

Cisco IP Contact Center Solution Reference Network Design Guide

November 2002

Corporate Headquarters

Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
<http://www.cisco.com>
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

Customer Order Number: 956527

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

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

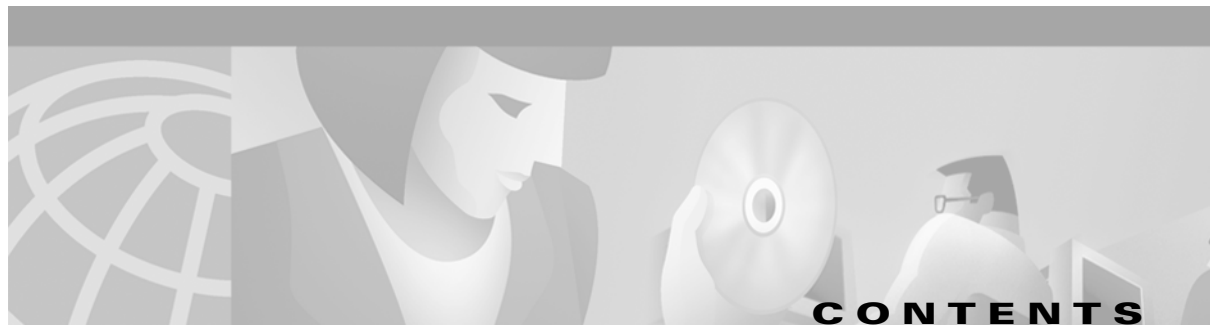
The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

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

CIP, the Cisco Arrow logo, the Cisco *Powered* Network mark, the Cisco Systems Verified logo, Cisco Unity, Follow Me Browsing, FormShare, iQ Breakthrough, iQ Expertise, iQ FastTrack, the iQ Logo, iQ Net Readiness Scorecard, Networking Academy, ScriptShare, SMARTnet, TransPath, and Voice LAN are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, Discover All That's Possible, The Fastest Way to Increase Your Internet Quotient, and iQuick Study are service marks of Cisco Systems, Inc.; and Aironet, ASIST, BPX, Catalyst, CCDA, CCDP, CCIE, CCNA, CCNP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, the Cisco IOS logo, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Empowering the Internet Generation, Enterprise/Solver, EtherChannel, EtherSwitch, Fast Step, IgaStack, Internet Quotient, IOS, IP/TV, LightStream, MGX, MICA, the Networkers logo, Network Registrar, *Packet*, PIX, Post-Routing, Pre-Routing, RateMUX, Registrar, SlideCast, StrataView Plus, Stratum, SwitchProbe, TeleRouter, and VCO are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

If other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0208R)



Preface ix

Purpose	ix
Scope	ix
Recommended Software Releases	ix
Audience	x
Organization	x
Obtaining Documentation	xi
World Wide Web	xi
Documentation CD-ROM	xi
Ordering Documentation	xi
Documentation Feedback	xii
Obtaining Technical Assistance	xii
Cisco.com	xii
Technical Assistance Center	xiii
Cisco TAC Web Site	xiii
Cisco TAC Escalation Center	xiii

CHAPTER 1

IPCC Architectural Overview 1-1

Cisco CallManager	1-1
Cisco IP Interactive Voice Response (IP IVR)	1-2
Cisco Intelligent Contact Management (ICM) Software	1-3
Basic IPCC Call and Message Flow	1-3
ICM Software Modules	1-4
IPCC Components, Terminology, and Concepts	1-6
IPCC Agent Desktop	1-6
Administrative Workstation	1-6
JTAPI Communications	1-7
ICM Routing Clients	1-10
Device Targets	1-10
Labels	1-10
Agent Desk Settings	1-11
Agents	1-11
Skill Groups	1-11

Directory (Dialed) Numbers and Routing Scripts	1-12
Agent Login and State Control	1-12
IPCC Routing	1-13
Translation Routing and Queuing	1-13
Reroute On No Answer (RONA)	1-15
Hybrid IP Telephony and IPCC Cisco CallManager Clusters	1-15
Queuing in an IPCC Environment	1-16
Transfers in an IPCC Environment	1-16
Dialed Number Plan	1-17
Route Request	1-18
Single-Step (Blind) Transfer	1-18
Consultative Transfer	1-19
Reconnect	1-19
Alternate	1-20
Non-ICM Transfers	1-20
Agent-to-Agent Transfers	1-20
Transferring from an IVR to a Specific Agent	1-21
Transfer Reporting	1-21
Combination or Multiple Transfers	1-21
Transfers of Conferenced Calls	1-21
PSTN Transfers (Takeback N Transfer, or Transfer Connect)	1-21
Call Admission Control	1-22
Gatekeeper Controlled	1-23
Locations Controlled	1-24

CHAPTER 2

Deployment Models 2-1

Single Site	2-2
Queuing	2-3
Transfers	2-3
Multi-Site with Centralized Call Processing	2-4
Centralized Voice Gateways	2-5
Distributed Voice Gateways	2-7
Multi-Site with Distributed Call Processing	2-9
Distributed Voice Gateways	2-10
Centralized Voice Gateways	2-12
Traditional ACD Integration	2-13

Traditional IVR Integration	2-14
Using PBX Transfer	2-15
Using PSTN Transfer	2-16
Using IVR Double Trunking	2-17
Using Cisco CallManager Transfer and IVR Double Trunking	2-17
Hybrid IP Telephony and IPCC System	2-18

CHAPTER 3

Voice Gateway Considerations for IPCC 3-1

Recommended Voice Gateways	3-2
Gateways That Are Not Recommended	3-2
Selection Criteria	3-3
Port Density	3-3
Media Resources	3-4
Transcoding	3-4
A-Law to Mu-Law Conversion	3-4
VoIP Interfaces	3-5
TDM Interfaces	3-5
PRI Interfaces	3-5
QSIG Interface	3-6
SS7	3-6
BRI	3-7
Features	3-7
Calling and Called Number Support	3-7
Calling Party Name	3-8
ANI/DNIS Delimiter	3-8
Presentation Indicator	3-9
ISDN Non-Facility Associated Signaling (NFAS)	3-9
Supplementary Services	3-9
QoS	3-10
Ringback on Transfer	3-10
DTMF Relay	3-10
Hookflash Transfer	3-10
Tone on Hold	3-11
Music on Hold	3-11
Cisco CallManager Redundancy	3-12
Call Preservation	3-12
ICM Reporting	3-13
Network Management	3-13

CHAPTER 4

Design Considerations for High Availability 4-1

Terminology and Conventions	4-1
Designing for High Availability	4-2
Data Network Design Considerations	4-5
Cisco CallManager and CTI Manager Design Considerations	4-7
Configuring ICM for CTI Manager Redundancy	4-10
IP IVR (CRS) Design Considerations	4-10
IP IVR (CRS) High Availability Using Cisco CallManager	4-12
IP IVR (CRS) High Availability Using ICM	4-12
Peripheral Gateway Design Considerations	4-12
Cisco CallManager Failure Scenarios	4-14
ICM Failover Scenarios	4-14
Scenario 1 - Cisco CallManager and CTI Manager Fail	4-15
Scenario 2 - Cisco CallManager PG Side A Fails	4-16
Scenario 3 - Only Cisco CallManager Fails	4-16
Scenario 4 - Only CTI Manager Fails	4-17
Understanding Failure Recovery	4-19
Cisco CallManager Service	4-19
IP IVR (CRS)	4-19
ICM	4-20
Cisco CallManager PG and CTI Manager Service	4-20
ICM Voice Response Unit PG	4-20
ICM Router and Logger	4-21
Real-Time Distributor	4-21
CTI Server	4-23
CTI OS Considerations	4-24
Other Considerations	4-24
Reporting	4-24

CHAPTER 5

Sizing IP Contact Center Resources 5-1

Sizing Call Center Resources	5-1
Design Tools - Erlang Calculators	5-3
Erlang-C	5-4
Erlang-B	5-4
Sizing Call Center Resources (Basic Example)	5-5
Sizing Agents	5-5

Sizing PSTN Trunks (Gateway Ports)	5-6
Sizing Trunks for Busy Hour Traffic	5-7
Sizing Trunks for Queued Calls	5-7
Calculating Total PSTN Trunks Required	5-8
Sizing IP IVR Ports	5-8
Sizing Call Center Resources (Front-End IP IVR for Call Treatment Example)	5-9
Sizing Additional PSTN Trunks for Initial IP IVR Call Treatment	5-9
Sizing Additional IP IVR Ports for Call Treatment	5-10
Sizing Call Center Resources (Agent Wrap-up Time Example)	5-11
Call Center Design Considerations	5-12

CHAPTER 6
Sizing IPCC Components and Servers 6-1

Sizing Considerations for IPCC	6-1
Core IPCC Components	6-1
Additional Sizing Variables	6-4
Peripheral Gateway Sizing Recommendations	6-7
Other ICM Applications	6-8
CTI Components	6-9
CTI OS	6-9
Cisco Agent Desktop and Cisco Supervisor Desktop	6-9
Summary	6-13

CHAPTER 7
IPCC Agent Desktop and Supervisor Desktop 7-1

Types of IPCC Agent Desktops	7-2
CTI Object Server (CTI OS) Toolkit	7-3
Cisco Agent Desktop and Cisco Supervisor Desktop	7-4
Agent Desktop	7-6
Cisco Supervisor Desktop	7-6
Cisco Supervisor Desktop Considerations and Guidelines	7-6
Required Switched Port Analyzer (SPAN) Port Configuration	7-7
Cisco CallManager Interfaces	7-8
Intelligent Contact Management (ICM) Interfaces	7-8
Packet Sniffing and Network Configuration	7-8
Catalyst Switch Capabilities	7-9
Verified Network Configurations	7-10

