical data storage. That data is then replicated to the HDS database from the Logger as it is written to the Logger database.

The Peripheral Gateways provide buffering (in memory and on disk) of the five-minute and reporting interval (15 or 30 minute) data collected by the system to handle network connectivity failures or slow network response as well as automatic retransmission of data when the network service is restored. However, physical failure of both Peripheral Gateways in a redundant pair can result in loss of the half-hour or five-minute data that has not been transmitted to the Central Controller. Use redundant Peripheral Gateways to reduce the chance of losing both physical hardware devices and their associated data during an outage window.

When agents log out, all their reporting statistics stop. The next time the agents log in, their real-time statistics start from zero. Typically, Central Controller failover does not force the agents to log out or reset their statistics; however, if the PG fails-over, their agent statistics are reset because the PIM and OPC processes that maintain these values in memory are restarted. If the CTI OS or CAD servers do not fail-over or restart, the agent desktop functionality is restored to its pre-failover state.

For further information, see the *Reporting Guide for Cisco IPCC Enterprise & Hosted Editions*.

CHAPTER 4

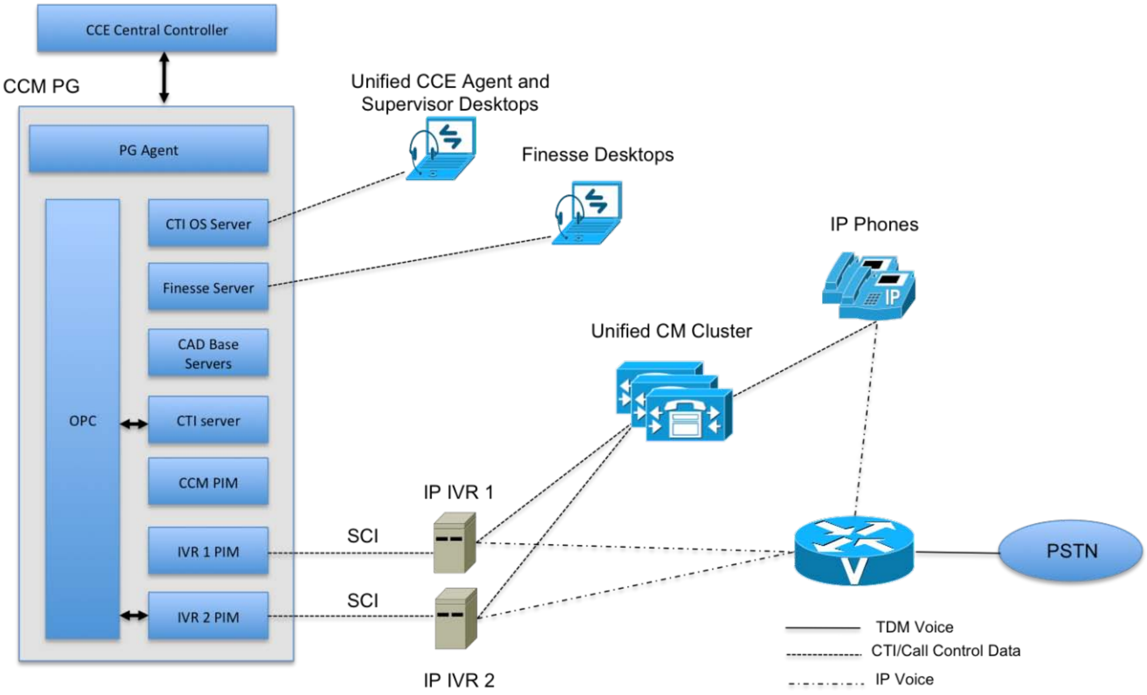
# Unified Contact Center Enterprise Desktop

The Cisco Unified Contact Center Enterprise (CCE) solution delivers a comprehensive set of desktop applications and services.

## Desktop Components

The desktop applications themselves typically run on Agent desktops, Supervisor desktops, Administration & Data Servers or Administration Client. Services supporting the desktop applications typically run on the Unified CCE Peripheral Gateway (PG) server. Within the Unified CCE deployment, there may be one or more PG systems, and for each PG there is one set of active desktop services, which includes the CTI Object Server (CTI OS) and the Cisco Agent Desktop Base Services for Cisco Agent Desktop (CAD) deployments. Figure 72 depicts the components within a Unified CCE deployment that support the various desktop applications.

**Figure 72** *Generic Unified CCE Desktop Components*



In the Unified CCE solution, the Peripheral Gateway may be deployed in either a simplex or duplex configuration. (Simplex mode is not supported for production environments.) Duplex configurations provide redundant desktop services for failover recovery support. These systems are typically identified as the primary, or A-side, and the backup, or B-side. For production deployments, a duplex configuration is required.

## 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 and Cisco Agent Desktop (CAD) solutions and is Cisco's latest version of the CTI implementation.

Communications from the desktop applications, such as agent state change requests and call control, are passed to the CTI OS server running on the Cisco Unified Peripheral Gateway. CTI OS serves as a single point of integration for CAD desktops, 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 CTI Server over TCP/IP and forwards call control and agent requests to CTI Server, which in turn forwards to the Open Peripheral Controller (OPC). From there, depending on the type of request, OPC will forward to either the CCM Peripheral Interface Manager (PIM) or to the Unified CCE Central Controller.

Requests initiated from the desktop application that affect the agent state are sent to the Unified CCE Central Controller, while requests initiated from the desktop application that affect call control are sent to the CCM PIM. The Unified CCE Central Controller monitors the agent state so that it knows when it can and cannot route calls to that agent and can report on that agent’s activities.

Call control flows from the agent desktop application to Cisco Unified Communications Manager (Unified CM). Unified CM then performs the requested call or device control. The desktop services located on the PG keep the agent desktop application synchronized with the agent's IP phone state.

CTI Toolkit desktop configuration and behavior information is also managed at the CTI OS server, 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 Client Interface Library (CIL) Software Development Kit (SDK) can utilize 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 login 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/stop the VoIP packet stream forwarding.

The CTI Object Server is typically installed in duplex mode, with two CTI OS servers running in parallel for redundancy, one on PG side-A and one on PG side-B. The CTI Toolkit Desktop applications randomly connect to one of the two servers and automatically fails-over to the alternate server if the connection to the original CTI OS server fails. CTI OS can also run in simplex mode with all clients connecting to a single server, but Cisco does not recommend this configuration. (Simplex mode is not supported for production environments.)

Agent capacity sizing for the PG is covered in the chapter on Sizing Unified CCE Components and Servers.

**Note** The CTI OS server interfaces to any desktop application built using the CTI Toolkit Software Development Kit. A single CTI OS server can support the use of both CAD and CTI Toolkit desktops concurrently.

## 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 co-resident on the Peripheral Gateway 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-BE and 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 failover 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.

Cisco Unified Contact Center Enterprise (Unified CCE) supports a variety of desktop application choices for agents and supervisors, as described in the following sections.

## Agent Desktops

An agent desktop application is a required component of a Unified CCE deployment. The contact center agent uses this application to perform agent state control (login, logout, ready, not ready, and wrap-up) and call control (answer, release, hold, retrieve, make call, transfer, and conference). In addition to these required features, the application can provide enhanced features that are useful in a contact center environment.

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

- Cisco Agent Desktop (CAD) — A packaged agent desktop solution supporting an embedded browser and scripted workflow automation.
- Cisco Finesse Desktop — 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 that 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 Unified IP Phone Agent — An agent desktop solution provided through the Cisco Unified IP Phone display.
- Cisco Agent Desktop Browser Edition (CAD-BE) — A browser-based agent application that supports many of the features of the CAD windows-based agent application with lower platform requirements.
- Agent Desktop Applications Offered through Cisco partners:
- 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.

- Prepackaged CRM integrations — CRM integrations are available through Cisco Unified CRM Technology Partners. They are based on the CTI Toolkit and are not discussed individually in this document.

## Agent Mobility

Within the Unified CCE deployment, the agent desktop application is not statically associated with any specific agent or IP phone extension. Agents and phone extensions (device targets) are configured within the Unified CCE configuration and associated with a specific Unified CM cluster.

When logging in from an agent desktop application, the agent is presented with a dialog box that prompts for agent ID or login name, password, and the phone extension to be used 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 logs out.

This mechanism enables an agent to work (or *hot-desk*) at any workstation. It also enables agents to take their laptops to any Cisco Unified IP Phone and log in from that device (assuming the phone has been configured in Unified CCE and in Unified CM to be used in the Unified CCE deployment). Agents can also log in to other phones using the Cisco Extension Mobility feature. For more information about Extension Mobility, see the Extension Mobility section of the Cisco Unified Communications Manager Features and Services Guide.

## Supervisor Desktops

In addition to the agent desktop application, a supervisor desktop application is also available. The contact center supervisor uses this application to monitor agent state for members within their team. The supervisor desktop also enables Silent Monitoring of agents during active calls.

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

- Cisco Supervisor Desktop (CSD) — A packaged supervisor desktop solution.
- CTI Toolkit Supervisor Desktop — A supervisor desktop application, built with the CTI Toolkit that supports customization and integration with other applications, customer databases, and Customer Relationship Management (CRM) applications.
- Cisco Finesse Supervisor Desktop — A fully browser-based supervisor application that extends the base Finesse Agent Desktop with supervisor capabilities.
- Supervisor Desktop Applications Offered through Cisco partners
- Prepackaged CRM integrations — CRM integrations are available through Cisco Unified CRM Technology Partners. They are based on the CTI Toolkit and are not discussed individually in this document.

# Desktop Solutions

Depending on the requirements of the contact center, a particular type of desktop might be better suited to the solution. Table 2 provides an abbreviated list of the functionality available in the various desktop applications. It is intended to provide a starting point to determine the desktop that best meets specific solution requirements. Further information is available for each of the Cisco desktops in the sections below and in their respective product specifications at Cisco.com.

**Table 2** ***Features Supported by Cisco Desktop Solutions***

| Desktop Functionality | Cisco Agent Desktop | Cisco Agent Desktop Browser Edition | CTI Toolkit | Cisco Unified CRM Connector for Siebel | IP Phone Agent | Cisco Finesse Desktop |
|---|---|---|---|---|---|---|
| Turn-key desktop applications | Yes | Yes | Yes | Yes | Yes | Yes |
| Custom desktop development using C++, .NET, and Java | | | Yes | | | |
| Custom development using standard web components (HTML, JavaScript) | | | | | | Yes |
| Desktop Security | Yes | | Yes | | | |
| Workflow Automation | Yes | Yes | | | | |
| Mobile (Remote) Agents | Yes | Yes | Yes | Yes | | |
| Siebel Integration | | | | Yes | | |
| Silent Monitoring | Yes | Yes | Yes | | Yes | Yes |
| Integrated Recording Capacity | Yes | Yes | | | Yes | |
| Monitor Mode Applications | | | Yes | | | |

| Desktop Functionality | Cisco Agent Desktop | Cisco Agent Desktop Browser Edition | CTI Toolkit | Cisco Unified CRM Connector for Siebel | IP Phone Agent | Cisco Finesse Desktop |
|---|---|---|---|---|---|---|
| Outbound Calls | Yes | | Yes | Yes | | |
| Microsoft Terminal Services Support | Yes | | Yes | | | |
| Citrix Presentation Server Support | Yes | | Yes | | | |
| Agent Mobility | Yes | Yes | Yes | | Yes | |
| IP Phone Solution (no soft desktop) | | | | | Yes | |
| Agent Greeting | | | Yes | | | |
| Specific capability or integration not offered by Cisco | | | | | | |

## 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

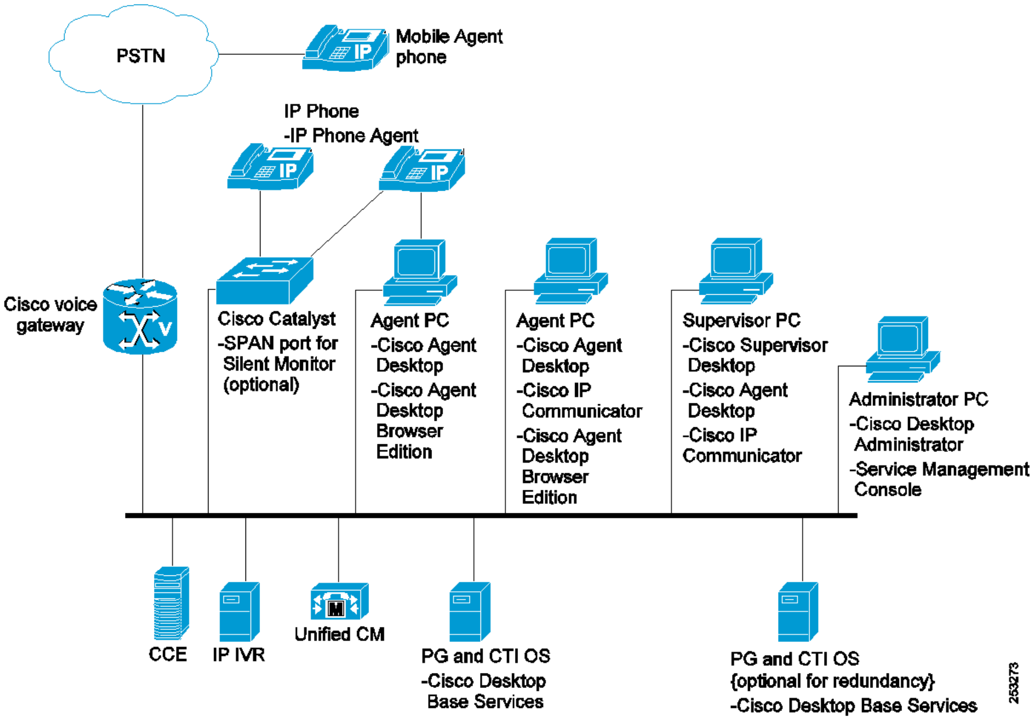
Supervisors can initiate a silent monitoring session with an agent on their team.

# CAD User Applications

CAD user applications include the following applications for contact center agents, supervisors, and administrators. (See Figure 73)

- Cisco Agent Desktop: Windows-based agent application
- Cisco Agent Desktop–Browser Edition (CAD-BE): Java-based version of Agent Desktop
- Cisco 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 73** *Cisco Agent Desktop System Configuration and Components*



## CAD Application Features

Table 3 compares some of the more important CAD features to assist users in selecting the appropriate agent application for their deployment.

**Table 3** *Comparison of Major CAD Features*

| Feature | CAD | CAD-BE | IPPA |
|---|---|---|---|
| Call control | Yes | Yes | n/a** |
| VPN/Mobile agent support | Yes | Yes | Yes |
| Chat / Unified Presence Integration | Yes | No | No |
| Supports Cisco IP Communicator | Yes | Yes | No |
| Team Messages | Yes | No | No |

| Feature | CAD | CAD-BE | IPPA |
|---|---|---|---|
| Supports Mobile Agent | Yes | Yes | n/a |
| Real Time Queue and Agent Displays | Yes | No | Yes |
| Supports Cisco Outbound Dialer | Yes | No | No |
| Integrated browser | Yes | Yes | n/a |
| Call event work flow automation | Yes | Yes (limited) | No |
| Agent state work flow automation | Yes | No | No |
| Supports thin client environment | Yes | No | n/a |
| Desktop monitoring and recording* | Yes | No | No |
| SPAN monitoring and recording* | Yes | Yes | Yes |
| Unified CM monitoring and recording* | Yes | Yes | Yes |

*For more detailed information about supported recording and monitoring, refer to *Configuring and Troubleshooting VoIP Monitoring*.

**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 Finesse Agent/Supervisor Desktop

Cisco Finesse Desktop is a browser-based agent desktop application that runs on Windows-based PCs using the Microsoft Internet Explorer browser. Finesse provides support for 7800/8800 series IP phones.

The Finesse Desktop application is divided between 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. The agent desktop may be configured to use Cisco and third-party gadgets using a layout management mechanism.

The Finesse server is deployed as a VMware virtual machine and runs on the Cisco VOS platform. The Finesse server provides the integration to CTI Server. Supported client operations are exposed by the Finesse server through a REST API that shields the application developer from many of the details surrounding the CTI Server wire protocol.

Finesse integration with Unified CCE includes the following:

- Authentication with Unified CCE over a direct connection to the Unified CCE AW Database. Finesse requires that the AW Database is configured to use Windows authentication and that the user configured for Finesse database access is a domain user.

- An active/active deployment model where both Finesse servers connect to the active CTI Server on the Agent PG.
- Redundancy through the standard Cisco VOS replication mechanism.

Finesse supervisor features extend the agent desktop with additional gadgets that are accessible to supervisors configured in Unified CCE. These features include the following:

- Team Performance gadget for viewing agent status
- Silent Monitoring

Cisco Finesse includes an administrative application that allows for the configuration of the following:

- CTI Server and AW Database connections
- Cluster settings for VOS replication
- Reason and Wrap Up codes
- Finesse layout configuration (media properties and gadget layout)

### Finesse Rest API

Finesse provides a REST API that enables client applications to access the supported server features. The REST API uses HTTP as the transport with XML payloads.

Developer documentation for the REST API may be found on the Cisco Developer Network.

## 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.

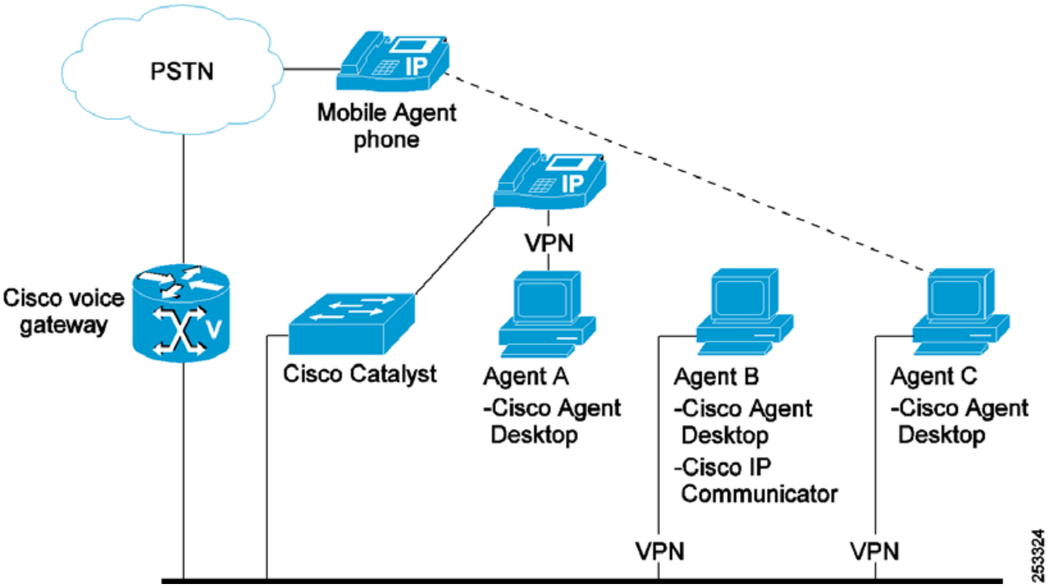
Agent Desktop supports Desktop, SPAN, and Unified CM monitoring and recording.

Figure 74 illustrates various ways agent desktops can be configured in a contact center.

- 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.*

**Figure 74** ***CAD Agents and Components***



## Cisco Agent Desktop Browser Edition

Cisco Agent Desktop—Browser Edition (CAD-BE) is a Java applet that runs in Microsoft Internet Explorer (Windows machines) or in Mozilla Firefox (Windows and Linux machines). CAD-BE interfaces with the BIPPA service for call control; BIPPA in turn interfaces with the CTI service. For all other features it communicates with the CAD services.

CAD-BE can use the abilities and features of any Java applet. This enables CAD-BE to have a small footprint on the agent desktop and still provide essential agent features.

Some limitations of CAD-BE beyond those listed Table 3 include:

- Desktop monitoring and recording is not supported. CAD-BE does support server monitoring and recording and Unified CM monitoring.
- The integrated browser is external to the CAD-BE window and is launched when CAD-BE is launched.
- The “click to dial from browser” feature is not supported.
- Agents cannot modify enterprise data.
- The dial pad does not support dial string formatting, phone books, or recent call list.
- Window behavior (for example, stealth or always on top) cannot be configured by the agent.
- For more information about CAD-BE features and capabilities, see the *Cisco Agent Desktop—Browser Edition User Guide*.

## Cisco Unified IP Phone Agent

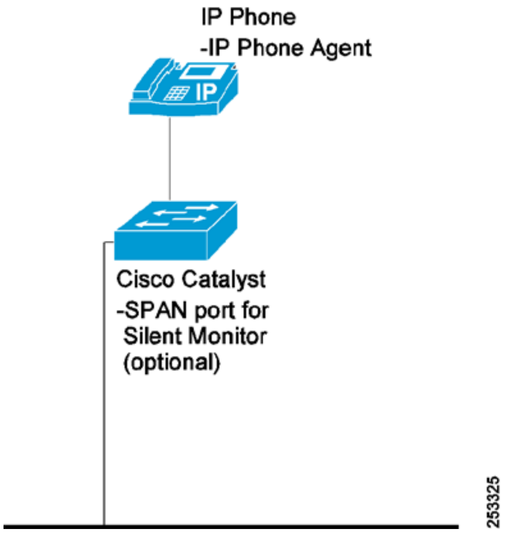
Cisco IP Phone Agent (IPPA) runs as an IP phone XML service. The agent is not required to have a PC. IPPA includes all the basic features required by a contact center agent, as well as advanced features such as reason codes, wrap-up data, and on-demand recording.

IPPA agents can be monitored and recorded using server monitoring, and monitored using Unified CM monitoring.

For more information about IPPA features and capabilities, see the *Cisco IP Phone Agent User Guide*.

The following figure illustrates the components used by IP Phone agents.

**Figure 75** ***Cisco 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 in order 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 *Cisco Supervisor Desktop User Guide*.

## 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 (CAD agents and CAD-BE agents).

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 *Cisco Desktop Administrator User Guide*.

## 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.

## CTI OS Desktop Toolkit Solution

The CTI OS Desktop Toolkit (CTI Toolkit) provides a Software Development Kit (SDK) for custom development of desktop applications. The CTI Toolkit supports C++, Java, and .NET development Client Interface Libraries (CILs) and provides sample applications for customization.

Additionally, the CTI Toolkit ships complete with pre-built, ready-to-run agent desktop, supervisor desktop, and call center monitoring applications. These applications can be used as-is or can be customized further to meet the particular needs of a call center.

The CTI Toolkit also offers advanced tools for integrating desktop applications with a database, Customer Relation Management (CRM) applications, or other contact center applications.

The CTI Toolkit solution offers a rich set of features for the contact center environment, including:

- Collaboration — A supervisor can text-chat directly with agents, and agents can text-chat with supervisors or other team members (if enabled). Interactive collaboration enables the contact center to communicate better, increase productivity, improve customer responsiveness, and coach or train agents.
- Secure Desktop Connection — Desktop security is provided between the agent/supervisor desktops and the CTI OS server.
- Silent Monitoring — A supervisor can initiate a silent monitoring session with an agent within their team.

## CTI OS Desktop Toolkit Software Development Kits and User Applications

The CTI Toolkit provides the following user tools and applications.

- C++ CIL API—A Windows software development kit for developing C++ CTI applications
- Java CIL API—A cross-platform library for developing Java CTI applications
- .NET CIL API—A Windows software development kit for developing custom .NET framework CTI applications
- COM CIL API—A set of COM Dynamic Link Libraries (COM DLL) for building a Visual Basic 6.0 CTI application
- ActiveX Controls—A set of Windows GUI controls for custom desktop development using Development environments that support ActiveX technology. For example Visual Basic 6.0
- CTI OS Runtime Callable Wrappers—A set of .NET assemblies that allows the use of COM CIL and ActiveX controls in native .NET applications
- CTI Toolkit Agent Desktop — A Windows Visual Basic application built on the COM CIL and Active-X controls, providing agent desktop functionality
- CTI Toolkit Supervisor Desktop—A Windows Visual Basic application built on the COM CIL and Active-X controls, providing supervisor desktop functionality
- CTI Toolkit Outbound Desktop—A Windows Visual Basic application built on the COM CIL and Active-X controls, supporting outbound call center campaigns in addition to standard agent desktop functionality
- CTI Toolkit Combo Desktop—A Windows agent and supervisor application based on the .NET CIL, which combines support for agent, supervisor, and outbound functionality

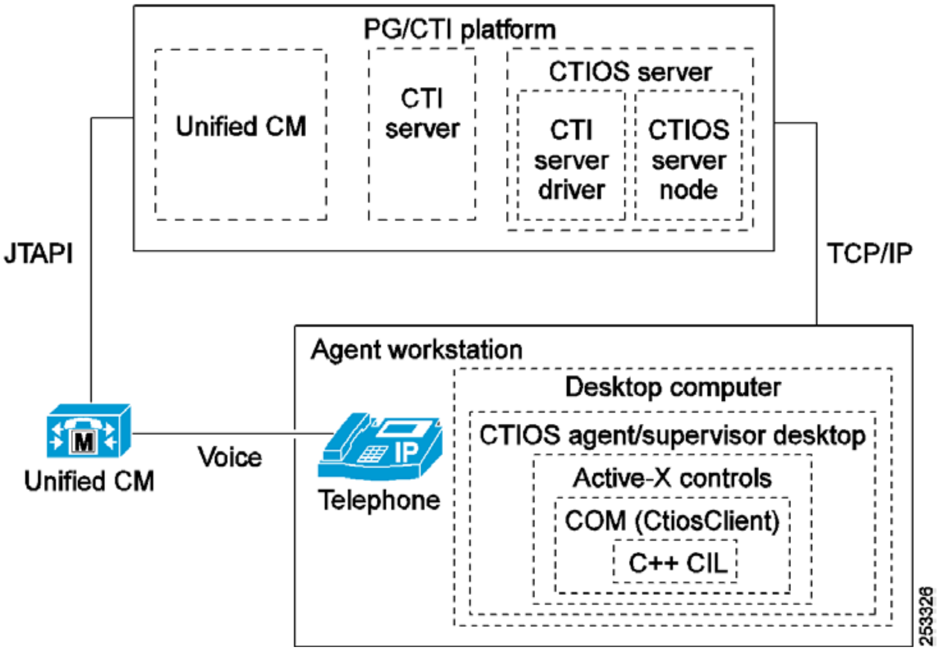
- CTI Toolkit All-Agents Monitor—A Windows Admin application based on the C++ CIL, providing call center agent status monitoring
- CTI Toolkit All-Calls Monitor—A Windows Admin application based on the C++ CIL, providing call center call status monitoring

**Note** The CTI Toolkit All-Agents and All-Calls Monitor applications can be used only if the number of skill groups per agent is less than 20.

**Note** The CTI OS Desktop Toolkit C++/COM CIL is the only APIs that support Agent Greeting Enable/Disable and Agent Greeting Recording.

Figure 76 illustrates the architecture of the CTI OS Desktop Toolkit. For more information regarding the CTI OS Desktop Toolkit, see the *CTI OS Developer's Guide for Cisco ICM/IPCC Enterprise and Hosted Editions*.

**Figure 76** ***CTI OS Desktop Toolkit Architecture***



### C++ CIL API

The CTI OS Desktop Toolkit C++ CIL provides a set of header files and static libraries for building C++ CTI applications using Microsoft Visual Studio .NET. The C++ CIL also supports a secure desktop connection between the agent PC and the CTI Object Server on the PG.

### Java CIL API

The CTI OS Desktop Toolkit Java CIL provides a powerful cross-platform library for developing Java CTI applications. It does not provide APIs for Agent Greeting Enable/Disable or Agent Greeting Recording.

### .NET CIL API

The CTI OS Desktop Toolkit .NET CIL provides native .NET class libraries for developing native .NET Framework applications. The .NET Combo Desktop is provided as a sample application built using the .NET CIL. It does not provide APIs for Agent Greeting Enable/Disable or Agent Greeting Recording.

### COM CIL API

The CTI OS Desktop Toolkit COM CIL provides a set of COM Dynamic Link Libraries for building Visual Basic 6.0 CTI applications. The CTI Toolkit Agent and Supervisor Desktops are provided as sample applications built with Visual Basic 6.0 and using the COM CIL.

### 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, Sybase PowerBuilder, and other applications supporting the OC96 ActiveX standard.

The ActiveX Controls include:

- 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 (Provides a new functionality to allow 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

### CTI Toolkit Agent Desktop

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

The CTI Toolkit Agent Desktop includes:

- Support for the agent to turn on/off (Enable/Disable) Agent Greeting playback on the routed incoming calls on their current logon session. In addition it allows the agent to record new Agent Greetings.
- Support for desktop monitoring, which captures the voice stream on the agent's IP phone to support the silent monitoring and call recording features.

### CTI Toolkit Supervisor Desktop

The CTI Toolkit Supervisor Desktop is a Microsoft Windows application that runs on a supervisor's desktop PC. The CTI Toolkit Supervisor Desktop interfaces with the CTI OS server for agent state change events and real-time statistics updates. The CTI Toolkit Supervisor Desktop provides the contact center supervisor with the ability to manage a team of agents. Supervisors are able to view real-time information about the agents in a team as well as interact with these agents. A supervisor can select an agent to change the agent's state, view information specific to that agent, silently monitor the agent's call, barge in or intercept the agent's call, or chat with the agent.

A supervisor may also receive emergency assistance requests from agents on their team through the supervisor desktop.

In Unified CCE, supervisors may also be configured to act as agents. When this is done, 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 application that runs on an agent's desktop PC and works with either a hardware IP phone or the Cisco IP Communicator software phone. The CTI Toolkit Outbound Desktop interfaces with the CTI OS server for call control and agent state change events. In addition to the standard set of agent controls present in the CTI Toolkit Agent Desktop, the Outbound Desktop provides a set of controls for managing outbound call campaigns. Outbound calls are automatically managed by Unified CCE, and the agent uses the additional controls to accept the next outbound call.

### CTI Toolkit Combo Desktop

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

The Combo Desktop integrates the functionality of the Toolkit Agent, Supervisor, and Outbound desktops into a single .NET application. The Combo Desktop source code is also provided as a starting point for custom desktop development using the Microsoft .NET Framework.

The CTI Toolkit Combo Desktop does not provide support for Agent Greeting Enable/Disable or recording new Agent Greetings.

#### CTI Toolkit All-Agents Monitor

The CTI OS Desktop Toolkit ships complete with a ready-to-run All-Agents Monitor application. This application provides a call center administrator with the ability to monitor agent login and state activity within the call center.

#### CTI Toolkit All-Calls Monitor

The CTI OS Desktop Toolkit ships complete with a ready-to-run All-Calls Monitor application. This application provides a call center administrator with the ability to monitor call activity within the call center.

## Cisco Unified CRM Connector for Siebel Solution

The Cisco Unified CRM Connector for Siebel is an installable component developed by Cisco that enables integration of the Cisco Unified CCE with the Siebel CRM Environment. In this solution, the Siebel Agent Desktop provides the agent state and call control interface. The Siebel Desktop uses the Cisco Unified CRM Connector for Siebel, which is built on top of the CTI OS Desktop Toolkit C++ CIL to communicate with the CTI Object Server.

The Cisco Unified CRM Connector for Siebel does not provide support for Agent Greeting Enable/Disable or Recording new Agent Greetings.

For more information about the capability of the Siebel eBusiness solution, see the Siebel website.

# Deployment Considerations

This section covers the deployment considerations.

## Citrix and Microsoft Terminal Services (MTS)

This section discusses deploying Cisco Agent Desktop and Cisco Toolkit Desktop in a Citrix or Microsoft Terminal Services (MTS) environment.

### Cisco Agent Desktop

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 recording) is not supported with Citrix terminal services. SPAN port monitoring must be used instead.
- Macros work only if they involve applications running on the Citrix server, and not those 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.

- Only Citrix 4.0 and 4.5 running on Windows 2000 Server or Windows 2003 Server are supported.

For implementation details, refer to *Integrating CAD into a Citrix MetaFrame Presentation Server or Microsoft Terminal Services Environment*.

### Cisco Toolkit Desktop

Cisco Unified CCE supports running CTI Toolkit Desktop within the Citrix and Microsoft Terminal (MTS) Services environments. When planning to use Citrix terminal services with the CTI Toolkit Desktop, take into account the following considerations:

- 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, refer to *Integrating CAD into a Citrix MetaFrame Presentation Server or Microsoft Terminal Services Environment*.

**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.

## Silent Monitoring

Silent monitoring enables supervisors to monitor the conversations of agents within their team. Supervisors are not able to participate actively in the conversation, and the agent(s) and caller(s) are unaware they are being monitored. Cisco Agent Desktop, CTI Object Server (CTI OS), and Cisco Finesse provide solutions support for silent monitoring. CAD Server-based monitoring supports Agent Desktops, IP Phone Agents, and Mobile Agents. Desktop monitoring supports only desktop agents. CTI OS releases 7.2 and later support two types of silent monitors: CTI OS based silent monitor and Unified CM silent monitor. Cisco Finesse supports Unified CM silent monitor only.

CTI OS based silent monitoring is accomplished by using one or more VoIP monitoring services located either on the agent’s desktop (desktop monitoring) or on a separate VoIP monitor server (server-based monitoring). CTI OS uses server-based silent monitoring to support mobile agents and desktop-based silent monitoring to support traditional (non-mobile) Unified CCE agents.

Unified CM accomplishes silent monitoring with a call between the supervisor's (monitoring) device and agent’s (monitored) device. The agent’s phone mixes and sends the agent’s conversation to the supervisor's phone, where it is played out to the supervisor. Unified CM silent monitoring can be initiated by any of the CTI OS supervisor desktops (out-of-the-box, Java, or .NET). Any Unified CCE agent desktop, including Siebel, can be silently monitored using Unified CM silent monitoring, provided the following requirements are met:

- The agent to be silently monitored is using a Cisco Unified IP Phone 7941, 7942, 7945, 7961, 7962, 7965, 7970, 7971, or 7975, and Cisco IP Communicator 7.0 or later.
- The contact center is using Cisco Unified CM 8.0 or higher. For details see the Unified CCE compatibility specifications.

- When CTI OS based silent monitor is used, the Cisco IP Phones, including Cisco IP Communicator, must be configured to use RTP streams (SRTP streams cannot be silently monitored).

Unified CM silent monitoring does not support mobile agents.

Unified CM silent monitoring supports a maximum of one silent monitoring session and one recording session for the same agent phone.

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

**Note** G.722 is used as the default codec for regions that are configured for G.711 on devices that support G.722. However, G.722 is not supported with silent monitoring and call recording based on CAD, CTI OS, or Unified CM. To disable this default, in Unified CM Administration go to **Enterprise Parameters** and set **Advertise G.722 Codec** to **disabled**.

### CTI Object Server

A given CTI OS Server can be configured to use either CTI OS based silent monitor or Unified CM silent monitor, or to disable silent monitoring. When supervisor desktops connect to the CTI OS Server, this configuration is downloaded. The supervisor desktop uses this information to invoke the configured type of silent monitor when the Start Silent Monitor button is pressed. The initial message from the supervisor desktop is used by the CTI OS Server to drive either the CTI OS or Unified CM silent monitor.

For details regarding the configuration of silent monitoring, system administrators can see the *CTI OS System Manager's Guide for Cisco ICM/IPCC Enterprise & Hosted Editions*.

Developers implementing either the CTI OS or Unified CM silent monitor should see the *CTI OS Developer's Guide for Cisco ICM/IPCC Enterprise & Hosted Editions*.

### Unified CM Silent Monitor

This section describes how CTI OS accomplishes silent monitoring when the CTI OS Server is configured to use the Unified CM silent monitor.

Unified CCE supports the silent monitoring functionality available in Unified CM 6.0 and higher.

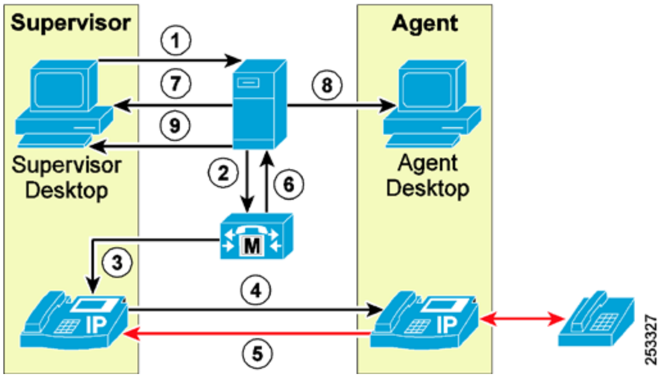
**Note** ASA does not currently support the type of call flow that the silent monitoring feature uses.

Figure 77 illustrates the following message flow, which occurs when the Unified CM silent monitor is initiated by the supervisor desktop:

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 CM.
3. Unified CM instructs the supervisor's phone to call the built-in-bridge in the agent’s phone.
4. The supervisor's phone places the call to the built-in-bridge in the agent’s phone.
5. The agent’s phone forwards a mix of the agent’s and customer's voice streams.
6. Call events for the silently monitored call are sent from Unified CM 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 77** ***Unified CM Silent Monitoring for Unified CCE***



Cisco Finesse provides silent monitoring as follows:

1. The supervisor application initiates silent monitoring by sending a REST request to the Finesse server.
2. The Finesse server sends the AgentSuperviseCall() message to Unified CCE to start silent monitoring.
3. Unified CCE sends the CallStartMonitor() message to Unified CM.
4. Unified CM instructs the supervisor’s phone to call the built-in-bridge in the agent’s phone.
5. The supervisor’s phone places the call to the built-in-bridge in the agent’s phone.
6. The agent’s phone forwards a mix of the agent and customer voice streams.
7. Unified CM 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.

Unified CM silent monitoring works the same as other call control functionality provided by Unified CM (such as conference, transfer, and so forth). When Unified CM is used for silent monitoring, a message is sent from the desktop, through Unified CCE, through Unified CM, and out to the phones where silent monitoring is executed.

The messaging through Unified CCE and Unified CM impacts Unified CCE performance. For further details regarding the impact of Unified CM silent monitoring on Unified CCE sizing, see the chapter on Sizing Unified CCE Components and Servers.

Unified CM silent monitoring is supported only for agents who are connected to Unified CCE on the LAN; it does not support mobile agents and mobile agents.

### CTI OS Based Silent Monitor

This section describes how CTI Object Server (CTI OS) accomplishes silent monitoring when the CTI OS Server component is configured to use the CTI OS Based silent monitor.

The silent monitoring solution provided by CTI OS in Release 7.0 and earlier was integrated in the Client Interface Library (CIL). The CIL had components to capture and forward voice packets as well as components to play back a stream of forwarded voice packets to the supervisor's sound card. This feature limited silent monitoring support to Unified CCE agent desktops deployed behind a Cisco IP Phone and Unified CCE supervisor desktops deployed on the supervisor's desktop.

Starting CTI OS release 7.1, two deployment types are supported: Citrix and Mobile Agent. In these two deployments, the CIL is not deployed where it has access to the voice stream. In Citrix, the CIL is located on the Citrix Server. Agents and supervisors use a Citrix client to run the desktop. When this is done, the desktop runs on the Citrix server. The Citrix client merely displays the UI of the desktop. Because it is the agent’s Citrix client that is deployed behind the IP phone, the CIL no longer has access to the voice path. Similarly, it is the supervisor's Citrix client that has the sound card. In this case, the CIL is running on the Citrix server and does not have access to the sound card.

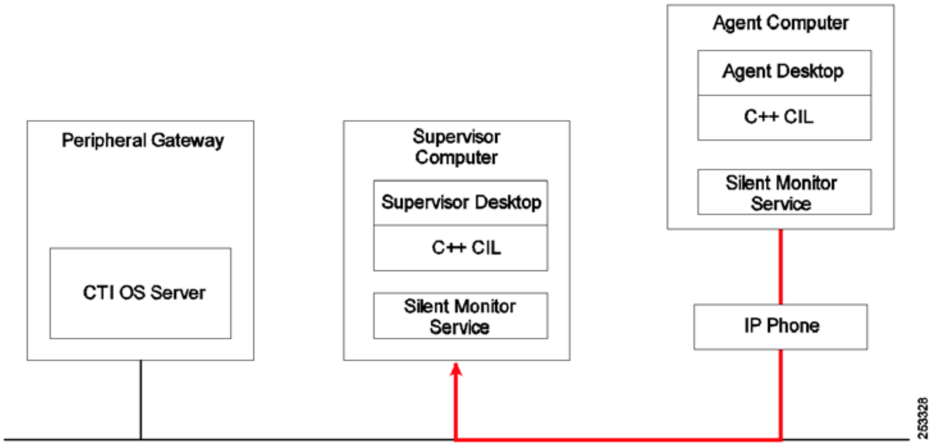
In Mobile Agent deployments, the CIL is deployed on an agent’s remote PC. When the agent uses an analog phone, the CIL does not have access to the voice stream.

To support these two deployment models, it was necessary to remove the silent monitor components from the CIL and put them on a separate service. This allows the service to be deployed where it has access to the agent’s voice stream or the supervisor's sound card.

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

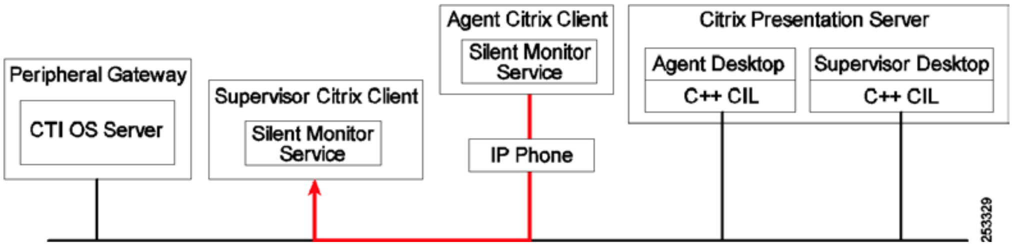
Figure 78 and Figure 79 illustrate deployments where the agent uses an IP phone. In these deployments, silent monitoring is configured the same way regardless of whether the agent is mobile or not.

**Figure 78** ***CTI OS Based Silent Monitoring for Cisco Unified CCE When a Mobile or Local Agent Uses an IP Phone***



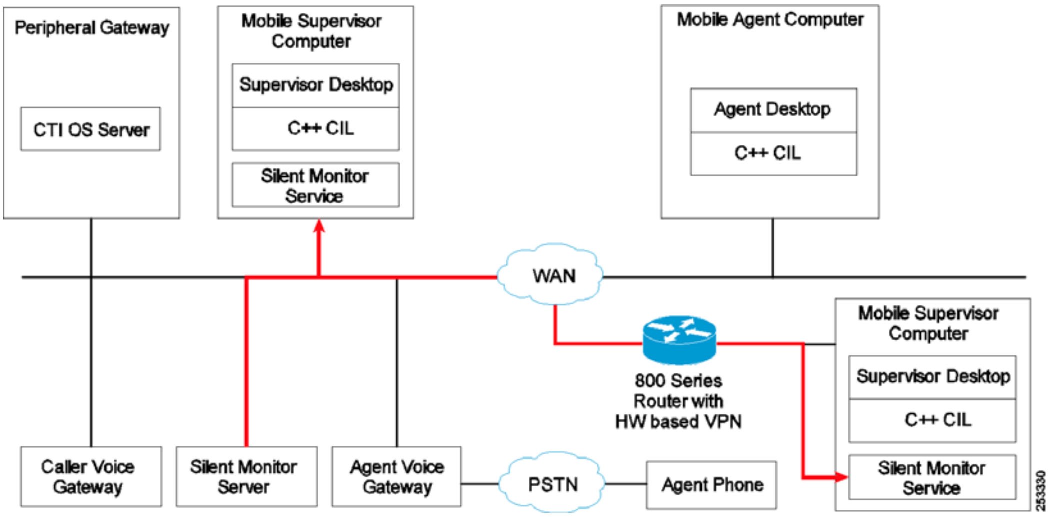
The deployment in Figure 78 is very similar to CTI OS Release 7.0 and earlier deployments. The only difference is that the silent monitoring service is running alongside the CIL to provide silent monitoring functionality.

**Figure 79** ***CTI OS Based Silent Monitoring for Cisco Unified CCE with Citrix When a Mobile or Local Agent Uses an IP Phone***



In the deployment model in Figure 79, the silent monitoring service is deployed on Citrix clients, where it has access to the agent's voice stream and the supervisor's sound card. The CIL makes a connection to the silent monitoring service and sends it instructions over a TCP connection in order to start and stop silent monitoring sessions.

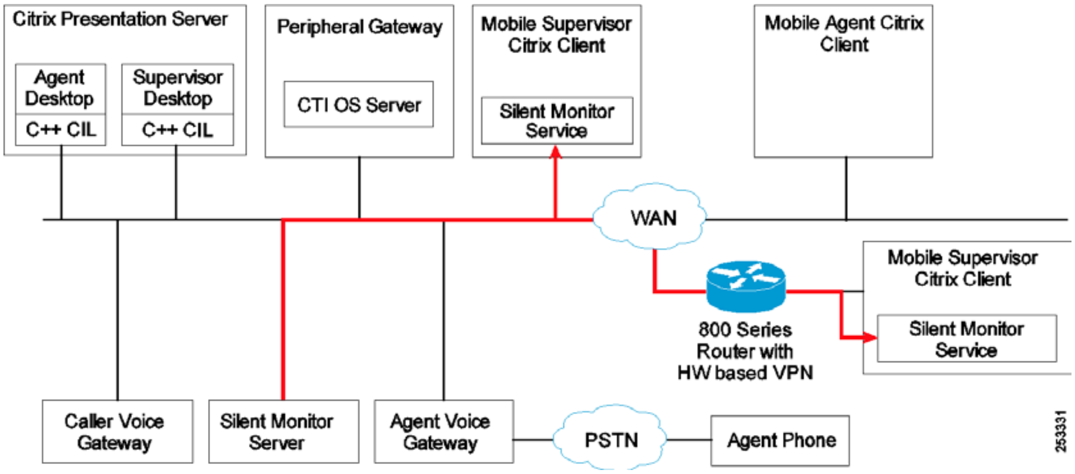
**Figure 80** ***Silent Monitoring for a Mobile Agent Using a PSTN Phone***



In the deployment model in Figure 80, one silent monitoring service is deployed on a switch's SPAN port in order to gain access to voice traffic passing through the agent gateway. Agents to forward their voice streams to the supervisor silent monitoring services use the silent monitoring service attached to the SPAN port.