CHAPTER 8

Bandwidth Provisioning and QoS Considerations 8-1

IPCC Network Components Overview	8-2
Single-Site Characteristics	8-3
Multi-Site Characteristics	8-4
Network Segmentation	8-5
Private Network	8-5
Visible (Public) Network	8-7
Signaling Access Network (PSTN Interfaces)	8-8
IPCC Network Bandwidth and QoS Overview	8-8
Traffic Flows from PG to ICM Central Controller	8-8
Latency and QoS Requirements for PG to ICM Central Controller	8-10
QoS Classification Implementation for PG to ICM Central Controller	8-11
Traffic Flows from ICM to ICM Central Controller	8-13
Latency and QoS Requirements for ICM to ICM Central Controller	8-14
Bandwidth and QoS Requirements for CTI Server to IPCC Desktop	8-15
Administrative Workstation (AW) Traffic Flows	8-15
Network Bandwidth Provisioning	8-16
Bandwidth Sizing Examples for CTI Server to Agent Desktop	8-17
Cisco Agent Desktop Call Control Bandwidth Usage	8-18
Cisco Agent Desktop Bandwidth for an Inbound Call	8-21
Bandwidth for Silent Monitor to Supervisor	8-22
Bandwidth for VoIP Monitor to RASCAL	8-22
Bandwidth for Cisco Supervisor Desktop to Servers	8-23
Module Interactions	8-24
Server Placement Recommendations	8-25
QoS Considerations	8-25
Cisco Agent Desktop Software Component Port Usage	8-26
Bandwidth Sizing Examples for PG to ICM Central Controller	8-26
ICM Inter-Server Network Traffic Volume	8-27

INDEX



Preface

This preface describes the purpose, scope, intended audience, and general organization of this *Cisco IP Contact Center Solution Reference Network Design Guide*. It also provides information on how to order documentation from Cisco Systems.

Purpose

This document provides guidelines, recommendations, and best practices to help you design an IP Contact Center (IPCC) solution for your enterprise using the Cisco Architecture for Voice, Video, and Integrated Data (AVVID). The IPCC solutions and design information presented in this document focus on high availability of contact center functions and features, increased agent productivity, higher customer satisfaction, and reduced costs.

Scope

This document describes the products and features used to build a Cisco IP Contact Center (IPCC) system, and it gives recommendations on how to combine those elements into an effective solution for your enterprise. However, this document does not contain specific implementation or configuration details for the products and features. For details about a particular product or feature, refer to the technical documentation available online at Cisco.com. (See [Obtaining Documentation](#), page xi.)



Note

Unless stated otherwise, the solution designs presented in this document require the minimum software releases listed in [Table 1](#), and the information presented here applies only to those releases.

Recommended Software Releases

[Table 1](#) lists the recommended software releases for the components (described later in this document) used in an IPCC solution.

For the latest software compatibility information, refer to the *Cisco CallManager Compatibility Matrix*, available at

http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/ccmcomp.htm

Table 1 Recommended Software Releases for IPCC Components

IPCC Component	Software Release
Intelligent Contact Manager (ICM)	ICM 4.6.2
Computer Telephony Integration Object Server (CTI OS)	4.6.2
Cisco Agent Desktop and Cisco Supervisor Desktop	4.2.1 and 4.4
Cisco CallManager	3.2(2c)
Cisco Customer Response Solution (CRS) — IP Interactive Voice Response (IP IVR) or Queue Manager	3.0(1)
Operating System	Windows 2000 Server SP 2

Audience


This document is intended for Cisco customers, partners, and systems engineers (SEs) who will be designing and implementing an IPCC solution in the enterprise environment.

Organization

This guide contains the chapters and information listed in the following table.


Note

Cisco strongly recommends that you carefully read chapters 1 and 2 before attempting to design an IPCC solution and before reading any other sections of this guide.

Chapter	Title	Description
1	IPCC Architectural Overview	Presents an overview of the Cisco IPCC architecture as well as basic concepts related to IPCC solutions.
2	Deployment Models	Describes the primary models used to deploy an IPCC solution and explains when to use each model. <div>  Note This guide makes frequent references to these deployment models. Cisco recommends that you read this chapter carefully and understand the main characteristics of each model. </div>
3	Voice Gateway Considerations for IPCC	Gives guidelines and recommendations for selecting the appropriate voice gateways for your IPCC deployment.
4	Design Considerations for High Availability	Presents factors to consider when designing your IPCC solution to provide high availability of system functions and features in a variety of possible failure scenarios.
5	Sizing IP Contact Center Resources	Explains how to size (provision or calculate) the major resources needed for your IPCC deployment. This chapter also includes extensive examples on calculating the required number of agents, gateway ports (PSTN trunks), and IP IVR ports for an IPCC deployment.

Chapter	Title	Description
6	Sizing IPCC Components and Servers	Provides guidelines and recommendations for sizing (provisioning) IPCC hardware and software components.
7	IPCC Agent Desktop and Supervisor Desktop	Presents several agent desktop and supervisor desktop options, along with guidelines for deploying each option.
8	Bandwidth Provisioning and QoS Considerations	Describes the basic network traffic flows between IPCC components and gives guidelines for provisioning the appropriate bandwidth and QoS features for those traffic flows.

Obtaining Documentation

The following sections explain how to obtain documentation from Cisco Systems.

World Wide Web

You can access the most current Cisco documentation on the World Wide Web at the following URL:

<http://www.cisco.com>

Translated documentation is available at the following URL:

http://www.cisco.com/public/countries_languages.shtml

Documentation CD-ROM

Cisco documentation and additional literature are available in a Cisco Documentation CD-ROM package, which is shipped with your product. The Documentation CD-ROM is updated monthly and may be more current than printed documentation. The CD-ROM package is available as a single unit or through an annual subscription.

Ordering Documentation

Cisco documentation is available in the following ways:

- Registered Cisco Direct Customers can order Cisco product documentation from the Networking Products MarketPlace:
http://www.cisco.com/cgi-bin/order/order_root.pl
- Registered Cisco.com users can order the Documentation CD-ROM through the online Subscription Store:
<http://www.cisco.com/go/subscription>
- Nonregistered Cisco.com users can order documentation through a local account representative by calling Cisco corporate headquarters (California, USA) at 408 526-7208 or, elsewhere in North America, by calling 800 553-NETS (6387).

Documentation Feedback

If you are reading Cisco product documentation on Cisco.com, you can submit technical comments electronically. Click **Leave Feedback** at the bottom of the Cisco Documentation home page. After you complete the form, print it out and fax it to Cisco at 408 527-0730.

You can e-mail your comments to bug-doc@cisco.com.

To submit your comments by mail, use the response card behind the front cover of your document, or write to the following address:

Cisco Systems
Attn: Document Resource Connection
170 West Tasman Drive
San Jose, CA 95134-9883

We appreciate your comments.

Obtaining Technical Assistance

Cisco provides Cisco.com as a starting point for all technical assistance. Customers and partners can obtain documentation, troubleshooting tips, and sample configurations from online tools by using the Cisco Technical Assistance Center (TAC) Web Site. Cisco.com registered users have complete access to the technical support resources on the Cisco TAC Web Site.

Cisco.com

Cisco.com is the foundation of a suite of interactive, networked services that provides immediate, open access to Cisco information, networking solutions, services, programs, and resources at any time, from anywhere in the world.

Cisco.com is a highly integrated Internet application and a powerful, easy-to-use tool that provides a broad range of features and services to help you to

- Streamline business processes and improve productivity
- Resolve technical issues with online support
- Download and test software packages
- Order Cisco learning materials and merchandise
- Register for online skill assessment, training, and certification programs

You can self-register on Cisco.com to obtain customized information and service. To access Cisco.com, go to the following URL:

<http://www.cisco.com>

Technical Assistance Center

The Cisco TAC is available to all customers who need technical assistance with a Cisco product, technology, or solution. Two types of support are available through the Cisco TAC: the Cisco TAC Web Site and the Cisco TAC Escalation Center.

Inquiries to Cisco TAC are categorized according to the urgency of the issue:

- Priority level 4 (P4)—You need information or assistance concerning Cisco product capabilities, product installation, or basic product configuration.
- Priority level 3 (P3)—Your network performance is degraded. Network functionality is noticeably impaired, but most business operations continue.
- Priority level 2 (P2)—Your production network is severely degraded, affecting significant aspects of business operations. No workaround is available.
- Priority level 1 (P1)—Your production network is down, and a critical impact to business operations will occur if service is not restored quickly. No workaround is available.

Which Cisco TAC resource you choose is based on the priority of the problem and the conditions of service contracts, when applicable.

Cisco TAC Web Site

The Cisco TAC Web Site allows you to resolve P3 and P4 issues yourself, saving both cost and time. The site provides around-the-clock access to online tools, knowledge bases, and software. To access the Cisco TAC Web Site, go to the following URL:

<http://www.cisco.com/tac>

All customers, partners, and resellers who have a valid Cisco services contract have complete access to the technical support resources on the Cisco TAC Web Site. The Cisco TAC Web Site requires a Cisco.com login ID and password. If you have a valid service contract but do not have a login ID or password, go to the following URL to register:

<http://www.cisco.com/register/>

If you cannot resolve your technical issues by using the Cisco TAC Web Site, and you are a Cisco.com registered user, you can open a case online by using the TAC Case Open tool at the following URL:

<http://www.cisco.com/tac/caseopen>

If you have Internet access, it is recommended that you open P3 and P4 cases through the Cisco TAC Web Site.

Cisco TAC Escalation Center

The Cisco TAC Escalation Center addresses issues that are classified as priority level 1 or priority level 2; these classifications are assigned when severe network degradation significantly impacts business operations. When you contact the TAC Escalation Center with a P1 or P2 problem, a Cisco TAC engineer will automatically open a case.

To obtain a directory of toll-free Cisco TAC telephone numbers for your country, go to the following URL:

<http://www.cisco.com/warp/public/687/Directory/DirTAC.shtml>

Before calling, please check with your network operations center to determine the level of Cisco support services to which your company is entitled; for example, SMARTnet, SMARTnet Onsite, or Network Supported Accounts (NSA). In addition, please have available your service agreement number and your product serial number.



IPCC Architectural Overview

The Cisco IP Contact Center (IPCC) solution consists of three primary Cisco software products:

- [Cisco CallManager, page 1-1](#)
- [Cisco IP Interactive Voice Response \(IP IVR\), page 1-2](#)
- [Cisco Intelligent Contact Management \(ICM\) Software, page 1-3](#)

In addition to these core components, several other Cisco products are required for a complete IPCC deployment:

- Cisco IP Phones
- Cisco IPCC Agent Desktop software
- Cisco voice gateways
- Cisco LAN/WAN infrastructure.

Once deployed, IPCC provides an integrated automatic call distribution (ACD), IVR, and computer telephony integration (CTI) solution.

The following sections discuss each of the software products in more detail and describe the data communications between each of these components. For more information on a particular product, refer to the specific product documentation available online at

<http://cisco.com/>

Cisco CallManager

Cisco CallManager, when combined with the appropriate LAN/WAN infrastructure, voice gateways, and IP Phones, provides the foundation for a Voice over IP (VoIP) solution. Cisco CallManager is a software application that runs on Intel Pentium-based servers running Microsoft Windows 2000 Server operating system software and Microsoft SQL Server relational database management software. The Cisco CallManager software running on a server is referred to as a Cisco CallManager server. Multiple Cisco CallManager servers can be grouped into a cluster to provide for scalability and fault tolerance. A Cisco CallManager cluster can contain up to eight Cisco CallManager servers (nodes), but only four of the Cisco CallManager servers can be call processing servers. Refer to the *Cisco IP Telephony Solution Reference Network Design Guide* for details on Cisco CallManager call processing capabilities and clustering options.

A single Cisco CallManager server (Release 3.1(x) or 3.2(x)) is capable of supporting up to 250 agents. In a fault-tolerant design, a typical Cisco CallManager cluster of seven servers (four subscriber nodes, two backup nodes, and one combined TFTP server and publisher node) is capable of supporting up to

1,000 agents. However, the number of agents and the amount of BHCA supported within a cluster varies and must be sized based upon the device weight rules and guidelines defined in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Typically when designing an IPCC solution, you first define the deployment scenario. After defining the deployment scenario, you can determine the sizing of the individual components within the IPCC design for such things as how many Cisco CallManager servers are needed within a Cisco CallManager 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 ICM software, how many IP IVR servers are needed, and so forth.

Cisco IP Interactive Voice Response (IP IVR)

In addition to being used as an interactive voice response (IVR) to prompt callers for input such as menu selections and account numbers, the IP IVR is used as a queuing point in an IPCC environment. When a call is ready to be routed to an agent, but no agents are available, the IP IVR queues the call and plays announcements to the caller. The control of what announcements to play is provided by the ICM via the Service Control Interface (SCI). When an agent becomes available, the ICM immediately instructs the IP IVR to transfer the call to the selected agent phone. The IP IVR then requests the Cisco CallManager to transfer the call to the selected agent phone.

Cisco IP IVR is a software application that runs on Intel Pentium-based servers running Microsoft Windows 2000 Server operating system software and Microsoft SQL Server relational database management software. Each IP IVR (version 3.0) server is capable of supporting up to 150 logical IP IVR ports. A single Cisco CallManager cluster can support up to four IP IVR servers for a total of 600 IP IVR ports per cluster.

The IP IVR has no physical telephony trunks or interfaces like a traditional IVR. The telephony trunks are terminated at the voice gateway. Cisco CallManager provides the call processing and switching to set up a G.711 Real-Time Transport Protocol (RTP) stream from the voice gateway to the IP IVR. The IP IVR communicates with the Cisco CallManager via Java Telephony Application Programming Interface (JTAPI). The IP IVR communicates with the ICM via Service Control Interface (SCI).

For deployments requiring more than 150 IP IVR ports (Cisco Customer Response Solution (CRS) release 3.0(1)), more than one IP IVR is required. Determining the number of IVR ports required is discussed in the chapter on [Sizing IP Contact Center Resources](#). For deployments requiring complete fault tolerance, a minimum of two IP IVRs is required. Details on IPCC fault tolerance can be found in the chapter on [Design Considerations for High Availability](#).

Cisco Internet Service Node (ISN) is another VoIP-based IVR offering. ISN can be used as a queue point for IPCC. Use of ISN is primarily intended to environments where time-division multiplexing (TDM) ACD and IPCC sites need to be integrated together and calls need to be transferred between the sites often. Design details of using ISN with IPCC are documented in a design guide specific to using ISN as an IPCC Queue Point. Therefore, no further discussion of ISN is made in this design guide.

Network IVRs and TDM IVRs can also be used in IPCC environments if they support an SCI interface to the Cisco Intelligent Contact Management (ICM) software. Design guidance for the appropriate usage of these IVR solutions is discussed in the chapter on [Deployment Models](#), but the focus of this chapter is based upon an IPCC deployment using Cisco IP IVR.

A lower-cost licensing option of the IP IVR is called the IP Queue Manager. The IP Queue Manager is almost identical to the IP IVR, except that the database, Java, and HTTP subsystems are not included in the software license. The IP Queue Manager provides an integrated mechanism for prompting and

collecting input from callers and for playing queuing announcements. The sizing for IP Queue Manager and IP IVR are the same. Because the IP IVR and IP Queue Manager are so similar, the remainder of this document will refer to the IP IVR only.

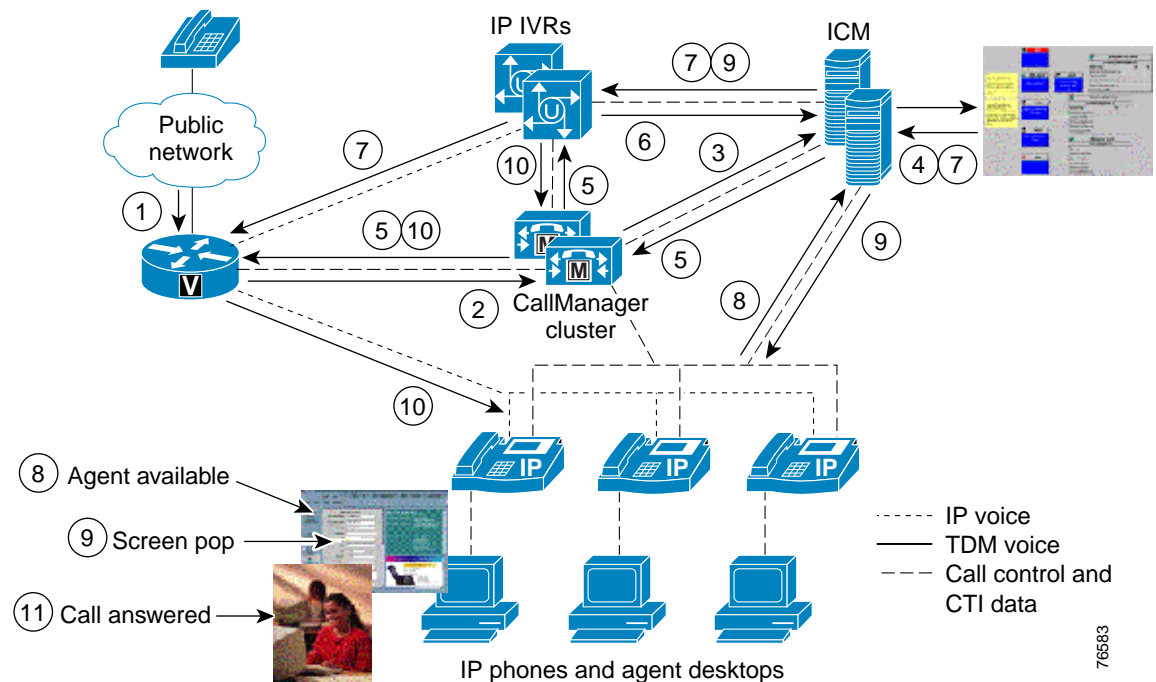
Cisco Intelligent Contact Management (ICM) Software

The Cisco ICM software provides call center features in conjunction with Cisco CallManager. Features provided by the ICM software include agent state management, agent selection, call routing and queuing, IVR control, screen pops, and call center reporting. ICM software runs on Pentium servers running Windows 2000 operating system software and SQL Server database management software. The supported Pentium servers can be single, dual, or quad Pentium CPU servers with varying amounts of RAM. This variety of supported servers allows the ICM software to scale and to be sized to meet the needs of the deployment requirements. Details on ICM server sizing can be found in the chapter on [Sizing IPCC Components and Servers](#).