Supervisors running locally are deployed the same as Unified CCE supervisors. Supervisors running remotely are also deployed the same as Unified CCE supervisors, but a Cisco 800 Series Router with hardware-based VPN is required in order for the supervisor to receive agent voice streams.

**Figure 81** *Silent Monitoring for a Mobile Agent Using a PSTN Phone with Citrix or Microsoft Terminal Services*



In the deployment model in Figure 81, one silent monitoring service is deployed on a switch's SPAN port in order to gain access to voice traffic passing through the agent gateway. Agents to forward their voice streams to the supervisor silent monitoring services use the silent monitoring service attached to the SPAN port. Mobile agents need to run only their Citrix clients. Agent desktops running on the Citrix server will connect to the silent monitoring server.

Supervisors running locally are deployed the same as Citrix Unified CCE supervisors. Supervisors running remotely are also deployed the same as Citrix Unified CCE supervisors, but a Cisco 800 Series Router with hardware-based VPN is required in order for the supervisor to receive agent voice streams.

In the two mobile agent deployments above (Figure 80 and Figure 81), calls whose voice traffic does not leave the agent gateway cannot be silently monitored. This includes agent-to-agent calls as well as agent consultations with other agents. The only calls that can be reliably monitored in this case are calls between agents and customers. This is because the mobile agent solution requires separate gateways for callers and agents to ensure that voice traffic is put on the network.

## Clusters

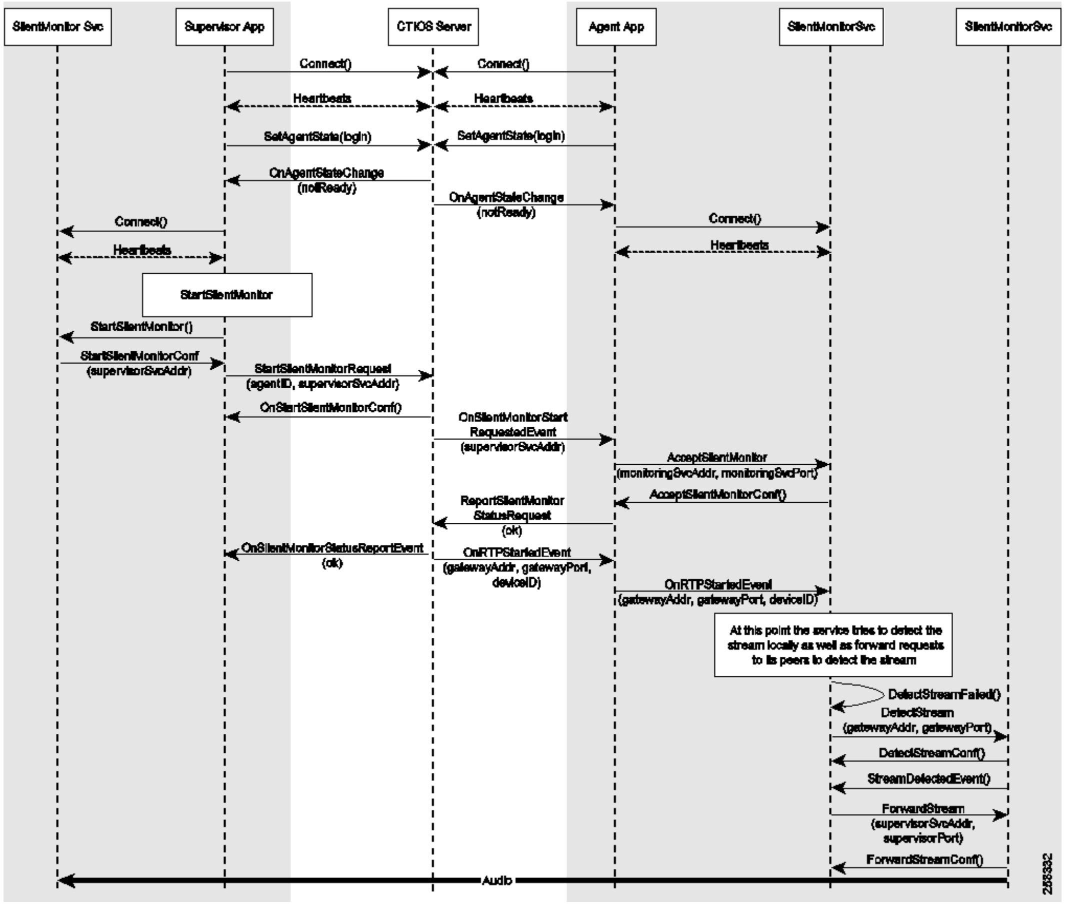
If a mobile agent’s login can be handled by one of two gateways, it is possible to cluster two and only two silent monitoring servers together to provide silent monitoring functionality regardless of the gateway that handles the call. A maximum of two silent monitoring servers are supported in a cluster (SPAN) based deployment. When a request to silently monitor the agent is received, the silent monitoring server that receives the request from the agent desktop will forward the request to its peer, and then both silent monitoring servers will attempt to detect the stream. Once the agent’s voice stream is detected, it is forwarded to the supervisor's silent monitoring service by the silent monitoring server that detected the stream.

For more information regarding deployment and configuration of the silent monitoring service, see the *CTI OS System Manager's Guide for Cisco ICM/IPCC Enterprise & Hosted Editions*.

## Message Flow

Figure 82 illustrates the messaging that occurs between the desktops, CIT OS Server, and silent monitoring services when a silent monitor session is initiated. Note that messaging between the desktops and the CTI OS Server has not changed from CTI OS Release 7.0.

**Figure 82** *Message Flow Between Desktops, CTI OS Server, and Silent Monitoring Service*


SilentMonitor Svc
Supervisor App
CTIOS Server
Agent App
SilentMonitorSvc
SilentMonitorSvc
Connect()
Connect()
Heartbeats
Heartbeats
SetAgentState(login)
SetAgentState(login)
OnAgentStateChange (notReady)
OnAgentStateChange (notReady)
Connect()
Connect()
Heartbeats
Heartbeats
StartSilentMonitor
StartSilentMonitor()
StartSilentMonitorConf (supervisorSvcAddr)
StartSilentMonitorRequest (agentID, supervisorSvcAddr)
OnStartSilentMonitorConf()
OnSilentMonitorStart RequestedEvent (supervisorSvcAddr)
AcceptSilentMonitor (monitoringSvcAddr, monitoringSvcPort)
AcceptSilentMonitorConf()
ReportSilentMonitor StatusRequest (ok)
OnSilentMonitorStatusReportEvent (ok)
OnRTPStartedEvent (gatewayAddr, gatewayPort, deviceID)
OnRTPStartedEvent (gatewayAddr, gatewayPort, deviceID)
At this point the service tries to detect the stream locally as well as forward requests to its peers to detect the stream
DetectStreamFailed()
DetectStream (gatewayAddr, gatewayPort)
DetectStreamConf()
StreamDetectedEvent()
ForwardStream (supervisorSvcAddr, supervisorPort)
ForwardStreamConf()
Audio
255332

## Connection Profiles

In mobile agent deployments, agent desktops learn where and how to connect to their silent monitoring server using a CTI OS connection profile. When an agent logs in, the agent desktop uses the following algorithm to determine where the silent monitoring service is located:

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 Citrix client's IP address.
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 their silent monitoring service:

1. If the desktop is running under Citrix, connect to the silent monitoring service running at the Citrix client's IP address.
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 in the connection profile, no attempt is made to connect to a silent monitoring service.

### CAD Silent Monitoring and Recording

The section describes Cisco Agent Desktop (CAD) silent monitoring.

### 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 phone to send all its network traffic along this connection to the software running on the PC. In this case, the packet sniffing software is able to see the voice traffic coming to and leaving from the agent’s phone. It will copy this traffic and send it to the supervisor 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 will be 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 will be 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 will be 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 Remote Silent Monitoring

This section covers Cisco Remote Silent Monitoring. Remote Silent Monitoring (RSM), which allows for the real-time monitoring of agents as a dial-in service.

The RSM solution consists of three components:

- VLEngine
- PhoneSim
- Callflow Script(s) for Unified CVP and IP IVR

For a further description of these components, see the *Cisco Remote Silent Monitoring Installation and Administration Guide*, available at cisco.com.

## Platform Considerations

The RSM solution is highly integrated part of a Cisco Unified Contact Center Enterprise environment. Because of this, the functioning of RSM requires resources from various other components of the platform as a whole. To properly integrate RSM, then, requires an understanding of its interactions with the rest of the environment so that capacity can be properly planned, provisioned, and managed.

In particular, RSM interacts mainly with the Unified CM cluster.

The RSM server has two tie-ins with each Unified CM cluster in the environment that it is configured to use:

**Simulated Phones**: The RSM PhoneSim component requires that a Cisco Unified IP Phone 7941 device entry be created on the Unified CM cluster for each of the simulated phones (or "simphones") 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 CM cluster will need to have at least of these 100 simphones. To the Unified CM cluster itself, these simphones appear as normal Cisco Unified IP Phone 7941 SIP phones; however, in reality they are homed to and controlled by PhoneSim instead of being an actual physical phone device.

When compared with the usage profile of a normal phone, the simphone usually puts a lighter load on the Unified CM cluster. This is because it exhibits only a small set of behaviors, consisting of:

- Registering with the Unified CM cluster when PhoneSim is started.
- 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 conversation the agent is having 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 that was created on the cluster.

When an agent is to be monitored, a JTAPI monitor request call is made from the RSM server to the Unified CM cluster that manages that agent’s phone. Also, while RSM is in use, a JTAPI CallObserver is kept attached to each simphone device. It 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, this will induce 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 Guide*, available at cisco.com.

### CTI OS Server

RSM makes a persistent "monitor-mode" connection to each CTI OS server it is configured to use. Through this connection certain platform events such as call start, call end, agent on hold, and so forth, are streamed in real-time.

Besides this, RSM will make an additional, short-lived "agent-mode" connection to possibly each CTI OS server when a supervisor dials in and authenticates. The purpose of this connection is to validate the supervisor's entered credentials by performing a corresponding login into CTI OS. Note that, if the built-in authentication mechanisms of the RSM call flow (for example, the checkCredentials API call) are not used, this connection is not made. If the login is successful, that supervisor's team membership is requested by the RSM server. Once returned, a logout is called and the connection is terminated.

Note that the total supervisor count in Unified CCE must be spread across CTI OS desktop users and RSM. For example, in a 2000 agent configuration, up to 200 agents can be supervisors. This means that the total supervisor count between CTI OS and RSM must not exceed 200.

CTI OS connections may be optionally encrypted (through the use of IP Sec configurations). However, this will induce a significant performance penalty on the server itself when higher agent loads are utilized. For more information about enabling CTI OS connection security, see the *Cisco Remote Silent Monitoring Installation and Administration Guide*, available at cisco.com.

### VRU

The RSM platform does not directly media-terminate inbound calls. Instead, supervisors dial into a Unified CVP or IP IVR-based VRU system, which runs call flow script logic that interacts with services hosted on the RSM server over HTTP. Thus, if a given RSM installation is to support up to 40 dialed-in supervisors, there must be a VRU present (as well as the necessary PRI/network resources) that can offer this same level of support.

Furthermore, a caller accessing RSM will often place a higher load on the VRU processor) and memory than a caller accessing some more traditional IVR-type call flow. This is because, in a more traditional IVR call flow, shorter, oftentimes cached or non-streamed prompts are played, separated by periods of caller input gathering and silence. With RSM, however, the predominant caller activity is monitoring an agent’s call, and to the VRU this looks like the playback of a long streaming audio prompt, which is an activity that requires a relatively high level of VRU processor involvement.

With Unified CVP deployments, supported VXML gateway models are listed in the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal (Unified CVP),* otherwise known as Unified CVP Bill of Materials (BOM), available at cisco.com.

When provisioning a VRU for use by RSM, a good rule of thumb is to count each RSM call as 1.3 non-RSM calls on a processor/memory-usage basis. So for a VRU that can normally handle 40 concurrent calls, plan for it to be able 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 IP IVR.

When RSM is used with CVP, the gateway 'IVR prompt streaming for HTTP' needs to be enabled. Note that this setting is not recommended for other CVP applications. Therefore RSM requires a dedicated VXML gateway. This gateway must not be used for other CVP applications. Also, 1GB of gateway memory is recommended for use with RSM. This will support up to 40 concurrent monitoring sessions per gateway.

### Agent Phones

Use of RSM to monitor an agent requires that that agent’s phone be a third generation of Cisco Unified IP Phones 79*x*1, 79*x*2, 79*x*5, 7970, or newer. This is because these phones include extra DSP resources in the form of a Built-in-Bridge (BiB). The BiB allows the phone to fork off a copy of the current conversation stream to the RSM server.

Cisco Unified Contact Manager provides for a maximum of one active monitoring session per agent because the agent’s phone can handle only one active monitoring session and one active recording session at any given time.

So, if a third-party recorder is recording the agent’s conversations, the agent can still be monitored by a supervisor using supervisor desktop or RSM. However, if both a RSM-based supervisor and a supervisor-desktop-based supervisor both tried to monitor the agent during the same time period, the request would fail with the last one to try because it would exceed the above-mentioned monitoring limit.

Note that RSM will set up only one monitoring session through Unified CM for a single monitored agent, even if two or more RSM users are requesting to monitor the agent's call at the same time. In this case, RSM forks the stream to cover all RSM users. This allows more than two RSM-based supervisors to monitor the same agent, for instance. However, if there are multiple RSM servers in the environment that monitor the same agent, they will each make a separate monitoring call to that agent.

If the monitoring call limit has been reached for a specific agent and a dialed-in supervisor then attempts to monitor this same agent, the supervisor's request will be denied through an audio prompt feedback from the system stating that the agent cannot be monitored.

## RSM Hardware Considerations

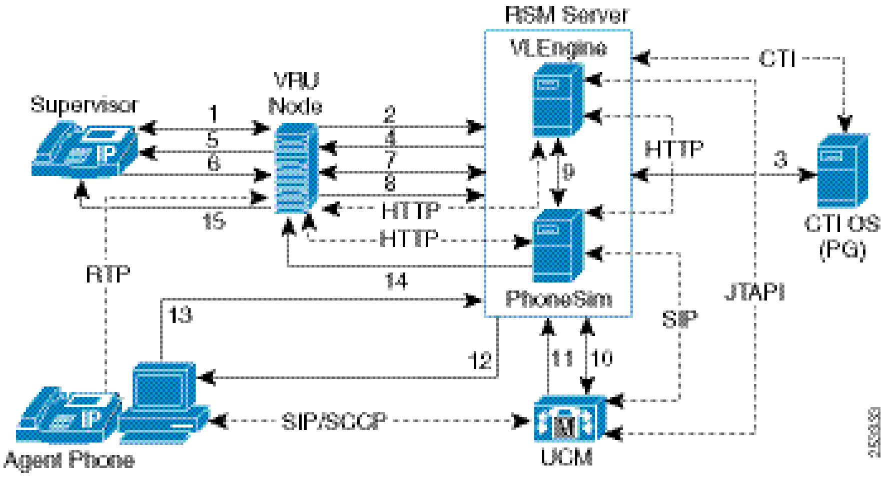
RSM is supported in installations where the number of agents in the enterprise is less than 8,000 and the number of maximum concurrent supervisors using the system is less than 80. In all supported RSM configurations, the VLEngine and PhoneSim components are installed on the same physical server.

For more information, see the RSM Requirements section of the *Cisco Remote Silent Monitoring Installation and Administration Guide*, available at cisco.com.

## RSM Component Interaction

Figure 83 illustrates the types of interactions that occur when a supervisor dials into an RSM-enabled platform and monitors an agent.

**Figure 83** ***Remote Silent Monitor Enabled Call Flow***



### RSM Call Flow

Figure 83 shows the following call flow steps:

1. Supervisor calls in, and the call is media-terminated on the VRU (Unified CVP or 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 their credentials.
2. After the user enters their credentials, the VRU makes a login request to RSM over HTTP.
3. The VLEngine component in RSM interacts with the CTI OS server to validate the authentication credentials.

4. VLEngine replies back to the VRU node by using HTTP with the authentication result.
5. If the supervisor is successfully authenticated, the script in the VRU will play the main menu prompt. From here, the supervisor will be allowed to monitor an agent.
6. The supervisor chooses to monitor a single agent from the main menu, and enters a Directory Number (DN) of an agent to be monitored.
7. The VRU checks with VLEngine if the given agent can be monitored. VLEngine then checks whether the agent with that DN is logged in, is in talking state, and is in the supervisor's team, using previously cached event feed information from the CTI OS server. If so, it replies back to the VRU node.
8. The VRU node then sends a monitor request to PhoneSim to monitor the entered DN.
9. VLEngine works internally using HTTP.
10. Following that, VLEngine sends a JTAPI request to Unified CM to monitor the agent’s phone, and it gets a JTAPI success response.
11. The PhoneSim component will then receive a SIP-based instruction from Unified CM for a simulated phone that it manages, to establish a monitoring call with the agent’s phone.
12. The chosen simulated phone establishes the monitoring call with the agent’s phone based on Unified CM's above request.
13. After the establishment of a monitoring call from RSM server to agent, the agent phone's Built-in-Bridge (BiB) forwards the call conversation to PhoneSim in the form of RTP packets.
14. In turn, PhoneSim strips the RTP headers and streams this data to the VRU node over HTTP as a response to the request made earlier in step 8.
15. The VRU then plays the data to the supervisor as if it were a streaming audio prompt.

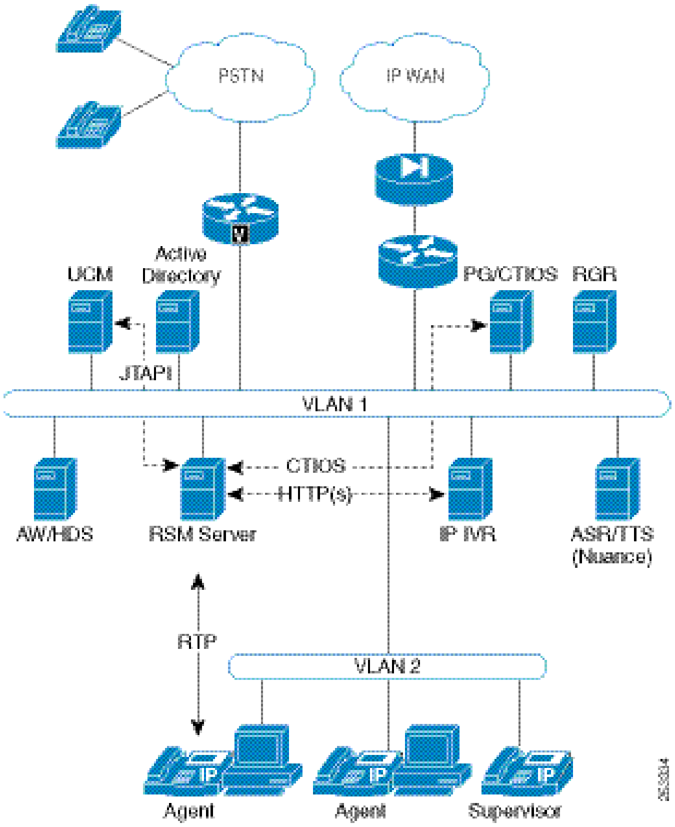
## RSM Deployment Models

This section illustrates some basic supported RSM deployments.

### Single Site

Figure 84 illustrates the basic network connectivity of RSM deployed within a typical single-site configuration.

**Figure 84** *Typical RSM VLAN Configuration*



As shown in Figure 84, supervisors may dial in through a VoIP phone as well as through the PSTN. The VRU that handles the supervisor's call is IP IVR in this case.

Typical RSM VLAN Configuration also illustrates the various protocol interfaces that RSM has into the rest of the system:

- **HTTP(S)**: As stated previously, HTTP is used as the carrier protocol for VRU-based requests into the RSM system. A request takes standard URL form and may look like one of the following URLs:

```
http://rsmserver:8080/vlengine/checkUserCredentials?supervisorID=1101&pin=1234&outputForm
at=plain
http://rsmserver:8080/vlengine/canMonitorAgentID?supervisorID=1101&agentID=1001&outputFor
mat=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 by using the GET method. The return data (as an HTTP response) is either plaintext or encapsulated in VoiceXML, depending on the API call being used and on the value specified for the outputFormat parameter (if available for that call).

- **CTI OS**: The RSM server makes several connections to CTI OS. One of these connections is for receiving platform events. (In the language of CTI OS, it is a monitor-mode connection.) The other(s) are what CTI OS calls agent mode connections and they are used to authenticate logging-in supervisors if the standard authentication facilitates are being utilized.

- **JTAPI**: The request to start monitoring an agent's phone is made through JTAPI. This requires a JTAPI application user to be defined on each Unified CM cluster in the environment, and to be associated to all agent phones.
- **RTP**: While a dialed-in supervisor is monitoring an agent, there will be a monitoring call in progress from the BiB (built-in-bridge) of that agent's phone to the RSM server. While the signaling data for this call is run through Unified CM (just like any other call), the RTP traffic will flow between the agent phone and the RSM server.

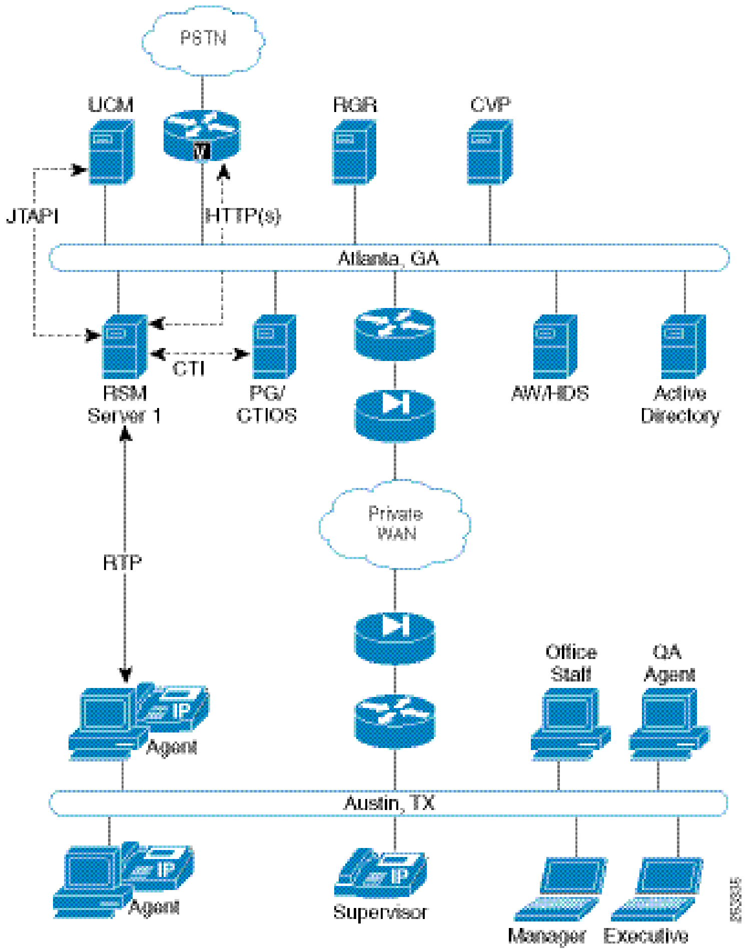
### Multisite WAN

The follow scenarios depict basic supported configurations for the RSM product in a multisite deployment.

#### Single Cluster, Single VRU

Figure 85 depicts a simple multisite setup involving a single Unified CM cluster and a single VRU.

**Figure 85** ***Multisite Deployment with a Single Unified CM Cluster and Single VRU***



In this case, the Unified CM and Unified CCE environment is co-located in Atlanta, and the Austin location contains the entire end-user population. The VRU is a VXML Gateway/voice gateway in Atlanta, controlled by a Unified CVP Call Server also in Atlanta.

The supervisor in Austin could possibly have two ways of dialing into the RSM system:

- Through the PSTN — Here the supervisor would dial an E.164 number, and the call would be hairpinned through the voice gateway. The Unified CVP RSM callflow application would handle the call as normal from that point.
- As a VoIP extension — In this case, Unified CM would have a trunk configuration set up to the VRU. The call would remain VoIP all the way through, and the call would likewise be handled by the Unified CVP RSM callflow application.

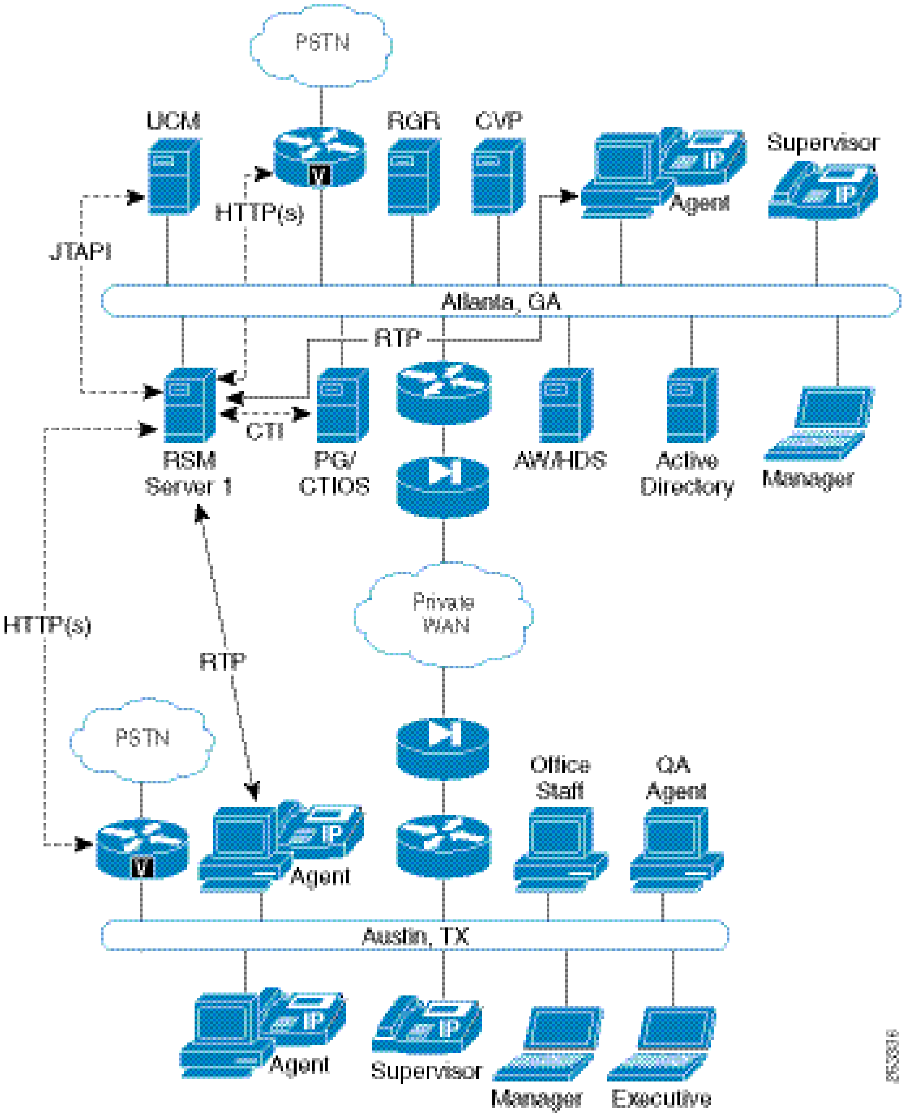
In this scenario, all RSM traffic is confined to the Atlanta site except:

- The RTP traffic of the agent being monitored (signified as a red dotted line)
- The actual supervisor call into the platform

**Single Cluster, Multiple VRUs**

Figure 86 depicts a multisite deployment with a single Unified CM cluster and multiple VRUs.

**Figure 86** ***Multisite Deployment with a Single Unified CM Cluster and Multiple VRUs***



This scenario is similar to the previous one, 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 could 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 still used as well as the Unified CVP Call Server. However, there are two VXML Gateways, one at each site. The environment is configured so that a supervisor dialing RSM from Austin will be routed to the RSM callflow application on the Austin VXML Gateway, while a supervisor dialing in from Atlanta will be routed to the Atlanta VXML Gateway.

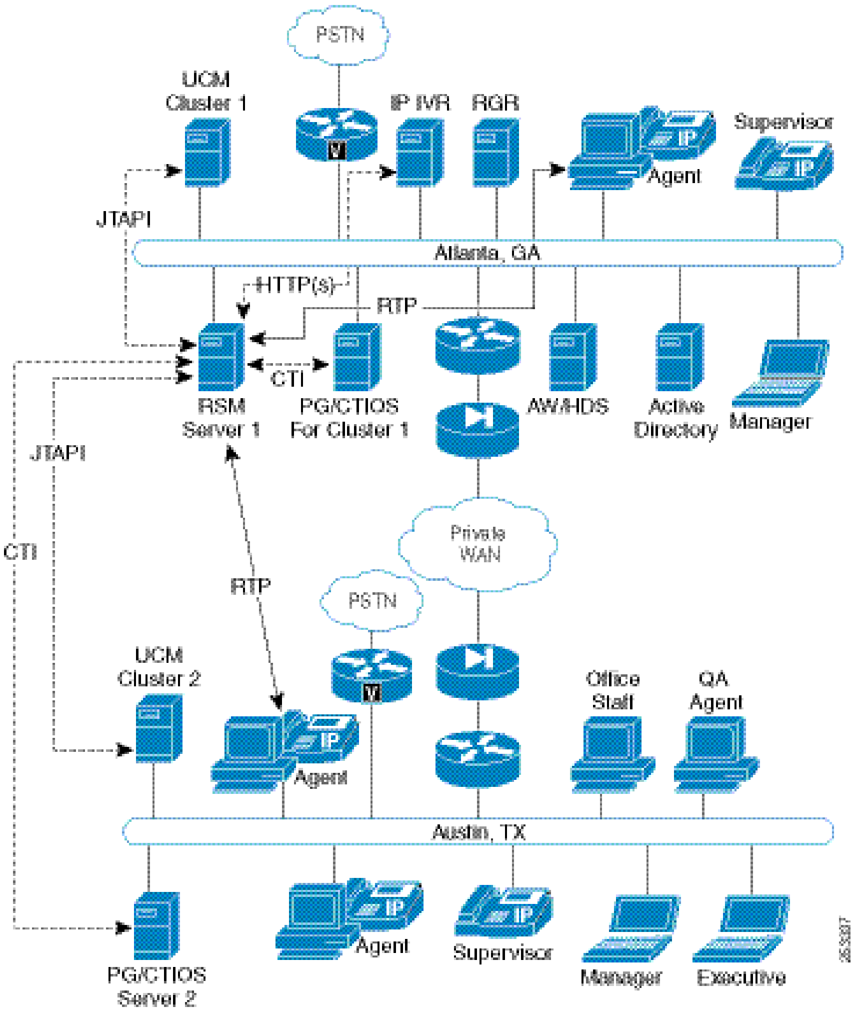
Because the Atlanta site houses the Unified CM and Unified CCE environment, all RSM-related JTAPI and CTI OS traffic is still confined there. However, the addition of a VXML Gateway at Austin will lead to HTTP-based traffic being streamed between the sites over the WAN. This traffic consists of relatively small requests from the gateway to the RSM server for services, and the RSM server's responses. The responses themselves can be sizeable, especially when it is the data for a monitored conversation.

Also, 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.

### Multiple Cluster, Single VRU

Multisite Deployment with Multiple Unified CM Clusters and a Single VRU depicts a multisite deployment with multiple Unified CM clusters and a single VRU.

**Figure 87** ***Multisite Deployment with Multiple Unified CM Clusters and a Single VRU***



This configuration includes a Unified CM cluster at both the Atlanta and Austin sites and a single 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 is linked to the CTI OS servers of both clusters in order to track all agents in the enterprise.

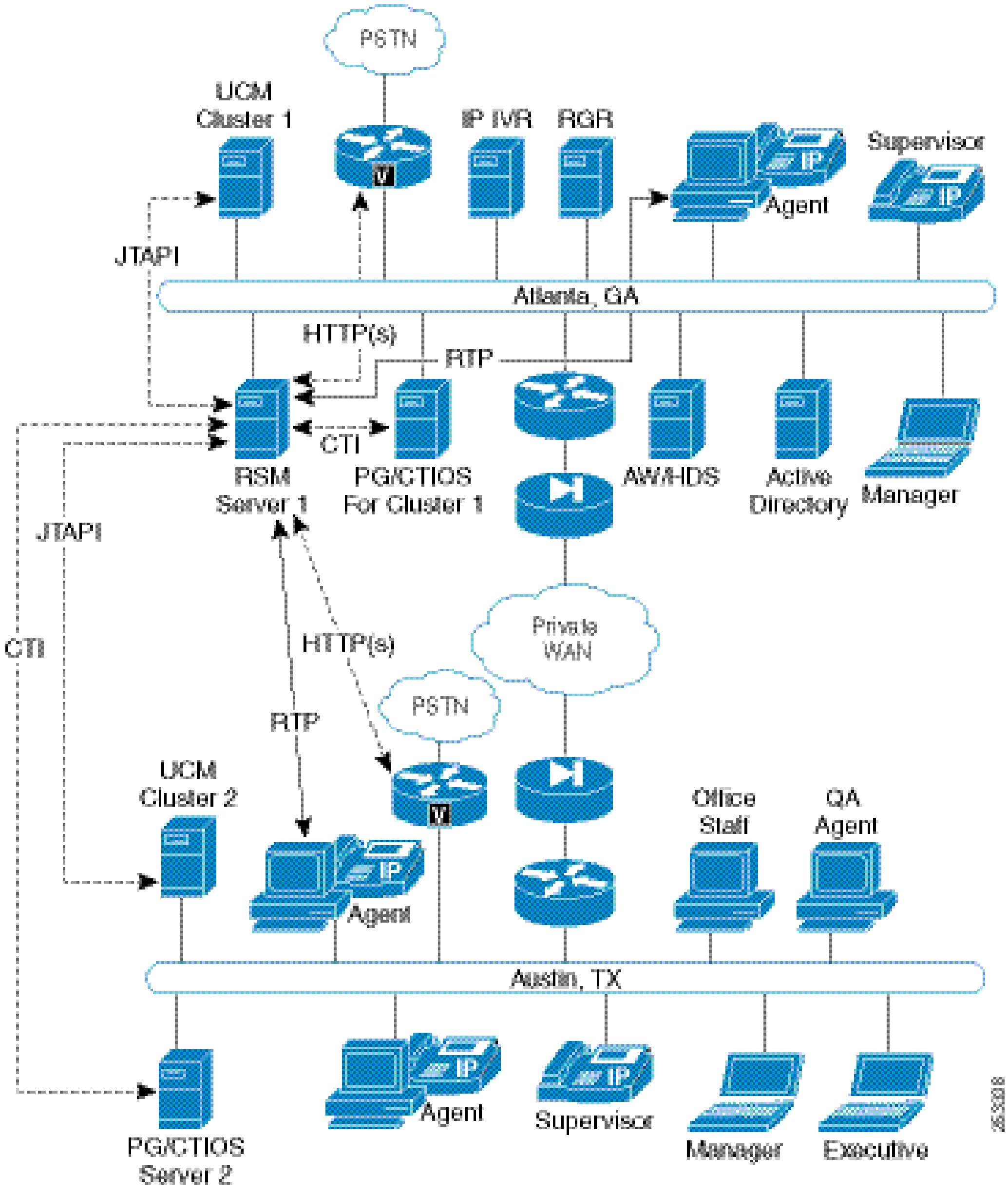
As IP IVR is in use, a supervisor call to the RSM callflow will be routed to, and media-terminated on, this IP IVR system over either the PSTN or IP WAN (as discussed previously). No VXML Gateway is involved in this configuration, and all RSM-related HTTP interaction is confined to the Atlanta site, between the RSM and IP IVR systems.

Because a Unified CM cluster now exists at the Austin site, several classes of data that RSM uses to track environment state and initiate agent monitoring requests (CTI OS and JTAPI traffic) are sent over the IP WAN.

### Multiple Cluster, Multiple VRUs

Multisite Deployment with Multiple Unified CM Clusters and Multiple VRUs depicts a multi-site deployment with multiple Unified CM clusters and multiple VRUs.

**Figure 88** ***Multisite Deployment with Multiple Unified CM Clusters and Multiple VRUs***



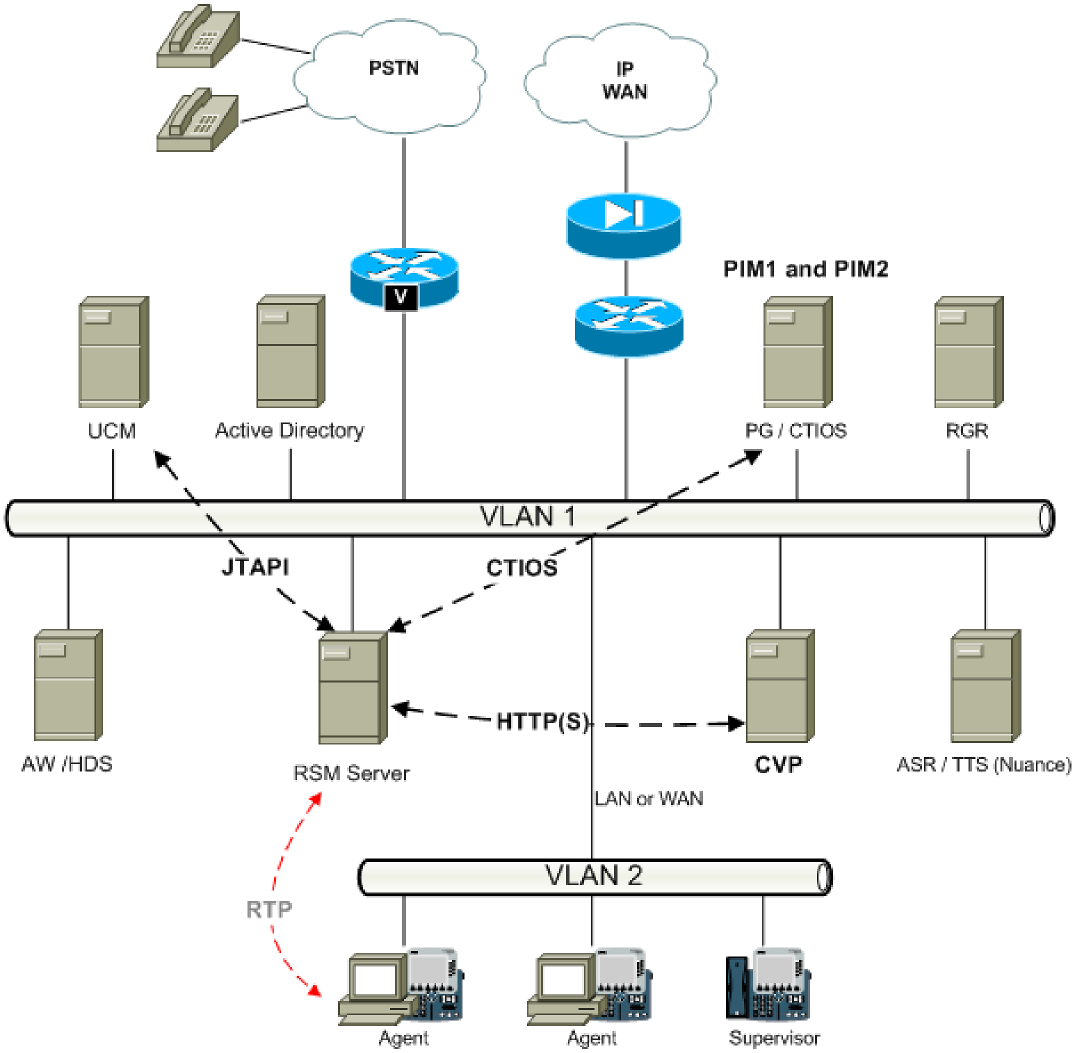
Multisite Deployment with Multiple Unified CM Clusters and Multiple VRUs  illustrates a Unified CM cluster as well as a Unified CVP VXML Gateway/voice gateway at each site. It is a combination of the previous deployment models, and it has the following characteristics:

- The Unified CVP Call Server controls the VXML Gateways at each site.
- Because there are agent phones at both sites, RTP data could 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. This means that a supervisor could dial in from either site (or anywhere in the world through PSTN) and listen to an agent in Atlanta or Austin.

### Single Cluster, Single PG/CTIOS with two UCM PIMs (agent expansion UCCE deployment configuration)

This diagram depicts a setup involving a single UCM cluster and single Agent PG/CTIOS server configured with two UCM PIMs.

**Figure 89** *Single UCM cluster and single Agent PG/CTIOS server configured with two UCM PIMs*



In this case, the single PG/CTIOS server is configured with two PIMs and each PIM points to a separate subscriber pair in the UCM cluster.

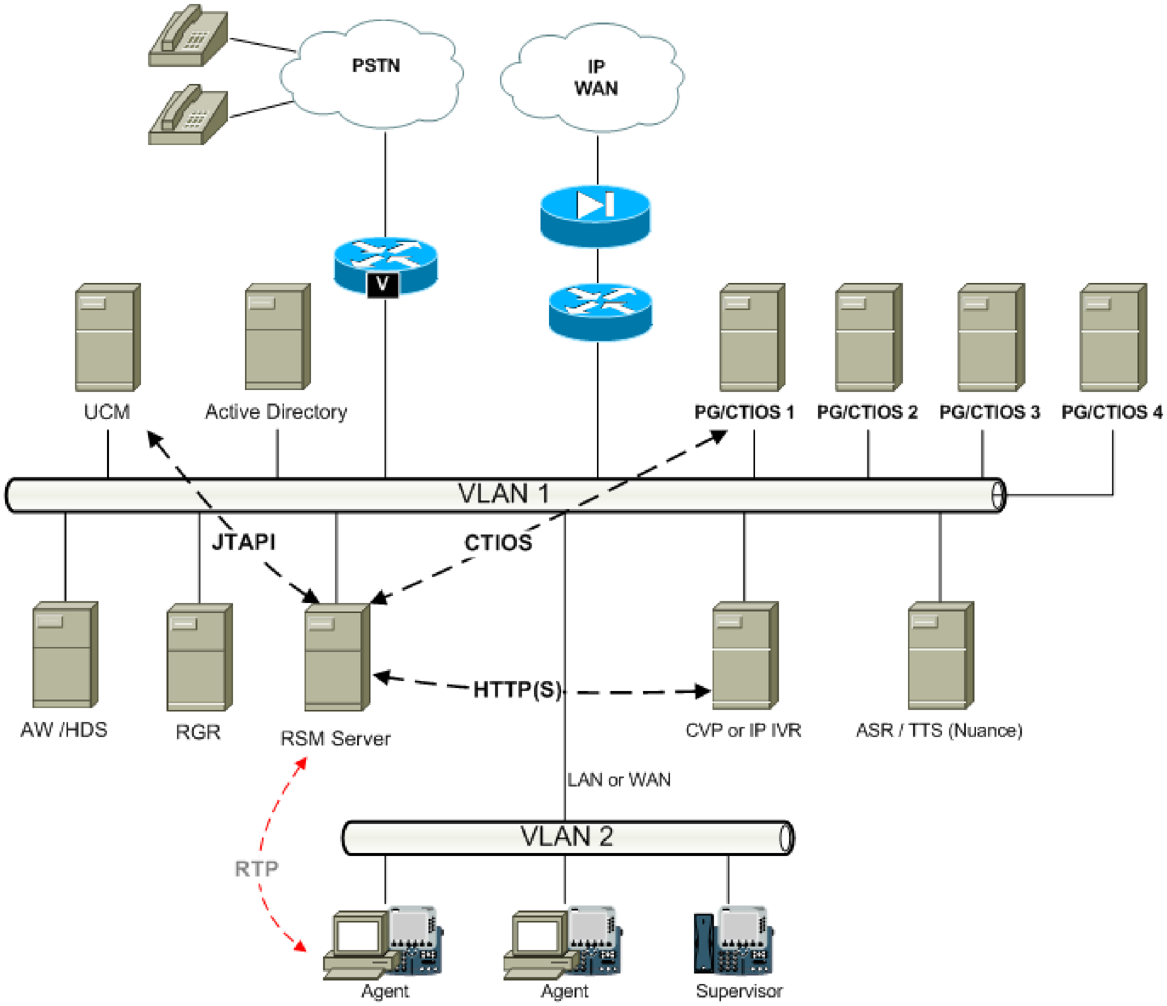
This is supported in UCCE/UCM 8.0 and higher, only when more than 2000 agents (up to 4000 agents) need to be configured in a single UCM cluster with a Unified CVP VRU. This expansion is not supported with IPIVR. See the “**Error! Reference source not found.**” section and Figure 129.

In this scenario RSM should be configured with two clusters with each specifying the corresponding UCM PIM Peripheral Number and its UCM Subscriber pair.

### Single Cluster, Multiple PG/CTIOS (agent expansion UCCE deployment configuration)

This diagram depicts a setup involving a single UCM cluster and multiple (up to 4) Agent PG/CTIOS servers

**Figure 90** *Single UCM cluster and multiple (up to 4) Agent PG/CTIOS servers*



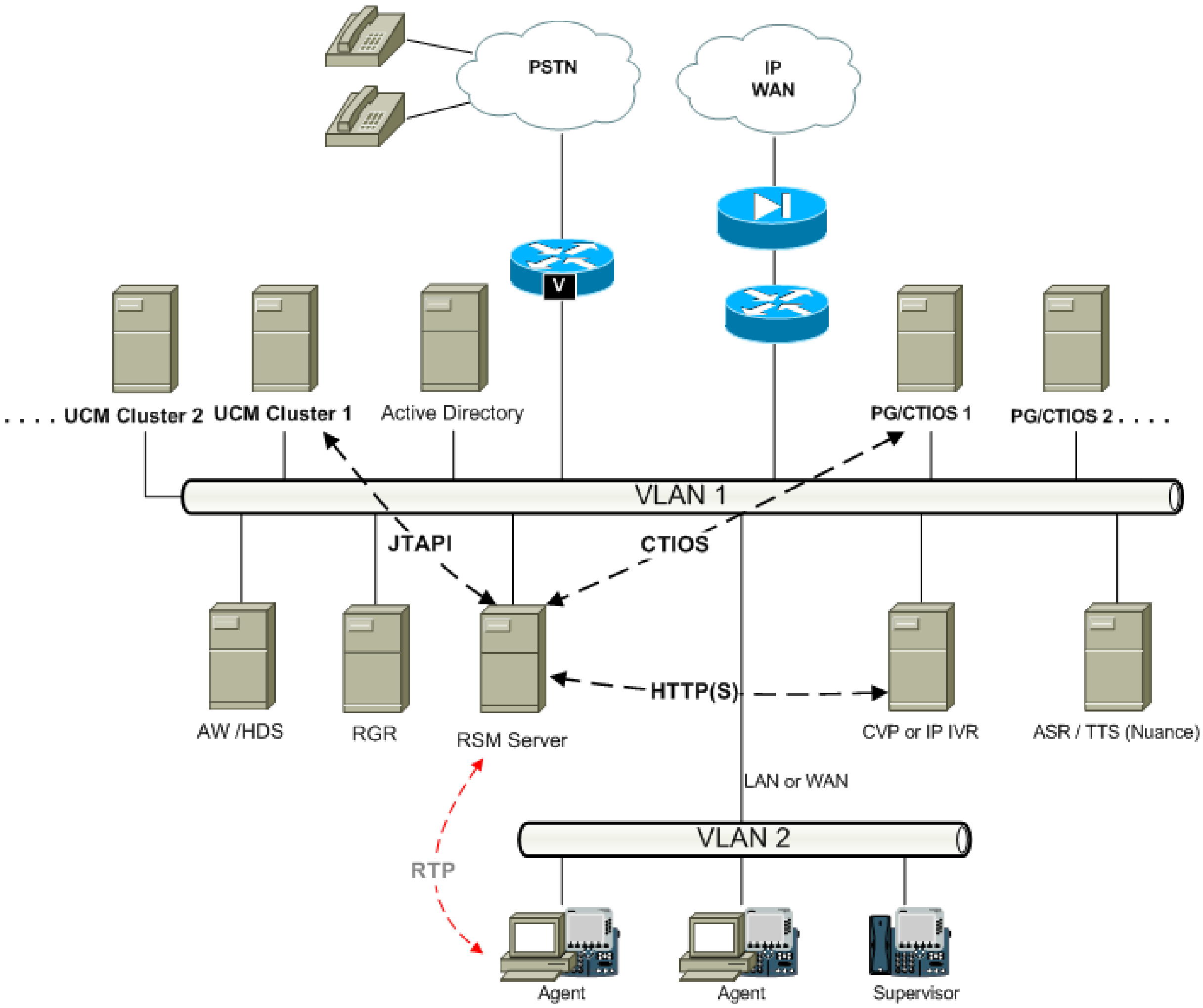
In this setup, a separate Agent PG/CTI OS server is deployed for each subscriber node pair (primary and backup). See the “**Error! Reference source not found.**” section and Figure 127.

In this deployment scenario a separate RSM cluster (in a single RSM instance) should be configured corresponding to each Agent PG/CTI OS server and its UCM subscriber pair

### Multiple Cluster, Multiple PG/CTIOS (agent expansion UCCE deployment configuration)

This diagram depicts a setup involving multiple UCM clusters and multiple Agent PG/CTIOS servers.

**Figure 91** ***multiple UCM clusters and multiple Agent PG/CTIOS servers***



In this method, a separate Agent PG/CTIOS server is deployed for each UCM cluster.

In this deployment scenario a separate RSM cluster (in a single RSM instance) should be configured corresponding to each Agent PG/CTIOS server and its UCM cluster subscriber pair.

### Bandwidth Requirements

As part of the network planning done before deploying the RSM solution, you should verify that the network infrastructure can support the bandwidth requirements of RSM.

The RSM solution has connectivity with multiple components in the larger Cisco environment (as the diagrams in the previous section demonstrate). Table 4 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 4** ***Bandwidth Requirements***

| RSM Peer | Purpose | Protocol(s) 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) | G711, chunked transfer mode encoding | 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 OS Server (PG) | Environment Events / Supervisor Logins | TCP (CTI OS) | Binary (CTI OS stream) | Minimal (< 1000 agents)<br>Moderate (> 1000 agents)<br>(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) | High (about 67 to 87 kbps per session) | < 400 ms avg. |

## Agent Phone Bandwidth Figures

Currently, the simulated phones on the RSM server support using only G.711 mu-law to monitor agent phones.

For bandwidth usage information, see the Cisco Voice Over IP - Per Call Bandwidth Consumption Tech Note.

There must be sufficient bandwidth 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 small branches on limited bandwidth or WAN connectivity. Regular Call Admission Control (CAC) and bandwidth calculations are applicable for monitoring calls.

Since G.711 is the only codec supported for monitoring calls between agent IP phone and RSM server (phonesim), use the Cisco TAC Voice Bandwidth Codec Calculator for additional bandwidth capacity planning.

## Agent Phone Transcoding Implications in G729 Environments

The monitoring call established between RSM's simulated phone (simphone) and Agent's phone is subject to regular call admission control (CAC) procedures. RSM server (phonesim) supports only G.711 codec and hence the simphones should be configured in a region (UCM Region) for G.711 only. Due to this, the monitoring call always negotiates a G.711 codec between RSM's simphone and Agent phone's BiB. If the Agent-Customer conversation is in G.711 then no transcoding is required. If the Agent-Customer conversation is in G.729 then transcoding is performed by Agent Phone's BiB itself. No additional transcoding resources or hardware needed in Voice Gateway. This BiB transcoding capability exists in both physical and soft-phone (IPC 7.0 or higher) models.

For more information, see the "Codec for Monitoring and Recording Calls" topic in "Monitoring and Recording" chapter of the *Cisco Unified Communications Manager Features and Services Guide*.

### Failover Redundancy and Load Balancing

Load balancing support is defined as the act of multiple RSM servers being associated together so that the incoming request load is distributed among them. The definition of failover is multiple RSM servers being associated together so that if one fails, the other(s) can act in its place. In the future, RSM will support load balancing and failover with both the Unified CVP and IP IVR VRUs. Currently, this support is not available in RSM 1.0. RSM 1.0 does, however, support the deployment of multiple standalone RSM servers within a single Unified CCE environment, and this concept is demonstrated in the advanced deployment scenarios described in this document.

Table 5 indicates how a failure of each of the various components affects a live supervisor call.

**Table 5** *Impact of Failures on a Supervisor Call*

| Component That Fails | Worst Possible Impact |
|---|---|
| VRU Node (IP IVR, Unified CVP) | Supervisor's call is terminated as any VRU failover occurs (depends). Supervisor may dial back in and log in again once VRU failover 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 will have the voice stream terminated and be 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) will result in a configurable delay of 3 to 5 seconds or so, 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, attempting to monitor an agent, and so forth), will fail, although the RSM callflow will still be answered because it is being hosted on the VRU node. |
| VLEngine or PhoneSIM software failure | Service automatically restarted via service wrapper. Supervisors with a request in-progress are given an error message and have a chance to 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, attempting to monitor an agent, and so forth), will fail, although the RSM callflow will still be answered because it is being hosted on the VRU node. |
| Unified CCE fails (CTI OS) | RSM will lose connectivity to the CTI OS server when the PG fails or is cycled. If connectivity to both CTI servers on a cluster fails, RSM will keep retrying both, connecting to the first server that is available. (The CIL's failover code is used for all of this.) When connectivity comes back up to a CTI server, the agent and call lists will be cleared and refreshed (to avoid "stale" agents). During this time, no new call events will be received, and the system will be working from an "out-of-date" agent and call list. Therefore some monitoring requests will fail, saying the agent is not talking when he or she is, and some monitoring requests will fail because the system would think 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 would work. |
| Unified CM fails (JTAPI) | Connectivity to one or more JTAPI providers will be 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 will fail-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 will periodically retry both and re-establish the connectivity to the first that comes up. Attempts to monitor agents (for example, monitorAgent calls) made during this time will fail until the JTAPI connection is re-established. |

### Host-Level Security

Incoming access to the RSM server can be restricted to only the necessary components by using the host-based Access Control List (ACL) functionality built into the Windows Server OS. In the most secure configuration, incoming access to the RSM system is permitted from the VRU systems. Built-in host-based access control can also be employed to allow limited access to other services if desired, such as remote administration mechanisms like Windows Remote Desktop and VNC.

Even though this is not required, a recommended ACL Configuration for a single-server RSM configuration would be 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)

### Cisco Security Agent

As part of the installation procedure, Cisco highly recommends that you install the Cisco Security Agent (CSA) software on the RSM system. This topic is covered in the Security Settings chapter of the *Cisco Remote Silent Monitoring Installation and Administration Guide*, available at cisco.com.

### Transport or Session Level Security

Because RSM maintains multiple connections to a number of components in the larger Cisco Contact Center environment, there is no simple answer to whether transport or session level security is supported or not. The follow notes describe RSM's support for this feature by protocol type:

**RSM to VRU (HTTP):** Currently there is no support for encryption of the HTTP-based data exchange between RSM and the VRU node.

**RSM to PG/CTI OS Server (CTI):** Because RSM makes use of the Java CIL, all CTI OS servers used by it 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 Security Settings chapter of the *Remote Silent Monitoring Configuration and Administration Guide*, available at cisco.com.

**RSM to UCM (JTAPI):** Like CTI OS traffic, JTAPI traffic may be encrypted through the use of IPSec transport mode encryption. For more information, see the Security Settings chapter of the *Remote Silent Monitoring Configuration and Administration Guide*, available at cisco.com.

**RSM to Agent Phone (RTP):** Currently there is no support for encryption of the RTP stream between Agent Phone (BiB) and RSM SimPhone. Secure RTP (SRTP) is not supported by RSM SimPhones.

### Support for Mobile Agent, IP Communicator, and Other Endpoints

Currently, the underlying Unified CM 8.0 monitoring functionality does not provide monitoring support for endpoints using any one of the following:

- Cisco Mobile Agent
- Second generation or older phones, such as the Cisco Unified IP Phones 7940 or 7960
- A media-terminated CTI OS Agent Desktop
- Monitoring of encrypted phone calls

Therefore, support for these products is also not available through RSM. For further information about this restriction, see the section on Silent Monitoring.

### Support for 6900, 8900, and 9900 Phone Models

6900, 8900 and 9900 phone models have Join, Join Across Lines (JAL), Direct Transfer, and Direct Transfer Across Lines (DTAL) features, which RSM does not support.

Since 6900 phones allow disabling of these features, RSM can support monitoring these phones if these features are disabled.

Since 8900 and 9900 do not allow for disabling of these features, RSM does not support monitoring of these phones.

## 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 or IP Phone Messenger (IPPM) 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.

Figure 92 and the description that follows explain how various components of CAD and Cisco Unified Presence interface with each other.

**Figure 92** ***Interface Between CAD and Cisco Unified Presence***

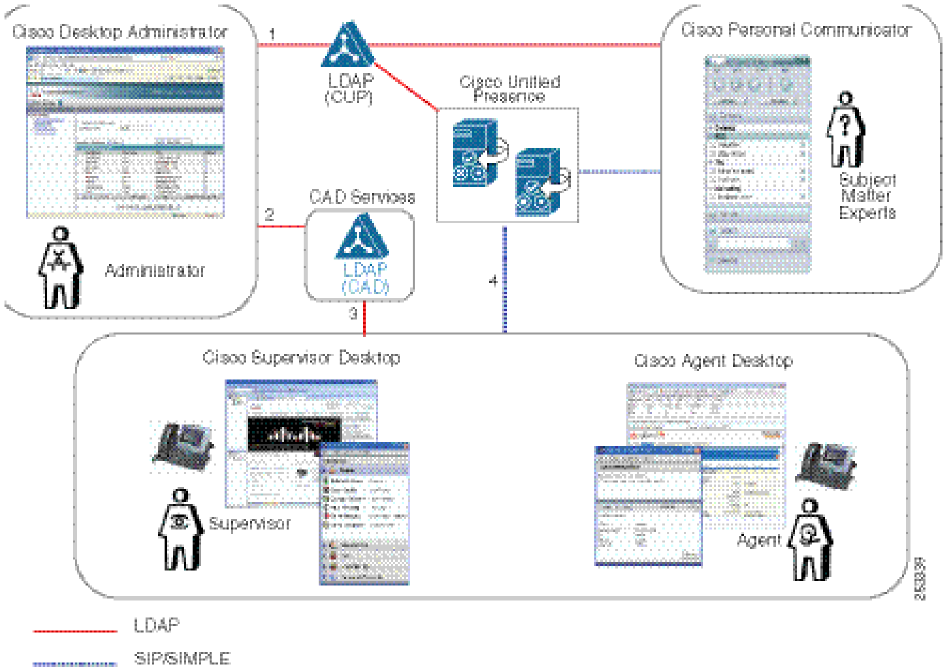


Figure 92 depicts the following sequence of events:

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 information about Cisco Unified Presence in the *Cisco Unified Communications SRND*.

## NAT and Firewalls

This section discusses deploying Cisco Agent Desktop (CAD), 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 regarding NAT and firewalls, see the Securing Cisco Unified Contact Center Enterprise chapter.

## Cisco Finesse and NAT

When Cisco Finesse is deployed in a network environment where two or more disjointed networks are interconnected using NAT, the Finesse clients and Finesse servers must all be located on the same network. In addition, the Finesse server must be located on the same network as the CTI Server on the Agent PG.

## 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, CAD-BE, 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 between the desktop applications as long as the firewall allows the required type of traffic through and the appropriate ports are opened. 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. Figure 93 shows the traffic types used between the CAD components.

For detailed port information, see the *Port Utilization Guide for Cisco ICM/IPCC Enterprise and Hosted Editions*.

**Figure 93** ***Communication Between CAD Components***

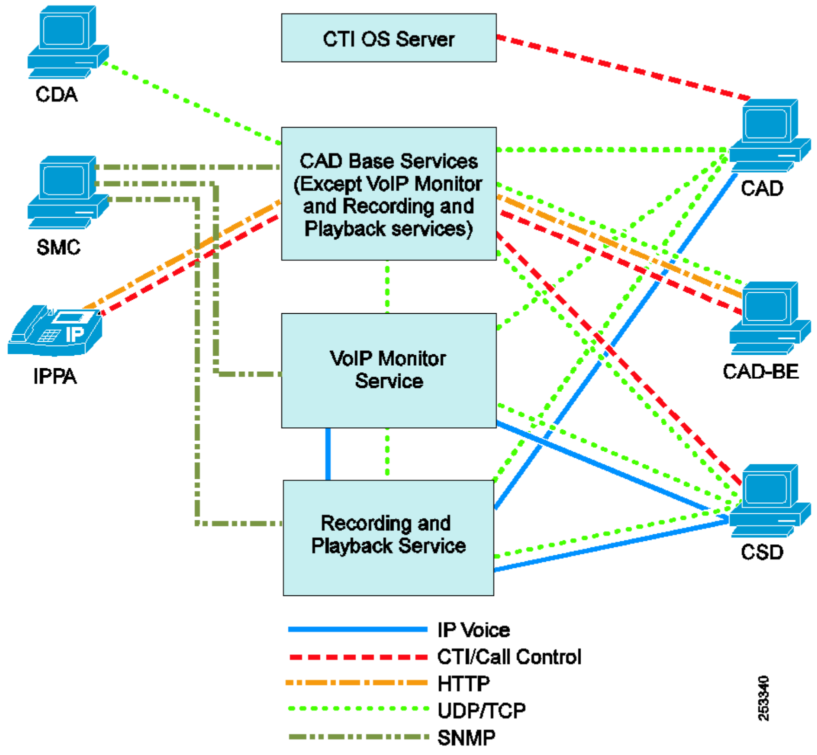

CTI OS Server
CDA
CAD Base Services (Except VoIP Monitor and Recording and Playback services)
CAD
SMC
IPPA
CAD-BE
VoIP Monitor Service
Recording and Playback Service
CSD
IP Voice
CTI/Call Control
HTTP
UDP/TCP
SNMP
253340

Figure 93 shows that IP voice streams are exchanged between the VoIP providers (CAD, the VoIP Monitor service, and the Recording & Playback service) and the VoIP requestors (CSD and the Recording & 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 CAD-BE, CSD, and 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 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 applet and the SMC servlet running on the CAD Base Services machine. HTTP is also the protocol used by the IPPA service and CAD-BE applet to communicate with the Browser and IP Phone Agent service.

The UDP/TCP traffic shown in the 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.

## CTI Toolkit Desktop and NAT

When the Cisco CTI Toolkit Desktop is deployed in a network environment where two or more disjointed networks are interconnected using NAT, then Unified CM, the physical IP Phone, the Cisco CTI OS Server, the Cisco CTI Toolkit Desktop, and the Cisco CTI OS IPCC Supervisor Desktop must all be on the same network.

## Coresidency of CTI OS and CAD Services on the PG

Cisco recommends that you install CTI OS and CAD Services (including VoIP Monitor and Recording) on the PG. This does reduce the supported maximum agent capacity on the PG. If the supported PG capacity numbers provided in the chapter on Sizing Unified CCE Components and Servers are not sufficient and you want to run these software components on separate servers to increase agent capacity on the PG, then prior approval from the Unified CCE product management team is required.

Legacy deployments where CTI OS or CAD Services were previously installed on separate servers are still supported. However, customers are encouraged to migrate CTI OS and CAD Services onto the PG. For more information regarding deployment configurations, see Chapter 2, “Deployment Models”.

## 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.

## Support for Mix of 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. If a mix is deployed, the sizing limitations of Finesse apply. Note that the Cisco Finesse supervisor application can monitor only Finesse agents, and the CTI OS supervisor application can monitor only CTI OS agents.

## Support for IP Phones and IP Communicator

CAD, CAD-BE, and the CTI Toolkit Desktop support the use of Cisco IP hardware phones and/or the Cisco IP Communicator software phone.

Some CAD agent application features (CAD, CAD-BE, and IPPA) require particular phone models, and some installations support either hardware phones or software phones but not both. For information about the exact phone models and IP Communicator versions supported, see the CAD documentation on cisco.com.

### IP Phones and Silent Monitoring

Silent Monitoring of agents is supported 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. Even analog phones can be used for this feature.

### IP Phones and Citrix or MTS

Both the Cisco IP hardware phones and Cisco IP Communicator are supported when using Citrix or MTS with either CAD or the CTI Toolkit desktops. In these environments, Cisco IP Communicator must be installed on the Agent desktop PC and cannot be deployed on the Citrix or MTS server.

### IP Phone Agent

The IP Phone XML service agent application supports only hardware IP phones because there is no desktop.

## 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’s 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 is configured as the primary or backup service for capturing voice streams. If desktop monitoring is configured as the primary method (with no secondary method), this information does not apply.

### Network Hubs

A network hub (including a "smart" hub) is not allowed between an agent’s hardware phone and PC when Desktop Monitoring is configured for the agent.

### Multiple Daisy-Chained Hardware Phones

There may be only a single hardware phone connected in series between the agent’s PC and the switch when Desktop Monitoring is configured for the agent.

### NDIS Compliance of NICs

The 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 using SRTP, the silent monitoring and recording feature will not work correctly. Although the voice streams can still be captured, they will not be decoded correctly. The end result is that speech will be unintelligible.

## High Availability and Failover Recovery

For detailed information about CAD, CTI Toolkit Desktop, and Cisco Finesse high availability, see the Design Considerations for High Availability chapter.

## Bandwidth and Quality of Service

For detailed information about CAD, CTI Toolkit Desktop, and Cisco Finesse bandwidth usage and QoS, see the Bandwidth Provisioning and QoS Considerations chapter.

## Desktop Latency

Agent and Supervisor desktops can be located remotely from the agent PG. Technically, the delay between the CTI OS server and CTI Toolkit Desktop clients, as well as between the CAD server and CAD/CSD desktop, could be very high because of high time-out values. However, large latency will affect the user experience and might become confusing or unacceptable from the user perspective. For this reason, Cisco recommends limiting the latency between the server and agent desktop to 400 ms round-trip time for CTI OS (preferably less than 200 ms round-trip time) and 200 ms round-trip time for CAD (preferably less than 100 ms round-trip time). Longer latencies up to a second are technically supported but will affect the agent experience negatively (for example, the phone will start ringing but the desktop will not be updated until a second later).

# References to Additional Desktop Information

The following additional information related to Cisco Agent Desktop and Cisco Supervisor Desktop is available at the listed URLs:

- *Cisco Unified Contact Center Enterprise (Unified CCE) Software Compatibility Guide*

  This document provides tables outlining Unified ICM/CCE Peripheral Gateway (PG) and Object Server (OS) support for versions of Cisco Agent Desktop, CTI OS Server, CTI OS Toolkit Desktop Clients, Data Collaboration Server (DCS), Siebel 6, and Siebel 7.

- *Voice-Over IP Monitoring Best Practices Deployment Guide for CAD*

  This document provides information about the abilities and requirements of Voice over IP (VoIP) monitoring for Cisco Agent Desktop (CAD). This information is intended to help you deploy VoIP monitoring effectively.

- *Integrating CAD Into a Citrix MetaFrame Presentation Server or Microsoft Terminal Services Environment*

This document helps guide a Citrix administrator through the installation of Cisco Agent Desktop applications in a Citrix thin-client environment.

- *Cisco CAD Service Information*

This document provides release-specific information such as product limitations, service connection types and port numbers, configuration files, registry entries, event/error logs, error messages, and troubleshooting.

CHAPTER 5

# Outbound Option for Cisco Unified Contact Center Enterprise and Hosted

## High-Level Components

Outbound Option for Cisco Unified Contact Center Enterprise and Hosted places outbound calls through a voice gateway. The Outbound Option Dialer is a software solution that does not require telephony cards for tone generation or tone/voice detection.

The Outbound Option solution involves the following processes:

- Campaign Manager and Import processes manage campaigns. They are always installed on the Side-A Logger and service only one customer instance.
- The Dialer process is responsible for dialing customers and connecting them with properly skilled agents or available IVRs. It reports the results of all contact attempts back to the Campaign Manager. All Dialer processes are managed by the central Campaign Manager. 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 also be co-resident on other servers in a Unified CCE deployment. See the Sizing Unified CCE Components and Servers chapter.

## Characteristics

The Outbound Option solution allows an agent to participate in outbound campaigns and inbound calls through a SCCP or SIP software dialer.

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 by using the Unified CCE script editor to control type of outbound mode and percentage of agents within a skill to use for outbound activity.
- Integrated reporting with outbound specific reporting templates.

## Best Practices

The following table shows the differences between SIP Dialer and SCCP Dialer.

**Table 6** ***SIP Dialer and SCCP Dialer Differences***

| SIP Dialer | SCCP Dialer |
|---|---|
| Use the voice gateway dial peers and CUSP routing policies for outbound call routing. | Use the Unified CM routing and dial plans for outbound call routing. |
| No need to configure Unified CM translation pattern to support Campaign ANI. | Need to configure Unified CM translation pattern to support Campaign ANI. |
| Perform CPA at gateway DSP resource. | Perform CPA at Unified CM dialer port. |
| CPA supports both G.711 and G.729 codecs. | CPA supports only G.711 codec. |
| No need to configure dialer port on Unified CM. | Need to configure dialer port on Unified CM. |
| Call Throttling supports 60 CPS per dialer. | Call Throttling supports 5 CPS per dialer. |
| Dialer need NOT be in proximity of voice gateway. | Dialer needs to be in proximity of voice gateway |
| Supports 1500 dialer ports. | Support 120 dialer ports. |
| Supports warm standby architecture. | Does not support warm standby architecture. |
| Requires one MR PIM for MR PG. | Requires two MR PIMs for duplex SCCP Dialers, and one MR PIM for simplex SCCP Dialer. |

| SIP Dialer | SCCP Dialer |
|---|---|
| Only connected outbound calls, which are transferred to agents or IVR, go through Agent PG and Unified CM. | All the outbound calls go through Agent PG and Unified CM. |

### Best Practices for Outbound Option

Follow these guidelines and best practices when implementing Outbound Option:

- Configure abandon to IVR in agent-based campaigns. This is often required to comply with telemarketing laws (for example, FTC/FCC regulations in the US and OfCom regulations in the UK).
- 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 Outbound Option. IP Communicator can introduce an additional delay in transferring customer calls to the agent.

### Best Practices for SCCP Dialer

The following guidelines and best practices are specific to implementing the SCCP Dialer:

Use a media routing PG and a media routing PIM for each SCCP Dialer. The Media Routing PG can be configured for multiple PIMs to support multiple SCCP dialers.

For high availability, deploy multiple SCCP dialers at a single Unified CM cluster. See Designing SCCP Dialer for High Availability . Deploy SCCP dialers in close proximity to the Unified CM cluster where the SCCP Dialers are registered.

Configure the Unified CM node to keep SCCP Dialer traffic localized to one subscriber as much as possible. See SCCP Dialer Throttling Considerations for Unified CM for more details.

Configure the same number of ports for SCCP Dialers at a specific peripheral.

Ensure proper Unified CM server sizing when installing SCCP Dialers. Unified SCCP Dialer places a large strain on Unified CM. See SCCP Dialer Throttling Considerations for Unified CM for more details.

Enable SCCP Dialer call throttling to prevent overloading the Unified CM server. See SCCP Dialer Throttling Considerations for Unified CM.

The Unified CM routing and dial plans are used for SCCP Dialer to place outbound calls. This allows calls to be placed using gateways that are deployed to leverage toll-bypass and lower local calling rates.

**Note** The SCCP dialer does not support CUBE for all releases of Unified CCE.

### Best Practices for SIP Dialer

The following guidelines and best practices are specific to implementing the SIP Dialer:

- Only T1 PRI and E1 PRI interfaces to the PSTN are supported for Outbound SIP Dialers.

- Use a media routing PG with one media routing PIM for duplex 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 duplex 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. See SIP Dialer Throttling Considerations for Voice Gateway and Cisco Unified SIP Proxy Server.
- The voice gateway dial peers and CUSP routing policies are used for SIP Dialers to place outbound calls. This allows 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 an IVR campaign or abandon to an IVR 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.

### Limitations

Cisco Finesse does not support the outbound option for agents.

# Functional Description

The Outbound Option Dialer is a software-only process that is co-resident on the Unified CM PG. The SIP Dialer process has communication sessions with voice gateways, Outbound Option Campaign Manager, CTI Server, and MR PIM. The Dialer process communicates with the Outbound Option Campaign Manager to retrieve outbound customer contact records and to report outbound call disposition (including live answer, answering machine, RNA, and busy). The Dialer process communicates with the voice gateway to place outbound customer calls. The Dialer process communicates with the CTI Server to monitor skill group activity and to perform third-party call control for agent phones. The SIP Dialer process communicates with the MR PIM to submit route requests to select an available agent.

The SCCP Dialer process has communication sessions with Unified CM, Outbound Option Campaign Manager, CTI Server, and MR PIM. The SCCP Dialer process communicates with the Outbound Option Campaign Manager to retrieve outbound customer contact records and to report outbound call disposition (including live answer, answering machine, RNA, and busy). The SCCP Dialer process communicates with Unified CM to place outbound customer calls and agent reservation calls from the dialer ports and thus has an impact on the Unified CM cluster. The SCCP Dialer process communicates with the CTI Server to monitor skill group activity and to perform third-party call control for agent phones. The SCCP Dialer process 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 enable it to run in full blended mode (an agent can handle inbound and outbound calls), in 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), or 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.

Multiple voice gateways and Unified SIP Proxy Server are used to achieve high availability for SIP Dialer deployment, while multiple SCCP dialers are used to achieve high availability for SCCP dialer deployment. The redundancy is also achieved with redundant SIP dialer. See Designing SIP Dialer for High Availability.

Outbound Option supports Call Progress Analysis configuration on a campaign basis. When this feature is enabled, 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 IVR-based campaigns. An IVR is generally configured in an agent-based campaign to allow for handling of overflow calls when all agents are busy. In a transfer to an IVR-based campaign, all of the calls are transferred to an IVR application after the outbound call is answered.

# Outbound Dialing Modes

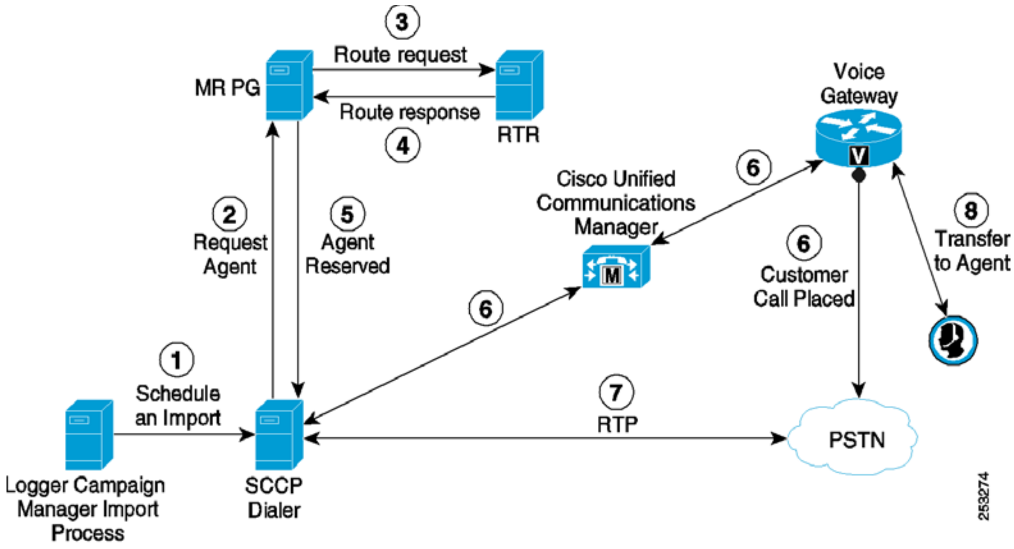
Outbound Option initiates calls using any of several modes, depending on the skill group:

- Predictive Mode—Dynamically calculates the number of lines to dial per agent in order 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 Description—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 SCCP Dialer call flow for predictive/progressive dialing proceeds as follows (Only one Unified CM PG, Generic PG, or System PG):

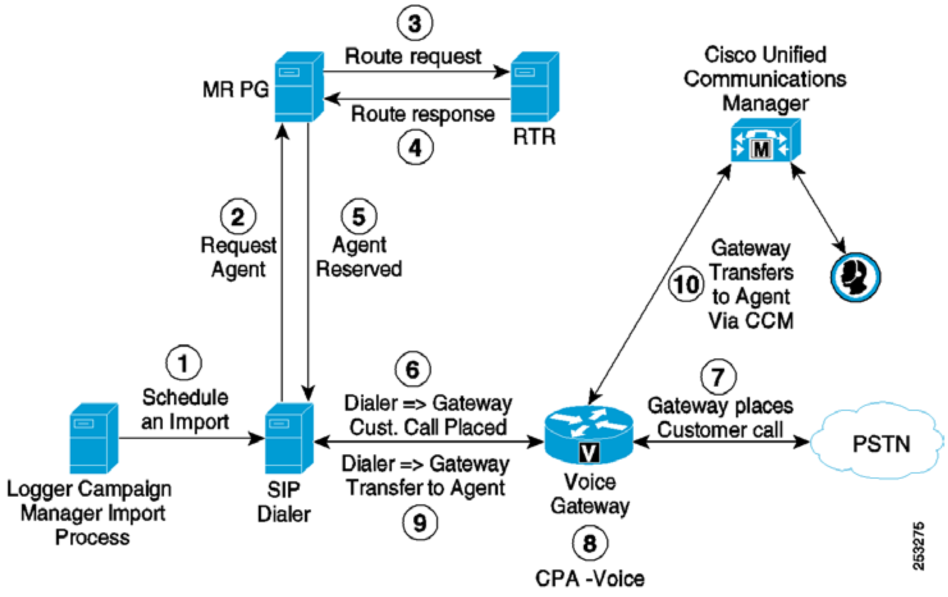
**Figure 94** ***SCCP Dialer Call Flow for Agent-Based Campaigns***



1. Import is scheduled and campaign starts. Customer records are delivered to Dialer.
2. The dialer process continually monitors peripheral skill group statistics from the CTI server for an available gent. 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, 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 for an available agent to the dialer. 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, and automatically places that virtual reservation call into answered state and then on hold.
6. The dialer initiates the customer call through Unified CM and the voice gateway.
7. If call progress analysis is configured, the dialer process will analyze the RTP stream to detect a live answer (or answering machine detection). 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. Similarly, if answering machine detection is enabled, the call can be transferred to the agent, to an IVR, or dropped.
8. The dialer auto-answers the transferred call for the agent by way of the CTI server so that the voice path between the customer and the agent can be quickly established. This releases 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 using the same message flow.

Figure 95 below, shows the SIP Dialer call flow with direct voice gateway deployment

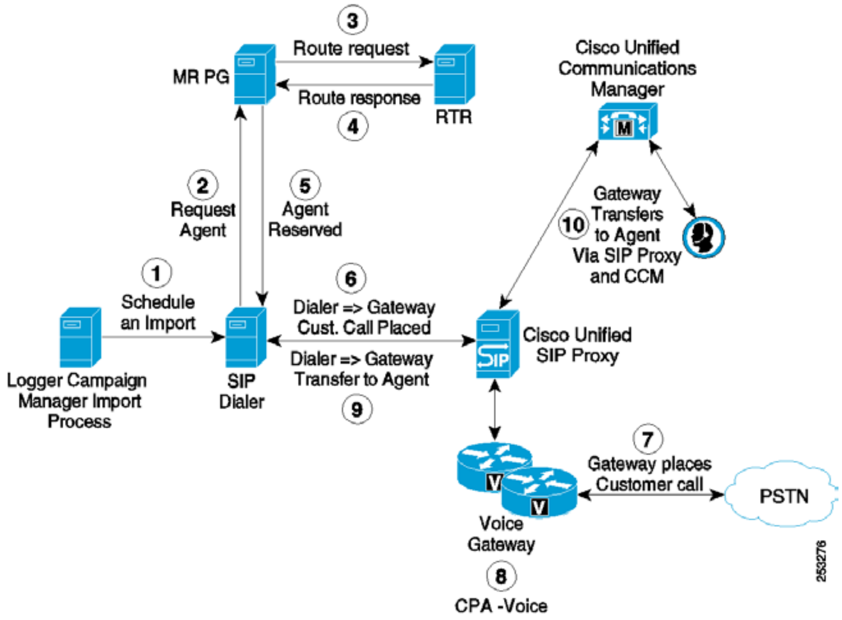
**Figure 95** ***SIP Dialer Call Flow for Agent-Based Campaigns—Direct Voice Gateway Deployment***



The SIP Dialer call flow with direct voice gateway 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, selects an available agent, reserves that agent, and then 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, and 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 Gateway places outbound calls to the customers, and Dialer is notified it is trying.
8. Call Progress Analysis is done at the gateway. Voice is detected, and Dialer is notified.
9. The Dialer asks the voice gateway to transfer the answered outbound call to the reserved agent by its agent extension.
10. The Gateway directs the answered outbound calls to the agents through Unified CM, using agent extensions and Unified CM host address. 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 quickly established.

**Figure 96** *SIP Dialer Call Flow for Agent-Based Campaigns – Unified SIP Proxy Server Deployment*



The SIP Dialer call flow in a Unified SIP Proxy Server deployment for predictive/progressive 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 Gateway places outbound calls to the customer
8. Call Progress Analysis is done at the gateway. Voice is detected, and Dialer is notified.
9. The Dialer asks the voice gateway to transfer the answered outbound call to the reserved agent by its agent extension.
10. The Gateway initiates the transfer to the Unified SIP Proxy Server, and the SIP Proxy forwards the invitations onto Unified CM. Unified CM forwards the call invitations to the agent’s phone. The dialer auto-answers the transferred call for the agent via the CTI server so that the voice path between the customer and the agent can be quickly established.

The message flows above describes 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 dialing, the agent will receive a customer record screen pop. If the agent wishes to place this call, the agent must click accept on the agent desktop. This generates a CTI event, which triggers the dialer to make a call to this customer.

## Call Flow Description - Transfer to IVR Campaign

CTI RP is not used for SIP Dialer because the Agent PG is not monitoring outbound calls during transfer to the IVR campaign and because using CTI RP would cause loss of ECC variables. SIP Dialer uses the MR routing interface instead to request a transferred label from the Router.

When using SCCP dialer for a Transfer-to-IVR campaign with IP IVR or Unified CVP, a CTI RP is used to request a transferred label from the Router; however, when using SIP dialer, CTI RP is not used, but the MR routing interface is used instead.

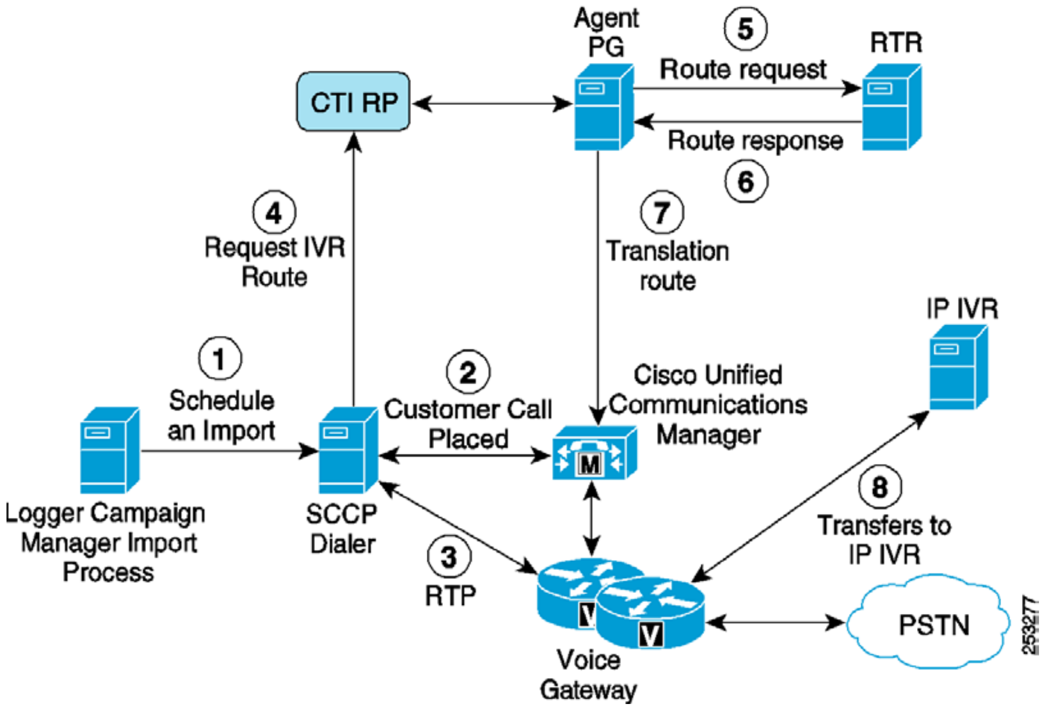
In an IVR-based campaign, the live call is transferred to an IVR system in an SCCP Dialer deployment according to the following process:

1. In this example, an unattended IVR campaign starts. Customer records are delivered to the Dialer.
2. The Dialer initiates a call to the customer.
3. The RTP stream is analyzed and voice is detected.
4. The Dialer requests an in-line transfer to a pre-configured route point.
5. The Unified CM PG requests a translation route for the router.
6. The router responds.
7. The response is translated and sent to Unified CM.
8. Unified CM transfers the call to the IVR.

SCCP Dialer Call Flow for IVR-Based Campaigns. The SIP Dialer call flow IVR-based campaigns with Unified SIP Proxy Server and IP IVR deployment proceeds as follows (Figure 97):

1. An unattended IVR 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 Gateway places the call.
4. Voice Gateway does Call Progress Analysis and detects live speech. The Dialer is notified.
5. The Dialer asks the MR PG where the IVR is.
6. MR PG forwards the request to the Router.
7. Routing Script identifies the IVR and notifies the MR PG.
8. The MR PG forward the route response to the Dialer
9. The Dialer notifies the voice gateway to transfer the call to the IVR
10. The Gateway initiates the transfer to the SIP Proxy, and the SIP Proxy forwards the invitation onto Unified CM. Unified CM forwards the call invitation to the IP IVR, and media is set up between the Gateway and the IP IVR.

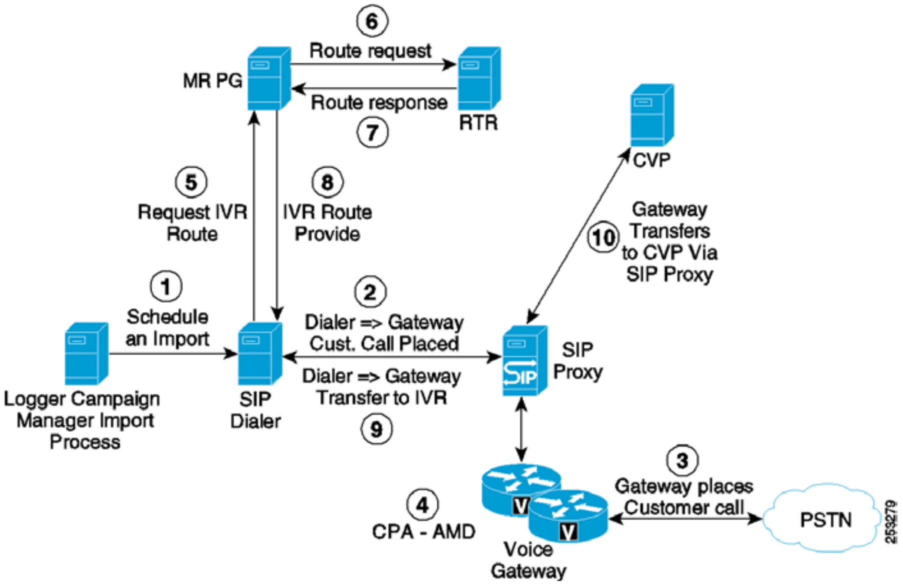
**Figure 97** *SIP Dialer Call Flow for IVR-Based Campaigns –Unified SIP Proxy Server and IP IVR Deployment*



The SIP Dialer call flow IVR-based campaigns with Unified SIP Proxy Server and Unified CVP deployment proceeds as follows ( SIP Dialer Call Flow for IVR-Based Campaigns –:

1. In this example, an unattended IVR campaign starts. Customer records are delivered to the Dialer.
2. The Dialer asks the SIP Proxy to forward an invite to an available gateway to start a call.
3. The Gateway places the call.
4. Voice Gateway does Call Progress Analysis and detects live speech. The Dialer is notified.
5. The Dialer asks the MR PG where the IVR is.
6. MR PG forwards the request to the Router.
7. Routing Script identifies the IVR and notifies the MR PG.
8. The MR PG forward the route response to the Dialer
9. The Dialer notifies the voice gateway to transfer the call to the IVR
10. The Voice Gateway sends its invitation to the SIP Proxy, which forwards it onto Unified CVP. The transfer is completed and media is set up between Unified CVP and the Voice Gateway.

**Figure 98** ***SIP Dialer Call Flow for IVR-Based Campaigns –Unified SIP Proxy Server and Unified CVP Deployment***


6
Route request
MR PG
Route response
7
RTR
CVP
5
Request IVR Route
8
IVR Route Provide
10
Gateway Transfers to CVP Via SIP Proxy
1
Schedule an Import
2
Dialer => Gateway Cust. Call Placed
Dialer => Gateway Transfer to IVR
9
Logger Campaign Manager Import Process
SIP Dialer
SIP Proxy
3
Gateway places Customer call
4
CPA - AMD
Voice Gateway
PSTN
253279

# Outbound Option for Cisco Unified Contact Center Enterprise & Hosted Deployment

This section describes deployment models for Outbound Option.

## Enterprise Deployment

Run Outbound Option on a Windows server that meets the minimum requirements specified in the latest version of the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*.

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.

The SIP or SCCP dialer must on the same server with MR-PG and Agent PG. The duplex MR-PGs and Agent PGs are required even when a simplex SCCP Dialer is installed.

The duplex Agent PG supports only duplex SIP Dialers, one is active and another is in warm standby mode. For duplex SIP Dialer installation, each SIP Dialer connects to the MR PIM on the same MR PG side (SideA or SideB).

For two SCCP dialers installed for an Agent PG, the two MR PIMs on one MR PG side (Side A or Side B) are active, and one connects to the SCCP Dialer on the same side (Side A) while the other connects to the SCCP Dialer on the other side (Side B).

When there are fewer than 120 dialer ports, installing one SCCP dialer or two for an Agent PG has both pros and cons. The installation with a single SCCP Dialer offers better Predictive/Progress dialing performance by achieving an agent pooling effect, but it does not provide redundancy. An installation with two SCCP Dialers, with dialer ports split across the two SCCP dialers (one SCCP dialer on PGA and one SCCP dialer on PGB), offers redundancy architecture but worse Predictive/Progress dialing performance.

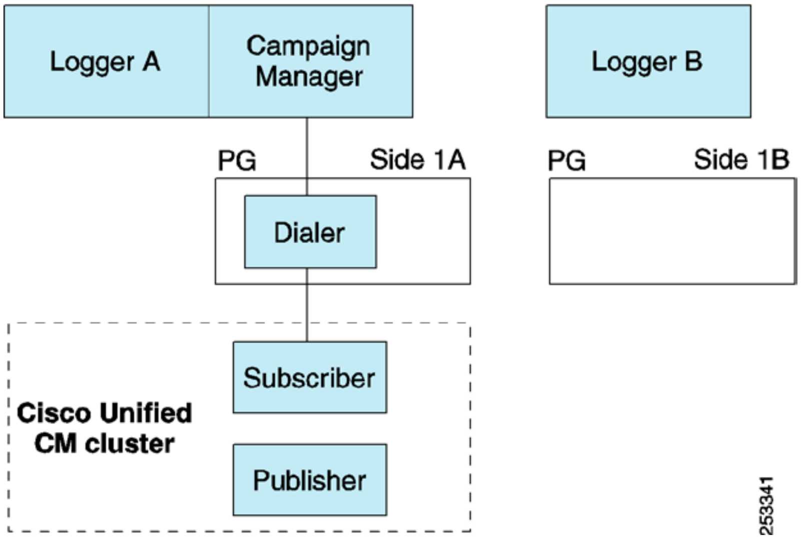
There is no upgrade path from the SCCP Dialer to SIP Dialer. Any new configuration and setup have to be created for the SIP Dialer only. It is possible to deploy both the SCCP and SIP Dialers in the same Unified CCE customer instance. The Campaign Manager is able to communicate with both the SCCP and SIP Dialers. However, the Campaign Manager performs only the warm standby feature for the SIP Dialer by knowing the dialer type. You can configure and set up only one dialer type, either SCCP dialer or SIP Dialer, for one Agent PG duplex.

The hybrid deployment model is usually used for upgrading one large customer so that the old SCCP Dialers can be removed and their outbound agents can be added to the new SIP Dialer gradually.

## Single SCCP Dialer Deployment

Figure 99 shows the installation of a single SCCP dialer. The SCCP Dialer is shown to be installed on side A of the duplex PGs. The single SCCP dialer configuration provides capacity for 120 ports. This deployment model is used when scaling and high availability are not factors.

**Figure 99** ***Single SCCP Dialer Deployment***



For Cisco Unified Contact Center Enterprise deployments, the SCCP Dialer and Media Routing PG processes run on the same physical server as the Agent PG. In a deployment with two SCCP dialers on a duplex PG pair, the Media Routing PG will have two PIMs because each dialer gets its own Media Routing PIM.

The connection between the SCCP Dialer and the Unified CM cluster consists of multiple Skinny Client Control Protocol (SCCP) sessions, one for each SCCP dialer port. The duplexed PGs (Side A and Side B) shown in Figure 99 are composed of a Generic PG (with Unified CCE PIM and a Unified IP IVR PIM), MR PG, CTI server, and CTI OS server process. The connection between the duplexed PG and the Unified CM cluster is the JTAPI link.

## Multiple SCCP Dialer Deployment

Figure 100 shows the deployment model for two dialers. Each dialer is associated with the Unified CM subscriber on its respective side and has all of its ports in one device pool for that subscriber. The configuration shown in Figure 100 provides 192 dialer ports. To scale upward, you can add more pairs of dialers (PG sides A and B) and subscribers, for up to four pairs (or eight dialers, PG sides, and subscribers) per Unified CM cluster (see Figure 101). The use of multiple dialers provides high availability for this deployment model. For more details on high availability, see Designing SCCP Dialer for High Availability.

**Figure 100** ***Multiple Dialer Deployment (Two Dialers)***

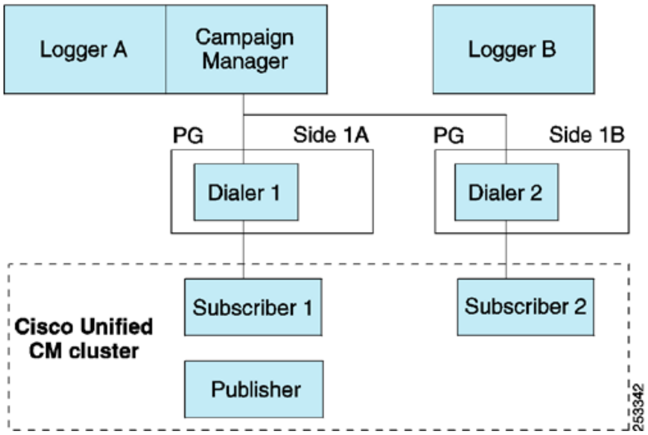
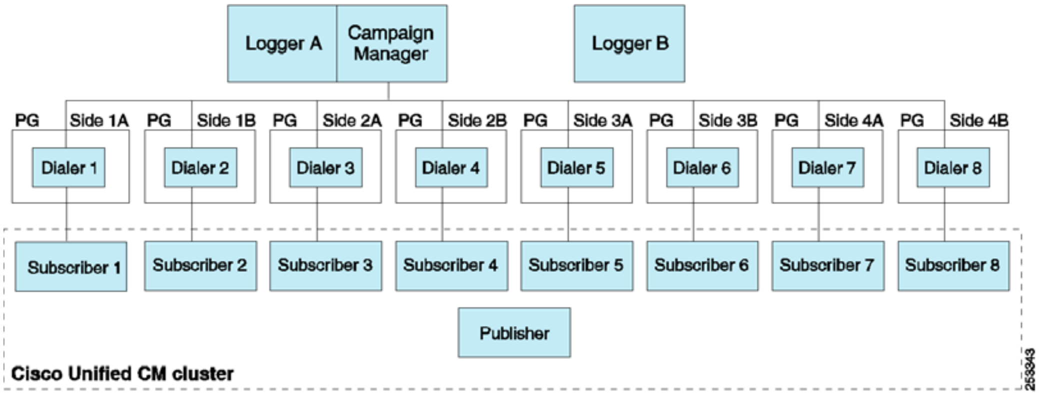


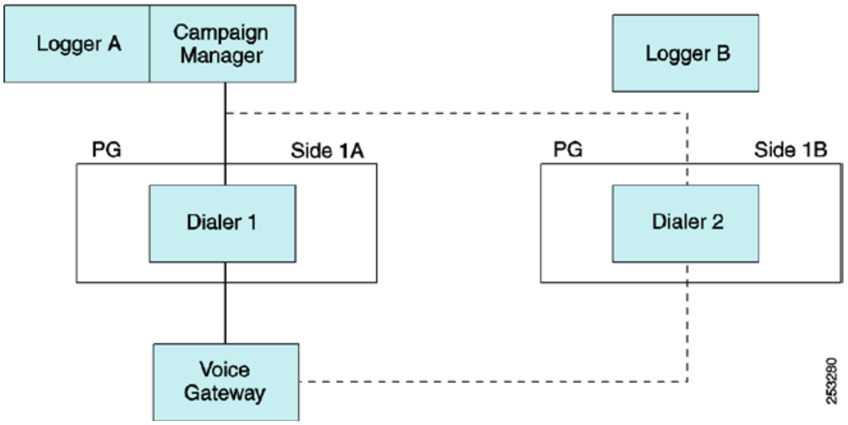
**Figure 101** ***Multiple SCCP Dialer Deployment (Eight Dialers)***



## Single Gateway Deployment for SIP Dialer

Figure 102 shows the installation of duplex SIP dialers with a single gateway. The Dialers are shown to be installed on side A and Side B of the duplex 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 102** ***Single Gateway Deployment for SIP Dialer***



The SIP Dialer architecture supports a single active SIP dialer per peripheral. Only one SIP Dialer needs to be configured. The two Dialers are installed on separate PG platforms, but each is installed using the same Dialer Name.

For Cisco Unified Contact Center Enterprise deployments, the SIP Dialer and Media Routing PG processes run on the same physical server as the Agent PG. In a deployment with duplexed SIP dialers on a duplex PG pair, the Media Routing PG will have one PIM because the each dialer gets its own Media Routing PIM on the same physical server.

The SIP Dialer uses the local static route file when it transfers outbound calls to Unified CVP, IP IVR, or outbound agents on Unified CM in a single gateway deployment when setting the “Sip Server Type” radio button to “Voice Gateway” in the Dialer setup dialog. Make sure the SIP Dialer uses the local static route file for the single gateway deployment.

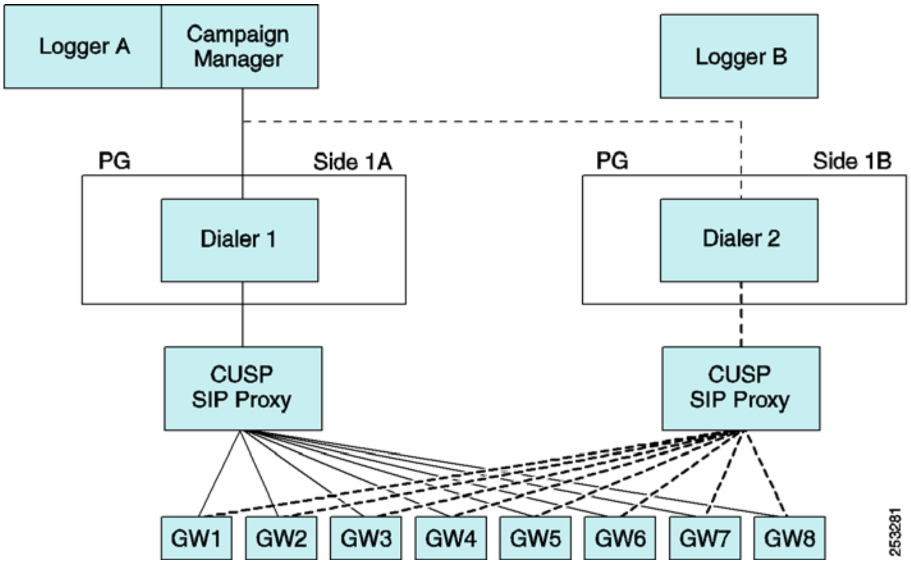
The SIP Dialer can use the Unified SIP Proxy Server for both placing outbound calls, and transferring outbound calls to Unified CVP, IP IVR, or outbound agents, on Unified CM in a single gateway deployment if the “Sip Server Type” radio button is set to “CUSP Server” in the Dialer setup dialog.

Codec configuration (G.729 versus G.711) has an impact on port capacity and CPU utilization of gateways. Configuring G.729 requires more DSP and CPU resources for gateways.

## Multiple Gateway Deployment for SIP Dialer

Figure 103 shows the deployment model for Unified SIP Proxy and eight voice gateways. The active Dialer will point to the Unified SIP Proxy Server. The proxy will handle load balancing and failover. The SIP Dialer supports Unified SIP Proxy on the Cisco 3845 Integrated Services Router. For more details on high availability, see Designing SIP Dialer for High Availability.

**Figure 103** ***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 identify where to send calls to Unified CVP, 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.

## 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 Chapter 2, “Deployment Models”). The Outbound Option high-availability model differs from that used in clustering over the WAN; therefore, when deploying cluster over the WAN, keep in mind that its benefits are for inbound traffic only.

## Distributed Deployment

A distributed deployment model involves a central Unified CCE system and Unified CM 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 Unified CM system with Outbound Options.

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 could be installed locally to customer phones, or each voice gateway could 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. See IPT: Multisite with Distributed Call Processing.

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. PRI (E1) trunks handed off at the India end by the service provider (the WAN cloud is usually built on SIP by the service provider, and they convert TDM to IP at the ingress/egress point (USA) and convert IP to TDM at the India end).

Option 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 US and India for various agent-based campaigns or transfer to an IVR campaign. The customers are in one country (for example, the US).

### Distributed Deployment Example for Agent-Based Campaign

The voice gateway and RGRA servers are distributed between two sites (Site 1 and 3) in the United States.

The Unified CM cluster is located at Site 2 (India) along with the Agent PG.

The MRPG/Dialer and Agent PGs are locally duplexed at Site 2.

The MRPG/Dialer will also be installed on the same Agent PG servers.

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 duplexed 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).

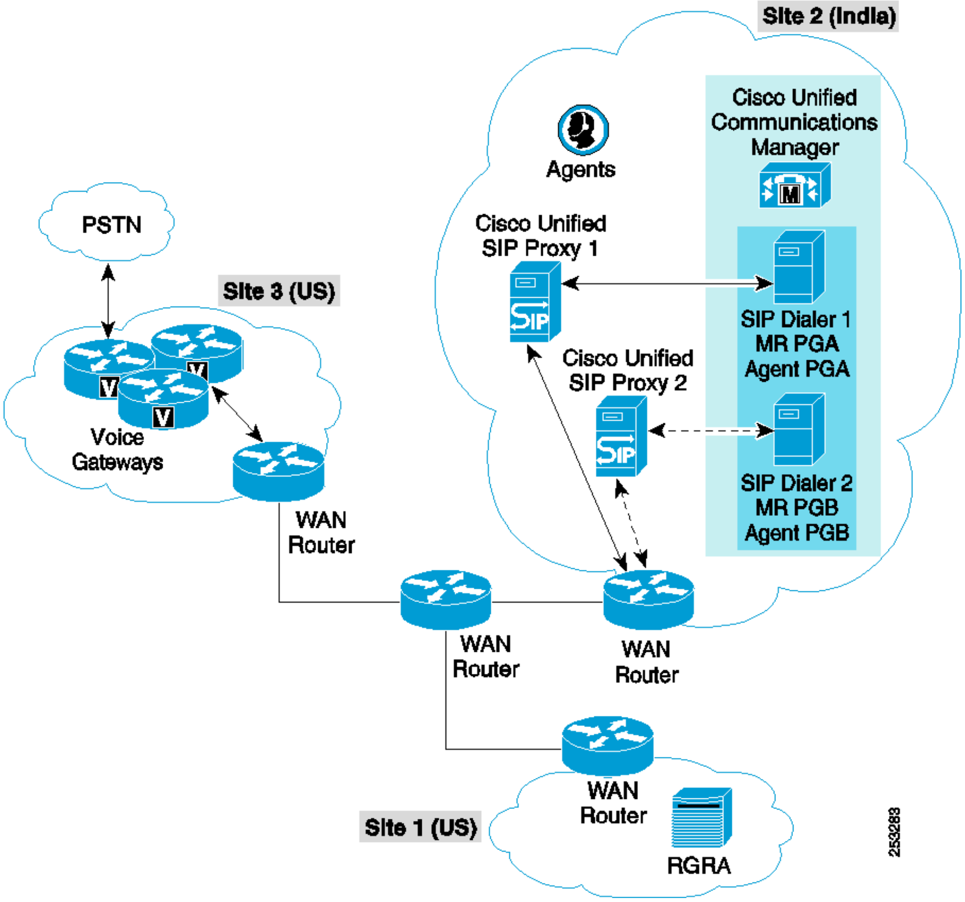
The Unified SIP Proxy servers provide n + 1 redundancy.

G.711 Codec outbound calls require a WAN bandwidth of 80 kbps per agent call.

G.729 Codec outbound calls require a WAN bandwidth of 26 kbps per agent call.

If the recording is enabled at the SIP Dialer, an alerting outbound call with G.711 Codec requires a WAN bandwidth of 80 kbps per agent call, and an alerting outbound call with G.729 Codec needs a WAN bandwidth of 26 kbps per agent call.

**Figure 104** *Distributed Deployment Example for Agent-Based Campaign*



## Distributed Deployment Example for Transfer-to-IVR Campaign – Unified -CVP

The voice gateway and RGRA servers are distributed between two sites (Site 1 and 3) in the United States.

The MRPG/Dialer and Agent PGs are locally duplexed at the Site 2.

The MRPG/Dialer will also be installed on the same Agent PG servers.

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 duplexed 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 duplexed at Site 3 to avoid the WAN SIP signaling traffic to transfer live outbound calls.

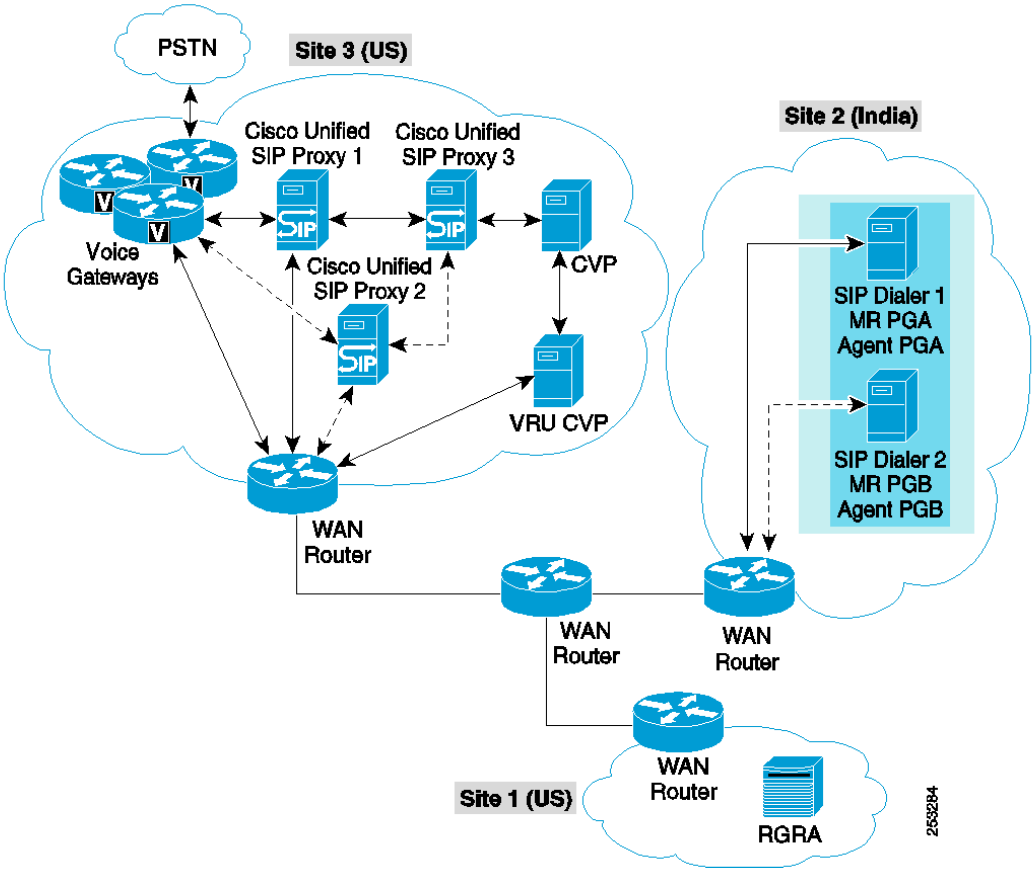
Each SIP Dialer connects to its own Unified SIP Proxy server at Site 3. Each Unified SIP Proxy Server controls the set of voice gateways at Site 3 (United States).

The Unified SIP Proxy servers provide n + 1 redundancy.

If recording is enabled at the SIP Dialer, the bandwidth requirements are as follows:

- An answered outbound call with G.711 Codec requires a WAN bandwidth of 80 kbps for the Call Progress Analysis time period.
- An answered outbound call with G.729 Codec requires a WAN bandwidth of 26 kbps for the Call Progress Analysis time period.
- An alerting outbound call with G.711 Codec requires a WAN bandwidth of 80 kbps per agent call, and an alerting outbound call with G.729 Codec needs a WAN bandwidth of 26 kbps per agent call.
- Outbound calls being queued or self-serviced at Unified IP IVR do not require WAN bandwidth.

**Figure 105** ***Distributed Deployment Example for Transfer-to-IVR Campaign – Unified CVP***



### Distributed Deployment Example for Transfer-to-IVR Campaign – IP IVR

The voice gateway and RGRA servers are distributed between two sites (Site 1 and 3) in the United States.

The Unified CM cluster is located at Site 3 (United States) along with the VRU PG.

The VRU PGs are locally duplexed at Site 3 (United States).

IP IVR is included at Site 3 (United States).

The MRPG/Dialer and Agent PGs are locally duplexed at Site 2.

The MRPG/Dialer will also be installed on the same Agent PG servers.

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 duplexed at Site 2 to avoid the WAN SIP signaling traffic to transfer live outbound calls.

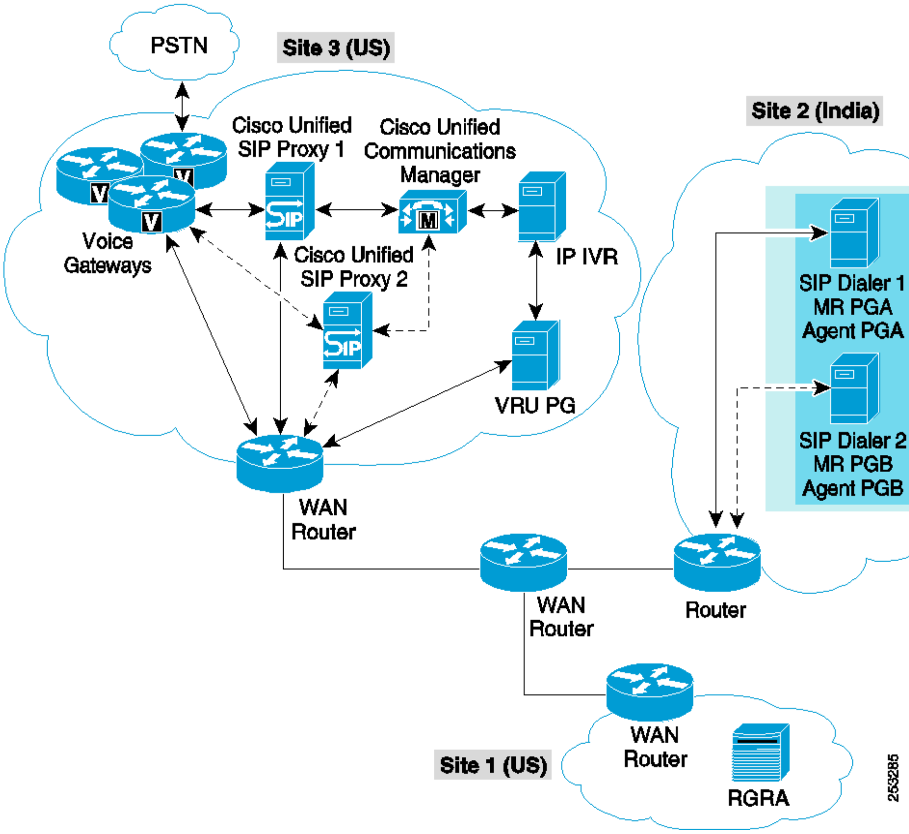
Each SIP Dialer connects to its own Unified SIP Proxy Server at Site 2. Each Unified SIP Proxy Server controls the set of voice gateways at Site 3 (United States).

The Unified SIP Proxy servers provide n + 1 redundancy.

If recording is enabled at the SIP Dialer, the bandwidth requirements are as follows:

- An answered outbound call with G.711 Codec requires a WAN bandwidth of 80 kbps for the Call Progress Analysis time period.
- An answered outbound call with G.729 Codec requires a WAN bandwidth of 26 kbps for the Call Progress Analysis time period.
- An alerting outbound call with G.711 Codec requires a WAN bandwidth of 80 kbps per agent call, and an alerting outbound call with G.729 Codec needs a WAN bandwidth of 26 kbps per agent call.
- Outbound calls being queued or self-serviced at Unified CVP do not require WAN bandwidth.

**Figure 106** ***Distributed Deployment Example for Transfer-to-IVR Campaign – IP IVR***



See the Bandwidth Provisioning and QoS Considerations chapter for further bandwidth requirements for Outbound Option.

## Voice Gateway Proximity for SCCP Dialer

Colocate Outbound Option SCCP Dialer with the Unified CCE PG and the Unified CM cluster (including the voice gateway). Because the SCCP Dialer supports only G.711 audio codec for customer calls for answering machine detection, you might have to allocate large blocks of WAN bandwidth. Even though the Dialer does not support G.729 audio codec for customer calls for answering machine detection, it is possible to support G.729 for the customer-to-agent portion of the call. This type of configuration is supported without requiring the use of transcoders.

In this deployment, the SCCP Dialer advertises G.729 capability (although the Dialer does not truly support G.729). This permits completion of the reservation call from the SCCP Dialer to the agent. The call from the SCCP Dialer to the customer must be G.711; however, the customer call is then transferred to the agent, and the call is renegotiated to G.729.

### Unified CCE Hosted Deployment

In a Unified CCE Hosted environment, only one SIP Dialer can be deployed on each of the Unified ICM Customer instances within a Unified ICM Complex. The Do-Not-Call List or Contact List files can be shared between instances.

See the *Outbound Option Guide for Cisco Unified Contact Center Enterprise and Hosted, Release 8.0(1)* for more information.

# Configuration of Outbound Option for Cisco Unified Contact Center Enterprise & Hosted

This section describes configuration considerations for Outbound Option.

## Blended Configuration

Outbound Option is capable of running campaigns in a fully blended fashion. Agents can handle inbound calls alternately with outbound calls. See the Sizing Unified CCE Components and Servers chapter for information regarding the MCS inbound capacity.

When sizing your deployment, do not use the maximum number of outbound agents allowed on a PG without also looking at the other key factors of expected hit rate, lines dialed per agent, and average handle times. An outbound campaign with a 10 second average handle time and dialing 10 lines per agent will be able to support only about 20 agents while fully occupying 240 ports on 2 SCCP dialers. However, with an average 100 second handle time dialing 3 lines per agent for a 30% hit rate, 240 ports could handle 100 agents. For sizing the Outbound Option for SCCP Dialer, use the Cisco Unified Communication Sizing Tool.

SIP Dialer targets the support of 1000 outbound agents for one PIM per PG. (Note that the number will be smaller when deploying mobile agents). In order to support this number of agents, the deployment must have at least 5 high-end gateways dedicated to outbound dialing.

SIP Dialer can support 1500 ports and 60 calls per second (cps). In order to achieve the rate of 60 cps the SIP Dialer has to support between 1000 and 2000 ports, depending on hit rates and handle times.

Each port is capable of dialing 2 calls per minute, assuming an average 30 seconds per call attempt, so 30 ports can handle 1 call per second for the Dialer. If the time to get all ports busy exceeds the average port busy time, then some number of ports will always be idle.

#### Dialer Ports Considerations

The following formula can be used to calculate the number of dialer ports that are required to achieve targeted call rate:

Num of Ports = target call rate * average call duration *(1 + hit rate %)

For example, given estimated average call duration for a 30 second outbound call with a 20% hit rate, the following table shows the ports required to achieve targeted call rate:

**Table 7** ***Ports for Targeted Call Rates***

| Target Call Rate | Ports Required |
|---|---|
| 10 | 360 |
| 20 | 720 |
| 30 | 1080 |
| 40 | 1440 |

### Voice Gateway Considerations

The most powerful gateway supports about 12 calls per second, even under the most favorable conditions, so 5 gateways would be able to support an aggregate spike of up to 60 calls per second when evenly distributed. But even this does not account for ports being tied up with agent or IVR calls after the transfer. So assuming a 50% transfer rate, 8 voice gateways (conservatively) would be required to support such as spike.

For the most current information about Unified CCE SIP dialer supported voice gateway models and releases, see the latest version of the compatibility matrix. For gateway sizing consideration, refer to published Cisco gateway performance data and UCCE sizing tool.

### Agent PG Considerations

The Unified CM PIM can support up to 15 calls per second. The PG can support 30 cps in a 2-PIM deployment, but each dialer is connected to a Peripheral/PIM.

If the voice hit rate for the campaign is 15%, then the PG could sustain the dialer dialing at a rate of 100 calls per second.

### Unified CM Considerations

The Unified CM 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 would need to be distributed across multiple subscribers using a Unified SIP Proxy Server.

### Cisco Unified SIP Proxy Considerations

**Table 8** ***Unified SIP Proxy Sizing Table***

| Hardware Model | Maximum Transaction Rate Per Second |
|---|---|
| NME-CUSP-522 | 100 |

Typical outbound call requires 2 transactions, if call is transferred to agent/IVR; or 1 transaction, if not transferred.

### Unified CVP Considerations

Calls will be distributed to Unified CVP using translation routes. Any load balancing across Unified CVPs will happen in the routing script.

Since 4 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.

### IP IVR Considerations

If the IP IVR is deployed, then all of its calls are front-ended through Unified CM. This will result in a higher call load on the Unified CM Subscribers. Since the Unified CM subscriber supports only 5 calls per second, it is likely that calls transferred to agents and the IVR will need to be distributed 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 will have the same impact on Unified CM as inbound calls. Maintain a 2-to-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 Outbound Option for SIP Dialer, use the Cisco Solution Sizing Tool.

## SCCP Dialer Throttling Considerations for Unified CM

SCCP Dialer Throttling is controlled by the field “Port Throttle” in the dialer configuration. Port Throttle indicates the number of ports to throttle in one second. For Cisco MCS-7845 and MCS-7835 servers, set this value to 5. With this setting, the SCCP Dialer will initiate calls on only five ports in one second of the campaign, and then the next five ports for the next one second, and so forth, until all 120 ports are utilized.

Setting the value to Port Throttle = 5 will allow dialing at a rate of 5 calls per second per Dialer, which gives Unified CM sufficient headroom to allow for other incoming traffic and even allow for some shared resources. It is a setting that works well for most situations. If your deployment requires a higher call rate, ensure that the call rate for all traffic for any one subscriber will not exceed 10 calls per second at any time. Also, make sure that traffic is not shared across subscribers.

Currently, a Unified CM subscriber node running on a dual-processor MCS-7845 server has a maximum capacity at 10 calls per second. Each SCCP Dialer is capable of dialing at a rate of 10 calls per second. If the solution is deployed in a way that allows for the Unified CM subscribers to be overloaded, then there is a risk of causing dropped customer calls and inefficient dialing.

The throttling mechanism is in each SCCP Dialer process, and it is not aware if another SCCP Dialer is sharing Unified CM resources. Therefore, if two SCCP Dialers share the same device pool or trunk, then there is a risk of dropped calls and inefficient dialing.

The Unified CM configuration must be designed and implemented to limit all traffic for a given Dialer to a distinct Unified CM subscriber node to prevent two SCCP Dialers from overwhelming any shared resources. This means that each SCCP Dialer requires separate device pools that point to one and only one subscriber. Each SCCP Dialer also needs its own calling search space, partition, translation pattern, and trunk configured on its Unified CM subscriber.

# Transferring to Unified CVP Using H.323 and MTP Resources

In cases where the customer is reached but no agents are currently available, or in cases where unattended campaigns are implemented, calls will be transferred to an IVR. If the solution design uses Unified CVP 4.x or earlier release with the H.323 protocol, then media termination point (MTP) resources are required when transferring calls to the IVR. To minimize MTP requirements, the trunks configured for calls transferred to Unified CVP must be separate from the trunks used for external gateways. With Unified CVP 7.0 and later releases, an MTP is not required.

# SIP Dialer Throttling Considerations for Voice Gateway and Cisco Unified SIP Proxy Server

SIP Dialer Throttling is controlled by the field “Port Throttle” in the dialer configuration. Port Throttle indicates the number of ports to throttle in one second. Setting the value to Port Throttle = 5 will allow SIP Dialer to dial outbound calls at a rate of 5 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 suggested 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 will be 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 CUSP is used by outbound deployments exclusively.

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 disenabled.

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:

Port Throttle = (Value for Gateway)

Use the following formula to calculate the Port Throttle if the gateway is shared by multiple SIP Dialers for outbound campaigns:

Port Throttle = (Value for Gateway) / (Number of SIP Dialers)

Use the following formula to calculate the Port Throttle if the gateway is shared by multiple Unified CCE components (Unified CM, Unified CVP, and SIP Dialer) for inbound/outbound calls:

Port Throttle = (Value for Gateway) * (Percentage of outbound calls) * (1 – Hit Rate)

## Multiple Gateway Deployment

Use the following formula to calculate the Port Throttle if the gateways are dedicated 100% for outbound campaigns:

Port Throttle = 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:

Port Throttle = (Total Values for Gateways) / (Number of SIP Dialers)

Use the following formula to calculate the Port Throttle, if the gateways are shared by multiple Unified CCE components (Unified CM, Unified CVP, and SIP Dialer) for inbound/outbound calls:

Port Throttle = (Total Values for Gateways) * (Percentage of outbound calls) * (1 – 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.

Weight = (Value for Gateway) / (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 receives 75 percent of the traffic and the 2800 Series gateway to receives 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)> lbtype weight
netmod(cusp-config-sg)> end server-group
```