Basic IPCC Call and Message Flow

Figure 1-1 shows the flow of a basic IPCC call. In this scenario, all of the agents are assumed to be "not ready" when the call arrives, so the call is routed by the ICM to the IP IVR. While the call is connected to the IP IVR, call queuing treatment (announcements, music, and so on) is provided. When an agent becomes available, the ICM directs the IP IVR to transfer the call to that agent's phone. At the same time the call is being transferred, the ICM pops the agent's desktop with any call data, such as automatic number identification (ANI), directory number (DN), and so forth.

Figure 1-1 Basic Call Flow



The call flow in [Figure 1-1](#) is as follows:

1. Call delivered from PSTN to voice gateway.
2. MGCP or H.323 Route Request sent to Cisco CallManager.
3. JTAPI Route Request sent to ICM.
4. ICM runs routing script. No available agent found, so IP IVR label returned from routing script.
5. ICM instructs Cisco CallManager to transfer call to IP IVR, and Cisco CallManager does as instructed.
6. IP IVR notifies ICM that call has arrived.
7. ICM instructs IP IVR to play queue announcements.
8. Agent becomes ready (completed previous call or just went ready).
9. ICM sends call data to selected agent screen and instructs the IP IVR to transfer the call to the agent phone.
10. IP IVR transfers the VoIP voice path to selected agent phone.
11. Call is answered by agent.

ICM Software Modules

The Cisco ICM software is a collection of modules that can all run on the same server or on separate servers (except for the Administrative Workstation, which requires its own server). The amount of software that can run on one server is primarily based upon 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 skills per agent, the number of IP IVR ports, the number of Run VRU Script nodes in an ICM routing script, what statistics agents need at their desktops, and so forth.

The core ICM software modules are:

- Router
- Logger
- Cisco CallManager Peripheral Interface Manager (PIM)
- IP IVR PIM
- CTI Server
- Administrative Workstation (AW)

The Router is the module that makes all routing decisions on how to route a call or customer contact. The Logger is the database module that stores contact center configuration and reporting data. The Cisco CallManager PIM is the module that interfaces to a Cisco CallManager cluster via the JTAPI protocol. The IP IVR PIM is the module that interfaces to the IP IVR via the Service Control Interface (SCI) protocol. The CTI Server is the module that interfaces to the IPCC Agent Desktop application.

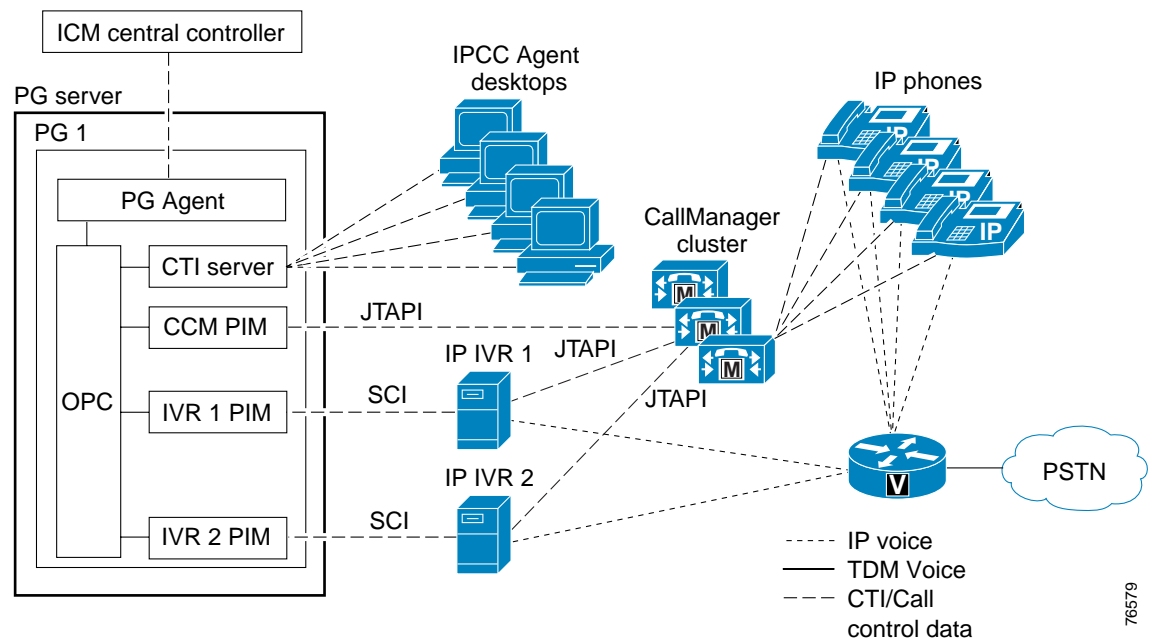
Each ICM software module can be deployed in a redundant fashion. When a module is deployed in a redundant fashion, we refer to the two sides as side A and side B. For example, Router A and Router B are redundant instances of the Router module (process) running on two different servers. Logically, there is only one Router.

Multiple instances of many of the software modules are possible. The only software modules that do not support multiple instances are the Router and Logger. The Router and Logger combined are often referred to as the ICM Central Controller. When the Router and Logger modules run on the same server,

the server is referred to as a *Rogger*. An IPCC deployment can have only one logical ICM Central Controller. If you want a fault-tolerant ICM Central Controller, then you can have side A and side B of the ICM Central Controller, but you will still have only one logical ICM Central Controller.

For each Cisco CallManager cluster in your IPCC environment, you need a Cisco CallManager PIM. For each Cisco CallManager PIM, you need one CTI Server to communicate with the desktops associated with the phones for that Cisco CallManager cluster. For each IP IVR, you need one IP IVR PIM. The server that runs the Cisco CallManager PIM, the CTI Server, and the IP IVR PIMs, is referred to as a Peripheral Gateway (PG). Often, the Cisco CallManager PIM, the CTI Server, and multiple IP IVR PIMs will run on the same server. This server is referred to as a "hybrid PG." Internal to the PG is a process called the PG Agent, which communicates from the PG to the ICM CC. Another internal PG process is the Open Peripheral Controller (OPC), which allows the other processes to communicate with each other and is also involved in synchronizing PGs in redundant PG deployments. Figure 1-2 shows the communication of the various PG software processes.

Figure 1-2 Communication of Peripheral Gateway Software Processes



In larger, multi-site (multi-cluster) environments, multiple PGs are usually deployed. **Each Cisco CallManager cluster requires a co-located PG.** When multiple Cisco CallManager clusters are deployed, the ICM software makes them all appear to be part of one logical enterprise-wide contact center with one enterprise-wide queue. The Cisco CallManager clusters at different sites do not all need to have the same number of Cisco CallManager servers. Each site is sized independently.

76579

IPCC Components, Terminology, and Concepts

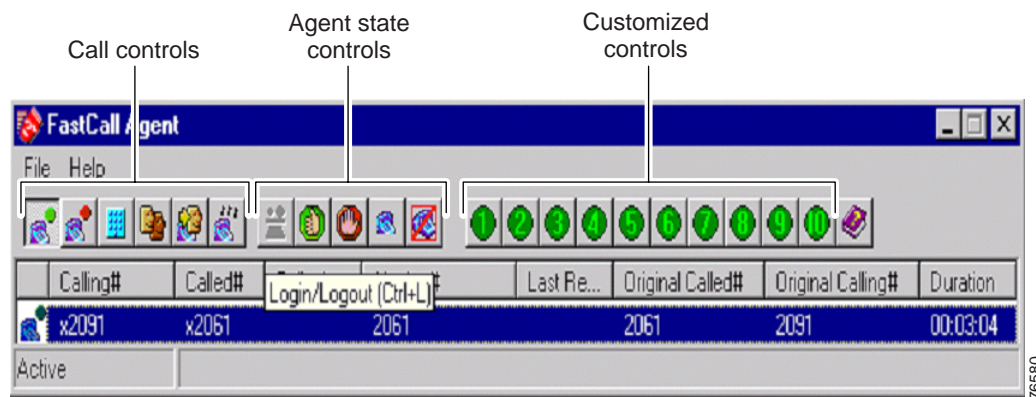
This section describes the major components and concepts employed in an IPCC solution.

IPCC Agent Desktop

The IPCC agent desktop application is the agent interface that allows the agent to perform agent state control (login, logout, ready, not ready, wrap up, and so on) and call control (answer, release, hold, retrieve, transfer, conference, make call, and so on). **All call control must be done via the agent desktop application.**

The agent desktop has a media-terminated "softphone" option that allows you to eliminate the need for a hardware IP Phone. (See [Figure 1-3](#). Do not confuse the agent desktop softphone option with other softphone applications such as the Cisco IP SoftPhone.) When using the agent desktop softphone option, a headset is connected to the PC, which will encode/decode the VoIP packets and send/receive those packets to/from the LAN.

Figure 1-3 IPCC Agent Desktop



There are two IPCC agent and supervisor desktop options available:

- Cisco Agent Desktop, an out-of-the-box agent desktop
- CTI Object Server (CTI OS) Toolkit, for agent desktops that need to be customized or integrated with other applications on the desktop or with customer databases such as a Customer Relationship Management (CRM) application

In addition to an agent desktop, a supervisor desktop is available with each of these options.

Details on desktop selection and design considerations are covered in the chapter on [IPCC Agent Desktop and Supervisor Desktop](#).

Administrative Workstation

The ICM Administrative Workstation (AW) provides a collection of administrative tools for managing the ICM configuration and monitoring the contact center performance. The two primary configuration tools on the AW are the Configuration Manager and the Script Editor. The Configuration Manager tool is used to configure the ICM database to add agents, add skill groups, assign agents to skill groups, add dialed numbers, add call types, assign dialed numbers to call types, assign call types to ICM routing

scripts, and so forth. The Script Editor tool is used to build ICM routing scripts. ICM routing scripts specify how to route and queue a contact (that is, the script identifies which agent should handle a particular contact).

The primary monitoring tool on the AW is Monitor ICM. Monitor ICM provides real-time and historical contact center reporting on agents, skill groups, and services.

Details on the use of these tools are provided in the *IPCC Administrator Guide* and the *IPCC Reporting Guide*, available online at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/icm46/ipcc/index.htm>

The AW is the only software module that is required to run on a separate server from all of the other IPCC software modules. An ICM installation supports an unlimited number of AWs and can be deployed co-located or remote from the ICM Central Controller. Each AW is independent of other AWs, and redundancy is handled by deploying multiple AWs.

Some AWs communicate directly with the ICM Central Controller, and they are called *distributor* AWs. An ICM deployment must have at least one distributor AW. Additional AWs (distributors or clients) are also allowed for redundancy (primary and secondary distributors) or for additional access by the AW clients in a site. At any additional site, at least one distributor and any number of client AWs can be deployed; however, client AWs should always be local to their AW distributor.

Client AWs communicate with a distributor AW to view and modify the ICM Central Controller database and to receive real-time reporting data. Distributor AWs off-load the Central Controller (the real-time call processing engine) from the task of constantly distributing real-time contact center data to the client AWs.

AWs can be installed with two software options:

- Historical Data Server (HDS)
- WebView Server

The Historical Data Server (HDS) option provides a replicated copy of the historical reporting data. **Distributor AWs must be installed with an HDS.** The WebView Server option provides browser-based reporting. This allows reporting to be done from any computer with a browser (as opposed to Monitor ICM, which requires you to use a client AW). **WebView Server must be installed on a distributor AW with HDS.** The net effect of adding the WebView Server is to turn the distributor AW into a web server. Up to 50 simultaneous web browser sessions are supported per WebView Server.

The reason for requiring the AW to run on a separate server for production systems is to ensure that complex reporting queries do not interrupt the real-time call processing of the Router and Logger processes. For lab or prototype systems, the AW (with the WebView Server option) can be installed on the same server as the Router and Logger. If the AW is installed on the same server as the Logger, then HDS is no longer required because a complete copy of the Logger database is already present on the server.

More details on the design and configuration of the AWs can be found in the ICM product documentation available online at Cisco.com.

JTAPI Communications

In order for Java Telephony Application Programming Interface (JTAPI) communication to occur between Cisco CallManager and external applications like the ICM and IP IVR, a JTAPI user ID and password must be configured within Cisco CallManager. Upon startup of the Cisco CallManager PIM or upon startup of the IP IVR, the JTAPI user ID and password are used to log in to the Cisco CallManager. This login process by the application (Cisco CallManager PIM or IP IVR) establishes the JTAPI communication between the Cisco CallManager cluster and the application. A single JTAPI user ID is

used for all communication between the entire Cisco CallManager cluster and the ICM. A separate JTAPI user ID is also required for each IP IVR server. In an IPCC deployment with one Cisco CallManager cluster and two IP IVRs, three JTAPI user IDs are required, one JTAPI user ID for the ICM application and two JTAPI user IDs for the two IP IVRs.

The Cisco CallManager software includes a module called the CTI Manager, which is the layer of software that communicates via JTAPI to applications like the ICM and IP IVR. Every node within a cluster can execute an instance of the CTI Manager process, but the Cisco CallManager PIM on the PG communicates with only one CTI Manager (and thus one node) in the Cisco CallManager cluster. The CTI Manager process communicates CTI messages to/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 the ICM. When a new call arrives at this voice gateway and needs to be routed by the ICM, node 1 sends an intra-cluster message to node 2, which will send a route request to the ICM to determine how the call should be routed.

Each IP IVR also communicates with only one CTI Manager (or node) within the cluster. The Cisco CallManager PIM and the two 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 Cisco CallManager PIM is redundant, only one side is active and in communication with the Cisco CallManager cluster. Side A of the Cisco CallManager PIM communicates with the CTI Manager on one Cisco CallManager node, and side B of the Cisco CallManager PIM communicates with the CTI Manager on another Cisco CallManager node. The IP IVR does not have a redundant side, but the IP IVR does have the ability to failover to another CTI Manager (node) within the cluster if its primary CTI Manager is out of service. Details on failover are covered in the chapter on [Design Considerations for High Availability](#).

The JTAPI communication between the Cisco CallManager and ICM includes three distinct types of messaging:

- Routing control
Routing control messages provide a way for the Cisco CallManager to request routing instructions from the ICM.
- Device and call monitoring
Device monitoring messages provide a way for the Cisco CallManager to notify the ICM about state changes of a device (IP Phone) or a call.
- Device and call control
Device control messages provide a way for Cisco CallManager to receive instructions from the ICM on how to control a device (IP Phone) or a call.

A typical IPCC call includes all three types of JTAPI communication within a few seconds. When a new call arrives, Cisco CallManager requests routing instruction from the ICM. For example, when the Cisco CallManager receives the routing response from the ICM, the Cisco CallManager attempts delivery of the call to the agent phone by instructing the phone to begin ringing. At that point, the Cisco CallManager notifies the ICM that the device (IP Phone) has started ringing, and that is what causes the agents answer button on their desktop application to become enabled. When the agent clicks the answer button, the ICM instructs Cisco CallManager to make the device (IP Phone) go off-hook and answer the call.

In order for the routing control communication to occur, the Cisco CallManager requires the configuration of a CTI Route Point. A CTI Route Point is associated with a specific JTAPI user ID, and this association enables the Cisco CallManager 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 the ICM JTAPI user ID, and this enables the Cisco CallManager to generate a route request to the ICM when a new call to that DN arrives.

In order for the IP Phones (devices) to be monitored and controlled, they also must be associated in Cisco CallManager with a JTAPI user ID. In an IPCC environment, the IP Phones are associated with the ICM JTAPI user ID. When an agent logs in from their desktop, the Cisco CallManager PIM requests the Cisco CallManager to allow the Cisco CallManager PIM to begin monitoring and controlling that device (IP Phone). Until the login has occurred, the Cisco CallManager does not allow the ICM to monitor or control that IP Phone. If the device has not been associated with the ICM JTAPI user ID, then the agent login request will fail.

Because the IP IVR also communicates with Cisco CallManager using the same JTAPI protocol, these same three types of communication also occur with the IP IVR. Unlike the ICM, the IP IVR provides both the application itself and the devices to be monitored and controlled.

The devices that the ICM monitors and controls are the physical IP Phones. The IP IVR does not have real physical ports like a traditional IVR. Its ports are logical ports (independent software tasks or threads running on the IP IVR application server) called CTI Ports. For each CTI Port on the IP IVR, there needs to be a CTI Port device defined in Cisco CallManager.

Unlike a traditional PBX or telephony switch, the Cisco CallManager does not select the 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 an IP IVR JTAPI user, the Cisco CallManager asks the IP IVR (via JTAPI routing control) which CTI Port (device) should handle the call. Assuming the IP IVR has an available CTI Port, the IP IVR will respond to the Cisco CallManager routing control request with the Cisco CallManager device identifier of the CTI Port that is going to handle that call.

When an available CTI Port is allocated to the call, an IP IVR workflow is started within the IP IVR. When the IP IVR workflow executes the accept step, a JTAPI message is sent to Cisco CallManager to answer the call on behalf of that CTI Port (device). When the IP IVR workflow wants the call transferred or released, it again instructs the Cisco CallManager what it would like done with that call. These scenarios are examples of device and call control performed by the IP IVR.

When a caller releases while interacting with the IP IVR, the voice gateway detects the caller release, notifies the Cisco CallManager via H.323 or Media Gateway Control Protocol (MGCP), which then notifies the IP IVR via JTAPI. When DTMF tones are detected by the voice gateway, it notifies the Cisco CallManager via H.245 or MGCP, which then notifies the IP IVR via JTAPI. These scenarios are examples of device and call monitoring performed by the IP IVR.

In order for the CTI Port device control and monitoring to occur, the CTI Port devices on Cisco CallManager must be associated with the appropriate IP IVR JTAPI user ID. If you have two 150-port IP IVRs, you would have 300 CTI Ports. 150 ports would be associated with JTAPI user IP IVR #1, and 150 would be associated with JTAPI user IP IVR #2.

While Cisco CallManager can be configured to route calls to IP IVRs on its own, in an IPCC environment, routing of calls to the IP IVRs should be done by the ICM (even if you have only one IP IVR and all calls require an initial IVR treatment). Doing so will ensure proper IPCC reporting. For deployments with multiple IP IVRs, this will also allow the ICM to load-balance calls across the multiple IP IVRs.

ICM Routing Clients

An ICM routing client is anything that can generate a route request to the ICM Central Controller. The Cisco CallManager PIM (representing the entire Cisco CallManager cluster) and each IP IVR PIM are routing clients. Routing clients generate route requests to the ICM Central Controller. The ICM Central Controller then executes a routing script and returns a routing label to the routing client. A redundant PIM is viewed as a single logical routing client, and only one side of a PIM is active at any point in time. In an IPCC deployment with one Cisco CallManager cluster (with any number of nodes) and two IP IVRs, three routing clients are required – the Cisco CallManager PIM and the two IP IVR PIMs.