# SIP Dialer Recording

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 will request the Voice Gateways to send the media stream to the SIP Dialer. The media stream could be 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 will be able to 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 9** ***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.

# Designing SCCP Dialer for High Availability

The Outbound Option with SCCP Dialer provides high availability through multiple dialers per Unified CM cluster. Calls are distributed evenly among the dialers. If a dialer fails, the calls are re-routed to the other dialers throughout the enterprise that are configured to support the remaining campaign contacts. The calls that were in progress on the failed dialer are marked for retry.

**Note** Campaign Manager and Import process components of Outbound Option are simplex components and must be co-located with the Logger (Side A).

It is normal practice to set up IP phones to be able to fail over to another Unified CM in case of a Unified CM node failure in which phones are distributed across the cluster. The Dialer is not a normal phone, and the ports for a dialer must not be distributed across multiple nodes within the cluster.

The dialer can tax the Unified CM node when starting a campaign or whenever resources are available (agents or IVR ports for transfer to an IVR campaign). If two dialers are configured to share the Unified CM as part of a distribution or node failure, a high-availability attempt can have a negative performance impact on the rest of the system. Each dialer has its own port throttling mechanism, and is not aware that another dialer may be sharing the same Unified CM. With two dialers competing, the subscriber might enter into a code-yellow condition.

The general rule in configuring the dialers for high availability is to do no harm. As part of this guideline, be aware that dialers significantly affect Unified CM performance, and therefore it is advisable to validate the deployment design by running the resource calculators.

# Designing SIP Dialer for High Availability

The 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 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 the connection is failed from the activated SIP dialer, it will activate the standby SIP dialer. The Campaign Manager will mark 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 could 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 be transferred to outbound agents or IVR , and leaves the outbound calls that have already be 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 the element in the Server group configuration. In the following configuration, one gateway of the elements in the cucm.example.com server group receives 50 percent of the traffic and the other two elements receive 25 percent. 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
```