The public switched telephone network (PSTN) can also function as a routing client. The ICM supports a software module called a Network Interface Controller (NIC), which allows the ICM to control how the PSTN routes a call. 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 the ICM. A detailed list of PSTN NICs and details on ICM pre-routing can be found in the standard ICM product documentation available at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/>

Other applications like the Cisco Media Blender, the Cisco Collaboration Server, and the Cisco E-Mail Manager can also function as routing clients to allow the ICM to become a multi-channel contact routing engine. Details of currently available multi-channel routing are covered in the ICM Web Option and Universal Queue documentation, available on Cisco.com.

Device Targets

Each IP phone must be configured in the ICM Central Controller database as a device target. Only one extension on the IP Phone may be configured as an ICM device target. Additional extensions can be configured on the IP Phone, but those extensions will not be known to the ICM software and thus no monitoring or control of those additional extensions is possible. The IPCC extension may be configured with only one line appearance and cannot be configured with call waiting, voice mail, or forward-no-answer. The ICM provides call treatment for Reroute On No Answer (RONA), therefore it is not necessary to configure call forwarding on ring-no-answer in the Cisco CallManager configuration for the IP Phones. Unless call center policy permits warm (agent-to-agent) transfers, the IPCC extension also should not be published or dialed by anyone directly, and only the ICM software should route calls to this IPCC phone extension.

At agent login, the agent ID and IP Phone extension are associated, and this association is released when the agent logs out. This feature allows the agent to log in at another phone and allows another agent to log in at that same phone. At agent login, the Cisco CallManager PIM requests Cisco CallManager to allow it to begin monitoring that IP Phone and to provide device and call control for that IP Phone. As mentioned previously, each IP Phone must be mapped to the ICM JTAPI user ID in order for the agent login to be successful.

Labels

Labels are the response to a route request from a routing client. The label is a pointer to the destination where the call is to be routed (basically, the number to be dialed by the routing client). Many labels in an IPCC environment correspond to the IPCC phone extensions so that Cisco CallManager and IP IVR can route or transfer calls to the phone of an agent who has just been selected for a call.

Often, the way a call is routed to a destination depends upon where the call originated and where it is being terminated. This is why IPCC uses labels. For example, suppose we have an environment with two regionally separated Cisco CallManager clusters, Site 1 and Site 2. An IP Phone user at Site 1 will typically just dial a four-digit extension to reach another IP Phone user at Site 1. In order to reach an IP Phone user at Site 2 from Site 1, users may need to dial a seven-digit number. To reach an IP Phone user at either site from a PSTN phone, users may need to dial a 10- digit number. From this example, we can see how a different label would be needed, depending upon where the call is originating and terminating.

Each combination of device target and routing client must have a label. For example, a device target in an IPCC deployment with a two-node Cisco CallManager cluster and two IP IVRs will require three labels. If you had 100 device targets (IP Phones), you would need 300 labels. If there are two regionally separated Cisco CallManager clusters, each with two IP IVRs and 100 device targets per site, then we would need 1200 labels for the six routing clients and 200 device targets (assuming we wanted to be able to route a call from any routing client to any device target). If calls are to be routed only to device targets at the same site as the routing client, then we would need only 600 labels (three routing clients to 100 device targets, and then doubled for Site 2).

Labels are also used to route calls to IP IVR CTI Ports. Details on configuring labels are provided in the *IPCC Installation Guide*, available on Cisco.com. A bulk configuration tool is available to simplify the configuration of the labels.

Agent Desk Settings

Agent Desk Settings provide a profile that specifies parameters such as whether auto-answer is enabled, how long to wait before re-routing a call for Ring No Answer, what DN to use in the re-routing, and whether reason codes are needed for logging out and going not-ready. Each agent must be associated with an agent desk setting profile in the ICM configuration. A single agent desk setting profile can be shared by many agents. Changes made to an agents desk setting profile while the agent is logged in are not activated until the agent logs out and logs in again.

Agents

Agents are configured within the ICM and are associated with one specific Cisco CallManager PIM (that is, one Cisco CallManager cluster). Within the ICM configuration, you also configure the password for the agent to use at login.

Skill Groups

Skill groups are configured within the ICM so that agents with similar skills can be grouped together. Agents can be associated with one or more skill groups. Changes made to an agents skill group association while they are logged in are not activated until the agent logs out and logs in again.

Skill groups are associated with a specific Cisco CallManager PIM. Skill groups from multiple PIMs can be grouped into Enterprise Skill Groups. Creating and using Enterprise Skill Groups can simplify routing and reporting in some scenarios.

Directory (Dialed) Numbers and Routing Scripts

In order for Cisco CallManager to generate a route request to the ICM, the Cisco CallManager must associate the DN with a CTI Route Point that is associated with the ICM JTAPI User. The DN must also be configured in the ICM. Once the ICM receives the route request with the DN, that DN is then mapped to an ICM Call type, which is then mapped to an ICM routing script.

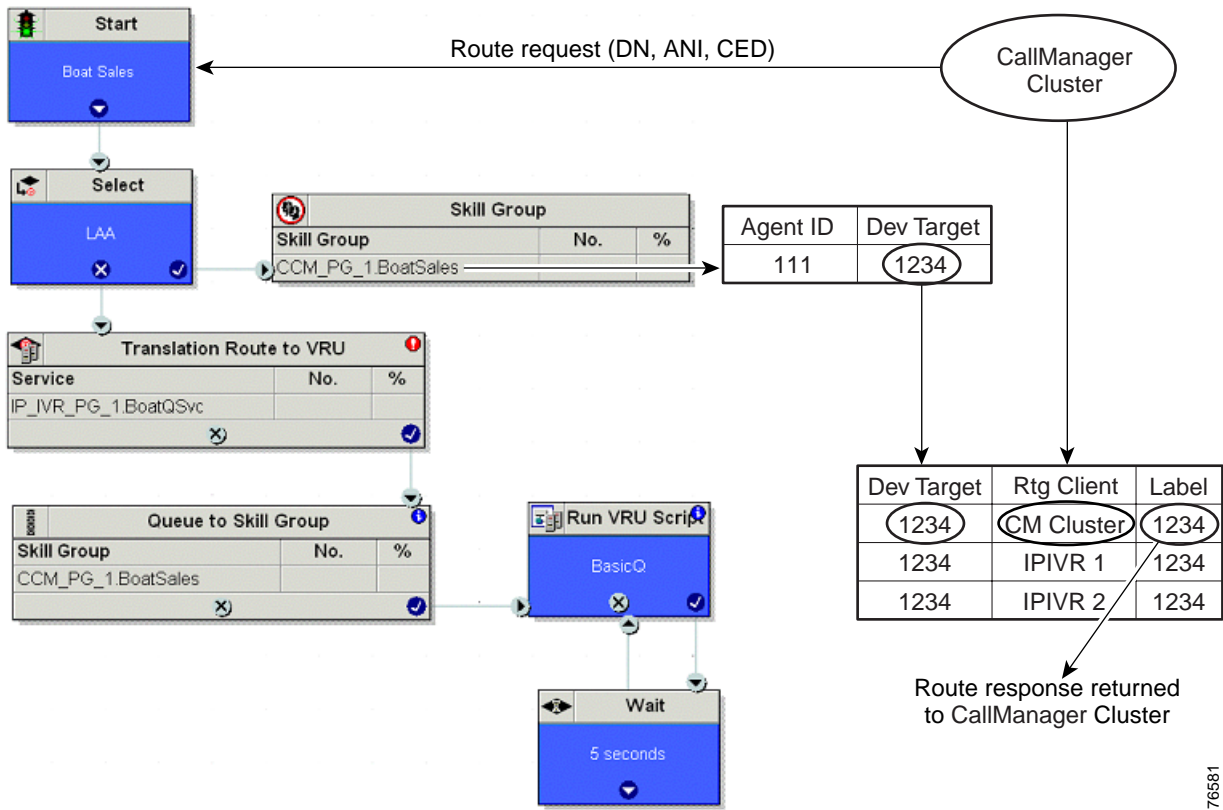
Agent Login and State Control

Agents log in to IPCC from their IPCC agent desktop application. When logging in, the agent is presented with a dialog box prompting them for their agent ID, password, and the IPCC phone extension to be used for this login session. It is at login time that the agent ID, their phone extension (device target), agent desk setting profile, skills, and desktop IP address are all dynamically associated. The association is released upon agent logout.

IPCC Routing

The example routing script in Figure 1-4 illustrates how the ICM routes calls. In this routing script, the Cisco CallManager PIM (or cluster) is the routing client. Upon receipt of the route request, the ICM maps the DN to a call type and then maps the call type to this routing script. In this routing script, the ICM router first uses a Select node to look for the Longest Available Agent (LAA) in the BoatSales skill group on the CCM_PG_1 peripheral (or cluster). The ICM router identifies that agent 111 is the LAA. Agent 111 is currently logged in from device target 1234 (Cisco CallManager phone extension 1234 in this scenario). The ICM router then determines the label to be returned, based upon the device target and routing client combination. The appropriate label is then returned to the routing client (Cisco CallManager cluster) so that the call can be routed properly to that IP Phone (device target).