```
failover-resp-codes 503
lbtype weight
ping
end server-group
```

If one of the gateways is overloaded or loses its WAN link to the PSTN network, the Unified SIP Proxy Server receives a SIP 503 response message. The “failover-resp-codes 503” configuration in the Server Group allows the Unified SIP Proxy Server to pick the next available gateway to resend an outbound call.

The Unified SIP Proxy Server supports the Hot Swappable Router Protocol (HSRP), which is a way to build redundancy into your network by allowing two Unified SIP Proxy servers to continuously test each other for connectivity and to take over if a Unified SIP Proxy Server fails.

Because the warm standby feature is already built into the Campaign Manager and SIP Dialer, and because configuring HSRP for the Unified SIP Proxy server adds undesirable complexity for Outbound Option, do not use the HSRP configuration for the Unified SIP Proxy servers dedicated for Outbound Option.

Server Group and Route Table configurations are duplicated for two duplex Unified SIP Proxy servers.

# Cisco Unified Mobile Agent

Mobiles agents are supported for outbound campaigns. However, only a nailed connection is supported. For more details regarding Cisco Unified Mobile Agent, see the Cisco Unified Mobile Agent chapter.

# References

For more information about Outbound Option, see the Outbound Option documentation.

CHAPTER 6

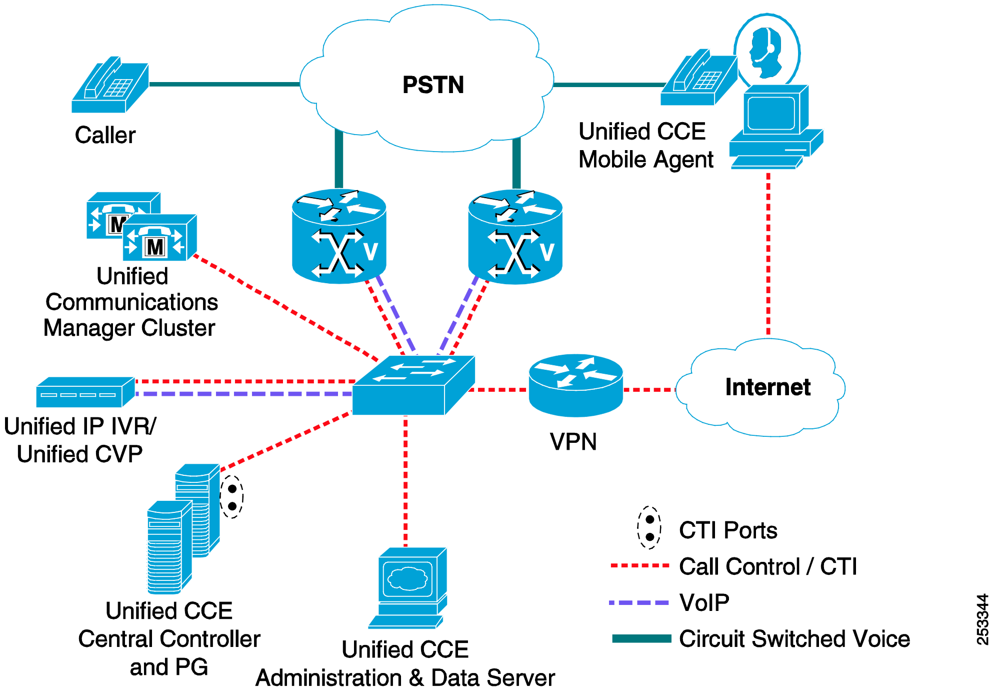
# Cisco Unified Mobile Agent

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 (Unified CM) JTAPI.

## Cisco Unified Mobile Agent Architecture

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 will then transmit the necessary subsequent call control via JTAPI to Unified CM for the two CTI ports to do whatever is needed to the media of the call. (See Figure 107)

**Figure 107** ***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.

Mobile agent functionality is supported with Unified CM 7.1(2) and later releases. Mobile Agent functionality is supported with both the System PG and Generic PG.

Queuing calls to mobile agents is supported with both Cisco Unified IP IVR and Unified CVP.

# 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:

1. 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. For more information about agent locations, see Agent Location and Call Admission Control Design.
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 ICM/CCE/CCH 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 will be 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 will be 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.

# 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 (Cisco Agent Desktop), and a phone number at which to call them. The administrator must preselect this CTI port DN based on the agent's location.
2. A call to the phone number supplied at mobile agent login is initiated from the remote CTI port statically associated (by the PG) with the local CTI port used at login. When the agent answers, the call is immediately placed on hold. Until this process completes, the agent is not considered logged in and ready.

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 processing is the same as for local agents.
4. 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.
5. The incoming call rings at the local CTI port used by the agent at login. The JTAPI gateway detects that the CTI port is ringing but does not immediately answer the call. The caller will hear ringing at this point.
6. The agent's desktop indicates 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 will be 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 is disconnected 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.

With a nailed connection, auto-answer is allowed.

A mobile agent nailed connection call can be terminated by the following two Unified CCM timers, and this termination can log out a nailed connection mobile agent:

- Maximum Call Duration timer (the default value is 720 minutes)
- Maximum Call Hold timer (the default value is 360 minutes)

To keep the mobile agent logged in, set the values for both these timers to 0, which makes the timer never expire. These timers can be configured from the Unified CCM Administration web page for the service parameters under the Cisco CallManager Service.

In a deployment with a firewall, if an agent in nailed connection mode is idle longer than the firewall H.323 Timeout value (which is typically 5 minutes), the media stream could be blocked by the firewall when the firewall H.323 timeout expires. To prevent this, increase the firewall H.323 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 *Release Notes for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.0(1)*.

# 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. That voice gateway may be registered with the same Unified CM cluster as the associated Agent PG or may be registered with another Unified CM cluster. In addition to using a phone, a Cisco Unified Mobile Agent must use an agent desktop application.

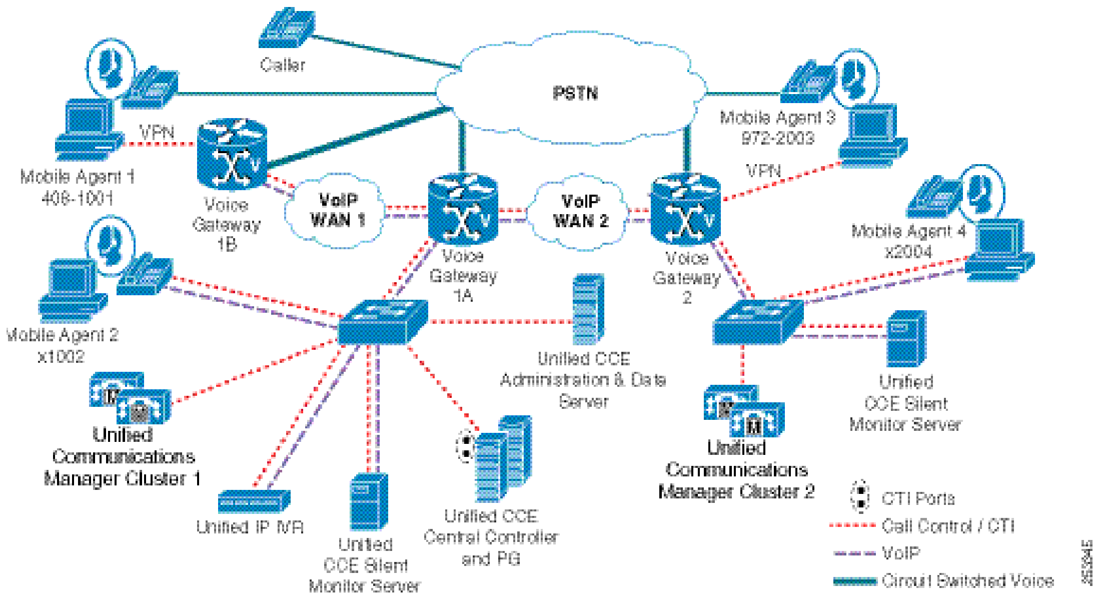
Any voice gateway supported by Unified CM and Unified CCE is supported for mobile agents. Caller (ingress) and mobile agent (egress) voice gateways can be configured with either H.323, MGCP, or SIP, and a combination of voice gateway types is also supported. The ingress and egress voice gateways can be the same voice gateway if supervisory silent monitoring is not required.

Cisco Unified Mobile Agents can also log in using a Cisco IP Phone. The IP Phone could be configured for SIP or SCCP, and a mixture is also allowed. This IP Phone may be registered with the same Unified CM cluster as the associated Agent PG, or it may be registered with another Unified CM cluster. Calls to mobile agents may also originate from SIP or SCCP IP Phones.

If an agent is using an IP Phone on the same cluster as the associated Agent PG, it is advantageous from the perspective of Unified CM performance for the agent to utilize Extension Mobility instead of the Mobile Agent feature. However, that IP Phone device would have to be associated with the JTAPI user, and there is a small performance hit on Unified CM for making that association.

In Figure 108, 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 108** ***Mobile Agent Call Scenarios***



Consider the following factors when designing a Mobile Agent solution:

- If you use SIP trunks, you must configure Media Termination Points (MTPs). This also applies if you use TDM trunks to interface with service providers. For detailed information, see the Mobile Agent Guide for Cisco Unified Contact Center Enterprise.
- Enabling the use of an MTP on a trunk will affect 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.

## 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 108 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 intercluster 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 Figure 108, 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 intercluster 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 Figure 108 are shown to be 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 Figure 108, 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 would be required at the remote site with the agent (egress) voice gateway if silent monitoring were required. For more details on silent monitoring, see CTI OS Silent Monitoring.

For additional information about call admission control design, see the call admission control information in the Cisco Unified Communications SRND.

## 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 would be 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 Unified Communications SRND*.

## 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 will be 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 will be 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 Unified Communications SRND*.

## 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 more details on silent monitoring, see CTI OS Silent Monitoring.

For additional information about codec design considerations, see the media resources information in the *Cisco Unified Communications SRND*.

## DTMF Considerations with Mobile Agent

MTP resources might be required for mobile agents who will be consulting an IVR 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 will automatically insert 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.

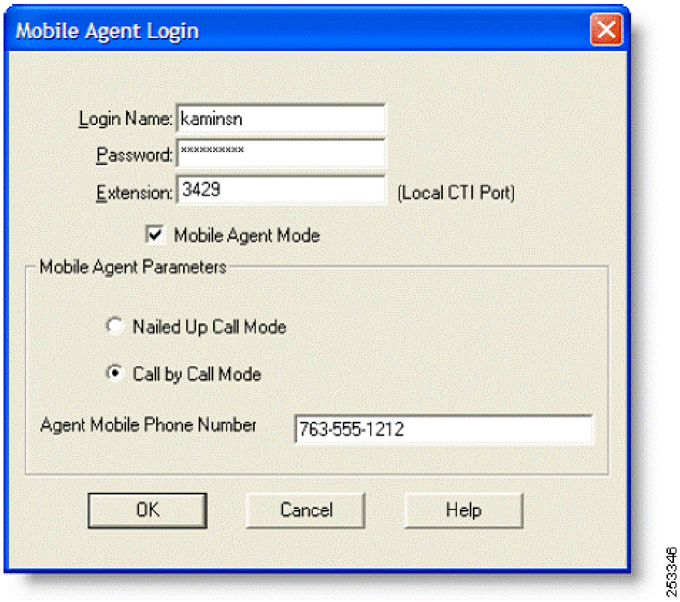
# Cisco Unified Mobile Agent Interfaces

IP Phone Agent (IPPA) is not an applicable agent interface for mobile agents. 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 109** ***Mobile Agent Login***



Mobile Agent Login

Login Name: kaminsn

Password: ***********

Extension: 3429 (Local CTI Port)

☑ Mobile Agent Mode

Mobile Agent Parameters

○ Nailed Up Call Mode

● Call by Call Mode

Agent Mobile Phone Number 763-555-1212

The phone number supplied must route to a VoIP endpoint (voice gateway, IP phone, or intercluster 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.

### CAD Silent Monitoring and Recording

The latest release of Cisco Supervisor Desktop (CSD) can silently monitor and record mobile agents using CAD SPAN port monitoring of the mobile agent voice gateway. However, Cisco Unified Mobile Agent does not support the use of Unified CM silent monitoring.

The CAD SPAN port monitor server provides a mechanism to access an agent’s RTP stream when desktop monitoring is not possible (primarily for CAD mobile agents, CAD 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 speaker(s). 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 CAD, 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 CAD 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 CAD 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 will be two one-way RTP streams flowing from the SPAN port monitor server to the CAD recording server.

CAD SPAN port monitoring of the agent voice gateway is somewhat different than CAD 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 CAD 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 CAD SPAN port monitoring software is searching for RTP packets to and from the agent voice gateway IP address and port.

A single CAD SPAN port monitor server can SPAN a network segment with both local agent Cisco IP Phones and multiple mobile agent voice gateways. The CAD 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 CAD, a single CAD deployment for a PG instance can support up to five CAD 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 CAD agents (which are statically associated in CAD administration to a SPAN port monitor server), mobile CAD agents are not mapped to a specific SPAN port monitoring server. Therefore, when a CAD 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 will be used to gain access to the RTP streams.

The CAD SPAN port monitor server must run separately from the agent PG, and one NIC must be connected to the SPAN port of a Cisco Catalyst switch in order to capture the RTP streams. A second NIC interface on the SPAN port monitor server is also required to communicate with other Unified CCE components such as the CSD and the CAD recording server. There is no redundancy for SPAN port monitor servers.

The CAD SPAN port monitor server supports both G.711 and G.729 RTP streams, but it cannot support encrypted RTP streams.

CAD SPAN port monitoring of the ingress (or customer) voice gateway is not supported. CAD 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.

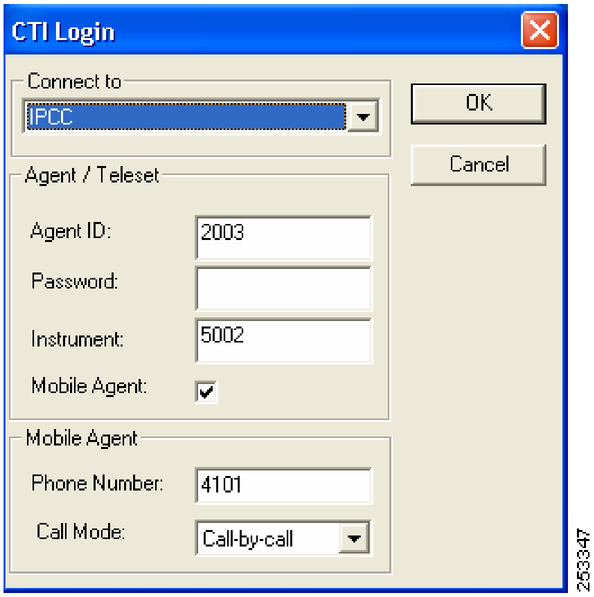
For more information about CAD supervisory silent monitoring and recording, see the Unified Contact Center Enterprise Desktop chapter.

## 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 mobile agent 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 110** ***CTI OS Login***



Connect to: IPCC

Agent ID: 2003

Password:

Instrument: 5002

Mobile Agent: ☑

Phone Number: 4101

Call Mode: Call-by-call

The phone number supplied must route to a VoIP endpoint (voice gateway, IP phone, or intercluster 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

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 server. 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 on how to configure SPAN-based silent monitoring, refer to the “Additional configuration for mobile agent environments” section in chapter 4 of the CTI OS System Manager Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted Release 9.0(1) .

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 duplex Unified CCE installations (which is a requirement for production deployments), the second server NIC interface is used for the private WAN connection and thus is not available for silent monitoring. Therefore, in duplex Unified CCE installations and as shown in Figure 108, a separate server 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. 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 were 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. For more details on desktop monitoring, see the Unified Contact Center Enterprise Desktop chapter.

Unlike CAD SPAN port monitoring, CTI OS SPAN port monitoring does not statically associate an agent with a specific SPAN port monitoring service.

## Cisco Finesse

Cisco Finesse does not support mobile agents.

## Customer Relationship Management Integrations

Customer Relationship Management (CRM) applications can be integrated with Unified CCE via CTI OS to allow an agent to log in through their CRM application, and they can be enhanced to allow an agent to have a mobile agent checkbook option and to supply a call mode and phone number. However, those integrated CRM interfaces must be enhanced in order to support using mobile agents. It is likely that a mobile agent could log in through the CTI OS agent desktop and then continue to use the integrated CRM agent interface as usual for call control and any further agent state control. However, this capability would have to be verified for each CRM integrated offering.

For more details on agent desktop options, see the Unified Contact Center Enterprise Desktop chapter.

# Cisco Unified Mobile Agent with Outbound Option for Cisco Unified Contact Center Enterprise and Hosted

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 CM 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 will be 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.

For more details about Outbound Option, see the Outbound Option for Cisco Unified Contact Center Enterprise and Hosted chapter.

# 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 will not impact call survivability. However, subsequent call control (transfer, conference, or hold) might not be possible after a failover. A mobile agent will be notified of a failover on their agent desktop, but they will have to log in again after a Unified CM or Unified CCE failover has occurred. For more details on Unified CM and Unified CCE failovers, see the Design Considerations for High Availability chapter.

# Cisco 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 CM and the Agent PG. 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. For more details about sizing a Unified CCE deployment with mobile agents, see the chapters Sizing Unified CCE Components and Servers and Sizing Cisco Unified Communications Manager Servers.

CHAPTER 7

# Securing Cisco Unified Contact Center Enterprise

This chapter describes the importance of securing the Cisco Unified Contact Center Enterprise (Unified CCE) solution and points to the various security resources available.

## 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. Once you have a good security policy, you can use many state-of-the-art Cisco technologies and products to protect your data center resources from internal and external threats and to ensure data privacy, integrity, and system availability.

As one of those applications in the Cisco Unified Communications network, Unified CCE security considerations at a high level are not very different than those of other applications making up a Cisco Unified Communications solution. Deployments of Unified CCE vary greatly and often call for complex network designs that require competence in all areas of Layer 2 and Layer 3 networking as well as voice, VPN, QoS, Microsoft Windows Active Directory, and so forth. While this chapter provides some guidance that may touch on these various areas, it is not meant to be 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. The SRNDs provide proven best practices for building a network infrastructure for Cisco Unified Communications. Among the SRNDs at this site are the following relevant documents relating to security and Cisco Unified Communications, so refer to them to successfully deploy a Unified CCE network:

- *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. Additional information is available from Microsoft on designing a new Active Directory logical structure, deploying Active Directory for the first time, upgrading an existing Windows environment to Windows Server 2003 or 2008 Active Directory, and restructuring your current environment to a Windows Active Directory environment. In particular, the *Designing and Deploying Directory and Security Services* section of the *Microsoft Windows Server 2003 Deployment Kit* can assist you in meeting all of 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.

- Server Hardening

  On top of support of a more hardened Windows Server 2003, you can configure the server automatically with security settings specifically designed for the application. Windows Server 2008 R2 does not support the Security Hardening feature in UCCE, as Windows Server 2008 R2 has default security settings on par with the security hardening settings provided by Cisco for Windows Server 2003.

- Host-Based Firewall

  Users wishing to take advantage of the Windows Firewall to protect from malicious users and programs that use unsolicited incoming traffic to attack servers can use the Windows Firewall Configuration Utility on servers or the Agent Desktop Installers to integrate with the firewall component of Windows Server 2003 SP2, Windows Server 2008 R2, and Windows XP SP3, respectively.

- Virus Protection

All servers must be running antivirus applications with the latest virus definition files (scheduled for daily updates). The *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)* contains a list of all the tested and supported antivirus applications.

- Intrusion Prevention

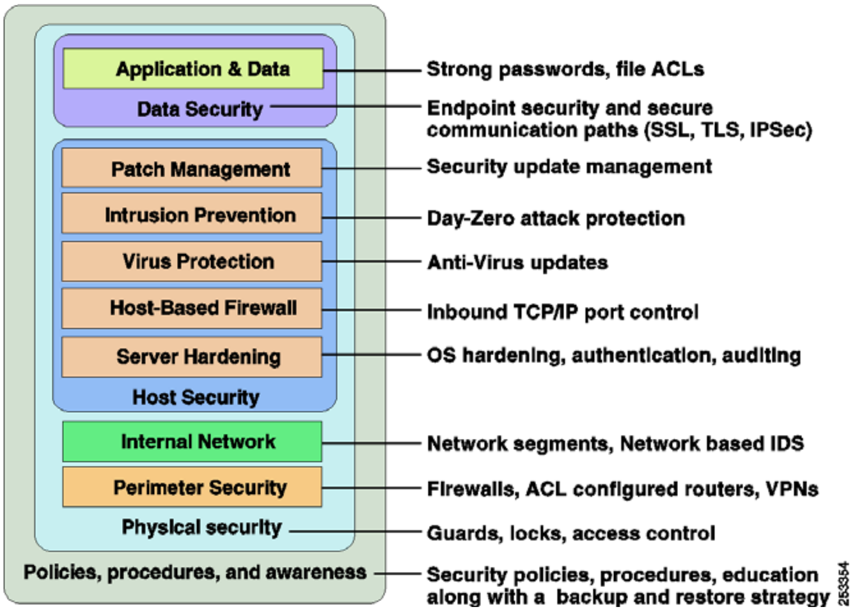
As an important defense layer in Windows Server 2003, the Unified CCE Cisco Security Agent (CSA) policy can be used to provide "day-zero" threat protection for servers. It helps to reduce operational costs by identifying, preventing, and eliminating known and unknown security threats. CSA is not supported on Windows Server 2008 R2.

- 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 to enforce the defense-in-depth paradigm illustrated in Figure 111. However, what Cisco cannot control or enforce is your enterprise policies and procedures for deploying and maintaining a secure Unified CCE solution.

**Figure 111** ***Defense-In-Depth***


Application & Data
Strong passwords, file ACLs
Data Security
Endpoint security and secure communication paths (SSL, TLS, IPSec)
Patch Management
Security update management
Intrusion Prevention
Day-Zero attack protection
Virus Protection
Anti-Virus updates
Host-Based Firewall
Inbound TCP/IP port control
Server Hardening
OS hardening, authentication, auditing
Host Security
Internal Network
Network segments, Network based IDS
Perimeter Security
Firewalls, ACL configured routers, VPNs
Physical security
Guards, locks, access control
Policies, procedures, and awareness
Security policies, procedures, education along with a backup and restore strategy
253354

# 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. All installations can be done on Windows Server 2003 or Windows Server 2008 R2 (for Release 8.5(2) or greater) Standard or Enterprise Edition. 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 chapter.

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 Unified Communications Solution Reference Network Design (SRND)* guide and other Unified CM product documentation.

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. Some examples of these enhancements include the support of Cisco's host-based intrusion prevention software (Cisco Security Agent) and default server hardening provided by the customized operating system or applications.

# Security Best Practices

As part of the Unified CCE 8.0 documentation set, Cisco has released a best-practices guide for the primary group of servers, which covers a number of areas pertaining to the new implementation in the release along with some general guidance for securing a Unified CCE deployment. The best-practices 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
- Intrusion Prevention (CSA)
- Microsoft Baseline Security Analysis
- Auditing
- Anti-Virus Guidelines and Guidelines
- Secure Remote Administration
- Additional Security Best Practices

For the most current security best practices, see the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)*.

The guidelines contained in this guide are based in part on hardening guidelines published by Microsoft, such as those found in the *Windows Server 2003 Security Guide*, as well as other third-party vendors' hardening guidelines. It also serves as a reference point for most of the security functionality in the product. The guide is also the installation guide 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 documents listed Other Security Documentation.

**Table 10** ***Other Security Documentation***

| **Security Topic** | **Document and URL** |
|---|---|
| Server staging and Active Directory deployment | *Staging Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)* |
| Cisco Security Agent | *Cisco Security Agent Installation/Deployment Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* |
| CTI OS encryption | *CTI OS System Manager's Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*<br>and<br>*Cisco CAD Installation Guide/Cisco Unified Contact Center Enterprise and Hosted Release 8.0* |
| SNMPv3 authentication and encryption | *SNMP Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* |

| Security Topic | Document and URL |
|---|---|
| Unified ICM partitioning (Database object/access control) | *Administration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*<br>**Note** Partitioning is supported only for Unified ICM Enterprise. It is not supported in Unified CCE, Unified ICM Hosted, or Unified Contact Center Hosted. |
| Feature Control (Software access control) | *Configuration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted* |
| Validating real-time clients | *Setup and Configuration Guide for Cisco Unified Contact Center Hosted* |

# 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 application servers 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 servers, 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 ICM/Contact Center Enterprise & Hosted*
- *Cisco Unified Contact Center Express Port Utilization Guide*
- *Cisco Unified Communications Manager TCP and UDP Port Usage Guide*

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.

## Topology

The deployment in Figure 112 represents the placement of firewalls and other network infrastructure components in a Unified CCE deployment. The design model in Figure 112 incorporates a parent Unified ICM system with legacy peripheral hosts and a child Cisco Unified Contact Center Enterprise with a Unified CM cluster. The following best practices apply to this type of deployment:

- 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 make use of 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 intra-server communications. The use of hardware off-load network cards is required to minimize the impact of encryption on the main CPU and to sustain the load level (including number of agents and call rate) that is supported with the Unified CCE system. See IPSec Deployment for a more detailed diagram and further information.
- Use Cisco IOS IPSec for site-to-site VPNs between geographically distributed sites, remote branch sites, or outsourced sites.

## 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, Release 8.x(y)*.

# Active Directory Deployment

This section describes the topology displayed in Figure 112. For more detailed Active Directory (AD) deployment guidance, consult the *Staging Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)*.

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.

## Parent/Child Deployments

The deployment of parent/child systems can be done on the same AD Domain or Forest, but they may also be deployed in totally disparate AD environments. The scenario where this deployment would be common is when the child Unified CCE system is housed at an outsourced contact center site. In this case, the Gateway PG that is a parent node would be a member of the parent AD domain. (Do *not* use Workgroup membership due to the administration limitations.) This type of deployment is common today for having remote branch offices with PGs that are added as members of the central site's domain to which the Routers, Loggers, and Distributors are members.

The topology shown in Figure 112 attempts to represent the AD Boundaries for each of the two AD domains involved in this deployment and to which domain the application servers are joined. The parent AD Domain Boundary is extended beyond the central data center site to include the Unified ICM Central Controllers and accompanying servers as well as 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 would have the Unified CCE servers as members. This may or may not be as part of an outsourcer’s corporate AD environment. Of course, it may also be a dedicated AD domain for Unified CCE.

## 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

The installation of Unified ICM or Unified CCE software requires that the AD Domain in which the servers 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 server is a member or on a trusted domain.

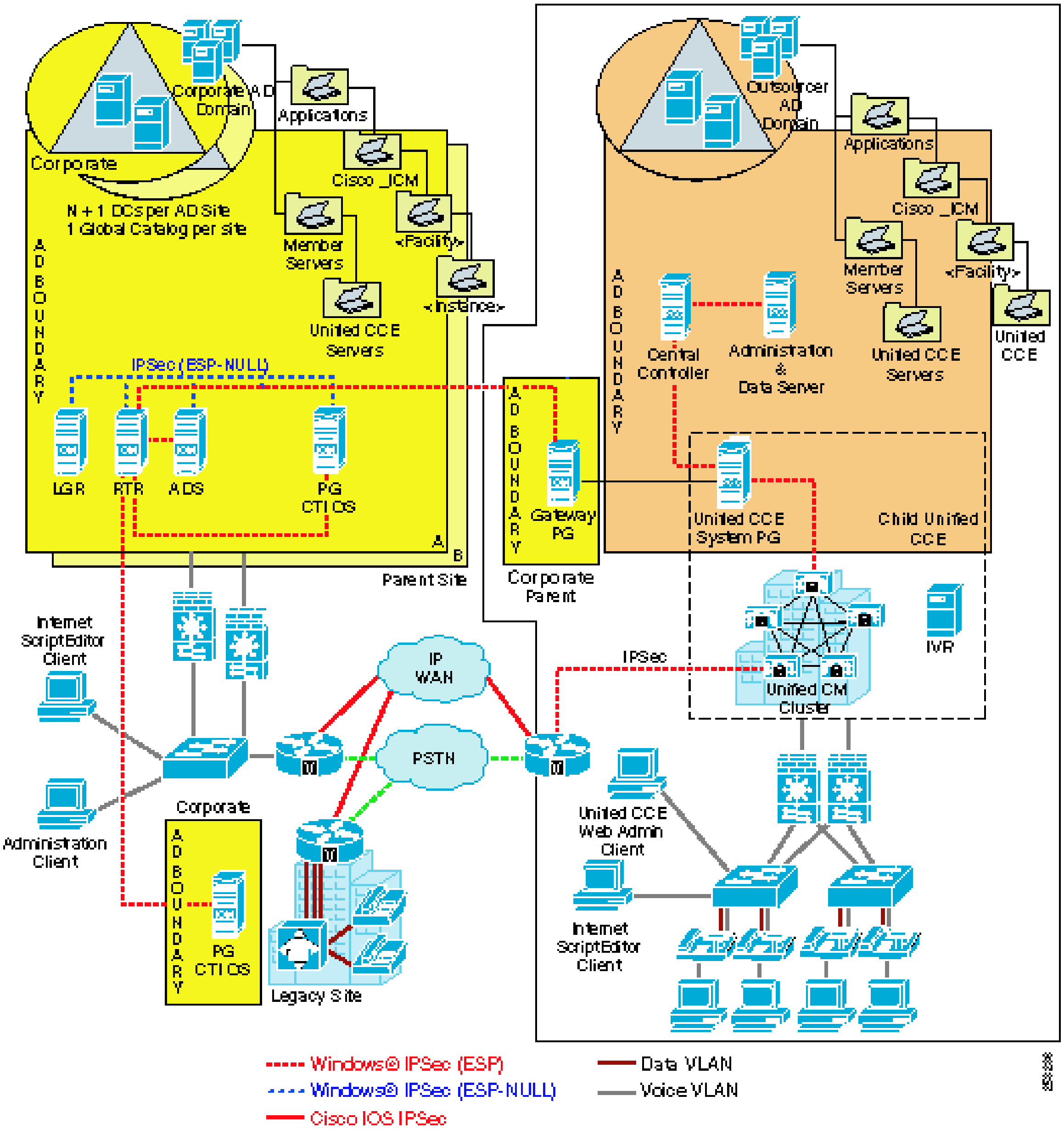
Do not confuse the creation of AD objects with Group Policy Objects (GPO). The Automated Security Hardening, which is provided following the standard Microsoft Security Template format, is *not* added to AD as part of the software installation through the configuration of a GPO. The security policy provided 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

As mentioned, there are certain AD objects that may be created by an administrator. The primary example in Figure 112 is represented by an OU container, Unified CCE Servers, which is manually added to contain the servers that are members of a given domain. These servers must be moved to this OU once they are joined to the domain. This ensures that some segregation is applied to control who can or cannot administer the servers (delegation of control) and, most importantly, which AD Domain Security Policy can or cannot be inherited by these application servers that are in the OU.

As noted before, Unified ICM/Unified CCE servers ship with a customized security policy that is modeled after the Microsoft Windows Server 2003 High Security policy. This policy can be applied at this server OU level through a Group Policy Object (GPO), but any differing policies must be blocked from being inherited at the Unified ICM/Unified CCE Servers' OU. Keep in mind that blocking inheritance, a configuration option at the OU object level, can be overridden when the Enforced/No Override option is selected at a higher hierarchy level. The application of group policies must follow a very well thought-out design that starts with the most common denominator, and those policies must be restrictive only at the appropriate level in the hierarchy. For a more in-depth explanation on how to deploy group policies properly, see the *Windows Server 2003 Security Guide*.

**Figure 112** ***Active Directory and Firewall Deployment Topology***



The following notes apply to Figure 112:

- Cisco_ICM organizational unit object hierarchies are created by the application setup.
- Unified ICM Servers and Unified CCE Servers organizational unit objects must be created by the AD administrators to separately apply custom Cisco Unified ICM Security Policies through a GPO if required.
- Flexible Single Master Operation servers must be distributed across Domain Controllers in the appropriate sites according to Microsoft guidelines.

# IPSec Deployment

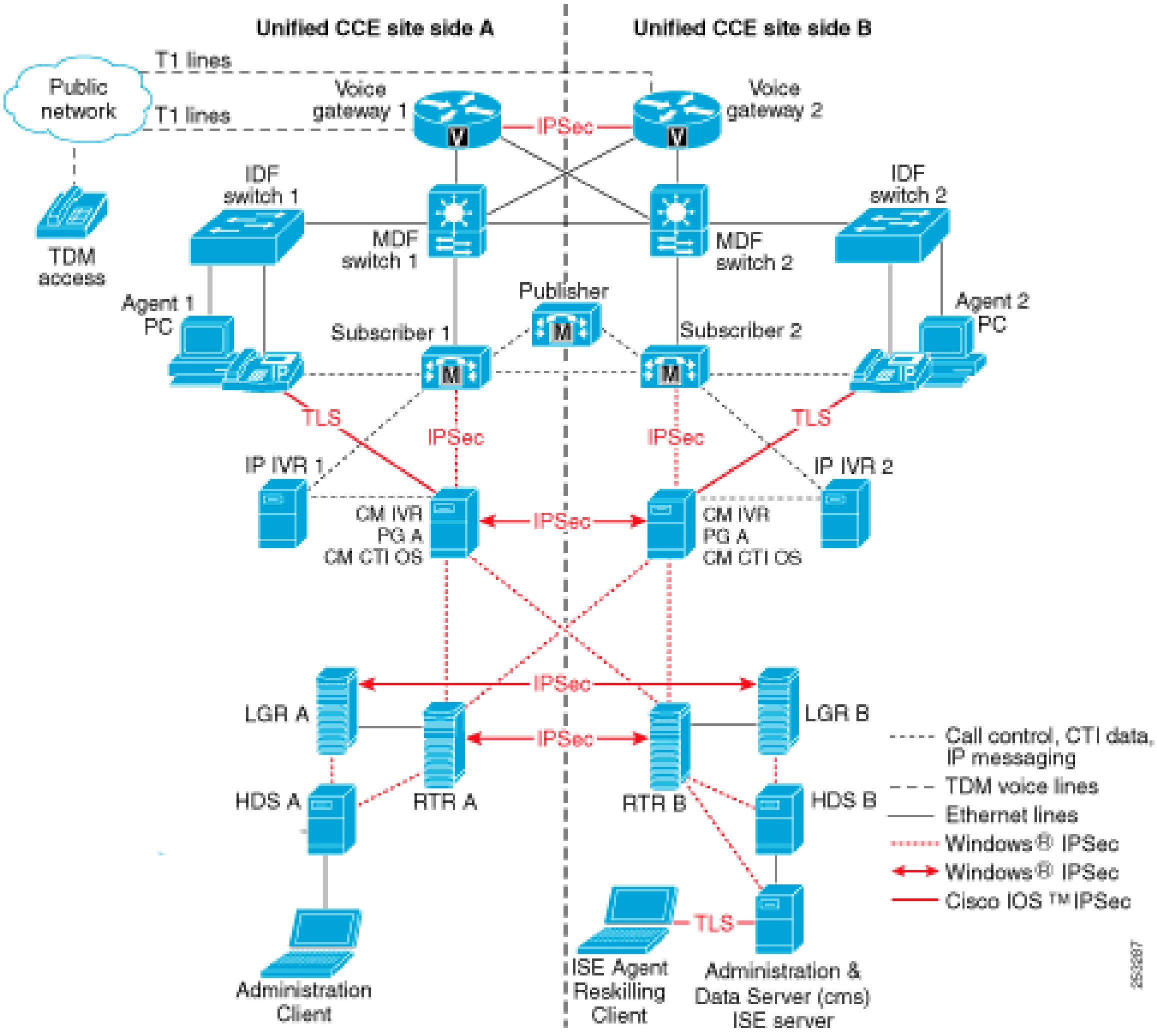
The Unified CCE solution relies on Microsoft Windows IPSec and/or Cisco IOS IPSec to secure critical links between application servers and sites. The solution can be secured either by deploying peer-to-peer IPSec tunnels between the servers and sites, or by deploying a more restrictive and preconfigured Network Isolation IPSec policy, or by using a combination of both. The peer-to-peer IPSec deployment requires manual configuration for each communication path that needs to be secured, using the tools provided by Microsoft. However, the Network Isolation IPSec policy can be deployed automatically on each server by using the Network Isolation IPSec utility, and it secures all communication paths to or from that server unless an exception is made. The Network Isolation IPSec utility is installed by default on all Unified CCE 8.0 servers.

For more details, see the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)*.

This guide not only lists the supported paths, but also information to help users deploy Windows IPSec, including appropriate settings and much more.

Figure 112 shows a number of connection paths where IPSec is supported. Figure 113 illustrates the guidelines provided in this chapter and shows the various server interconnections that must be secured with either Windows IPSec or Cisco IOS IPSec. The diagram also shows a number of paths that support TLS. More information about TLS support can be found in the section on Endpoint Security.

**Figure 113** *IPSec Deployment Example*



# Host-Based Firewall

By providing host firewall protection on the innermost layer of your network, Windows Firewall, a new security component introduced in Microsoft Windows Server 2003 with Service Pack 1 (SP1), can be an effective part of your defense-in-depth security strategy. Unified CCE supports the deployment of Windows Firewall on the application servers. The *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)* 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 Configuration Utility (CiscoICMfwConfig) uses a configuration file (CiscoICMfwConfig_exc.xml) to determine which ports, applications, or services must be enabled in the Windows Firewall. When deploying Cisco Security Agent (CSA) in managed mode, hence requiring communication with a CSA Management Center (MC), you must change this file to add the default TCP port used for the MC to connect to the CSA Agent. This must be done before running the Configuration Utility. Add the following line to the configuration file Ports XML element as needed:

```
<Ports>
..
<Port Number="5401" Protocol="TCP" Name="ManagedCSA" />
</Ports>
```

The Windows Firewall may also be configured afterwards by directly adding the port exception using the Windows Firewall Control Panel Applet or from the command line by using the following commands:

```
netsh firewall add portopening protocol = TCP port = 5401 name = ManagedCSA mode = ENABLE
scope = ALL profile = ALL
```

For more information about the Windows Firewall, see the *Windows Firewall Operations Guide*.

# Configuring Server Security

## Unified Contact Center Security Wizard

The Unified Contact Center Security Wizard allows easy configuration of the security features defined above, namely: Automated OS Security Hardening, SQL 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 8.0. The *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)* 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 *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*) and the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal (Unified CVP)*, as well as the Cisco Unified CCX and Unified CM product documentation for the applications supported. These documents are available on cisco.com.

Deploy only the supported applications for your environment, otherwise a software conflict might arise, especially when an application such as the Cisco Security Agent is installed on the Unified CCE systems.

## Configuration Guidelines

Antivirus applications have numerous configuration options that allow very granular control of what and how data must be scanned on a server.

With any antivirus product, configuration is a balance of scanning versus the performance of the server. 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 will be for installing an antivirus application within a particular environment. See the *Security Best Practices Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.x(y)* and your particular antivirus product documentation for more detailed configuration information about a Unified ICM environment.

The following list highlights some general best practices:

- 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 servers.
- 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 might not be required.
- Due to the higher scanning overhead of heuristics scanning over traditional antivirus scanning, use this advanced scanning option only at key points of data entry from untrusted networks (such as email and Internet gateways).
- Real-time or on-access scanning can be enabled, but only on incoming files (when writing to disk). This is the default setting for most antivirus applications. Implementing on-access scanning on file reads will yield a higher impact on system resources than necessary in a high-performance application environment.
- While on-demand and real-time scanning of all files gives optimum protection, this configuration does have the 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, Release 8.x(y)*.
- 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, Release 8.x(y)* listed in the previous item.

# Intrusion Prevention

## Cisco Security Agent

Cisco Security Agent provides threat protection for servers, also known as endpoints. It identifies and prevents malicious behavior, thereby eliminating known and unknown ("day zero") security risks and helping to reduce operational costs. The Cisco Security Agent aggregates and extends multiple endpoint security functions by providing host intrusion prevention, distributed firewall capabilities, malicious mobile code protection, operating system integrity assurance, and audit log consolidation (in managed mode), all within a single product.

**Note** Unlike antivirus applications, Cisco Security Agent analyzes behavior rather than relying on signature matching, but both remain critical components to a multilayered approach to host security. (Do *not* consider Cisco Security Agent a substitute for antivirus applications.)

- Deploying Cisco Security Agent on Unified CCE components involves obtaining a number of application-compatible agents and implementing them according to the desired mode.
- The Cisco Security Agent Policy provided for Unified CCE is limited to servers and may not be deployed on Agent Desktops. Customers may choose to deploy the Cisco Security Agent product in their enterprise and modify the default desktop security policies in the Management Center to allow legitimate application activity on their desktop endpoints, including that of the Agent Desktop software deployed.

## Agents Modes

The Cisco Security Agent can be deployed in two modes:

- Standalone mode — A standalone agent can be obtained directly from the Cisco Software Center for each voice application and can be implemented without communication capability to a central Cisco Security Agent Management Center (MC).
- Managed mode — An XML export file specific to the agent and compatible with each voice application in the deployed solution, can be downloaded from the same location and imported into an existing Cisco Security Agent Management Center.

The Cisco Security Agent Management Center incorporates all management functions for agents in core management software that provides a centralized means of defining and distributing policies, providing software updates, and maintaining communications to the agents. Its role-based, web browser manage-from-anywhere access makes it easy for administrators to control thousands of agents per MC.

Cisco Unified ICM, Unified CCE, and Customer Voice Portal Agents are available online.

Cisco Security Agent can reside on the same server with only those supported applications listed in the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)* or the installation guides for the Cisco Security Agent you are installing. For more details about downloading the CSA software and the installation of Cisco Security Agent, see the *Cisco Security Agent Installation/Deployment Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*.

- **C**isco does not test or support other intrusion prevention products by vendors such as Sygate, McAfee, and so forth. Such products are capable of blocking legitimate application functionality if they incorrectly identify that application as a security threat. Just as it is the case with Cisco Security Agent, these products must be specifically 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 application servers running the standard Windows Operating System, not the customized Cisco Unified Communications operating system (CIPT OS).

Follow Microsoft’s guidelines regarding when and how to apply their updates. All Contact Center customers must separately assess all security patches released by Microsoft and install those deemed appropriate for their environments. For information about tracking Cisco-supported operating system files, SQL Server, and security files, refer to *Cisco IP Telephony Operating System, SQL Server, Security Updates*.

The Security Patch and Hotfix Policy for Unified CM specifies that any applicable patch deemed Severity 1 or Critical must be tested and posted to cisco.com within 24 hours as Hotfixes. All applicable patches are consolidated and posted once per month as incremental Service Releases.

## Automated Patch Management

Unified CCE servers (except for the applications installed on the CIPT OS) support integration with Microsoft's Windows Server Update Services, whereby customers control which patches can be deployed to those servers and when the patches can be deployed.

Selectively approve updates and determine when they get deployed on production servers. The Windows Automatic Update Client (installed by default on all Windows hosts) can be configured to retrieve updates by polling a server that is running Microsoft Window Update Services in place of the default Windows Update Web site.

For more configuration and deployment information, see the 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, Release 8.x(y)*.

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 desktops both support TLS encryption to the server. This encryption protects agent login and CTI data from snooping. A mutual authentication mechanism was implemented for the CTI OS server and client to agree on a cipher suite used 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

Figure 114 shows the encryption implementation's use of X.509 certificates on the agent desktops as well as on the servers. The implementation supports the integration with a Public Key Infrastructure (PKI) for the most secure deployment. By default, the application will install and rely on a self-signed certificate authority (CA) used to sign client and server requests. However, Cisco supports integration with a third-party CA. This is the preferred method due to the increased security provided by a corporate managed CA or external authority such as Verisign.

**Figure 114** *Secure Agent Desktops (Certificate-Based Mutual Authentication)*

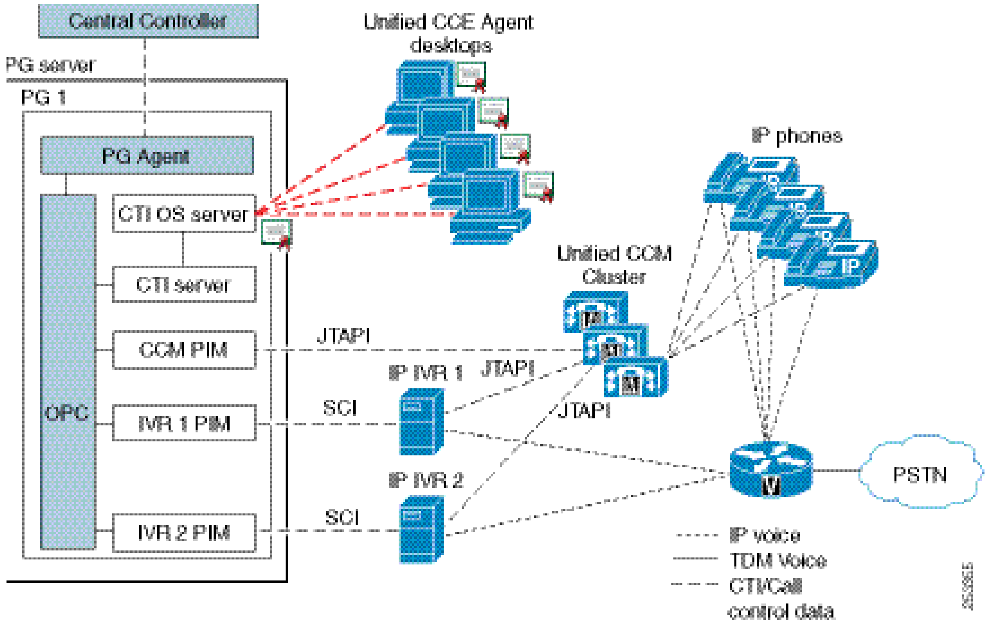
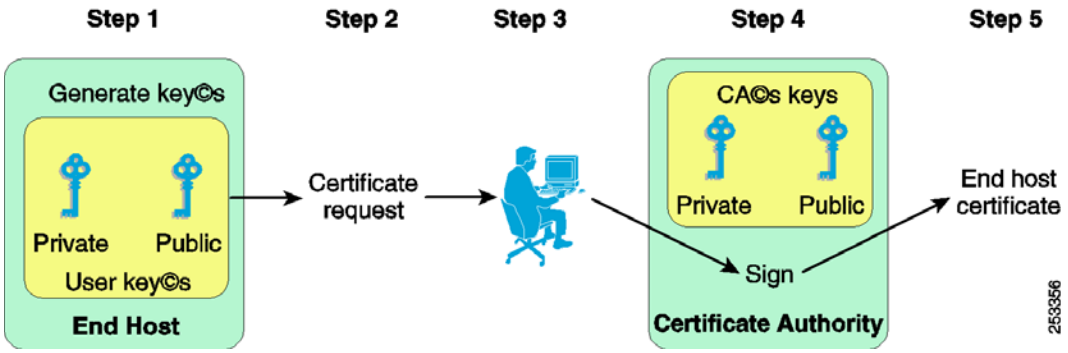


Figure 115 shows the Certificate Authority enrollment procedure to generate certificates used by the agent and the servers. The agent desktop certificate enrollment process is manual, requiring the creation of certificate signing requests (CSRs) at each endpoint, which are then transferred to the certificate authority responsible for signing and generating the certificates.

**Figure 115** *Certificate Authority Enrollment Procedure*



## Unified IP Phone Device Authentication

When designing a Unified CCE solution based on Unified CM Release 7.*x* or 8.0, customers may choose to implement device authentication for the Cisco Unified IP Phones. Unified CCE 8.0 was tested with Unified CM'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

## Unified IP Phone Media Encryption

Media Encryption may be used with Unified CCE; however, it prevents the use of the silent monitoring feature. Also, if you are deploying a recording system, contact the recording system vendor to verify support for recording in an environment with Secure Real-Time Transport Protocol (SRTP).

## 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 8

# 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 IVRs.

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.

In addition to the terms listed in this section, see Cisco Unified CCE Resource Calculators, which defines the specific terms used for the input and output of the Unified CCE Resource Calculator, the Cisco contact center sizing tool.

Also, for more details on various contact center terms and concepts discussed in this document, see the Unified CCE product documentation available online at cisco.com.

### Busy Hour or Busy Interval

A busy interval could 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 agent 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 IVR 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 could 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 UCCE 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 UCCE Call Router in order 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 UCCE Call Router. A routing request is made to the UCCE 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 should 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 UCCE 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 will be lower than or equal to the corresponding CPS rate reported by the UCCE 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. Assuming 10% of the calls on average generate two routing requests, the CPS reported by Call Router will be equal to 11 CPS. In this case, the UCCE platform would need 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 IVR ports.

### Talk Time

Talk time is 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:

Traffic in Erlangs = (Number of calls in the busy hour * AHT in sec) / 3600 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:

BHT = (BHCA * AHT seconds) / 3600, or

BHT = (BHCA * 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 IVR 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, IVR 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 would be 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 will be 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 will be 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 will be queued, the average time calls will 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 will have 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. (See Sizing Contact Center Agents, IVR Ports, and Gateways or Trunks (Inbound Contact Center) for examples on how to calculate the number of trunks and gateway ports needed.)

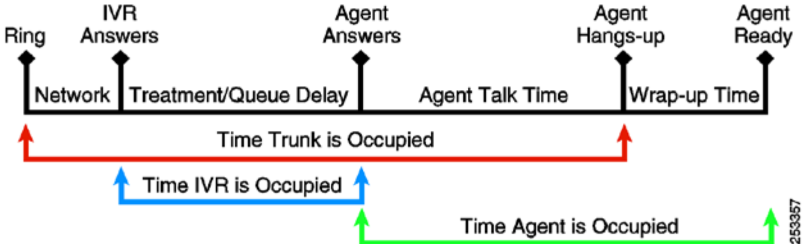
## Contact Center Resources and the Call Timeline

The focus of this chapter 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. Figure 116 shows the main resources used and the occupancy (hold/handle time) for each of these resources.

**Figure 116** ***Inbound Call Timeline***



Ring delay time (network ring) must be included, if calls are not answered immediately. This delay could be a few seconds on average, and it must be added to the trunk average handle time.

There are various tools and resources available for sizing the entire Unified CCE system. Using your contact center traffic data and service level requirements as input to the Unified CCE Resource Calculator, the calculator can generate output that you can feed into the Cisco Unified Communications Sizing Tool (available to Cisco employees and partners with proper login authentication at cisco.com).

# 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 IVR 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 remaining sections of this chapter help explain the differences between the Erlang-B and Erlang-C traffic models in simple terms, and they list 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.

Cisco has chosen to develop its own telephony sizing tool, called Cisco Unified CCE Resource Calculator. The version discussed here is designed to size contact center resources. Basic examples are included later in this chapter to show how to use the Cisco Unified CCE Resource Calculator. Additional examples are also included to show how to use the tool when some, but not all, of the input fields are known or available.

Before discussing the Cisco Unified CCE Resource Calculator, the next two sections present a brief description of the generic Erlang models and the input/output of such tools (available on the internet) to help the reader who does not have access to the Cisco Unified CCE Resource Calculator or who chooses to use other non-Cisco Erlang tools.

## 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 will be 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

# Cisco Unified CCE Resource Calculators

Cisco is continually enhancing the Cisco Unified Communications Resource Calculators, which currently include the following calculators:

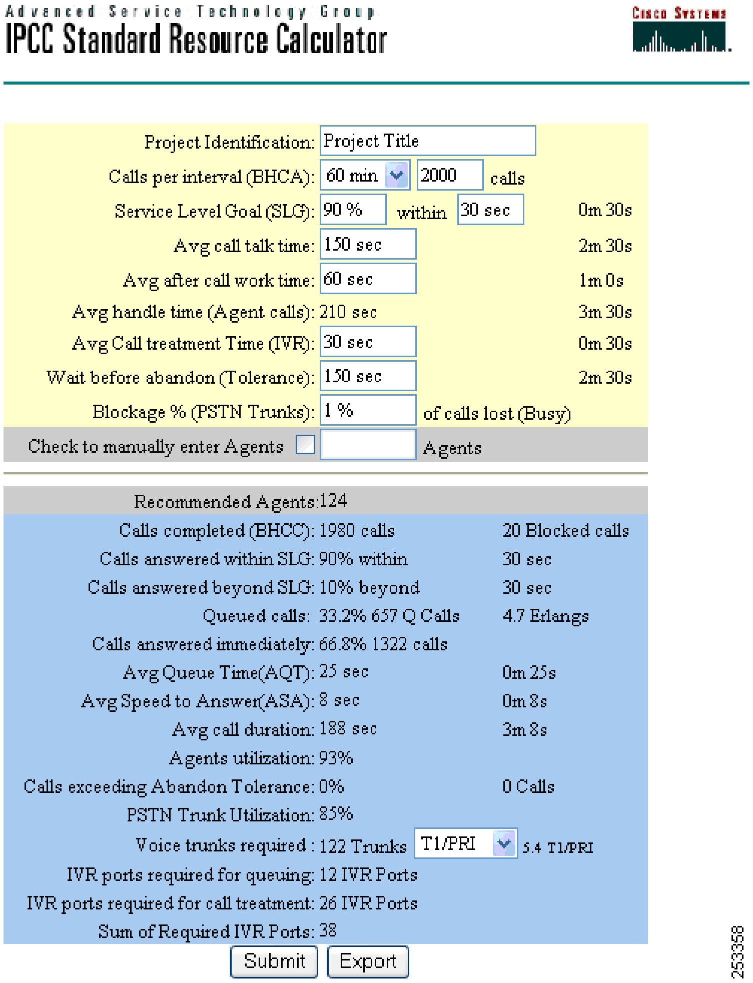
- Standard Unified CCE Resource Calculator (known as the “IPCC Standard Resource Calculator”)— Designed to provide outputs for a single contact center with a single trunk group. It allows the user to vary the number of agents.
- Advanced Unified CCE Resource Calculator — Includes all the calculations of the Standard calculator and adds the ability to allocate traffic between multiple trunk groups to include calls going to a self-service IVR. It also has inputs for confidence and growth factors.
- Unified IP IVR Self-Service Calculator — A standard Erlang B calculator for determining the number of ports required on a self-service IVR. It has inputs for up to five port groups or separate IVRs.

See the latest versions of these calculators and their associated user guides.

The Cisco Unified CCE Resource Calculators are accessible to Cisco internal employees and Cisco partners. These tools are based on industry Erlang traffic models. Other Erlang traffic calculators available on the Web can also be used for sizing various contact center resources.

Figure 117 is a snapshot of the current Standard Unified CCE Resource Calculator, followed by a definition of each of the input and output fields, how to use them, and how to interpret them.

**Figure 117** ***Cisco Unified CCE Resource Calculator***



Advanced Service Technology Group

IPCC Standard Resource Calculator

| | | | |
|---|---|---|---|
| Project Identification: | Project Title | | |
| Calls per interval (BHCA): | 60 min | 2000 calls | |
| Service Level Goal (SLG): | 90 % | within 30 sec | 0m 30s |
| Avg call talk time: | 150 sec | | 2m 30s |
| Avg after call work time: | 60 sec | | 1m 0s |
| Avg handle time (Agent calls): | 210 sec | | 3m 30s |
| Avg Call treatment Time (IVR): | 30 sec | | 0m 30s |
| Wait before abandon (Tolerance): | 150 sec | | 2m 30s |
| Blockage % (PSTN Trunks): | 1 % | of calls lost (Busy) | |
| Check to manually enter Agents □ | | Agents | |

| | | |
|---|---|---|
| Recommended Agents: | 124 | |
| Calls completed (BHCC): | 1980 calls | 20 Blocked calls |
| Calls answered within SLG: | 90% within | 30 sec |
| Calls answered beyond SLG: | 10% beyond | 30 sec |
| Queued calls: | 33.2% 657 Q Calls | 4.7 Erlangs |
| Calls answered immediately: | 66.8% 1322 calls | |
| Avg Queue Time(AQT): | 25 sec | 0m 25s |
| Avg Speed to Answer(ASA): | 8 sec | 0m 8s |
| Avg call duration: | 188 sec | 3m 8s |
| Agents utilization: | 93% | |
| Calls exceeding Abandon Tolerance: | 0% | 0 Calls |
| PSTN Trunk Utilization: | 85% | |
| Voice trunks required : | 122 Trunks | T1/PRI 5.4 T1/PRI |
| IVR ports required for queuing: | 12 IVR Ports | |
| IVR ports required for call treatment: | 26 IVR Ports | |
| Sum of Required IVR Ports: | 38 | |

## Standard Unified CCE Resource Calculator Input Fields (What You Must Provide)

When using the Cisco Standard Unified CCE Resource Calculator (known as the “IPCC Standard Resource Calculator”), you must provide the following input data:

### Project Identification

A description to identify the project or customer name and the specific scenario for this calculation. It helps to distinguish the different scenarios run (exported and saved) for a project or a customer proposal.

### Calls Per Interval (BHCA)

The number of calls attempted during the busiest interval, or busy hour call attempts (BHCA). You can choose the interval to be 60 minutes (busy hour), 30 minutes (half-hour interval), or 15 minutes. This choice of interval length allows the flexibility to calculate staffing requirements more accurately for the busiest periods within one hour, if desired. It can also be used to calculate staffing requirements for any interval of the day (non-busy hour staffing).

### Service Level Goal (SLG)

The percentage of calls to be answered within a specified number of seconds (for example, 90% within 30 seconds).

### Average Call Talk Time

The average number of seconds a caller will be on-line after an agent answers the call. This value includes time talking and time placed on hold by the agent, until the call is terminated. It does not include time spent in the IVR for call treatment or time in queue.

### Average After-Call Work Time

The average agent wrap-up time in seconds after the caller hangs up. This entry assumes that agents are available to answer calls when they are not in wrap-up mode. If seated agents enter into another mode (other than the wrap-up mode) where they are unavailable to answer calls, then this additional time must be included (averaged for all calls) in the after-call work time.

### Average Call Treatment Time (IVR)

The average time in seconds a call spends in the IVR before an attempt is made to send the call to an agent. This time includes greetings and announcements as well as time to collect and enter digits (known as prompt and collect, or IVR menuing) to route the call to an agent. It does not include queuing time if no agents are available. (This queuing time is calculated in the output section of the calculator.) The call treatment time must not include calls arriving at the IVR for self-service with no intention to route them to agents. Self-service IVR applications must be sized separately using an Erlang-B calculator.

### Wait Before Abandon (Tolerance)

This field is the amount of time in seconds that a contact center manager expects callers to wait in queue (tolerance) for an agent to become available before they abandon the queue (hang up). This value has no effect on any of the output fields except the abandon rate (number of calls abandoned).

### Blockage % (PSTN Trunks)

This field is also known as Grade of Service, or the percentage of calls that will receive busy tone (no trunks available on the gateway) during the busy hour or interval. For example, 1% blockage means that 99% of all calls attempted from the PSTN during the interval will have a trunk port available on the gateway to reach the IVR or an agent.

### Check to Manually Enter Agents

After checking this box, the user may manually enter the number of agents. If the number of agents entered is too far from the calculated (recommended) number, the calculator will show an error message. The error will appear any time the number of calls queued reaches 0% or 100%.

## Standard Unified CCE Resource Calculator Output Fields (What You Want to Calculate)

The Standard Unified CCE Resource Calculator calculates the following output values based on your input data:

### Recommended Agents

The number of seated agents (calculated using Erlang-C) required to staff the contact center during the busy hour or busy interval.

### Calls Completed (BHCC)

The busy hour call completions (BHCC), or the number of expected calls completed during the busy hour. It is the number of calls attempted minus the number of calls blocked.

### Calls Answered Within Target SLG

The percentage of calls that are answered within the set target time entered in the Service Level Goal (SLG) field. This value is the calculated percentage of calls answered immediately if agents are available. It includes a portion of calls queued if no agents are available within the SLG (for example, less than 30 seconds). It does not include all queued calls because some calls will be queued beyond the SLG target.

### Calls Answered Beyond SLG

The percentage of calls answered beyond the set target time entered in the Service Level Goal (SLG) field. For example, if the SLG is 90% of calls answered within 30 seconds, the calls answered beyond SLG would be 10%. This value includes a portion of all calls queued, but only the portion queued beyond the SLG target (for example, more than 30 seconds).

### Queued Calls

The percentage of all calls queued in the IVR during the busy hour or interval. This value includes calls queued and then answered within the Service Level Goal as well as calls queued beyond the SLG. For example, if the SLG is 90% of calls answered within 30 seconds and queued calls are 25%, then there are 10% of calls queued beyond 30 seconds, and the remaining 15% of calls are queued and answered within 30 seconds (the SLG).

### Calls Answered Immediately

The percentage of calls answered immediately by an agent after they receive treatment (if implemented) in the IVR. These calls do not have to wait in queue for an agent. As in the preceding example, if 25% of the calls are queued (including those beyond the target of 30 seconds), then 75% of the calls would be answered immediately.

### Average Queue Time (AQT)

The average amount of time in seconds that calls will spend in queue waiting for an agent to become available during the interval. This value does not include any call treatment in the IVR prior to attempting to send the call to an agent.

### Average Speed of Answer (ASA)

The average speed of answer for all calls during the interval, including queued calls and calls answered immediately.

### Average Call Duration

The total time in seconds that a call remained in the system. This value is the sum of the average talk time, the average IVR delay (call treatment), and the average speed of answer.

### Agents Utilization

The percentage of agent time engaged in handling call traffic versus idle time. After-call work time is not included in this calculation.

### Calls Exceeding Abandon Tolerance

The percentage (and number) of calls that will abandon their attempt during the interval, based on expected Tolerance time specified in the input. If this output is zero, it means that all queued calls were answered by an agent in less than the specified abandon time (longest queued call time is less than the abandon time).

### PSTN Trunk Utilization

The occupancy rate of the PSTN trunks, calculated by dividing the offered load (Erlangs) by the number of trunks.

### Voice Trunks Required

The number of PSTN gateway trunks required during the busy interval, based on the number of calls answered by the voice gateway and the average hold time of a trunk during the busy interval. This value includes average time of call treatment in the IVR, queuing in the IVR (if no agents are available), and agent talk time. This calculated number assumes all trunks are grouped in one large group to handle the specified busy hour (or interval) calls. If several smaller trunk groups are used instead, then additional trunks would be required, therefore smaller groups are less efficient.

### IVR Ports Required for Queuing

The number of IVR ports required to hold calls in queue while the caller waits for an agent to become available. This value is based on an Erlang-B calculation using the number of queued calls and the average queue time for those calls.

### IVR Ports Required for Call Treatment

The number of IVR ports required for calls being treated in the IVR. This value is based on an Erlang-B calculation using the number of calls answered and the average call treatment time (average IVR delay).

### Total IVR Ports Requirement

The total number of IVR ports required if the system is configured with separate port groups for queuing and treatment. Pooling the ports for treatment and queuing results in fewer ports for the same amount of traffic than if the traffic is split between two separate IVR port pools or groups. However, configure the number of ports required for queuing in a separate group, with the ability to overflow to other groups if available.

### Submit

After entering data in all required input fields, click **Submit** to compute the output values.

### Export

Click **Export** to save the calculator input and output in the format of comma-separated values (CSV) to a location of your choice on your hard drive. This CSV file could be imported into a Microsoft Excel spreadsheet and formatted for insertion into bid proposals or for presentation to clients or customers. Multiple scenarios could be saved by changing one or more of the input fields and combining all outputs in one Excel spreadsheet by adding appropriate titles to the columns to reflect any changes in the input. This format makes comparing results of multiple scenarios easy to analyze.

# Sizing Contact Center Agents, IVR Ports, and Gateways or Trunks (Inbound Contact Center)

The contact center examples in this section illustrate how to use the Unified CCE Resource Calculator in various scenarios, along with the impact on required resources for inbound contact centers. The first example in this section is a basic call flow, where all incoming calls to the contact center are presented to the voice gateway from the PSTN. Calls are routed directly to an agent, if available; otherwise, calls are queued until an agent becomes available.

## Basic Contact Center Example

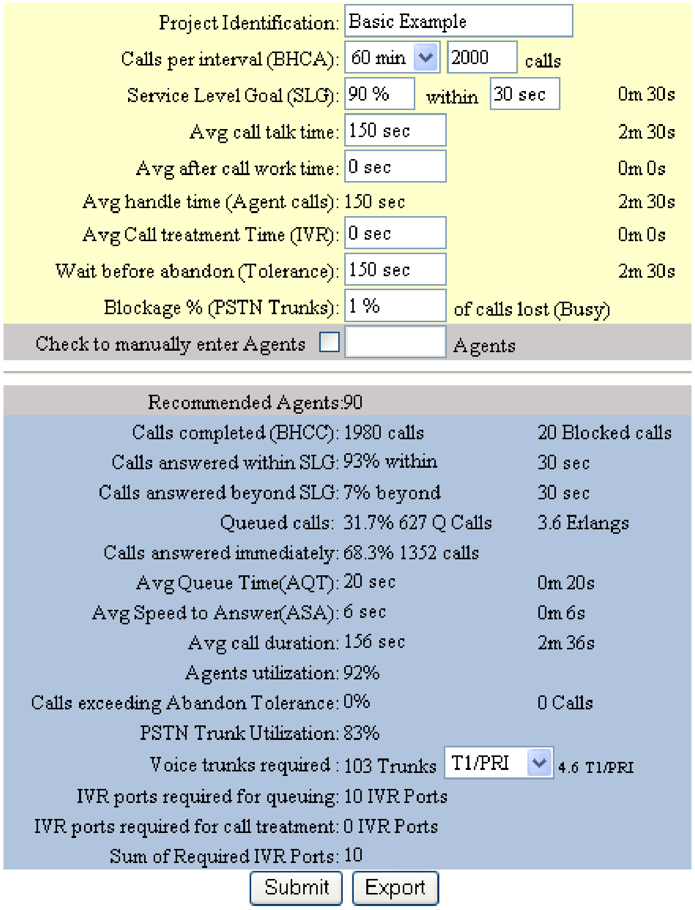
This example forms the basis for all subsequent examples in this chapter. After a brief explanation of the output results highlighting the three resources (agents, IVT ports, and PSTN trunks) in this basic example, subsequent examples build on it by adding different scenarios, such as call treatment and agent wrap-up time, to demonstrate how the various resources are impacted by different call scenarios.

This basic example uses the following input data:

- Total BHCA (60-minute interval) into the contact center from the PSTN to the voice gateway = 2,000.
- Desired service level goal (SLG) of 90% of calls answered within 30 seconds.
- Average call talk time (agent talk time) = 150 seconds (2 minutes and 30 seconds).
- No after-call work time (agent wrap-up time = 0 seconds).
- No call treatment (prompt and collect) is implemented initially. All calls will be routed to available agents or will be queued until an agent becomes available.
- Wait time before caller hangs up (tolerance) = 150 seconds (2 minutes and 30 seconds).
- Desired grade of service (percent blockage) for the PSTN trunks on the voice gateway = 1%.

After entering the above data in the input fields, clicking **Submit** at the bottom of the calculator results in the output shown in Figure 118.

**Figure 118** ***Basic Contact Center Sizing Example***



| | | | |
|---|---|---|---|
| Project Identification: | Basic Example | | |
| Calls per interval (BHCA): | 60 min | 2000 calls | |
| Service Level Goal (SLG): | 90 % | within 30 sec | 0m 30s |
| Avg call talk time: | 150 sec | | 2m 30s |
| Avg after call work time: | 0 sec | | 0m 0s |
| Avg handle time (Agent calls): | 150 sec | | 2m 30s |
| Avg Call treatment Time (IVR): | 0 sec | | 0m 0s |
| Wait before abandon (Tolerance): | 150 sec | | 2m 30s |
| Blockage % (PSTN Trunks): | 1 % | of calls lost (Busy) | |
| Check to manually enter Agents □ | | Agents | |

| | | |
|---|---|---|
| Recommended Agents: | 90 | |
| Calls completed (BHCC): | 1980 calls | 20 Blocked calls |
| Calls answered within SLG: | 93% within | 30 sec |
| Calls answered beyond SLG: | 7% beyond | 30 sec |
| Queued calls: | 31.7% 627 Q Calls | 3.6 Erlangs |
| Calls answered immediately: | 68.3% 1352 calls | |
| Avg Queue Time(AQT): | 20 sec | 0m 20s |
| Avg Speed to Answer(ASA): | 6 sec | 0m 6s |
| Avg call duration: | 156 sec | 2m 36s |
| Agents utilization: | 92% | |
| Calls exceeding Abandon Tolerance: | 0% | 0 Calls |
| PSTN Trunk Utilization: | 83% | |
| Voice trunks required : | 103 Trunks T1/PRI | 4.6 T1/PRI |
| IVR ports required for queuing: | 10 IVR Ports | |
| IVR ports required for call treatment: | 0 IVR Ports | |
| Sum of Required IVR Ports: | 10 | |

Notice that the output shows 1980 calls received and processed (completed) by the voice gateway, out of the total of 2000 calls attempted from the PSTN. This is because we have requested a provisioning of 1% blockage from our PSTN provider, which results in 20 calls (1%) being blocked by the PSTN (and receiving busy tone) out of the total 2000 calls.

### Agents

The result of 90 seated agents is determined by using the Erlang-C function imbedded in the Unified CCE Resource Calculator, and calls will be queued to this resource (agents).

Notice that, with 90 agents, the calculated service level is 93% of calls answered within 30 seconds, which exceeds the desired 90% requested in the input section. Had there been one less agent (89 instead of 90), then the 90% SLG would not have been met.

This result also means that 7% of the calls will be answered beyond the 30 second SLG. In addition, there will be 31.7% of calls queued; some will queue less than 30 seconds and others longer. The average queue time for queued calls is 20 seconds.

If 31.7% of the calls will queue, then 68.3% of the calls will be answered immediately without delay in a queue, as shown in the output in Figure 118.

### IVR Ports Required for Queuing

In this basic example, the Unified IP IVR is being used as a queue manager to queue calls when no agents are available. The calculator shows the percent and number of calls queued (31.7%, or 627 calls) and the average queue time (20 seconds).

These two outputs from the Erlang-C calculation are then used as inputs for the imbedded Erlang-B function in the calculator to compute the number of IVR ports required for queuing (10 ports in this example).

### PSTN Trunks (Voice Gateway Ports)

Similarly, the calculator uses Erlang-B to calculate the required number of voice gateway ports (PSTN trunks) based on the call load (answered calls) and the calls that have to queue when no agents are available.

Total trunks required to carry this total traffic load is 103 trunks.

This calculation does not include trunks that might be needed for call scenarios that require all calls to be treated first in the IVR before they are presented to available agents. That scenario is discussed in the next example.

## Call Treatment Example

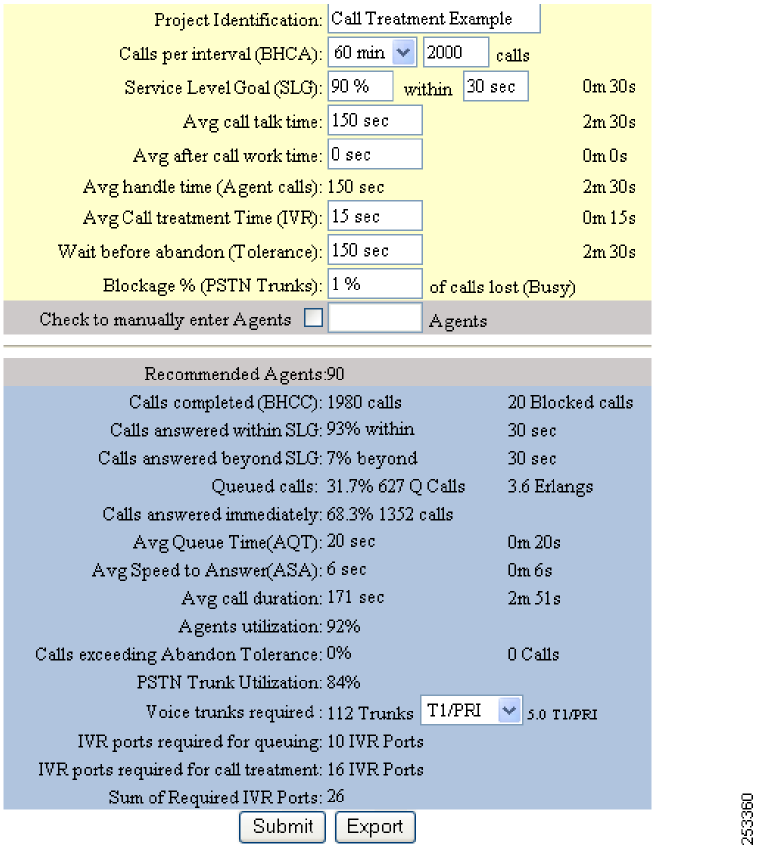
This example builds on the basic example in the preceding section. Again, all incoming calls to the contact center are presented to the voice gateway from the PSTN, then calls are immediately routed to the Unified IP IVR for call treatment (such as an initial greeting or to gather account information by using prompt-and-collect) before they are presented to an agent, if available. If no agents are available, calls are queued until an agent becomes available.

The impact of presenting all calls to the Unified IP IVR is that the PSTN trunks are held longer, for the period of the call treatment holding time. More Unified IP IVR ports are also required to carry this extra load, in addition to the ports required for queued calls.

Call treatment (prompt and collect) in this example appears not to impact the number of required agents because the traffic load presented to the agents (number of calls, talk time, and service level) is assumed not to have changed. In reality, adding call treatment such as collecting information input from the callers to identify them to agents using a CTI-Pop screen will reduce the average time a caller spends with an agent, thus saving valuable resources, providing more accurate selection and routing of appropriate agent, and improving customer service.

Using a 15-second call treatment and keeping all other inputs the same, Figure 119 shows the number of PSTN trunks (112) and Unified IP IVR ports (16) required in addition to the existing 10 ports for queuing.

**Figure 119** ***Call Treatment in IVR***



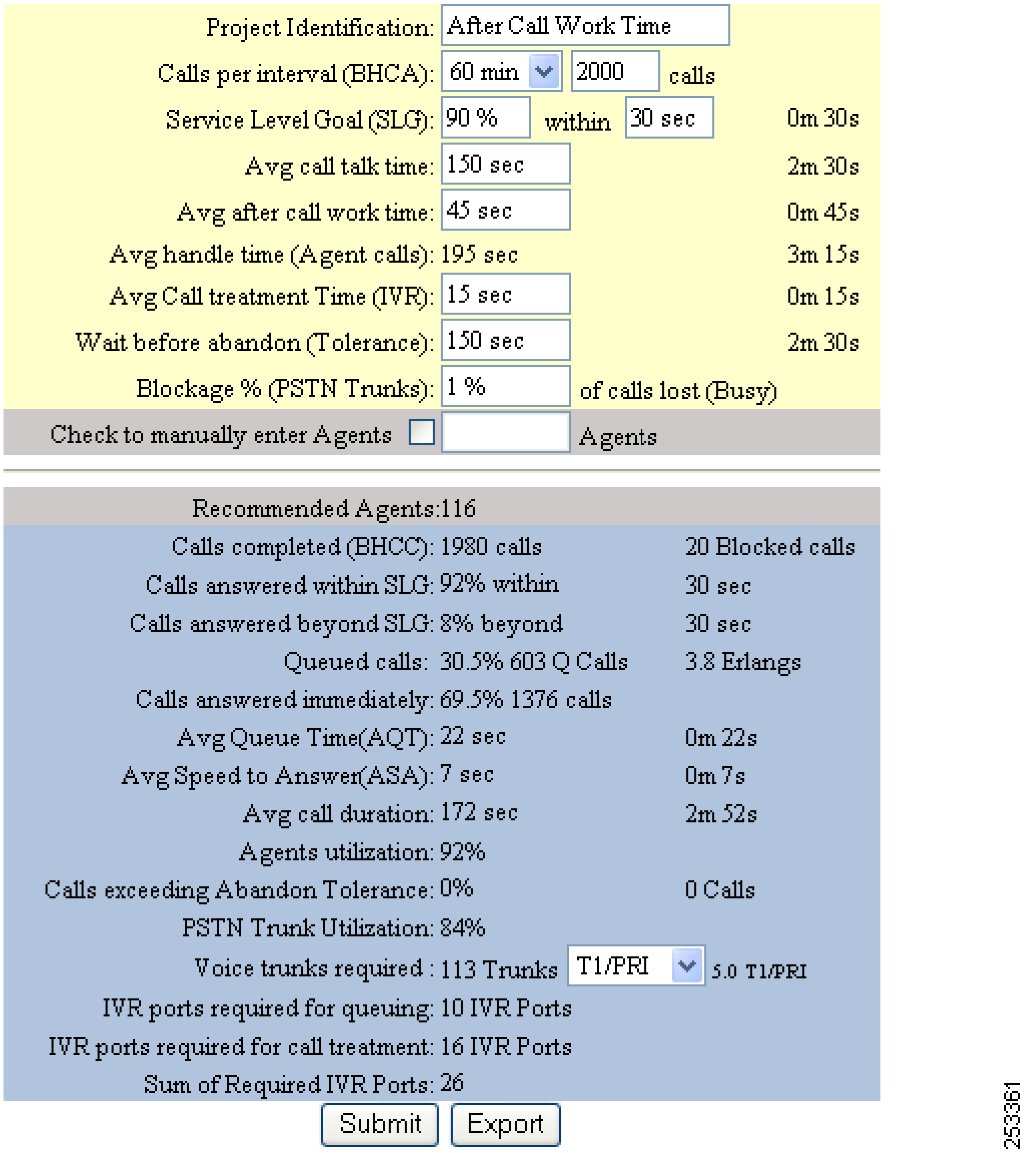
|  |  |  |  |
|---|---|---|---|
| Project Identification: | Call Treatment Example |  |  |
| Calls per interval (BHCA): | 60 min ▾ | 2000 calls |  |
| Service Level Goal (SLG): | 90 % | within 30 sec | 0m 30s |
| Avg call talk time: | 150 sec |  | 2m 30s |
| Avg after call work time: | 0 sec |  | 0m 0s |
| Avg handle time (Agent calls): | 150 sec |  | 2m 30s |
| Avg Call treatment Time (IVR): | 15 sec |  | 0m 15s |
| Wait before abandon (Tolerance): | 150 sec |  | 2m 30s |
| Blockage % (PSTN Trunks): | 1 % | of calls lost (Busy) |  |
| Check to manually enter Agents □ |  | Agents |  |
| Recommended Agents: | 90 |  |  |
| Calls completed (BHCC): | 1980 calls |  | 20 Blocked calls |
| Calls answered within SLG: | 93% within |  | 30 sec |
| Calls answered beyond SLG: | 7% beyond |  | 30 sec |
| Queued calls: | 31.7% 627 Q Calls |  | 3.6 Erlangs |
| Calls answered immediately: | 68.3% 1352 calls |  |  |
| Avg Queue Time(AQT): | 20 sec |  | 0m 20s |
| Avg Speed to Answer(ASA): | 6 sec |  | 0m 6s |
| Avg call duration: | 171 sec |  | 2m 51s |
| Agents utilization: | 92% |  |  |
| Calls exceeding Abandon Tolerance: | 0% |  | 0 Calls |
| PSTN Trunk Utilization: | 84% |  |  |
| Voice trunks required : | 112 Trunks | T1/PRI ▾ | 5.0 T1/PRI |
| IVR ports required for queuing: | 10 IVR Ports |  |  |
| IVR ports required for call treatment: | 16 IVR Ports |  |  |
| Sum of Required IVR Ports: | 26 |  |  |

## After-Call Work Time (Wrap-up Time) Example

Using the previous example, we now add an average of 45 seconds of work time (wrap-up time) after each call. We can then use the Unified CCE Resource Calculator to determine the number of agents required to handle the same traffic load (see Figure 120).

After-call work time (wrap-up time) begins after the caller hangs up, so trunk and Unified IP IVR resources are not impacted and will remain the same, assuming all other input remains the same. Assuming the SLG and traffic load also remain the same, additional agents would be required only to service the call load and to compensate for the time agents are in the wrap-up mode.

**Figure 120** ***After-Call Work Time***



| | | | |
|---|---|---|---|
| Project Identification: | After Call Work Time | | |
| Calls per interval (BHCA): | 60 min ▾ | 2000 calls | |
| Service Level Goal (SLG): | 90 % | within 30 sec | 0m 30s |
| Avg call talk time: | 150 sec | | 2m 30s |
| Avg after call work time: | 45 sec | | 0m 45s |
| Avg handle time (Agent calls): | 195 sec | | 3m 15s |
| Avg Call treatment Time (IVR): | 15 sec | | 0m 15s |
| Wait before abandon (Tolerance): | 150 sec | | 2m 30s |
| Blockage % (PSTN Trunks): | 1 % | of calls lost (Busy) | |
| Check to manually enter Agents □ | | Agents | |

| | | |
|---|---|---|
| Recommended Agents: | 116 | |
| Calls completed (BHCC): | 1980 calls | 20 Blocked calls |
| Calls answered within SLG: | 92% within | 30 sec |
| Calls answered beyond SLG: | 8% beyond | 30 sec |
| Queued calls: | 30.5% 603 Q Calls | 3.8 Erlangs |
| Calls answered immediately: | 69.5% 1376 calls | |
| Avg Queue Time(AQT): | 22 sec | 0m 22s |
| Avg Speed to Answer(ASA): | 7 sec | 0m 7s |
| Avg call duration: | 172 sec | 2m 52s |
| Agents utilization: | 92% | |
| Calls exceeding Abandon Tolerance: | 0% | 0 Calls |
| PSTN Trunk Utilization: | 84% | |
| Voice trunks required : | 113 Trunks | T1/PRI ▾ 5.0 T1/PRI |
| IVR ports required for queuing: | 10 IVR Ports | |
| IVR ports required for call treatment: | 16 IVR Ports | |
| Sum of Required IVR Ports: | 26 | |

Note that trunks and IVR ports remained virtually the same, except that there is one additional trunk (113 instead of 112). This slight increase is not due to the wrap-up time, but rather is a side effect of the slight change in the SLG (92% instead of 93%) due to rounding calculations for the required 116 agents due to wrap-up time.

## Agent Staffing Considerations

In calculating agent requirements, make the following adjustments to factor in all the activities and situations that make agents unproductive or unavailable:

### Agent Shrinkage

Agent shrinkage is a result of any time for which agents are being paid but are not available to handle calls, including activities such as breaks, meetings, training, off-phone work, unplanned absence, non-adherence to schedules, and general unproductive time.

### Agent Shrinkage Percentage

This factor will vary and must be calculated for each contact center. In most contact centers, it ranges from 20% to 35%.

### Agents Required

This number is based on Erlang-C results for a specific call load (BHCA) and service level.

### Agents Staffed

To calculate this factor, divide the number of agents required from Erlang-C by the productive agent percentage (or 1 minus the shrinkage percentage). For example, if 100 agents are required from Erlang-C and the shrinkage is 25%, then 100/.75 yields a staffing requirement of 134 agents.

# Contact Center Design Considerations

Consider the following design factors when sizing contact center resources:

- Compute resources required for the various busy intervals (busy hours), such as seasonal busy hours and average daily busy hour. Many businesses compute the average of the 10 busiest hours of the year (excluding seasonal busy hours) as the busy-hour staffing. Retail business contact centers will add temporary staff based on seasonal demands such as holiday seasons. Run multiple interval calculations to understand daily staff requirements. Every business has a different call load throughout the day or the week, and agents must be staffed accordingly (using different shifts or staffing levels). Customer Relationship Management (CRM) and historical reporting data help to fine-tune your provisioning computations to maintain or improve service levels.
- When sizing IVR ports and PSTN trunks, it is better to over-provision than to under-provision. The cost of trimming excess capacity (disconnecting PSTN lines) is much cheaper than lost revenue, bad service, or legal risks. Some governmental agencies are required to meet minimum service levels, and outsourced contact centers might have to meet specific service level agreements.
- If the contact center receives different incoming call loads on multiple trunk groups, additional trunks would be required to carry the same load using one large trunk group. You can use the Erlang-B calculator to size the number of trunks required, following the same methodology as in the Call Treatment Example. Sizing of required trunks must be done for each type of trunk group.
- Consider marketing campaigns that have commercials asking people to call now, which can cause call loads to peak during a short period of time. The Erlang traffic models are not designed for such short peaks (bunched-up calls); however, a good approximation would be to use a shorter busy interval, such as 15 minutes instead of 60 minutes, and to input the expected call load during the busiest 15 minutes to compute required agents and resources. Using our Basic Contact Center Example a load of 2000 calls in 60 minutes (busy interval) requires 90 agents and 103 trunks. We would get exactly the same results if we used an interval of 15 minutes with 500 calls (¼ of the call load). However, if 600 of the calls arrive during a 15-minute interval and the balance of the calls (1400) arrive during the rest of the hour, then 106 agents and 123 trunks would be required

instead to answer all 600 calls within the same service level goal. In a sales contact center, the potential to capture additional sales and revenue could justify the cost of the additional agents, especially if the marketing campaign commercials are staggered throughout the hour, the day, and the various time zones.

- Consider agent absenteeism, which can cause service levels to go down, thus requiring additional trunks and Unified IP IVR queuing ports because more calls will be waiting in queue longer and fewer calls will be answered immediately.
- Adjust agent staffing based on the agent shrinkage factor (adherence to schedules and staffing factors, as explained in Agent Staffing Considerations).
- Allow for growth, unforeseen events, and load fluctuations. Increase trunk and IVR capacity to accommodate the impact of these events (real life) compared to Erlang model assumptions. (Assumptions might not match reality.) If the required input is not available, make assumptions for the missing input, run three scenarios (low, medium, and high), and choose the best output result based on risk tolerance and impact to the business (sales, support, internal help desk, industry, business environment, and so forth). Some trade industries publish contact center metrics and statistics, such as those shown in Table 11, available from web sites such as benchmarkportal.com. You can use those industry statistics in the absence of any specific data about your contact center (no existing CDR records, historical reports, and so forth).
- Agent Greeting incurs in a greater cost to CVP and the VXML gateway servers. For more information, see the CVP SRND. For the cost and impact on UCCE see, Sizing Unified CCE Components and Servers.
- IP-IVR port utilization increases when you enable the Whisper Announcement feature. CUCM conferencing resource utilization increases when you enable Agent Greeting for mobile agent. Additional messaging on PGs such as setting up a conference call for Agent Greeting and Whisper Announcement also affects performance.
- Whisper Announcement with IP-IVR in Parent/Child has no impact on agent sizing, and incurs great impact on the IP-IVR. For more information, see the SRND for Unified CCX and Unified IP-IVR.
- Whisper Announcement with CVP has no impact on agent sizing, and only incurs a small impact on the CVP and the VXML gateway sizing. For more information, see the CVP SRND.

**Table 11** ***eBusiness Best Practices for All Industries, 2001***

| **Inbound Contact Center Statistics** | **Average** | **Best Practices** |
|---|---|---|
| 80% calls answered in? (seconds) | 36.7 | 18.3 |
| Average speed of answer (seconds) | 34.6 | 21.2 |
| Average talk time (minutes) | 6.1 | 3.3 |
| Average after-call work time (minutes) | 6.6 | 2.8 |
| Average calls abandoned | 5.5% | 3.7% |

| **Inbound Contact Center Statistics** | **Average** | **Best Practices** |
|---|---|---|
| Average time in queue (seconds) | 45.3 | 28.1 |
| Average number of calls closed on first contact | 70.5% | 86.8% |
| Average TSR occupancy | 75.1% | 84.3% |
| Average time before abandoning (seconds) | 66.2 | 31.2 |
| Average adherence to schedule | 86.3% | 87.9% |
| Cost per call | $9.90 | $7.12 |
| Inbound calls per 8-hour shift | 69.0 | 73.9 |
| Percentage attendance | 86.8% | 94.7% |

[1] Special Executive Summary; Principal Investigator, Dr. Jon Anton; Purdue University, Center for Customer-Driven Quality.

Use the output of the Unified CCE Resource Calculator as input for other Cisco configuration and ordering tools that may require as input, among other factors, the number of IVR ports, number of agents, number of trunks, and the associated traffic load (BHCA).

CHAPTER 9

# Sizing Unified CCE Components and Servers

## Sizing Unified CCE Components and Servers

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 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 best design practices focusing on scalability and capacity for Unified CCE deployments. The design considerations, best practices, 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.

## Sizing Tools

Sizing tools are available online.

The sizing tools are available to Cisco internal employees and Cisco partners only, and proper login authentication is required.

## Sizing Considerations for Unified CCE

This section discusses Unified CCE sizing considerations:

# 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 (Unified CM) capacity and sizing of Cisco Unified Communications components, see the Sizing Cisco Unified Communications Manager Servers chapter and to the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

Additionally, because of varying agent and skill group capacities, consider proper sizing of the Agent PG, including CTI OS and Cisco Agent 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 Figure 121, Figure 122, and Table 12 specifies capacity for inbound calls only.

The information presented in Figure 121, Figure 122, and Table 12 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 servers 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 CPU servers. This information is meant as a guide for determining when Unified CCE software processes can be co-resident within a single server and when certain processes need their own dedicated server. Table 12 assumes that the deployment scenario includes two fully redundant servers that are deployed as a duplexed pair.

**Note** The Cisco Unified Contact Center solution does not provide a quad-processor Cisco MCS Unified CM appliance at this time. For the most current server specifications, see the latest version of the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*.

# Operating Conditions

The sizing information presented in this chapter is based on the following operating conditions:

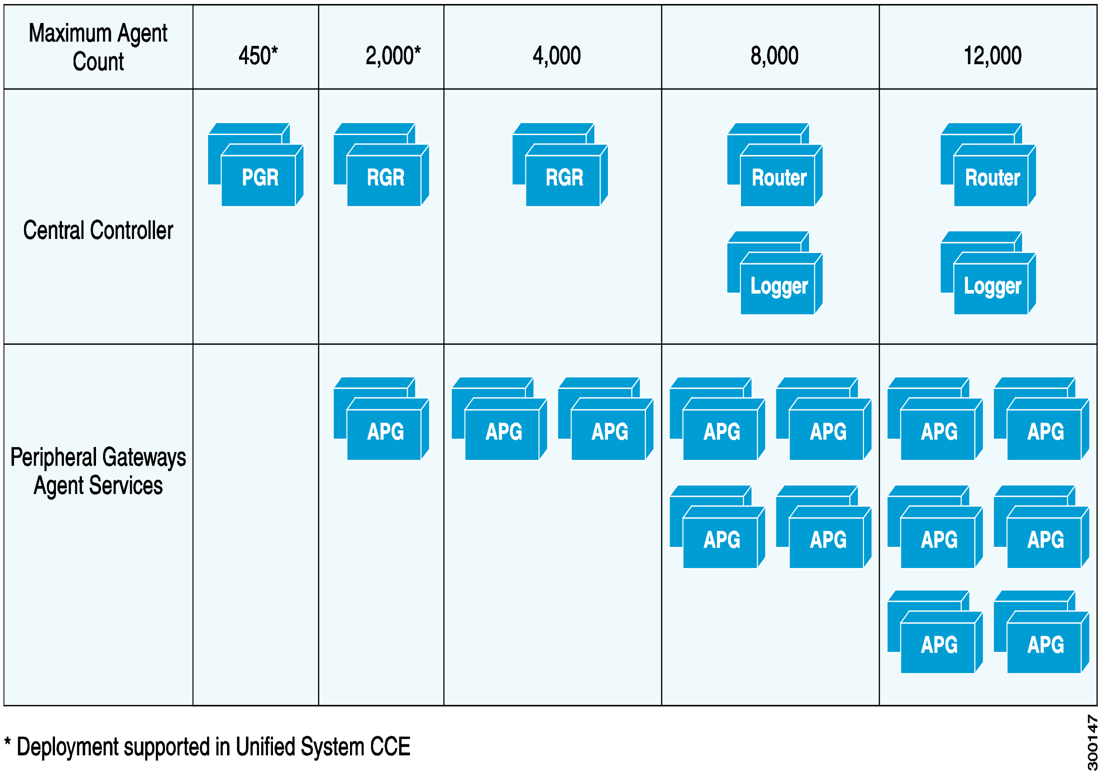
- Maximum of 30 busy hour call attempts (BHCA) per agent.
- Five skill groups per agent.
- The total number of agents indicated in the following tables and figures consists of 90% agents and 10% supervisors. For example, if a table or figure indicates 100 agents, the assumption is that there are 90 agents 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 IVR call.
- Five Extended 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 chapter:

- The number of agents indicates the number of logged-in agents.
- Server types:
  - APG = Agent Peripheral Gateway
  - PGR = Progger
  - RGR = Rogger

**Figure 121** *Minimum Servers Required for Release 8.0(x) Unified CCE Deployments with CTI OS Desktop*



| Maximum Agent Count | 450* | 2,000* | 4,000 | 8,000 | 12,000 |
|---|---|---|---|---|---|
| Central Controller | PGR | RGR | RGR | Router<br>Logger | Router<br>Logger |
| Peripheral Gateways Agent Services |  | APG | APG<br>APG | APG<br>APG<br>APG<br>APG | APG<br>APG<br>APG<br>APG<br>APG<br>APG |

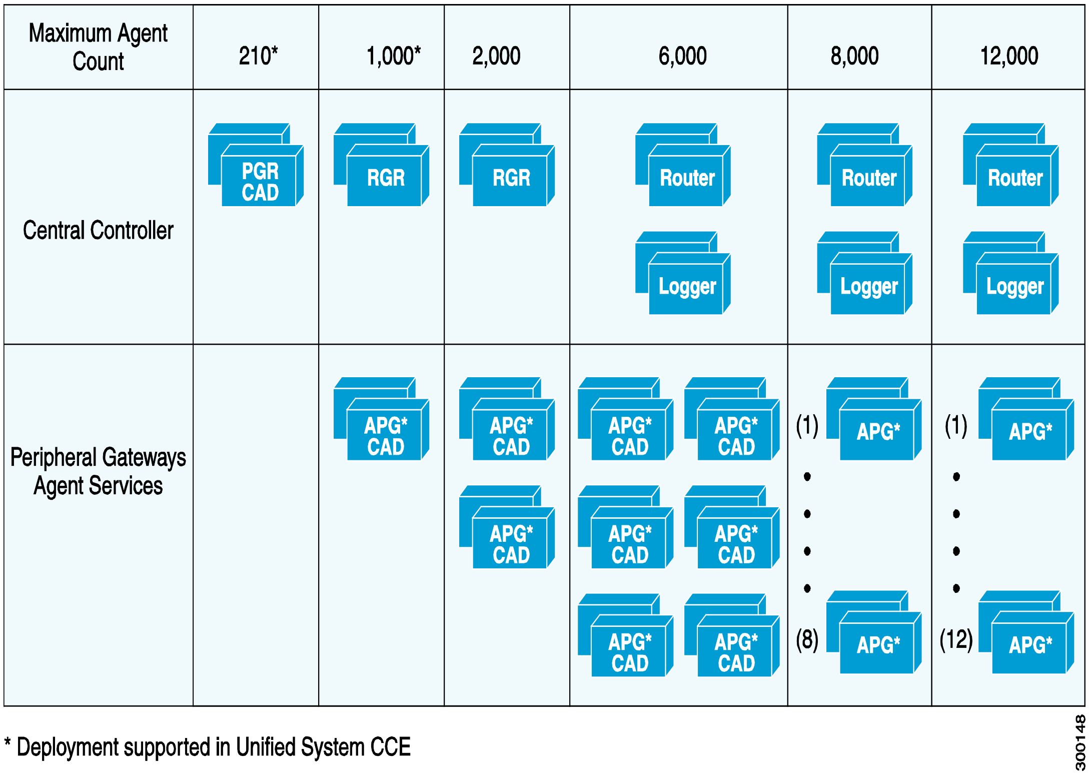
* Deployment supported in Unified System CCE

The following notes apply to Figure 121:

- Sizing is based on the information listed under Operating Conditions.
- Voice Response Unit (VRU), Administration & Data Server, and Unified CM components are not shown.
- For more information, see Peripheral Gateway and Server Options

**Note** The terms Rogger and Central Controller are used interchangeably throughout this chapter.

**Figure 122** ***Minimum Servers Required for Release 8.0(x) and Later Unified CCE Deployments with Cisco Agent Desktop***



| Maximum Agent Count | 210* | 1,000* | 2,000 | 6,000 | 8,000 | 12,000 |
|---|---|---|---|---|---|---|
| Central Controller | PGR CAD | RGR | RGR | Router<br>Logger | Router<br>Logger | Router<br>Logger |
| Peripheral Gateways Agent Services | | APG* CAD | APG* CAD<br>APG* CAD | APG* CAD<br>APG* CAD<br>APG* CAD<br>APG* CAD<br>APG* CAD<br>APG* CAD | (1) APG*<br>•<br>•<br>•<br>•<br>(8) APG* | (1) APG*<br>•<br>•<br>•<br>•<br>(12) APG* |

* Deployment supported in Unified System CCE

The following notes apply to Minimum Servers Required for Release 8.0(x) and Later Unified CCE Deployments with Cisco Agent Desktop:

- Sizing is based on the information listed under Operating Conditions.
- Voice Response Unit (VRU), Administration & Data Server, and Unified CM components are not shown.
- For more information, see Peripheral Gateway and Server Options
- The maximum values apply to all CAD clients (CAD, IPPA, and CAD-BE).

**Table 12** ***Sizing Information for Unified CCE Components and Servers***

| **Component** | **Server Class** | **Maximum Agents (local agents without agent greeting)** | **Notes** |
|---|---|---|---|
| **Progger: Peripheral Gateway, Router, and Logger** | | | Cannot be co-resident with an Administration & Data Server. In addition, the Progger cannot have additional Agent Peripheral Gateways.<br>Logger database is limited to 14 days of historical data. |
| | MCS-30-004-Class | CTI OS: 100<br>CAD: No | With MCS-30-004-Class servers:<br>• Maximum number of simultaneous queued calls equals half the number of agents.<br>• Outbound:<br>(Maximum PG agent capacity) – (4 x (number of SCCP dialer ports))<br>**or**<br>(Maximum PG agent capacity) – (1.33 x (number of SIP dialer ports))<br>• To determine the maximum Progger agent capacity, see the Progger inbound Agent entry in this table. The capacity depends on your ICM software release. |
| | MCS-40-005-Class | CTI OS: 450<br>CAD: 297 | With MCS-40-005-Class servers:<br>• Maximum number of simultaneous queued calls equals half the number of agents.<br>• Outbound:<br>(Maximum PG agent capacity) – (4 x (number of SCCP dialer ports))<br>**or**<br>(Maximum PG agent capacity) – (1.33 x (number of SIP dialer ports))<br>• To determine the maximum Progger agent capacity, see the Progger inbound Agent entry in this table. The capacity depends on your ICM software release. |

| Component | Server Class | Maximum Agents (local agents without agent greeting) | Notes |
|---|---|---|---|
| **Router** | MCS-30-004-Class | 500 | |
| **Router and Logger** | MCS-40-005-Class | 4,000 | |
| **Logger** | MCS-40-006-Class | 6,000 | MCS-30-00x-Class servers are not supported. |
| **Logger** | GEN-50-006-Class<br>GEN-50-007-Class | 12,000 | |
| **Router** | MCS-40-005-Class | 8,000 | MCS-30-00x-Class servers are not supported. |
| **Router** | MCS-40-015-Class<br>MCS-40-016-Class | 12,000 | |
| **Logger** | GEN-50-005-Class | 8,000 | MCS-30-00x-Class servers are not supported. |
| **Administration & Data Server** | | | For the most current hardware specifications for Administration & Data Server, see the latest version of the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*. |
| **Agent PG (Inbound Only)** | MCS-30-004-Class | CTI OS: 450<br>CAD: 297 | For more information about the various Agent PG deployment options, see the section Peripheral Gateway and Server Options. |

| **Component** | **Server Class** | **Maximum Agents (local agents without agent greeting)** | **Notes** |
|---|---|---|---|