Figure 1-4 Routing Script Example



Translation Routing and Queuing

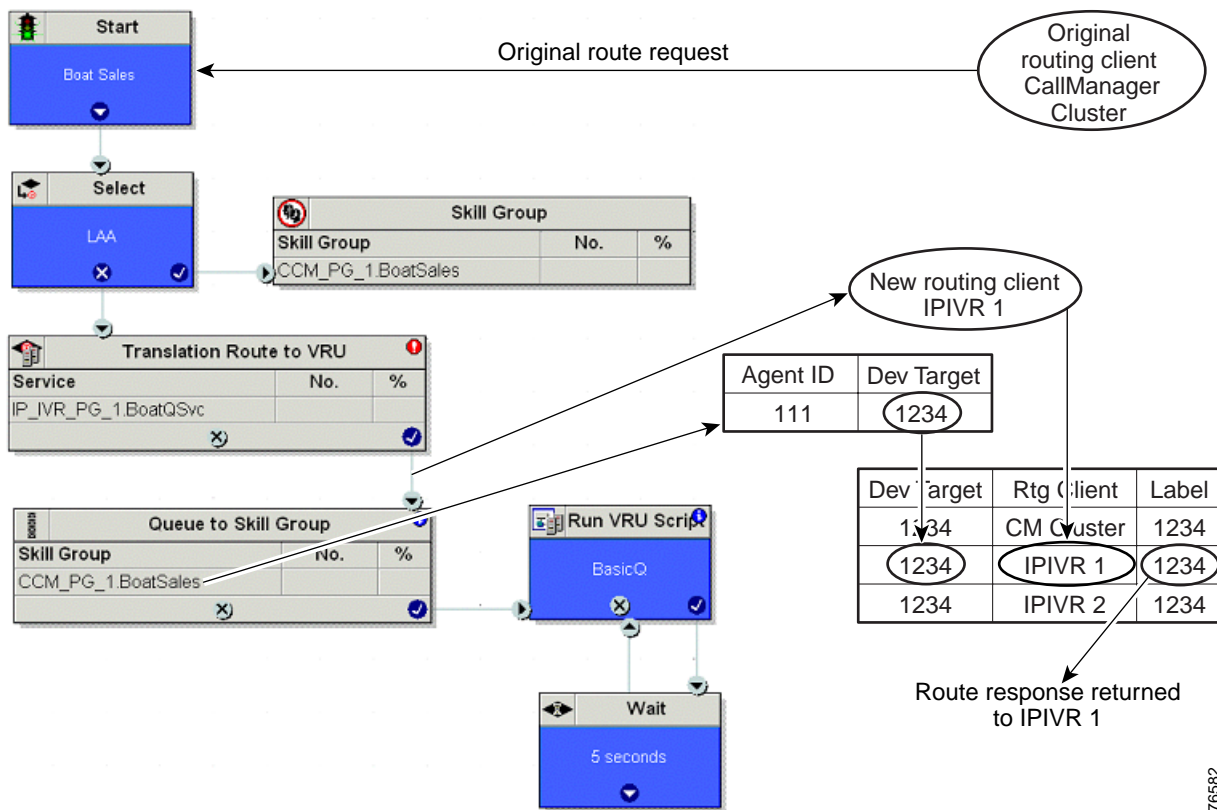
If no agents are available, as agent 111 was in the previous example, then the router exits the Select node and transfers the call to an IP IVR to begin queuing treatment. The transfer is completed using the “Translation Route to VRU” node. The Translation Route to VRU node returns a unique translation route label to the original routing client, the Cisco CallManager cluster. The translation route label will equal a DN configured in Cisco CallManager. In Cisco CallManager, that DN is mapped to a CTI Route Point that is associated with the JTAPI user for the IP IVR where the call is being transferred.

Cisco CallManager and IP IVR will execute the JTAPI routing control messaging to select an available CTI Port.

When the call is successfully transferred to the IP IVR, the IP IVR translation routing application first sends a request instruction message to the ICM via the SCI between the IP IVR and the ICM. The ICM identifies the DN as being the same as the translation route label and is then able to re-associate this call with the call that was previously being routed. The ICM then re-enters the routing script for this call that was previously being run. The re-entry point is the successful exit path of the Translation Route to VRU node. (See Figure 1-5.) At this point the routing client has changed from the Cisco CallManager cluster to IPIVR1.

While the call was being transferred, the routing script was temporarily paused. After the transfer to the IP IVR is successfully completed, the IP IVR becomes the routing client for this routing script. Next the routing script queues the call to the BoatSales skill group and then instructs the IP IVR to run a specific queue treatment via the Run VRU Script node. Eventually agent 111 becomes available, and as in the previous example, the label to be returned to the routing client is identified based upon the combination of device target and routing client. Note that the routing client is now the IP IVR. The label returned (1234) when agent 111 becomes available causes the IP IVR to transfer the call to agent 111 (at extension 1234).

Figure 1-5 Translation Routing and Queuing



For each combination of Cisco CallManager cluster and IP IVR, a translation route and a set of labels is required. For example, if a deployment has one Cisco CallManager cluster and four IP IVRs, then four translation routes and sets of labels are required.

For deployments with multiple IP IVRs, the ICM routing script should select the IP IVR with the greatest number of idle IP IVR ports and then translation route the call to that specific IP IVR. If there are no IP IVR ports available, then the script should execute a Busy node. If there are a high number of calls executing busy nodes, then it will be important to resize your IP IVR port capacity.

Reroute On No Answer (RONA)

When a call is routed to an agent but the agent fails to answer the call within a configurable amount of time, the Cisco CallManager PIM for the agent who did not answer will change that agent's state to "not ready" (so they don't get more calls) and launch a route request to find another agent. Any call data will be preserved and popped onto the next agent's desktop. If no agent is available, the call can be sent back to the IP IVR for queuing treatment again. Again, all call data will be preserved. The routing script for this RONA treatment should set the call priority to high so that the next available agent is selected for this caller. The RONA timer and the DN to be used to specify a unique call type and routing script for RONA treatment is configured in the agent desk setting.

Hybrid IP Telephony and IPCC Cisco CallManager Clusters

It is possible for a Cisco CallManager cluster to support IP Phones with both normal IP Telephony (office) extensions and IPCC (call center) extensions. When running hybrid Cisco CallManager clusters with both IP Telephony and IPCC extensions, it is important to realize that sometimes the most recent Cisco CallManager software release will not immediately be supported in IPCC deployments until testing is completed later. It is also important to note that many contact center environments have very stringent maintenance windows. Because of these software and environmental limitations, it may sometimes be advantageous to separate Cisco CallManager clusters for IP Telephony extensions from the Cisco CallManager clusters for IPCC extensions. It is important to consider the environment where IPCC is being deployed to determine whether a separate Cisco CallManager cluster is advantageous.

Combined IP Telephony and IPCC Extensions on the Same IP Phone

It is possible to have multiple extensions on an IP Phone. In an IPCC environment, at least one of those extensions must be dedicated to IPCC and be configured with only a single line appearance, no voice mail, and no call forwarding. Cisco recommends that the bottom-most extension on the IP Phone be used as the IPCC extension so that, when the user lifts the handset, the IPCC extension is not selected by default. Other extensions on the IP Phone may have multiple lines or appearances and voice mail. However, it is important to note that there is no control over, or visibility of, those IP Telephony extensions from the IPCC Agent Desktop. Cisco recommends that any IP Telephony extensions be forwarded to voice mail prior to logging into the IPCC, so that agents are not interrupted by IP Telephony calls while they are working on IPCC calls. Also, prior to making any outbound calls on the IP Telephony extension, Cisco recommends that the agent change to a "not ready" state so that calls are not routed to them while they are on the phone. The ICM has no visibility to the state of the other extensions on this phone.

Queuing in an IPCC 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 IPCC deployment, it is important to consider how queuing and requeuing are going to be handled.

Call queuing in an IPCC deployment requires usage of an IVR platform that supports the SCI interface to the ICM. The Cisco IP IVR is one such platform. Cisco also offers another IVR platform, called Internet Service Node (ISN), that can be used as a queuing point for IPCC deployments. Considerations for deployments with ISN are provided in the chapter on [Deployment Models](#). Traditional IVRs can also be used in IPCC deployments. Considerations for deployments with traditional IVRs are also provided in the chapter on [Deployment Models](#).

In an IPCC 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 the ICM via the SCI interface. The Run VRU Script node in an ICM routing script is the component that causes the ICM to instruct the IVR to play a particular queuing treatment.

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

Transfers in an IPCC Environment

Transfers are a commonly used feature in contact centers, therefore it is very important to consider all of the possible transfer scenarios desired for your IPCC installation. This section explains basic transfer concepts. The transfer scenarios themselves are discussed later within the chapter on [Deployment Models](#).

Transfers involve three parties – the original caller, the transferring agent, and the target agent. The original caller is the caller that made the original call that was routed to the transferring agent. The transferring agent is the agent requesting the transfer to the target agent. The target agent is the agent receiving the transfer from the transferring agent. This terminology is used throughout this document when referring to the different parties.



Note

Cisco recommends that all call control (answer, release, transfer, conference, and so on) be done from the agent desktop application.

When a transferring agent wants to transfer a call to another skill group or agent, the transferring agent clicks on the transfer button on their IPCC Agent Desktop. A dialog box allows the transferring agent to enter the dialed number of a skill group or agent. An alphanumeric dialed number string (such as "sales" or "service") is also valid. The transferring agent also selects whether this transfer is to be a single-step (blind) transfer or a consultative transfer. (Single-step transfer is the default.) The transferring agent then clicks OK to complete (single-step) or initiate (consultative) the transfer. The transfer request message flows from the transferring agent desktop to the CTI Server and then to the Cisco CallManager PIM.

Any call data that was delivered to the transferring agent or added by the transferring agent is sent along with the transfer request to the Cisco CallManager PIM.