| | MCS-40-005-Class | CTI OS: 2,000, 4,000<br>CAD: 1000 | **VRU Ports:**<br>Additional VRU PGs can be deployed to accommodate a greater number of VRU ports.<br>**Mobile Agents:**<br>Use the following calculations to determine mobile agent capacity:<br>– Each mobile agent for a nailed connection (nailed-up configuration): Equals 1.73 local agents for Release 8.0(*x*)<br>– Each mobile agent for a call-by-call basis: Equals 2.4 local agents for Release 8.0(*x*)<br>Support for 4,000 agents is limited to multiple Unified CM PIMs and multiple CTI OSs:<br>•2 to 10 Unified CM PIMs<br>•2 to 10 CTI OSs (Number of CTI OSs has to be equal to number of Unified CM PIMs)<br>•No VRU PIMs<br>**Note**: The Cisco Media Blender is not supported when installed on a PG system. |
| **Voice Response Unit (VRU) PG** | | | Use the number of ports instead of agent count.<br>Average of 5 Run VRU Script Nodes per call. |
| | MCS-30-004-Class | 1200 ports | Maximum of 4 PIMs; maximum of 10 cps. |
| | MCS-40-005-Class | 9600 ports | Maximum of 10 PIMs; maximum of 40 cps. |

| **Component** | **Server Class** | **Maximum Agents (local agents without agent greeting)** | **Notes** |
|---|---|---|---|
| **Agent PG with Outbound Voice (Includes Dialer and Media Routing PG)** | | [Maximum inbound agent capacity] - 4 * [Number of SCCP Dialer ports]<br>Or<br>[Maximum inbound agent capacity] – 1.33 * [Number of SIP Dialer ports] | 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.<br>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 2 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 you campaigns.<br>Example: Agent PG with CAD and 10 SCCP Dialer ports.<br>Available inbound CAD agents = 1000 - (4*10) = 960.<br>**Note**: The Cisco Media Blender is not supported when installed on a PG system. |
| **Silent Monitor Server** | MCS-30-004-Class | 20 (Simultaneous recording sessions) | |
| | MCS-40-005-Class | 40 (Simultaneous recording sessions) | |
| **Cisco Unified Web and E-Mail Interaction Manager** (Agent PG with Media Blender, collaboration includes Media Routing PG) | MCS-40-005-Class | 250 (all media) | Media Routing (MR) PG co-residency requires the MCS-40-005class server.<br>See EIM/WIM documentation for the most current information. |

| Component | Server Class | Maximum Agents (local agents without agent greeting) | Notes |
|---|---|---|---|
| **Cisco Unified Web and E-Mail Interaction Manager** (Media Blender optional with MR PG) | MCS-40-005-Class | | MCS-30-00x class is not supported.<br>See EIM/WIM documentation for the most current information. |
| **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 *Hardware and System Software Specification for Cisco Unified CVP*. |
| **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 Cisco Unified Intelligence Center Bill of Materials |

[1] In addition to the MCS models listed in Table 12, there are other server models based on Intel Xeon Nehalem quad-core technology that can be used for Unified CCE deployments. For further details, see the latest version of the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*.

# Additional Sizing Factors

Many variables in the Unified CCE configuration and deployment options can affect the hardware requirements and capacities. This section describes the major sizing variables and how they affect the capacity of the various Unified CCE components. In addition, Table 15 summarizes the sizing variables and their effects.

### 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 will decrease 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 will impact the performance of most Unified CCE server components, including Unified CM clusters. For the impact of agents on the performance of Unified CM components, see the Sizing Cisco Unified Communications Manager Servers chapter.

### Average Skill Groups per Agent

The number of skill groups per agent (which is independent of the total number of skills per system) has significant effects on the CTI OS and Finesse servers, the Agent PG, and the Call Router and Logger. Limit the number of skill groups per agent to 5 or fewer, when possible, and that you periodically remove unused skill groups so that they do not affect system performance. You can also manage the effects on the CTI OS server by increasing the value for the frequency of statistical updates. Table 13 shows examples of how the number of skill groups per agent can affect the capacity of the Unified CCE system. The numbers in Table 13 are based on the information listed in the section on Operating Conditions, and it shows capacity per CTI OS instance.

**Table 13** ***Sizing Effects Due to Number of Skill Groups per Agent (12,000 Agents)***

The numbers in this table are subject to specific hardware and software requirements. Refer to the *Hardware and System Software Specification (Bill of Materials) for Cisco Unified ICM / Contact Center Enterprise & Hosted* for details on 12,000 agent support.

| | System | Generic PG Limits | | | |
|---|---|---|---|---|---|
| Avg Configured SG Per Agent | Max Concurrent Agent Per System | Max Concurrent Agent Per PG | Max Configured SG Per PG | Max Configured VRU Ports Per PG | Max Configured VRU PIMs Per PG |
| 5 | 12000 | 2000 | 3000 | 1000 | 5 |
| 10 | 11038 | 1832 | 3000 | 1000 | 5 |
| 15 | 10078 | 1663 | 3000 | 1000 | 5 |
| 20 | 9116 | 1495 | 3000 | 1000 | 5 |
| 25 | 8156 | 1326 | 3000 | 1000 | 5 |
| 30 | 7194 | 1158 | 3000 | 1000 | 5 |
| 35 | 6234 | 989 | 3000 | 1000 | 5 |
| 40 | 5272 | 820 | 3000 | 1000 | 5 |
| 45 | 4312 | 652 | 3000 | 1000 | 5 |
| 50 | 3350 | 484 | 3000 | 1000 | 5 |

**Table 14** *Sizing Effects Due to Number of Skill Groups per Agent (8,000 Agents)*

| | System | Generic PG Limits | | | |
|---|---|---|---|---|---|
| AvgSG/Agent | Max Agent Per System | Max Agent Per PG | Max Skillgroups Per PG | Max VRU Ports Per PG | Max VRUPIMs Per PG |
| 5 | 8000 | 2000 | 3000 | 1000 | 5 |
| 10 | 7342 | 1832 | 3000 | 1000 | 5 |
| 15 | 6683 | 1663 | 3000 | 1000 | 5 |
| 20 | 6024 | 1495 | 3000 | 1000 | 5 |
| 25 | 5364 | 1326 | 3000 | 1000 | 5 |
| 30 | 4705 | 1158 | 3000 | 1000 | 5 |
| 35 | 4047 | 989 | 3000 | 1000 | 5 |
| 40 | 3389 | 820 | 3000 | 1000 | 5 |
| 45 | 2730 | 652 | 3000 | 1000 | 5 |
| 50 | 2072 | 484 | 3000 | 1000 | 5 |

**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 small number of agents.

**Note** Supervisors can monitor only agents within their own team, all of whom must be configured on the same peripheral.

**Note** You can add a maximum of 50 agents per team.

**Note** You can add maximum of 10 supervisors per team.

A Unified CCE 8.*x* system can support a maximum of 50 agents per supervisor with assumptions below. If a particular environment requires more than 50 agents per supervisor, then use the following formula to ensure that there will be 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), 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 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') will be, Y = 90% * 1 + 10% * 12 = 2.1.

(Note that the number of configured statistics in 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 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 updates per second

- The maximum number of agents per supervisor must not exceed 200 for any given configuration, still holding updates per sec to a max of 31 with above formula.

### CTI OS Monitor Mode Applications

A CTI OS Monitor Mode application can impact 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 as well as 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 have an effect on 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 will impact performance of most Unified CCE server components. An increase in the number of transfers and conferences will increase the load on the system and, thus, decrease 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 IVR will handle all calls initially (call treatment) and direct the callers to agents after a short queuing period, or whether the agents will handle calls immediately and the IVR will queue only unanswered calls when all agents are busy. The answer to this question determines very different IVR sizing requirements and affects the performance of the Call Router/Logger and Voice Response Unit (VRU) PG. Required VRU ports can be determined using the Cisco Unified CCE Resource Calculator. (See Cisco Unified CCE Resource Calculators for more information.)

### 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:

Translation route pool = 20 * (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 will increase significantly. The delay time between replaying Run VRU scripts also has an impact.

### Reporting

Real-time reporting can have a significant effect on Logger, Progger, and Rogger processing due to database access. A separate server is required for an Administration & Data Server to off-load reporting overhead from the Logger, Progger, and Rogger.

### IVR Script Complexity

As IVR script complexity increases with features such as database queries, the load placed on the IVR server and the Router also increases. There is no good rule of thumb or benchmark to characterize the Unified IP IVR performance when used for complex scripting, complex database queries, or transaction-based usage. Test complex IVR 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 IVR 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 and must be factored into the sizing requirements as stated in Table 13.

### 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 very 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.

### Extended Call Context (ECC)

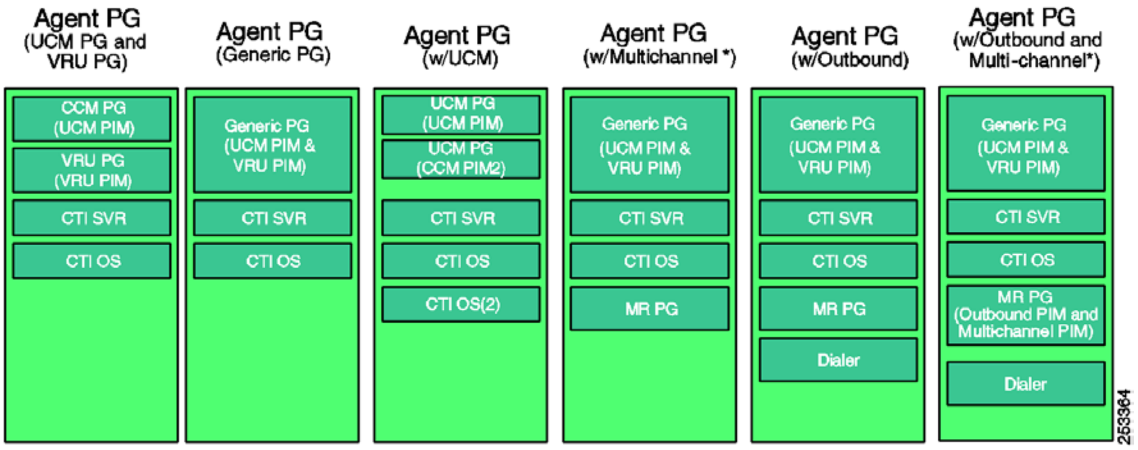
The ECC usage impacts PG, Router, Logger, and network bandwidth. There are many ways that ECC can be configured and used. The capacity impact will vary based on ECC configuration; handled these on a case-by-case basis.

# Peripheral Gateway and Server Options

A Unified CCE Peripheral Gateway (PG) translates messages coming from the Unified CM 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.

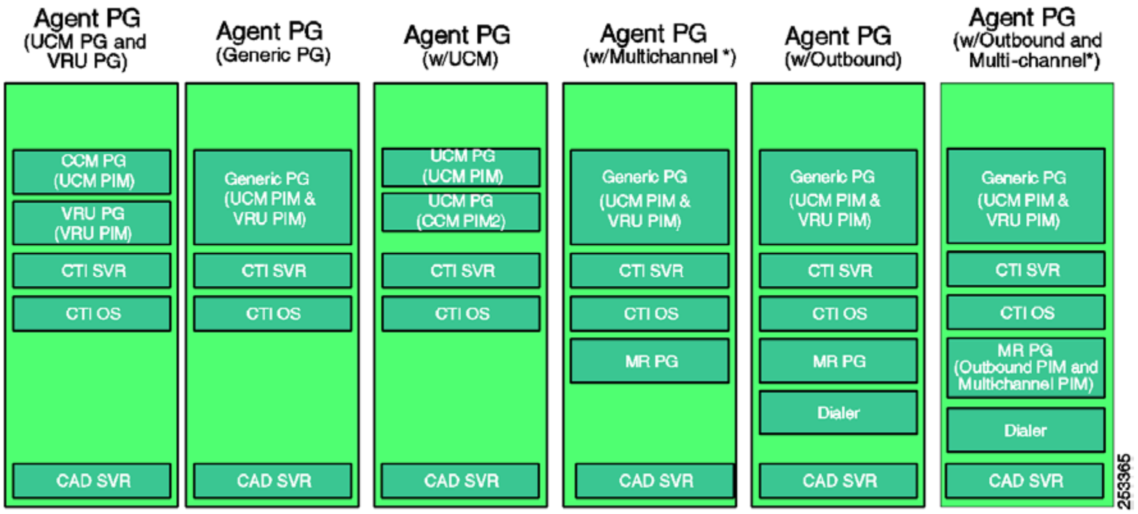
Figure 123 and Figure 124 illustrate various configuration options for the Agent PG with CTI OS and Cisco Agent Desktop. Table 15 lists PG and PIM sizing guidelines.

**Figure 123** *Agent PG Configuration Options with CTI OS*


Agent PG (UCM PG and VRU PG)
CCM PG (UCM PIM)
VRU PG (VRU PIM)
CTI SVR
CTI OS
Agent PG (Generic PG)
Generic PG (UCM PIM & VRU PIM)
CTI SVR
CTI OS
Agent PG (w/UCM)
UCM PG (UCM PIM)
UCM PG (CCM PIM2)
CTI SVR
CTI OS
CTI OS(2)
Agent PG (w/Multichannel *)
Generic PG (UCM PIM & VRU PIM)
CTI SVR
CTI OS
MR PG
Agent PG (w/Outbound)
Generic PG (UCM PIM & VRU PIM)
CTI SVR
CTI OS
MR PG
Dialer
Agent PG (w/Outbound and Multi-channel*)
Generic PG (UCM PIM & VRU PIM)
CTI SVR
CTI OS
MR PG (Outbound PIM and Multichannel PIM)
Dialer
253364

* 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).

**Figure 124** *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).

**Table 15** ***PG and PIM Sizing Guidelines***

| **Sizing Variable** | **Guidelines Based on Unified CCE Software Releases 8.0(*x*)** |
|---|---|
| Maximum number of PGs per Unified CCE | 250 |
| Maximum number of PG types per server platform | Up to two PG types are permitted per server, provided that any given server is limited to the maximum agent and VRU port limitations outlined in Table 12. |
| Maximum number of Unified CM PGs per server | Only one Unified CM PG, Generic PG, or System PG is allowed per physical server |
| Maximum number of Unified CM PIMs per PG | 10 Unified CM PIMs with associated Agents can be configured per PG, provided that any given server is limited to the maximum agent and VRU port limitations outlined in Table 12.<br>However, if any VRU PIM is configured, a generic PG cannot support configuration of more than 1 CM PIM.<br>See Peripheral Gateway Design Considerations for more information. |
| Can PGs be remote from Unified CCE? | Yes |
| Can PGs be remote from Unified CM? | No |
| Maximum number of IVRs controlled by one Unified CM | See the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*. |
| Maximum number of CTI servers per PG | 1 |
| Can PG be co-resident with Cisco MCS Unified CM appliance? | No |

# Cisco Agent Desktop Component Sizing

For details on the components and architecture of the Cisco Agent Desktop, see the Unified Contact Center Enterprise Desktop.

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.

## 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 (refer to the Compatibility Matrix for details).

## 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 computers 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 server (an MCS-30-003-Class server or equivalent). Each dedicated VoIP Monitor Service can support up to 400 phones if a 100 Megabit NIC is used to connect to the switch or 1,000 phones if a Gigabit NIC is used.

## 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.

Table 16 summarizes the raw Recording and Playback Service capacity.

**Table 16** ***Capacity of Recording and Playback Service***

| Recording and Playback Service Type | Maximum Simultaneous Recordings |
|---|---|
| Co-resident | 32 |
| Dedicated | 80 |

# 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 UCCE 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, as does having two agent peripherals on the same PG. When sizing the PG, the sizing calculator factors in Agent Greeting weight.

## Communications Manager

When Agent Greeting and/or Mobile Agent and IP-IVR are in use, the number of agents supported by a UCM subscriber is impacted.

The Unified Communications 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, it is recommended that for each Mobile Agent (when Agent Greeting is enabled); indicate a conference (in place of Agent Greeting) 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 CVP SRND 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 Unified Communications Sizing Tool.

# 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.

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 * (the total number of CPUs on the system)).
- Monitoring Memory
  - % Committed Bytes; this value must remain less than (0.8 * (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 * (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 * (the physical size 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 servers 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 8,000 agents and 216,000 BHCA. For smaller deployments, cost savings can be achieved with careful planning and co-resident Unified CCE components (for example, Progger, Rogger, and Agent PG).

Additionally, pay careful attention to the sizing variables that will 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 and Proggers. While new versions will scale far higher, the Cisco Agent Desktop Monitor Server is still limited in the number of simultaneous sessions that can be monitored by a single server when monitoring and recording are required.

CHAPTER 10

# Sizing Cisco Unified Communications Manager Servers

This chapter discusses the concepts, provisioning, and configuration of Cisco Unified Communications Manager (Unified CM) clusters when used in a Unified CCE deployment. Unified CM clusters provide a mechanism for distributing call processing across a converged IP network infrastructure to support Cisco Unified Communications, facilitate redundancy, and provide feature transparency and scalability.

This chapter covers only the Unified CCE operation with Unified CM clusters and proposes reference designs for implementation.

The information in this chapter builds on the concepts presented in the *Cisco Unified Communications SRND*. Some duplication of information is necessary to clarify concepts relating to Unified CCE as an application supported by the Unified CM call processing architecture. However, the foundational concepts are not duplicated here; become familiar with them before continuing with this chapter.

This chapter documents general best practices and scalability considerations for sizing the Unified CM servers used with your Unified CCE deployments. Within the context of this document, scalability refers to Unified CM server and/or 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 Unified Communications SRND*.

## Cluster Sizing Concepts

Before attempting to size a Unified CM 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 will be 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 will share a Unified CM cluster with a Unified CCE deployment (such as 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 Unified CM cluster(s). Many factors impact the sizing of a Unified CM 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 (can be zero if Unified CVP is used for call treatment, self service, and queuing)
- Number of voice gateway ports and the BHCA rate on those VoIP endpoints
- Type of outbound dialer (SCCP or SIP dialer)
- Number of outbound agent phones, outbound dialing mode, and BHCA rate per phone
- Number of outbound dialer ports, number of IVR 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 protocol(s) 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, route lists, and so forth
- Amount of media resources needed for functions such as transcoding, conferences, encryption, and so forth
- Co-resident applications and services such CTI Manager, E-911, and Music on Hold
- Unified CM release (sizing will vary per release)
- Desired hardware server model (sizing will vary per hardware server model)

Other factors can affect cluster sizing, but the above list shows the most significant factors in terms of resource consumption.

The general process to sizing a Unified CM cluster is to estimate the resource consumption (CPU, memory, and I/O) for each of these factors and then to choose hardware that will satisfy the resource requirements.

It is important to gather information with regard to the factors listed above before attempting to size a cluster with any accuracy. To assist in calculating these numbers that relate to Unified CCE, use the Unified CCE Resource Calculators discussed in the Sizing Contact Center Resources chapter.

The next section describes the tools for sizing Cisco Unified CM deployments.

# Cisco Unified Communications Sizing Tool

To size Cisco Unified Contact Center Enterprise servers, use the Cisco Unified Communications Sizing Tool (Unified CST) available at http://tools.cisco.com/cucst/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 CM, 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 CM clusters with Unified CCE.

- A cluster may contain a mix of server platforms, but this is strongly discouraged except in migration or upgrade scenarios. All primary and failover (backup) server pairs must be of the same type. All servers in the cluster must run the same Unified CM software release and service pack.
- Within a cluster, you may enable a maximum of 8 servers with the Cisco Call Manager Service, including backup servers. Additional servers may be used for more dedicated functions such as TFTP, publisher, music on hold, and so forth.
- In a deployment with IP IVR, each Unified CM cluster (four primary and four backup subscriber servers) can support up to approximately 2,000 Unified CCE agents. This limit assumes that the BHCA call load and all configured devices are spread equally among the eight call processing servers with 1:1 redundancy. (See Unified CM Redundancy for redundancy schemes.) Each of the eight Unified CM servers (MCS-7845 High Performance Servers) would support a maximum of 250 agents. In a failover scenario, the primary server would support a maximum of 500 agents. These capacities can vary, depending on your specific deployment. All deployments must be sized by using the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool. Some deployments with more than 500 agents per pair of subscriber nodes or 2000 agents per cluster might be supported, depending on the output of the Unified CM Capacity Tool or the Unified Communications Sizing Tool.
- In a deployment with Unified CVP (no IP IVR), each Unified CM cluster (four primary and four backup subscriber servers) can support up to about 4,000 Unified CCE agents. This limit assumes that the BHCA call load and all configured devices are spread equally among the eight call processing servers with 1:1 redundancy. (See Unified CM Redundancy for redundancy schemes.) Each of the eight Unified CM servers (MCS-7845-H2/I2 or later High Performance Servers) would support a maximum of 500 agents. In a failover scenario, the primary server would support a maximum of 1,000 agents. These capacities can vary, depending on your specific deployment. All deployments must be sized by using the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool.

When sizing the Unified CM Cluster to support Contact Center solutions for the appropriate number of CTI resources, remember to also account for additional configured phones from non-logged in agents, additional applications like Call Recording, Attendant Console, PC-clients which remotely control the device, and other 3rd-Party applications which consume additional CTI resources. Multiple lines on the same device prior to Unified CM 7.1(3) also require additional CTI resources.

Unified CM 7.1(3) (or later) has been enhanced to support more CTI resources. Consider upgrading when multiple lines and/or multiple applications (e.g. Contact Center and Recording) will be used concurrently.

Those CTI resources follow the same CTI rules as described in the Unified Communications SRND. For more details, refer to the Unified Communications SRND. All deployment must be sized by using the Cisco Unified Communications 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. Any administrative work on Unified CM will impact call processing and CTI Manager activities if there are any devices registered with the publisher.
- Do not use a publisher as a failover or backup call processing server unless you have fewer than 150 agent phones and the installation is not mission critical or is not a production environment. The Cisco MCS-7825 server is the minimum server supported for Unified CCE deployments. (Refer to Table 17 for more details.) Any deviations will 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 subscriber servers 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.
- If you require more than one Unified CCE JTAPI user (CTI Manager) and more than one primary Unified CM subscriber, then 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 server if possible.
- 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 additional CTI Manager best practices, see the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.
- If you have a mixed cluster with Unified CCE and general office IP phones, group and configure each type on a separate server if possible (unless you need only one subscriber server). For example, all Unified CCE agents and their associated devices and resources (gateway ports, CTI ports, and so forth) would be on one or more Unified CM servers, and all general office IP phones and their associated devices (such as gateway ports) would be on other Unified CM servers, as long as cluster capacity allows. If you use the Cisco Unified CM Capacity Tool, you have to run it multiple times with the specific device configuration for each primary Unified CM server because the tool assumes all devices are equally balanced in a cluster. In this case, use the 1:1 redundancy scheme.
- (See Unified CM Redundancy for details.) If you use the Unified Communications Sizing Tool instead, you do not have to run it multiple times because this tool supports deployments with dedicated Unified CM 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 Unified CM cluster.
- Configure all CTI route points associated with the Unified CCE Peripheral Gateway (PG) JTAPI user to register with the subscriber node running the CTI Manager instance that is communicating with that Unified CCE PG.

- The Cisco Unified CM Capacity Tool and the Unified Communications Sizing Tool do not currently measure CTI Manager impact on each server separately. However, CTI Manager does place an additional burden on the subscriber node running that process. The Cisco Unified CM Capacity Tool and the Unified Communications Sizing Tool report the resource consumption based on these nodes. The actual resource consumption on the other Cisco Unified CM nodes might be slightly lower.
- Count devices that are associated with a Unified CCE PG JTAPI user, but are not used by a call center agent, as an agent device, because the PG will still be notified of all device state changes for that phone even though it is not being used by an agent. If a device is unlikely to be used regularly by a call center agent, do not associate the device with the Unified CCE PG JTAPI user in order to increase cluster scalability.
- For deployments requiring large numbers of IVR ports, use Unified CVP instead of IP IVR. IP IVR ports place a significant call processing burden on Unified CM, while Unified CVP does not. Thus, Unified CCE deployments with Unified CVP will allow more agents and higher BHCA rates per cluster. Size all deployments by using the Unified Communications Sizing Tool.
- In deployments with multiple IP IVRs, associate those servers with different CTI Managers on different subscriber nodes in order to better balance call processing across the cluster.
- Unified CM CPU resource consumption varies, depending on the trace level enabled. Changing the trace level from Default to Full on Unified CM can increase CPU consumption significantly under high loads. Changing the tracing level from Default to No tracing can decrease CPU consumption significantly at high loads, but this configuration is not supported by Cisco Technical Assistance Center.
- Under normal circumstances, place all servers from the Unified CM 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, you must follow the specific guidelines for clustering over the IP WAN as described in both IPT: Clustering Over the WAN in this guide and the section on Clustering over the IP WAN in the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

For the most current information about Unified CM and Unified CCE supported releases, see the latest version of the *Cisco Unified Communications Manager Compatibility Matrix*.

For additional Unified CM clustering guidelines, see the *Cisco Unified Communications Solution Reference Network Design (SRND)* guide.

# Unified CM Servers

Unified CM clusters utilize various types of servers, depending on the scale, performance, and redundancy required. They range from non-redundant, single-processor servers to highly redundant, multiprocessor servers.

Unified CM is supported only on specific hardware platforms. For a list of the currently supported hardware configurations, see the documentation on Cisco MCS 7800 Series Unified CM appliances.

In order to size a Unified CM deployment, you must use the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool, both of which indicate the number of Unified CM servers needed for each type of platform.

Avoid deploying only one Unified CM subscriber for mission-critical contact center deployments and for deployments with more than 150 agents. Table 17 lists the maximum number of agents in a system where only one Unified CM subscriber server is deployed, with the Unified CM publisher used as backup.

**Table 17** ***Capacity When Deploying Only One Unified CM Subscriber Server***

| Server Type | Maximum Number of Agents |
|---|---|
| MCS-7825 | 100 |
| MCS-7835 or MCS-7845 | 150 |

The Cisco MCS-7815 and MCS-7816 servers are not supported with Unified CCE deployments, but lab or demo setups can use these servers. They are, however, a supported Unified CM platform for Cisco Unified Communications deployments only.

# Unified CM Redundancy

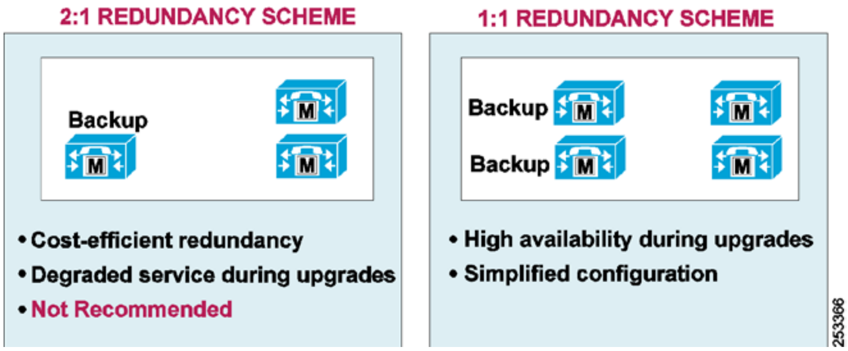
With Unified CM, you can choose from the following two redundancy schemes:

- 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 CM deployments with Unified CCE.

- Figure 125 illustrates these two 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 Unified Communications Solution Reference Network Design (SRND) Guide*.

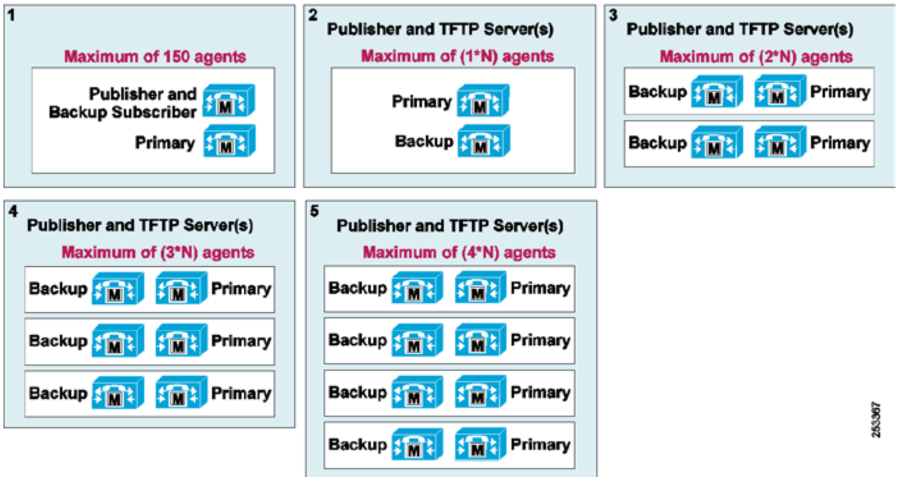
**Figure 125** ***Basic Redundancy Schemes***


2:1 REDUNDANCY SCHEME
Backup
M
M
M
• Cost-efficient redundancy
• Degraded service during upgrades
• Not Recommended
1:1 REDUNDANCY SCHEME
Backup
Backup
M
M
M
M
• High availability during upgrades
• Simplified configuration
253366

In Figure 126 the five 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 CM 8.*x* and Unified CVP (not IP IVR), N is equal to 1000. For deployments with IP IVR, N is equal to 500. For other types of deployments, use the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool. In all types of deployments, the exact number of servers will depend on the hardware platforms chosen or required, as determined by the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool.

**Figure 126** ***1:1 Redundancy Configuration Options***



# Load Balancing for Unified CM

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. Normally (as in the 2:1 redundancy scheme) a backup server has no devices registered unless its primary is unavailable.

With load balancing, you can move up to half of the device load from the primary to the secondary subscriber by using the Unified CM redundancy groups and device pool settings. In this way, you can reduce by 50% the impact of any server becoming unavailable.

To plan for 50/50 load balancing, calculate the capacity of a cluster without load balancing and then distribute the load across the primary and backup subscribers based on devices and call volume. To allow for failure of the primary or the backup, the total load on the primary and secondary subscribers must not exceed that of a single subscriber. For example, if deploying Unified CVP and Unified CM 8.0 or later releases, MCS-7845-H3 servers have a total server limit of 1,000 Unified CCE agents. In a 1:1 redundancy pair, you can split the load between the two subscribers, configuring each subscriber with 500 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 server.

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.

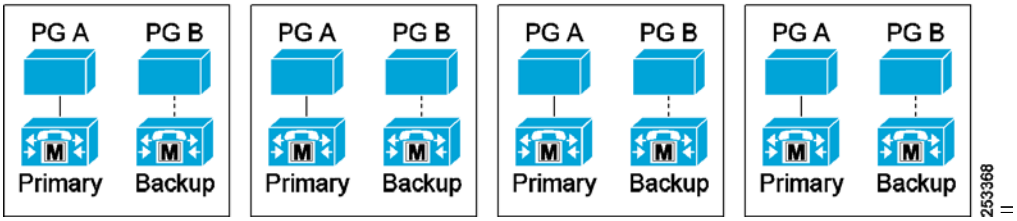
For additional information about general call processing topics such as secondary TFTP servers and gatekeeper considerations, see the *Cisco Unified Communications Solution Reference Network Design (SRND) Guide*.

# Deployment of Agent PG in a Unified CM Cluster

Agent PGs can be deployed in a Unified CM cluster in either of the following ways:

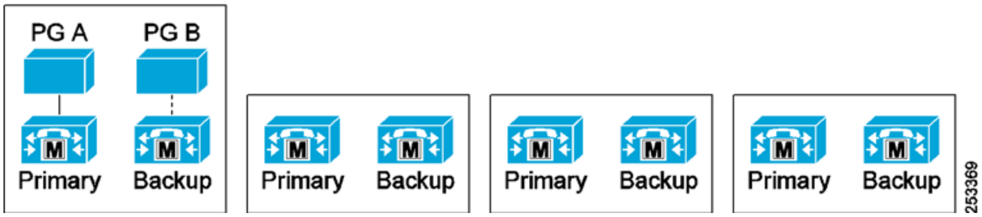
- The first method is to deploy an Agent PG for each pair of Cisco Unified CM subscriber nodes. In this case, each Unified CM subscriber node runs the CTI Manager service, and each Agent PG connects to a CTI Manager running on its corresponding Unified CM subscriber pair. The following diagram shows an example where four primary Unified CM subscribers are required and four backup Unified CM subscribers are deployed to provide 1:1 redundancy.

**Figure 127** *Deploy Agent PG for Each Pair of Unified CM Subscriber Nodes*



- Another possible method is to deploy a single Agent PG for the entire Cisco Unified CM cluster. This type of deployment requires a single pair of Cisco Unified CM subscriber nodes running CTI Manager. Spread agent phone registration among all the Cisco Unified CM subscriber nodes, including the Unified CM subscribers running the CTI Manager service. The following diagram shows an example where four primary Unified CM subscribers are required and four backup Unified CM subscribers are deployed to provide 1:1 redundancy.

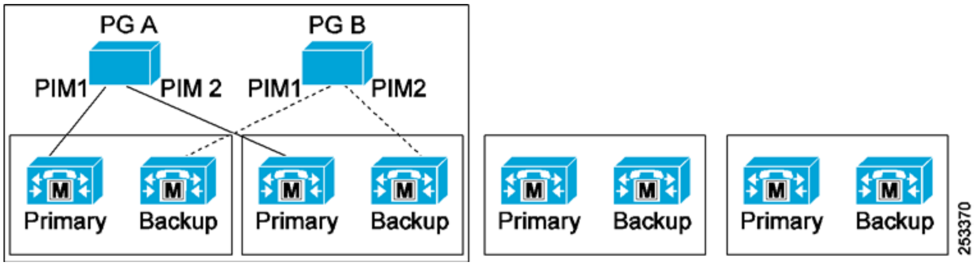
**Figure 128** *Deploy Single Agent PG for Entire Unified CM Cluster*



One benefit of this model is the reduction of the server count for the PG. Another benefit is that there is a single PIM for the entire Unified CM cluster. This makes it possible to create teams that span across many Unified CM subscribers, thus allowing supervisors, for example, to monitor agent phones registered to any Unified CM subscriber node in the Unified CM cluster. However, the resource utilization on the Unified CM cluster might be slightly higher in this type of deployment. Use the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool to size the Unified CM servers for a particular deployment

A variation of this type of deployment is available with Unified CM 8.0 or later releases, when Unified CVP only is deployed. (This model is not supported when IP IVR is deployed.) Up to 4,000 agents can be supported in a single Unified CM cluster in this case. When deploying more than 2,000 agents, a minimum of 2 CTI Manager pairs are required. One Agent PG with two PIMs could be deployed, with each PIM configured with a separate pair of Unified CM subscribers running the CTI Manager service and each PIM configured with up to 2,000 agents. Spread agent phone registration among all the Cisco Unified CM subscriber nodes, including the Unified CM subscribers running the CTI Manager service. The following diagram shows an example where four primary Unified CM subscribers are required, and four backup Unified CM subscribers are deployed to provide 1:1 redundancy.

**Figure 129** ***Use Unified CVP for Unified CM 8.0 or later***



# Upgrading Unified CM

The 1:1 redundancy scheme allows upgrades with only the failover periods impacting the cluster. The 1:1 redundancy scheme enables you to upgrade the cluster using the following method:

Step 1. Upgrade the publisher server.

Step 2. Upgrade dedicated TFTP and music-on-hold (MoH) servers.

Step 3. Upgrade the backup subscribers one at a time. This step will affect some users if 50/50 load balancing is implemented.

Step 4. Fail over the primary subscribers to their backups, and stop the Cisco CallManager Service on the primaries. All users are on primaries and are moved to backup subscribers when the Cisco CallManager Service is stopped. CTI Manager is also stopped, causing the Peripheral Gateway (PG) to switch sides and inducing a brief outage for agents on that particular node.

Step 5. Upgrade the primaries, and then re-enable the Cisco CallManager Service.

With this upgrade method, there is no period (except for the failover period) when devices are registered to subscriber servers that are running different versions of the Unified CM software. This factor can be important, because the Intra-Cluster Communication Signaling (ICCS) protocol that communicates between subscribers can detect a different software version and shut down communications to that subscriber.

The 2:1 redundancy scheme allows for fewer servers in a cluster, but it can potentially result in an outage during upgrades. This is not a preferred scheme for Unified CCE deployments, although it is supported if it is a customer requirement and if possible outage of call processing is not a concern to the customer.

The 2:1 redundancy scheme enables you to upgrade the cluster using the following method. If the Cisco CallManager Service does not run on the publisher database server, upgrade the servers in the following order:

Step 1. Upgrade the publisher database server.

Step 2. Upgrade the Cisco TFTP server if it exists separately from the publisher database server.

Step 3. Upgrade servers that have services related only to Unified CM (music on hold, Cisco IP Media Streaming Application, and so forth) running on them. Make sure that you upgrade only one server at a time. Make sure that the Cisco CallManager Service does not run on these servers.

Step 4. Upgrade each backup server, one server at a time.

Do not oversubscribe the backup server or servers during the upgrade. The number of agent phones registered to the backup server during the upgrade must not exceed the maximum capacity indicated by the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool. Perform the upgrade during off-peak hours when the call volume is low.

Step 5. Upgrade each primary server that has the Cisco CallManager Service running on it. Remember to upgrade one server at a time. During the upgrade of the second primary subscriber, there will be some outage for users and agents subscribed on that server, until the server is upgraded. Similarly, when you upgrade the fourth primary subscriber, there will be some outage for users and agents subscribed on that server, until the server is upgraded.

# Cisco 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 CM cluster, even if the voice gateway used to call the mobile agent is registered with a different Unified CM cluster. The agent desktop is configured with the IP address of the PG and/or CTI server to which it is associated.

For specific Unified CM node and cluster sizing for Unified CCE deployments, the Cisco Unified CM Capacity Tool or the Unified Communications Sizing Tool must be used. When sizing the Unified CM 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 CM 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 CM Capacity Tool. As an alternative, or when using the Cisco Unified Communications 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.

For more details on Cisco Unified Mobile Agent architecture, deployment models, and Unified CCE sizing, see the Cisco Unified Mobile Agent chapter.

CHAPTER 11

# Bandwidth Provisioning and QoS Considerations

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 between remote components over the WAN, including recommendations on how to apply proper Quality of Service (QoS) to WAN traffic flows. For a more detailed description of the Unified CCE architecture and various component internetworking, see the Architecture Overview chapter.

Cisco Unified CCE has traditionally been deployed using private, point-to-point leased-line network connections for both its *private* (Central Controller or Peripheral Gateway, side-to-side) as well as *public* (Peripheral Gateway to Central Controller) WAN network structure. Optimal network performance characteristics (and route diversity for the fault-tolerant failover mechanisms) are provided to the Unified CCE application only through dedicated private facilities, redundant IP routers, and appropriate priority queuing.

Enterprises deploying networks that share multiple traffic classes, of course, 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.

Provided that the required latency and bandwidth requirements inherent in the real-time nature of this product are satisfied, Cisco supports Unified CCE deployments in a convergent QoS-aware public network as well as in a convergent QoS-aware private network environment. This chapter presents QoS marking, queuing, and shaping recommendations for both the Unified CCE public and private network traffic.

Historically, two QoS models have been used: Integrated Services (IntServ) and Differentiated Services (DiffServ). The IntServ model relies on the Resource Reservation Protocol (RSVP) to signal and reserve the desired QoS for each flow in the network. Scalability becomes an issue with IntServ because state information of thousands of reservations has to be maintained at every router along the path. DiffServ, in contrast, categorizes traffic into different classes, and specific forwarding treatments are then applied to the traffic class at each network node. As a coarse-grained, scalable, and end-to-end QoS solution, DiffServ is more widely used and accepted. Unified CCE applications are not aware of RSVP, and the QoS considerations in this chapter are based on DiffServ.

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.

# Unified CCE Network Architecture Overview

Unified CCE is a distributed, resilient, and fault-tolerant network application that relies heavily on a network infrastructure with sufficient performance to meet the real-time data transfer requirements of the product. A properly designed Unified CCE network is characterized by proper bandwidth, low latency, and a prioritization scheme favoring specific UDP and TCP application traffic. These design requirements are necessary to ensure both the fault-tolerant message synchronization of specific duplexed Unified CCE nodes (Central Controller and Peripheral Gateways) as well as the delivery of time-sensitive system status data (routing messages, agent states, call statistics, trunk information, and so forth) across the system. Expeditious delivery of PG data to the Central Controller is necessary for accurate call center state updates and fully accurate 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 (H.323, MGCP, SCCP, or TAPI/JTAPI), according to the endpoints involved in the call. Call control functions include those used to set up, maintain, tear down, or redirect calls. For Unified CCE, control traffic includes routing and service control messages required to route voice calls to peripheral targets (such as agents, skill groups, or services) and other media termination resources (such as Unified IP IVR ports) as well as the real-time updates of peripheral resource status.

- Data traffic

  Data traffic can include normal traffic such as email, web activity, and CTI database application traffic sent to the agent desktops, such as screen pops and other priority data. Unified CCE priority data includes data associated with non-real-time system states, such as events involved in reporting and configuration updates.

This chapter focuses primarily on the types of data flows and bandwidth used between a remote Peripheral Gateway (PG) and the Unified CCE Central Controller (CC), on the network path between sides A and B of a PG or of the Central Controller, and on the CTI flows between the desktop application and CTI OS and/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 is 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 Unified Communications Solution Reference Network Design (SRND) Guide*.

Data traffic and other mission-critical traffic will vary 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.

## Network Segments

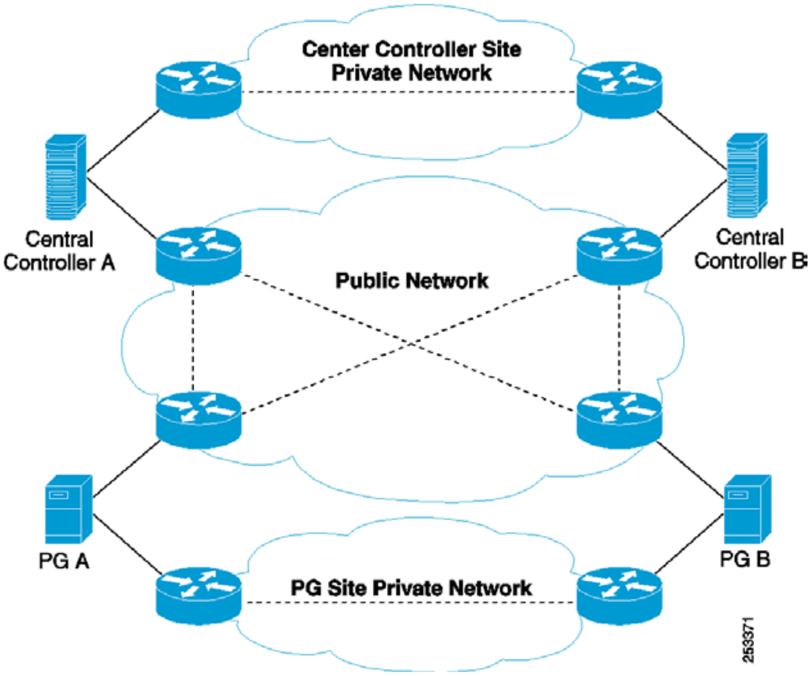
The fault-tolerant architecture employed by Unified CCE requires two independent communication networks. The *private network* (using a separate path) carries traffic necessary to maintain and restore synchronization between the systems and to allow clients of the Message Delivery Subsystem (MDS) to 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 Hosted Unified CCH/Unified CCE architecture. The signaling access network is not addressed in this chapter.

Figure 130 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 130** ***Example of Public and Private Network Segments for a Unified CCE System***



The following notes apply to Figure 130:

- 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 any single point of failure is truly isolated between public and private networks. Additionally, public network WAN segments traversing a routed network must be deployed so that PG-to-CC (Central Controller) route diversity is maintained throughout the network. Be sure to avoid routes that result in common path selection (and, thus, a common point of failure) for the multiple PG-to-CC sessions (see Figure 130).

## IP-Based Prioritization and QoS

For each of the WAN links in Figure 130, 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 provide 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. For details, see Where to Mark Traffic.

**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 recommended. Microsoft Windows Packet Scheduler is not well supported or suited to Unified UCCE, and support will be removed in future versions. 802.1p markings are not widely used, nor should they be required when DSCP markings are available.

## 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, and each end looks for analogous heartbeats from the other. If either end misses 5 heartbeats in a row (that is, if a heartbeat is not received within a period that is 5 times the period between heartbeats), then the side detecting this condition assumes something is wrong and the application closes the socket connection. At that point, a TCP Reset message is typically generated from the closing side. Loss of heartbeats can be caused by various factors, such as: the network failed, the process sending the heartbeats failed, the machine on which the sending process resides is shut down, the UDP packets are not properly prioritized, and so forth.

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 2003 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 is 100 milliseconds between the duplexed sides, meaning that 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, meaning that the circuit failure threshold is 2 seconds in this case.

As part of the Unified CCE QoS implementation, the UDP heartbeat is replaced by a TCP keep-alive message in the public network connecting a Central Controller to a Peripheral Gateway. In Unified CCE 7.*x* and later releases, a consistent heartbeat or keep-alive mechanism is enforced for both the public and 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 in that case causes the server/client side to terminate. It operates by sending probe packets (namely, keep-alive packets) across a connection after the connection has been idle for a certain period, and the connection is considered down if a keep-alive response from the other side is not heard. Microsoft Windows 2003 and Windows 2008 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, meaning that a failure can be detected after 2 seconds, as was the case 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 have, in some cases, little correspondence to the TCP connections.
- The use of UDP heartbeats creates deployment complexities in a firewall environment. The dynamic port allocation for heartbeat communications makes it necessary to open a large range of port numbers, thus defeating the original purpose of the firewall device.

## 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 recommendations:

- Configure the HSRP hold time (plus the associated processing delay) to be lower than five times the heartbeat interval (100 ms on the private network and 400 ms on the public network) in order to avoid 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 failover times can exceed the threshold by which network outage detection is made, thus causing the enterprise system to complete a failure and recovery phase. If primary and secondary designations are made in the HSRP configuration and the primary path router fails to the secondary side, HSRP will subsequently reinstate the primary path when possible, thereby leading to a second private network outage detection.

For this reason, configured HSRP convergence delays that approach 500 ms for the private network and 2 seconds for the public network are best *not* configured with primary and secondary designations to avoid the start-path reinstatement mentioned above. On the other hand, convergence delays that can be configured below the detected threshold (which thus render an HSRP failover to be 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 the private network traffic to the public network link, or vice versa.
- The bandwidth requirement for Unified CCE must be guaranteed anytime with HSRP, otherwise the system behavior is unpredictable. For example, if HSRP is initially configured to do load sharing, ensure that sufficient bandwidth for Unified CCE remains on the surviving links in the worst-case failure situations.

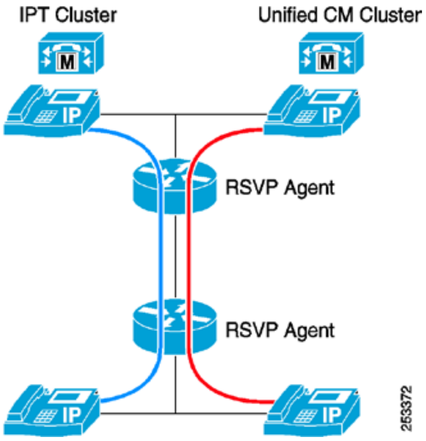
## RSVP

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 CM 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 CM, multiple phone calls can traverse the same RSVP controlled link even if the calls are controlled by multiple Unified CMs.

Figure 131 shows a scenario in which two different Unified CM clusters provide service to phones at the same remote site. This may occur if a Unified CM 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 CM to the network routers.

**Figure 131**



For more information about Unified CM RSVP, see the *Cisco Unified Communications SRND*.

## 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 and they can project bandwidth requirements far more accurately. See Bandwidth Requirements for Unified CCE Public and Private Networks 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 4 peripherals, each taking 10 calls per second, will generally be configured to have 320 kbps of bandwidth. The 1,000 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 in order 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 will be 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 in order 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 recommended 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 2008, 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 18, Table 19, and Table 20. 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 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.

## How to Mark Traffic

The default Unified CCE QoS markings are set in compliance with Cisco Unified Communications recommendations but can be overwritten if necessary. Table 18, Table 19, and Table 20 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 Unified Communications Solution Reference Network Design (SRND) Guide*.

**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 18** ***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<br>TCP port:<br>• 40003 + (i# * 40) for DMP high-priority connection on A<br>• 41003 + (i# * 40) for DMP high-priority connection on B<br>UDP port: 39500 to 39999 for UDP heartbeats if QoS is not enabled on Unified CCE | 200 ms | AF31 / 3 |

| Priority | Server-Side IP Address and Port | One-Way Latency Requirement | DSCP / 802.1p Marking |
|---|---|---|---|
| Medium | IP address: Router's high-priority public IP address<br>TCP port:<br>• 40017 + (i# * 40) for DMP high-priority connection on A<br>• 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<br>TCP port:<br>• 40002 + (i# * 40) for DMP low-priority connection on A<br>• 41002 + (i# * 40) for DMP low-priority connection on B | 5 seconds | AF11 / 1 |

**Table 19** ***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<br>TCP port: 41005 + (i# * 40) for MDS high-priority connection<br>UDP port: 39500 to 39999 for UDP heartbeats if QoS is not enabled on Unified CCE | 100 ms (50 ms preferred) | AF31 / 3 |
| Medium | IP address: Router's non-high-priority private IP address<br>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<br>TCP port:<br>• 41004 + (i# * 40) for MDS low-priority connection<br>• 41022 + (i# * 40) for CIC StateXfer connection<br>• 41021 + (i# * 40) for CLGR StateXfer connection<br>• 41023 + (i# * 40) for HLGR StateXfer connection<br>• 41020 + (i# * 40) for RTR StateXfer connection | 1000 ms | AF11/1 |

**Table 20** ***PG Private Network Traffic Markings (Default) and Latency Requirements***

| Priority | Server-Side IP Address and Port | One-Way Latency Requirement | DSCP / 802.1p Markin g |
|---|---|---|---|
| High | IP address: PG high-priority private IP address<br>TCP port:<br>• 43005 + (i# * 40) for MDS high-priority connection of PG no.1<br>• 45005 + (i# * 40) for MDS high-priority connection of PG no.2<br>UDP port: 39500 to 39999 for UDP heartbeats if QoS is not enabled on Unified CCE | 100 ms (50 ms preferred) | AF31/3 |
| Medium | IP address: PG's non-high-priority private IP address<br>TCP port:<br>• 43016 + (i# * 40) for MDS medium-priority connection of PG no.1<br>• 45016 + (i# * 40) for MDS medium-priority connection of PG no.2 | 1000 ms | AF11/1 |
| Low | IP address: PG's non-high-priority private IP address<br>TCP port:<br>• 43004 + (i# * 40) for MDS low-priority connection of PG no.1<br>• 45004 + (i# * 40) for MDS low-priority connection of PG no.2<br>• 3023 + (i# * 40) for OPC StateXfer of PG no.1<br>• 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.

## Configuring QoS on Unified CCE Router and PG

The QoS setup on the Unified CCE Router and PG is necessary only if the marking will be done in the Unified ICM and will be trusted by the network. For details, see the *Installation Guide for Cisco Unified /CCE Enterprise & Hosted Editions*.

## Configuring QoS 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_Queuing
```

```
    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 failover and the high-priority and low-priority private Unified CCE traffic respectively require 200 kbps and 100 kbps, the configuration would be:

```
policy-map Converged_Link_Queuing
    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 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.

Finally, the queuing policy is applied to the outgoing interface:

```
interface mod/port
    service-policy output ICM_Public_Queuing
```

### 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
```

## 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 for this purpose, see the *Administration Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*.

# 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 Cisco 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

Table 21 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 21** ***Worksheet for Calculating Private Network Bandwidth***

| **Component** | **Effective BHCA** | **Multiplication Factor** | **Recommended Link** | **Multiplication Factor** | **Recommended 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 Table 21. 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 would be 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, would be 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, would be 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, would be 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

Table 22 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 22** ***Example Calculation for a Combined Dedicated Private Link***

| Component | Effective BHCA | Multiplication Factor | Recommended Link | Multiplication Factor | Recommended 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 | |
| | | **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 = 2,550,000 bps
- Router/Logger high-priority bandwidth queue of 264,000 bps
- PG high-priority bandwidth queue 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 would be 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 Unified CCE Clustering Over the WAN

For details about Unified CCE clustering over the WAN, see IPT: Clustering Over the WAN.

Bandwidth must be guaranteed across the highly available (HA) WAN for all Unified CCE private, public, CTI, and Unified CM intra-cluster 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.

In addition to the bandwidth requirements for the private and public networks, this section adds 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 CM intra-cluster communication signaling (ICCS).

### Unified IP IVR or Unified CVP PG to Unified IP IVR or Unified CVP

At this time, 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 bandwidth needed for this communication. Bandwidth consumed between the Unified CCE Central Controller and Unified IP IVR or Unified CVP PG is very 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 login authentication) through the Cisco Steps to Success Portal.

If the Unified IP IVR or Unified CVP PGs are split across the WAN, total bandwidth required would be double what the tool reports: once for Unified CCE Central Controller to Unified IP IVR or Unified CVP PG and once for Unified IP IVR or Unified CVP 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. Use a bandwidth queue to guarantee availability for this worst case.

For this model, the following simple formula can be used to compute worst-case bandwidth requirements:

- With no Extended Call Context (ECC) or Call Variables:

BHCA * 20 = bps

- With ECC and/or Call Variables

BHCA * (20 + ((Number of Variables * Average Variable Length) / 40) = 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 bps = 400 kbps

### Unified CM Intra-Cluster Communication Signaling (ICCS)

The bandwidth required for Intra-Cluster Communication Signaling (ICCS) between Unified CM subscriber nodes is significantly higher when Unified CCE is deployed, due to the number of call redirects and additional CTI/JTAPI communications encompassed in the intra-cluster communications. The following formulae may be used to calculate the required bandwidth for the ICCS and database traffic between Unified CM subscriber nodes when they are deployed with Unified CCE:

- Unified CM releases prior to 6.1
  - Intra-Cluster Communications Signaling (ICCS)

    BHCA * 200 = bps

    This is the bandwidth required between each site where Unified CM subscribers are connected to voice gateways, agent phones, and agent PGs.
  - Database and other communications
  - 644 kbps for each subscriber remote from the publisher
- Unified CM 6.1 and later releases
  - Intra-Cluster Communications Signaling (ICCS)

    Total Bandwidth (Mbps) = ((Total BHCA) / 10,000) [2.25 + (0.006 Delay)],
    where Delay = Round-trip-time delay in msec

    This is the bandwidth required between each Unified CM subscriber that is connected to voice gateways, agent phones, and Agent PGs.
  - Database and other communications

    1.544 Mbps for each subscriber remote from the publisher

    The BHCA value to use for the ICCS formulae above is the total BHCA for all calls coming into the contact center.

These bandwidth requirements assume proper design and deployment based on the recommendations contained throughout this document. Inefficient design (for example, if ingress calls to Site 1 are treated in Site 2) will cause additional intra-cluster communications, possibly exceeding the defined bandwidth requirements.

## Bandwidth Requirements for Gateway PG to System PG

This section provides some basic guidelines for provisioning bandwidth for the connection between the gateway PG and the system PG.

## Bandwidth Requirements for Unified CCE Gateway PG to Central Controller

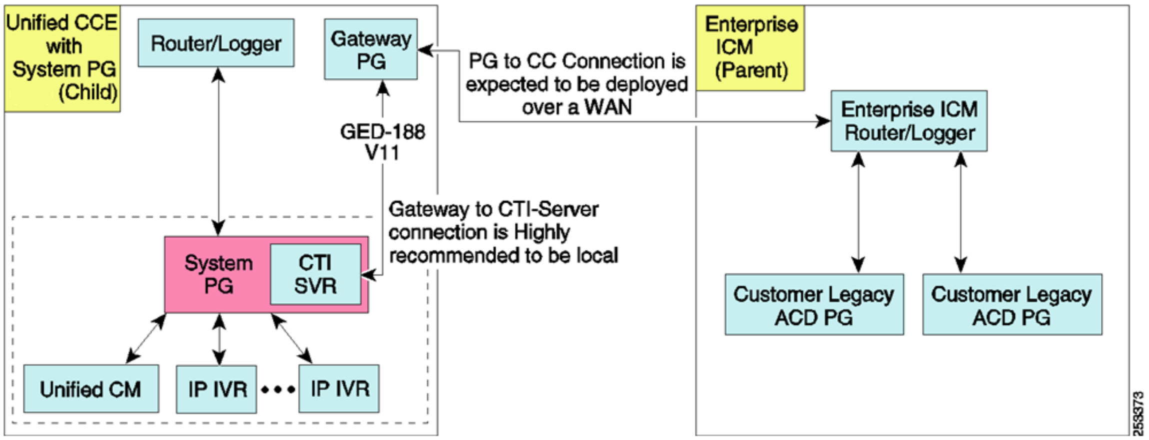
No special considerations are necessary for the PG-to-CC connection over other TDM PGs.

If agent reporting is not used, then uncheck the **Enable Agent Reporting** checkbox in the Agent distribution tab of the PG explorer to avoid sending unnecessary data over the link. For more information, see Bandwidth and Latency Requirements.

## Bandwidth Requirements for Unified CCE Gateway PG to System PG

Figure 132 illustrates the connection between the parent PG/PIM and the child system PG.

**Figure 132** ***Connection Between Gateway PG and System PG***


Unified CCE with System PG (Child)
Router/Logger
Gateway PG
PG to CC Connection is expected to be deployed over a WAN
Enterprise ICM (Parent)
Enterprise ICM Router/Logger
GED-188 V11
Gateway to CTI-Server connection is Highly recommended to be local
System PG
CTI SVR
Customer Legacy ACD PG
Customer Legacy ACD PG
Unified CM
IP IVR
IP IVR
253373

**Note** Do *not* deploy the gateway PG remote from the system PG that it is monitoring.

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). A Route Request with no data, for example, can be a very small message. If all call variables and ECC variables are populated with large values, this will drastically affect the size of the message.
- Call scenarios can cause great variation in the number of messages per call that are transmitted over the line. A simple call scenario might cause 21 messages to be transmitted over the line. More complex call scenarios involving queuing, hold retrieves, conferences, or transfers will add greatly to the number of messages per call that are transmitted over the line.
- The more skill groups to which an agent belongs, the more messages are transmitted 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 with no hold, retrieve, conference, or transfer) with a single skill group will typically generate 21 messages; plan for a minimum of approximately 2700 bytes of required bandwidth for it.

In a basic call flow, there are four places where call variables and ECC data can be sent. Thus, if you use call data and/or ECC variables, they will all be sent four times during the call flow. Using a lot of call data could easily increase (by double, triple or more) the 2700 bytes of estimated bandwidth per call.

**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 = IIIIIIIII (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.

### Basic Call Flow Example

Assume a call rate of 300 simple calls per minute (5 calls per second) and 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 = 13,500 bytes per second = 108 kbps of required bandwidth

Note that a more complex call flow or a call flow involving call data could easily increase this bandwidth requirement.

## Auto configuration

If auto configuration is used, it is possible that the entire agent, skill group, and route-point configuration could be transmitted from the child PG to the parent PG. If not much bandwidth is available, it could take considerable time for this data to be transmitted.

Table 22 lists the approximate number of bytes (worst case) that are transmitted 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 will be transmitted. Note that the values in Table 22 are worse-case estimates that assume transmitting only one item per record, with each field having the maximum possible size (which is extremely unlikely).

**Table 23** ***Bytes Transmitted per Data Item Under Worst-Case Conditions***

| Data Item Transmitted | Size |
|---|---|
| Agent | 500 bytes |
| Call type | 250 bytes |
| Skill group | 625 bytes |
| Device (route point, device target, 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 the amount of configuration data transmitted could be estimated as follows:

100 agents * 500 bytes = 50,000 bytes

10 call types * 250 bytes = 2,500 bytes

5 skill groups * 625 bytes = 3,125 bytes

20 route points * 315 bytes = 6,300 bytes

50,000 + 2,500 + 3,125 + 6,300 = 61,925 bytes

The total amount of data (approximate maximum) transmitted for this configuration is 61,925 bytes.

## Best Practices and 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. (See the Note under Bandwidth Calculation for Basic Call Flow.)

- Use the MAPVAR = IIIIIIIII and MAPECC = IIIIIIIII peripheral configuration parameters.

  If you do not use the MAPVAR and MAPECC option (which means that the settings default to MAPVAR = BBBBBBBBBB and MAPECC = BBBBBBBBBB), 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 an IVR 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 IP IVR deployments for transfer to an IVR campaign, and the IP IVR is separated from the Outbound Option Dialer servers by a WAN. Provide IP IVR with its own Unified CM cluster to reduce the WAN traffic.
- When deploying a SIP Dialer solution for transfer to an IVR campaign, and the Cisco Unified SIP Proxy servers for the SIP Dialers are separated from the Outbound Option Dialer servers by a WAN.
- 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 with H.323. 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 would be:

Hit Call Signaling Bandwidth = (12,500 bytes/call) (8 bits/byte) = 100,000 bits per call = 100 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 could be the result of a busy ring no-answer, an invalid number, and so forth. The average non-hit call signaling bandwidth usage would be:

Non-Hit Signaling Call Bandwidth = (6,200 bytes/call) (8 bits/byte) = 49,600 bits per call = 49.6 Kb per call

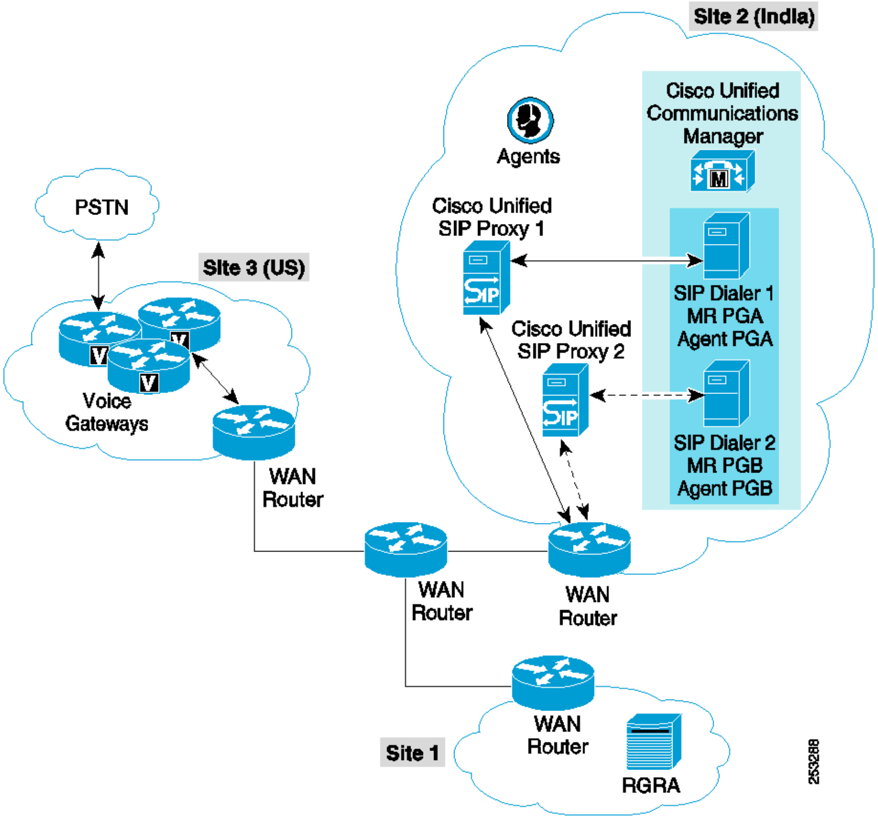
Codec Bandwidth = 80 Kbps per call for G.711 Codec,

or 26 Kbps per call for G.729 Codec

## Agent-Based Campaign – No SIP Dialer Recording

Figure 133 shows an example of the distributed Outbound SIP Dialer deployment for an agent-based campaign.

**Figure 133** ***Distributed Outbound SIP Dialer Deployment for an Agent-Based Campaign***



The average WAN bandwidth usage in this case is:

WAN Bandwidth = Calls Per Second *

(Hit Rate * (Codec Bandwidth + Hit Call Signaling Bandwidth)

+ (1 – Hit Rate) * Non-Hit Call Signaling Bandwidth)

= 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, the bandwidth usage is:

60 * (20% * (80 + 100) + (1 – 20%)*49.6) = 4540.8 kbps = 4.5 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, the bandwidth usage is:

60 * (20% * (26 + 100) + (1 – 20%)*49.6) = 3856.8 kbps = 3.9 Mbps

## Agent-Based Campaign – SIP Dialer Recording

The average WAN bandwidth usage in this case would be:

WAN Bandwidth = Calls Per Second *

(Codec Bandwidth

+ Hit Rate * Hit Call Signaling Bandwidth

+ (1 - Hit Rate) * Non-Hit Call Signaling Bandwidth)

= 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, the bandwidth usage is:

60 * (80 + 20% *100 + (1 – 20%)*49.6) = 8380.0 kbps = 8.4 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, the bandwidth usage is:

60 * (26 + 20% *100 + (1 – 20%)*49.6) = 5140.8 kbps = 5.1 Mbps

## Transfer-To-IVR Campaign – No SIP Dialer Recording

Figure 134 and Figure 135 show examples of the distributed Outbound SIP Dialer deployment for transfer to an IVR campaign.

**Figure 134** *Distributed Outbound SIP Dialer Deployment for Transfer to an IVR Campaign Using Cisco Unified CVP*

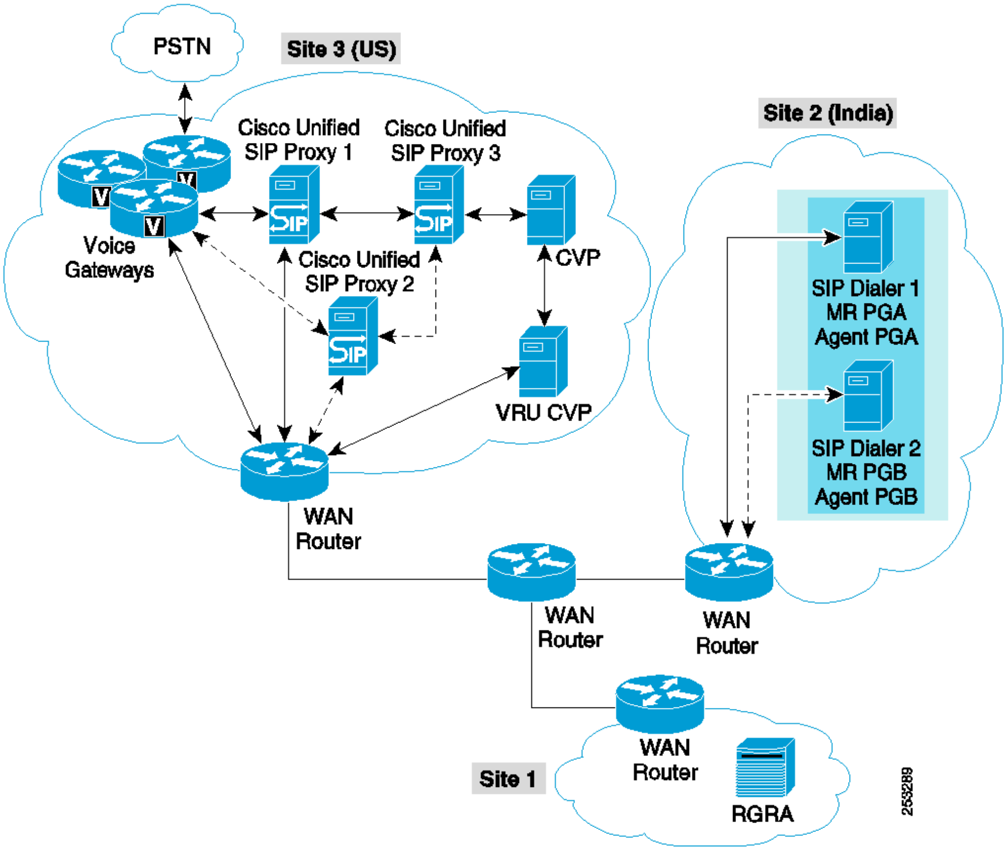
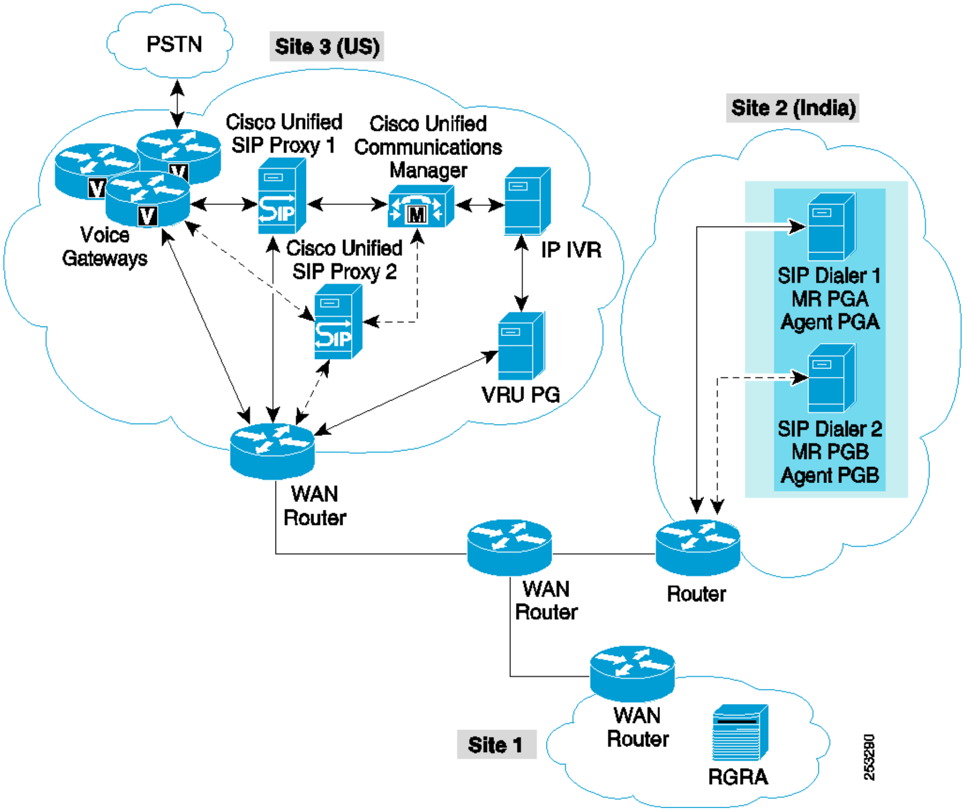


**Figure 135** ***Distributed Outbound SIP Dialer Deployment for Transfer to an IVR Campaign Using Cisco Unified IP IVR***



The average WAN bandwidth usage in this case would be:

WAN Bandwidth = Calls Per Second * Hit Rate * Hit Call Signaling Bandwidth + Calls Per Second * (1 - Hit Rate) * Non-Hit Call Signaling Bandwidth = 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 kbps = 3.6 Mbps

## Transfer-To-IVR Campaign – SIP Dialer Recording

The average WAN bandwidth usage in this case would be:

WAN Bandwidth = Calls Per Second * (Codec Bandwidth

+ Hit Rate * Hit Call Signaling Bandwidth

+ (1 - Hit Rate) * Non-Hit Call Signaling Bandwidth)

= 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, the bandwidth usage is:

60 * (80 + 20% *100 + (1 – 20%)*49.6) = 8380.0 kbps = 8.4 Mbps

**Example 7:** 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.729 codec, the bandwidth usage is:

60 * (26 + 20% *100 + (1 – 20%)*49.6) = 5140.8 kbps = 5.1 Mbps

## Distributed SCCP Dialer Deployment

Call control signaling uses H.323 over the WAN between the voice gateway and the Unified CM to which the SCCP dialers are connected. The typical H.323 outbound call flow uses an average of 4,000 bytes per call that is transferred to an outbound agent. The average hit call signaling bandwidth usage would be:

Hit Call Signaling Bandwidth = (4,000 bytes/call) (8 bits/byte) = 32,000 bits per call = 32 Kb per call

The typical H.323 outbound call flow uses about 6,200 bytes per call that is disconnected by the outbound dialer. Those outbound calls could be the result of a busy ring no-answer, an invalid number, and so forth. The average non-hit call signaling bandwidth usage would be:

Non-Hit Signaling Call Bandwidth = (2,000 bytes/call) (8 bits/byte) = 8,000 bits per call = 8 Kb per call

Codec Bandwidth = 80 Kbps per call for G.711 Codec,

or 26 Kbps per call for G.729 Codec

Figure 136 shows an example of the distributed Outbound SCCP Dialer deployment for an agent-based campaign.

**Figure 136** ***Distributed Outbound SCCP Dialer Deployment for an Agent-Based Campaign***

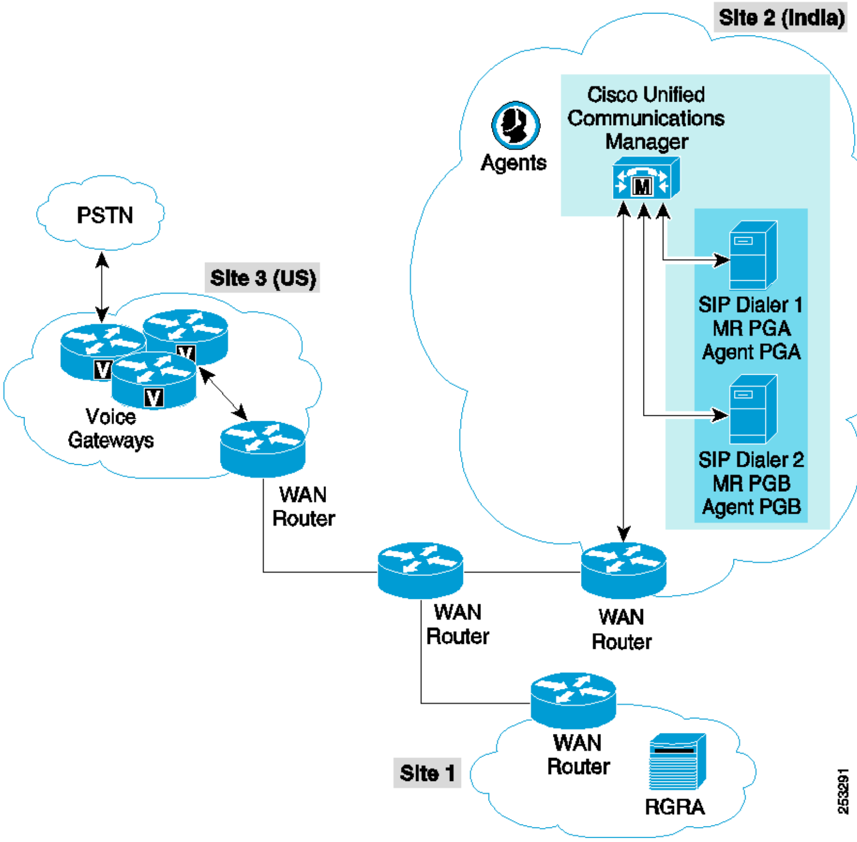


Figure 137 and Figure 138 show examples of the distributed Outbound SCCP Dialer deployment for transfer to an IVR campaign.

**Figure 137** *Distributed Outbound SCCP Dialer Deployment for Transfer to an IVR Campaign Using Cisco Unified CVP*

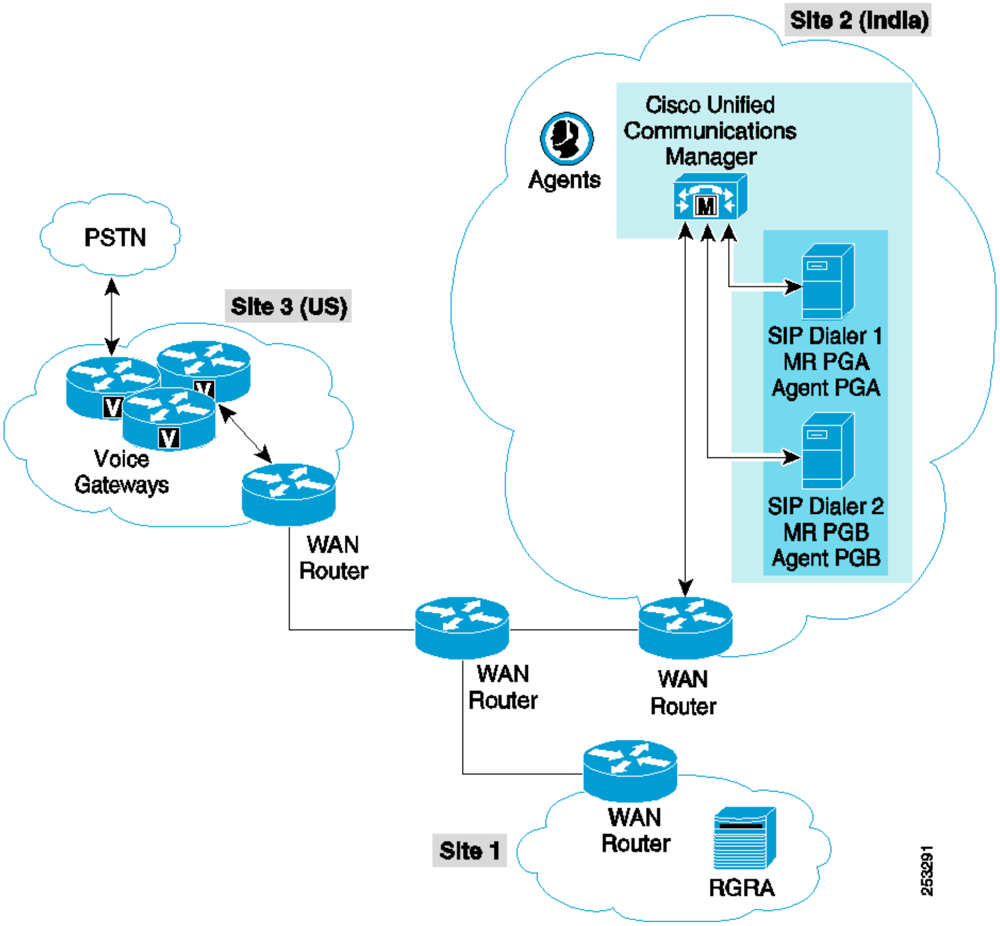
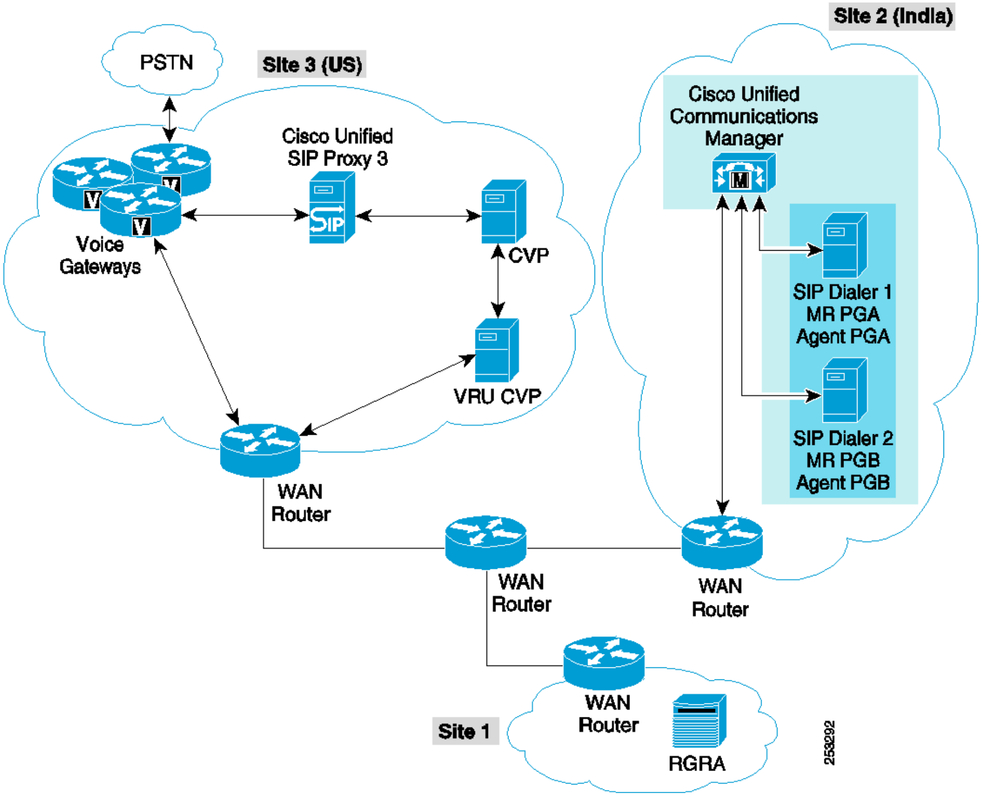


**Figure 138** *Distributed Outbound SCCP Dialer Deployment for Transfer to an IVR Campaign Using Cisco Unified IP IVR*



The average WAN bandwidth usage in this case would be:

WAN Bandwidth = Calls Per Second * Number of SCCP Dialers *

(Codec Bandwidth

+ Hit Rate * Hit Call Signaling Bandwidth

+ (1 - Hit Rate) * Non-Hit Call Signaling Bandwidth)

= Kbps

**Example 8:** For two SCCP Dialers with call throttling of 5 cps, a 20% hit rate for the agent campaign, and a WAN link with G.711 codec, the bandwidth usage is:

5*2 * (80 + 20% *100 + (1 – 20%)*49.6) = 1396.8 kbps = 1.4 Mbps

**Example 9:** With call throttling of 60 cps on the SCCP Dialer, a 20% hit rate for the transfer-to-IVR campaign, and a WAN link with G.729 codec, the bandwidth usage is:

5 * 2 * (26 + 20% *100 + (1 – 20%)*49.6) = 856.8 kbps = 0.86 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 will be 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.

Table 23 shows the type of messaging of each CTI OS application.

**Table 24** ***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.

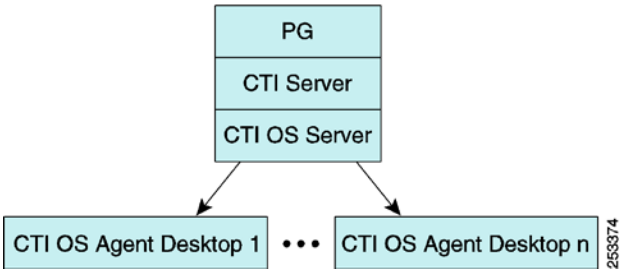
Silent Monitoring of an agent consumes roughly the same network bandwidth as an additional voice call. If a single agent requires bandwidth for one voice call, then the same agent being silently monitored would require bandwidth for two concurrent voice calls.

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 that examines the CTI OS Server-to-CTI OS Desktop bandwidth, as illustrated in Figure 139. It calculates Total bandwidth, agent bandwidth, and supervisor bandwidth requirements with CTI OS Security turned on or off.

**Figure 139** ***CTI OS Server-to-CTI OS Desktop Communication***



## Best Practices and Options 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 will decrease the traffic sent to the agents. The statistics cannot be specified on a per-agent basis at this time. For more information about agent statistics, see the *CTI OS System Manager's Guide for Cisco Unified ICM/Contact Center Enterprise & Hosted*.

### 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 would have no statistics traffic at all 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.

A remote supervisor or selected agents might still be able to log statistics by using a different connection profile with statistics enabled, if more limited statistics traffic is acceptable for the remote site.

In the case where Unified Mobile Agents have their skill group statistics turned off but the supervisor would like 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 would be 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. If we assume 10 agents at the 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 his agent team belongs), then this approach would reduce skill-group statistics traffic by 90% if only the supervisor has statistics turned on to observe the two skill groups but agents have statistics turned off.

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 would be required.

In the case where there are multiple remote locations, assuming only supervisors need to see the statistics, it would be 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 Cisco Finesse Desktop

This section addresses the traffic and bandwidth requirements between the Finesse Agent Desktop and the Finesse server. These requirements are important in provisioning the network bandwidth and QoS required between the agents and the Finesse 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.

## Finesse Client/Server Traffic Flows and Bandwidth Requirements

The network bandwidth requirements increase linearly as a function of the number of active agents signed in to the Finesse server. The traffic between the Finesse client and server can be categorized as follows:

- HTTP request/response traffic to load the Finesse desktop application
- HTTP request/response traffic for REST requests to the Finesse server
- HTTP request/response traffic for XMPP notifications via the BOSH protocol.

Table 25 shows the type of messaging of each Finesse user type.

**Table 25** ***Messaging Type By Finesse User Type***

| Application Name | Message Types |
|---|---|
| Finesse Agent Desktop | Agent state changes |
| | Call control |
| | Call status information |
| Finesse Supervisor | Agent state changes |
| | Call control |

| Application Name | Message Types |
|---|---|
| | Call status information |
| | Monitoring agent states |
| | Silent monitoring |

### Finesse Desktop Bandwidth Usage

The bandwidth requirement for Cisco Finesse is a linear relationship based on the following:

- The number of agents active on the Finesse server
- The number of configured call and ECC variables

Table 26 shows the average bandwidth requirements on a per agent basis. This information is derived from bandwidth testing and extrapolation of bandwidth data. Because there are many variables that can affect bandwidth, a configuration that results 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 Table 26, the Finesse Desktop can run without delays in message passing.

The most expensive operation from a network perspective is the agent/supervisor login. This operation involves the web page load and includes the CTI login and display of the initial agent state. After the desktop web page loads, the required bandwidth is significantly smaller. The data presented is based on the load generated by the default Finesse desktop layout. Additional custom gadgets increase the network bandwidth requirement based on the content contained in the custom gadget.

The login bandwidth is split between the following scenarios:

- No web browsing caching involved: This scenario typically only occurs on the first Finesse login from a specific client PC and is presented to provide the worst case scenario. This value is the minimum value to ensure that  2000 users can log in over a 5-minute period.
- Login with browser caching:  This scenario is the typical case for a user login.  This value is the minimum value to ensure that 2000 users can log in over a 5-minute period.
- Cached login with full user login completed within 10 seconds.

The bandwidth of the RTP streams for the call, recording, or monitoring sessions from the user’s phone device is not included.

**Table 26** ***Average Bandwidth Requirements for Cisco Finesse Agent***

| Activity | Average Download Bandwidth (Kilobytes per second) | Average Upload Bandwidth (Kilobytes per second) |
|---|---|---|
| Agent Login (no cache) | 66.0 | 2.4 |
| Agent login (cached) | 24.6 | 1.6 |

| Activity | Average Download Bandwidth (Kilobytes per second) | Average Upload Bandwidth (Kilobytes per second) |
|---|---|---|
| Agent login (cached) 10-second duration | 720 | 52.8 |

### Finesse Supervisor Bandwidth Usage

Table 27 shows the average bandwidth requirements on a per supervisor basis. The additional bandwidth is included to account for the supervisor monitoring of agent state changes.

**Table 27** *Average Bandwidth Requirements for Cisco Finesse Supervisor*

| Activity | Average Download Bandwidth (Kilobytes per second) | Average Upload Bandwidth (Kilobytes per second) |
|---|---|---|
| Supervisor login<br>(no cache) | 91.1 | 2.8 |
| Supervisor login (cached) | 36.5 | 2.0 |
| Supervisor login (cached) 10-second duration | 1120 | 58.4 |

## 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 CAD 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 CAD 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 CAD software:

- Cisco Supervisor Desktop
- Recording and Playback service

Cisco Supervisor Desktops will send silent monitoring requests when the supervisor wishes to listen to an agent’s call in real-time or listen to a call that was recorded earlier. The Recording and Playback service will send recording requests when a supervisor or agent wishes 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 CAD 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. It is responsible for processing silent monitoring requests only for the agent logged into the CAD application on the desktop. It captures voice packets sent to the phone or IP Communicator software phone associated with the logged-in agent. The phone must be a Cisco Unified IP Phone 7910, 7940, 7960, or 7970 connected in series with the agent desktop on the network. These phones are supported because they contain an additional 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.

By default, this service is active on all agent desktops when the application is started. After initial installation of the CAD 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** Cisco Unified IP Phone Agents who do not have a desktop must be configured to use a VoIP Monitor service for the silent monitoring feature.

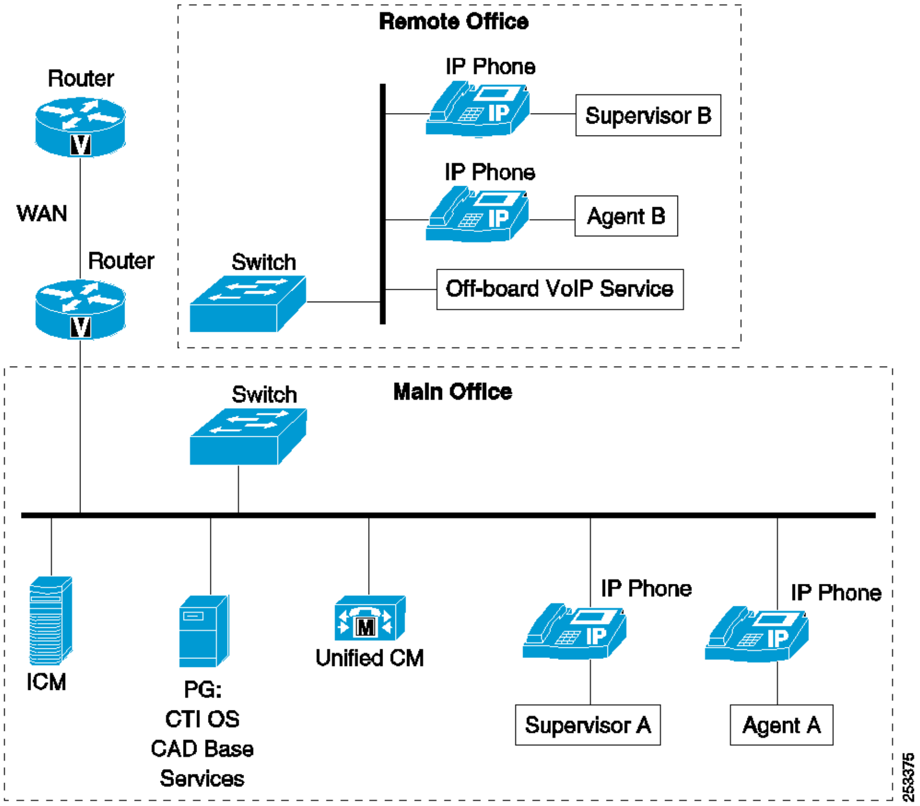
The Recording and Playback service may also provide the two streams representing a phone call when a supervisor plays back a recorded agent call. In this case, the streams have already been stored on disk from an earlier recording session. The Recording and Playback service reads the raw data files from 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 CAD base services. Additional VoIP services and a second Recording and Playback service may be installed on other boxes.

Figure 140 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 140** ***VoIP Monitor Service at Main and Remote Sites***



When you locate the requestors and providers, you can determine where the bandwidth will be required for the silent monitoring feature. The following notes regarding bandwidth apply:

- Although an administrator is able to 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 in order 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 CAD 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 will have one stream going to the phone and one stream coming from the phone, the VoIP provider will have 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 Unified Communications Solution Reference Network Design (SRND) Guide*.

## 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 CAD 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 1,000 CAD 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 base CAD services running on server machines. In addition, the agent desktop application communicates with the CTI server through the CTI OS server for call control actions and state changes. Table 28 lists the types of messages for each application.

**Table 28** ***Messaging Type By CAD Application***

| **Application Name** | **Message Types** |
|---|---|
| Cisco Agent Desktop | Login/logoff |
| | Agent state changes |
| | Call Control |

| Application Name | Message Types |
|---|---|
| | Call status information |
| | Desktop monitoring and recording |
| | Chat messages |
| | Team performance messages |
| | Report generation |
| | Real-time data refresh |
| CTI OS Supervisor Desktop | Login/logoff |
| | 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 |

| Application Name | Message Types |
|---|---|
| | SNMP Get messages |

### Cisco Agent Desktop Bandwidth Usage

CAD agents are able to log in and log 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.

Table 29 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 will be able to run without delays in message passing.

The following configuration parameters affect bandwidth and apply to both Table 29 and Table 30:

- 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 that 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 29** ***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 |

| Number of Agents | Average Download Bandwidth (Kilobytes per second) | Average Upload Bandwidth (Kilobytes per second) |
|---|---|---|
| 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/logoff, 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.

Table 30 uses the same basic configuration parameters used to determine the bandwidth numbers in Table 29. In addition, Table 30 takes into account the fact that the Team Skill Statistics report is being viewed and refreshed.

**Table 30** *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 |
| 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 |

| Number of Agents | Average Download Bandwidth (Kilobytes per second) | Average Upload Bandwidth (Kilobytes per second) |
|---|---|---|
| 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.

## Best Practices and Recommendations for 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 servers (off-board).

### VoIP Monitor Server

A single VoIP Monitor Service supports up to 114 simultaneous silent monitoring sessions. Additional 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 may exist on a single server.

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 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, you must move that service off-board to another server. A single VoIP Monitor Service supports the network traffic of 400 agent phones if you use a 100 Megabit NIC to connect to the switch. A single VoIP Monitor Service supports the network traffic of 1000 agent phones if you use 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 you must use a dedicated 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 server. 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 a call's RTP packets to RAW files and stores these files for playback 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.

## Bandwidth Requirements for an Administration & Data Server and Reporting

For more information about the bandwidth requirement for Cisco Unified Intelligence Center, see the latest version of the *Cisco Unified Intelligence Center Bill of Materials*.

## Bandwidth Requirements for Cisco EIM/WIM

The bandwidth requirement for Cisco EIM/WIM integrations are documented in EIM/WIM SRND document.

## 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 Table 31. 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 31** ***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 12

# Cisco Unified Contact Center Management Portal

## Cisco 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 multitenant provisioning platform that overlays the Cisco Unified Contact Center Enterprise (Unified CCE), Cisco Unified Communications Manager (Unified CM), 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 folder(s).

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 7 or Internet Explorer 8 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, or ConAPI) 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 Release 8.0 (*x*) 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 Unified CCE physical server 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 server. 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[1].
- Dual Server. In this mode, the front-end Unified CCMP components are installed on one server (the Web Server) and back-end components on another (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[1].

**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 CM 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[2].

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 & 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

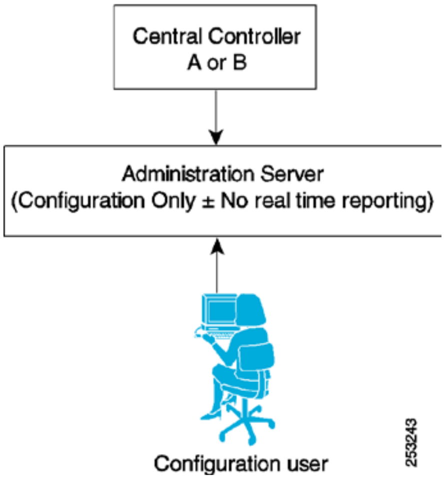
The Administration & Data Servers are classified into roles based on the system configuration and the call load that they can handle. Beginning with Unified CCE 8.0, there is a new type (role) of Administration & Data Server. This new role is known as a Configuration-Only Administration Server. This new role was designed specifically for use with Unified CCMP in a hosted deployment when a lightweight Administration & Data Server running on Virtual Machines is desirable. (See the Deployment Models chapter for details about the other Administration & Data Server roles.)

## 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. (See Figure 141)

This deployment role allows Contact Center Hosted using CCMP to configure a specific Unified CCE Customer Instance through the ConAPI interface. For historical and real-time data, the customer can use a shared AW-HDS-DDS.

**Figure 141** ***Configuration-Only Administration Server***



### Systems Exceeding Published limits

If your requirements exceed the capacity limits outlined in this document and detailed in the *Hardware and System Software Specification (Bill of Materials) for Cisco Unified ICM / Contact Center Enterprise & Hosted*, 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 row labeled *Provisioning Operations per Hour* in the *Configuration Limits and Scalability Constraints, Unified ICM, Unified CC* table.

For details on additional resource capacity criteria and required networking connectivity within geographically dispersed deployments, see the *Hardware & System Software Specification (Bill of Materials) for Cisco Unified ICM/Contact Center Enterprise & Hosted, Release 8.5(x)*.

## Software Compatibility

Unified CCMP 8.0(*x*) is backward compatible with Unified CCE versions starting with Unified CCE 7.1. Always install the latest version of Unified CCMP to get the latest feature set.

## 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 one of the following two product solution options:

- Cisco Unified Intelligence Center (Unified IC) — The next-generation advanced reporting platform available with Unified CCE 8.0.
- Exony VIM Analytics — The OLAP-based operational analysis, performance management, and data mining platform available under the Cisco Solution+ reseller agreement. This platform shares the Unified CCMP user, roles, and folder hierarchies.

## 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.

# Glossary

## 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 |
| ConAPI | Configuration Application Programming Interface |
| 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

| | |
|---|---|
| DCA | Dynamic Content Adapter |
| 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 | Extended 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 |
| IPCC | Cisco IP Contact Center |
| 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 |
| LLA | Longest available agent |
| LSPAN | Local switched port analyzer |

## M

| | |
|---|---|
| MAC | Media access control |
| Mbps | Megabits per second |
| MC | Management center |
| MCS | Media Convergence Server |
| 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 |
| Progger | Peripheral gateway, router, and logger |
| 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 |
| Rogger | 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

| | |
|---|---|
| S1, S2, S3, and S4 | Severity levels for service requests |
| SAA | Service assurance agent |
| SCCE | System Contact Center Enterprise |
| 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

| | |
|---|---|
| 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 |