Dialed Number Plan

The Cisco CallManager PIM then attempts to match the dialed number against an entry in the Dialed Number Plan. The ICM Dialed Number Plan (DNP) is currently administered via the Bulk Configuration tool on the ICM AW. Entries in the DNP are entered per peripheral (PIM), and all DNP entries for a particular PIM are downloaded to the PIM upon PIM startup. Updates and additions to the DNP are also sent to the PIM dynamically, and they take effect immediately and are used for the next call being transferred. In order for the ICM to route this transfer and have all call data move with the transfer and be saved for cradle-to-grave reporting, a dialed number match must be found in the DNP for the PIM where this agent is currently logged in.

Within the DNP, fuzzy (wildcard) matching of dialed number strings is allowed. The DNP is not the same as the Dialed Number table used by the ICM router and managed via the AW Configuration Manager tool. The ICM router maps dialed numbers to call types, and call types are mapped to ICM routing scripts. This is how a specific dialed number is mapped to a routing script in the ICM router. Details on administration of editing dialed numbers, call types, and routing scripts can be found in the *IPCC Administrator Guide*, available at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/icm46/ipcc/index.htm>

For help with designing a dial plan for your IPCC deployment, consult your Cisco Systems Engineer (SE).

Dial Plan Type

Entries in the Dialed Number Plan must be configured with a dial plan type. There are six predefined (via a list box) DNP types that correspond to the types specified in the agent desk settings profile. In order for a call or transfer to proceed any further, the DNP type for that call must be allowed in the agent desk setting profile used by the transferring agent. Because the Cisco CallManager *calling search spaces* override any desk settings, it is best to allow all dial plan types in the agent desk settings.



Note

Changes to the agent desk settings profile do not take place until the agent logs out and logs in again.

Post Route

Entries in the Dialed Number Plan must also be configured to indicate whether a post-route is required. For dialed numbers to be used in transfer scenarios, Cisco recommends that the post-route option be set to **Yes** for transfers. When this field is set to **Yes**, the dialed number to be used for the route request needs to be supplied in the dialed number column of the dialed number plan editor.

Route Request

Assuming a match is found in the DNP for the transfer, the DNP type is allowed for the transferring agent, and the post-route option is set to yes, the PIM logic will then generate a route request to the ICM central controller using the dialed number specified in this same DNP entry.

Upon receipt of the route request, the ICM router matches the dialed number to a call type and executes the appropriate routing script to find an appropriate target agent for the call. Within the routing script, any of the call data collected so far could be used in the intelligent routing of the call. The ICM router will determine what device target (phone extension and desktop) the agent is logged into and will then return the label that points to that device target to the Cisco CallManager PIM.

At this point there are numerous scenarios that can occur, depending upon the type of transfer being performed.

Single-Step (Blind) Transfer

A blind transfer is used when the transferring agent does not need to speak with the target agent. After a blind transfer is specified in the transfer dialog box on the agent desktop, a DN is entered and the transferring agent clicks the initiate transfer button. The desktop then sends the transfer request to the Cisco CallManager PIM. Assuming a match is found in the DNP, the DNP type is valid, and post-route is selected, the Cisco CallManager PIM generates the route request to get a routing label and then instructs the Cisco CallManager to perform a single step transfer (without any further action from the transferring agent). The transferring agent will see the call disappear from their desktop and they will transition to the next agent state (wrap-up, ready, or not ready), depending on the agent desk settings for the transferring agent. While the call is being placed to the target agent, the original caller is temporarily placed on hold. When the target agent's phone begins ringing, the original caller hears the ringing (assuming auto-answer is not enabled). The target agent receives a screen pop with all call data, and the answer button on their agent desktop are enabled when the phone begins ringing. Upon answering the call, the target agent is speaking with the original caller and the transfer is then complete. If the target agent does not answer, then RONA (reroute on no answer) call rerouting logic will take over.

If auto-answer is enabled, the original caller and the target agent do not hear any ringing; the call is just connected between the original caller and the target agent.

If the agent is transferring the call to a generic (skill-group) DN to find an available agent with a particular skill but no such agent is currently available, then the ICM routing script should be configured to translation route the call to an IP IVR for queuing treatment. The call would still be released from the transferring agent desktop almost immediately. Any call data collected by the transferring agent would automatically be passed to the IVR. The caller will not hear any ringback tones because the IP IVR CTI Port will answer immediately. When the target agent becomes ready, the ICM will instruct the IVR to transfer the call, and the ICM will pop the agent desktop with all call data.

If the agent has transferred the call to a number that is not within the ICM Dialed Number Plan, then the caller will be transferred anyway. The destination for the transferred call depends upon the number that was dialed and what is configured in the Cisco CallManager dial plan. Transfers not using the dialed number plan are not recommended because of agent roaming restrictions, call data not following the call, and reporting limitations.

Consultative Transfer

Some parts of the message flow for a consultative transfer are similar to the message flow for a blind transfer. When the Cisco CallManager PIM receives the label from the ICM router indicating where to transfer the call, the Cisco CallManager PIM tells the Cisco CallManager to initiate a consultative transfer to the number specified in the label. The Cisco CallManager places the original caller on hold and makes a consultative call to the number specified in the label. The caller generally hears tone on hold while the transfer is being completed.

When the target agent phone begins ringing, the Cisco CallManager generates a consult call confirmation message and a device ringing message.

The consult call confirmation message causes the Cisco CallManager PIM to notify the transferring agents desktop that the call is proceeding, and it enables the transfer complete button. The transferring agent can hear the target agent's phone ringing (assuming auto-answer is not enabled for the target agent). At any time after this, the agent can click the transfer complete button to complete the transfer (before or after the target answers their phone).

The device ringing message causes the Cisco CallManager PIM to pop the target agent's desktop with call data and to enable their answer button (assuming auto-answer is not enabled). When the target agent clicks the answer button (or auto-answer is invoked), a voice path between the transferring agent and target agent is established (assuming the transferring agent has not clicked the transfer complete button).

Generally the transferring agent will not click the transfer complete button before the target agent answers because the probable reason they used consultative transfer was that they wanted to talk with the target agent before completing the transfer. However, the transferring agent can click on the transfer complete button at any time after it is enabled.

If the agent is transferring the call to a generic DN to find an available agent with a particular skill but no such agent is currently available, then the ICM routing script should be configured to route the call to an IVR for queuing. In this scenario, the transferring agent would hear the IP IVR queue announcements. The transferring agent could press the transfer complete button at any time to complete the transfer. The caller would then begin hearing the IP IVR queuing announcements. Upon availability of an appropriately skilled agent, the IP IVR transfers the call to this target agent and pops any call data onto their screen.

If the agent is transferring the call to a number that is not in the ICM Dialed Number Plan and a number that is not valid on the Cisco CallManager, the transferring agent will hear the failed consultation call and will be able to reconnect with the original caller, as explained in the section on [Reconnect](#), page 1-19.

Reconnect

During the consultation leg of a consultative transfer, the transferring agent can reconnect with the caller and release the consult call leg. To do so, the agent simply clicks the reconnect button. This action causes the agent desktop to instruct the Cisco CallManager PIM to instruct Cisco CallManager to release the consultation call leg and to reconnect the agent with the original caller.

This is basically the process an agent should use when they need to make a consultation call but do not plan to complete the transfer. After a call is successfully reconnected, the transferring agents desktop functionality will be exactly the same as before they requested the transfer and reconnected. Therefore, the transferring agent can later request another transfer, and there is no limit to the number of consult calls an agent can make.

Consultative transfers and reconnects are all done from the agent desktop and use the single Cisco CallManager extension that is associated with the IPCC. The IPCC system does not support allowing the transferring agent to place the original caller on hold and then use a second extension on

their hardware phone to make a consultation call. The hardware phone offers a button to allow this kind of transfer, but it is not supported in an IPCC environment. If an agent transfers a call in this way, any call data will be lost because the ICM did not route the call.

Alternate

Alternate is the ability for the agent to place the consultation call leg on hold and then retrieve the original call leg while in the midst of a consultative transfer. The agent can then alternate again to place the original caller back on hold and retrieve the consultation call leg. An agent can do this as many times as they would like.

When the transferring agent has alternated back to the original caller, the only call controls (buttons) that are enabled are Release and Alternate. The Transfer (Complete) and Reconnect controls will be disabled. The Alternate control will alternate the transferring agent back to talking with the consult party. When the agent has alternated back to the consultation leg, the Release, Alternate, Transfer, and Reconnect call controls will be enabled. The Alternate control will alternate the transferring agent back to talking with the original caller. The Transfer control will complete the transfer, The Reconnect button will drop the consulted party and reconnect the agent with the original caller.

Non-ICM Transfers

Transfers to numbers not in the DNP or to numbers configured in the DNP with post-route set to **No** are allowed but do *not* result in an ICM-routed call. In these scenarios, the PIM simply sends a call transfer request directly to Cisco CallManager and uses the dialed number from the transfer dialog on the agent desktop. Call data is lost if the ICM does not route the call. Cisco recommends that any dialed numbers for transfers have a match in the DNP, that it be marked for post-route, and that it have a DNP type that is allowed for the transferring agent (based on their agent desk settings).

Agent-to-Agent Transfers

If the transfer is to a specific agent, then the agent requesting the transfer must enter the agent ID into the transfer dialog box. The DNP entry matching the dialed number (agent ID) must have DNP type equal to PBX. This causes the PIM to place the dialed number (agent ID) into the CED field before it sends the route request to the ICM router. In the script editor, use the agent-to-agent routing node and specify the CED field as the location of agent ID so that the ICM router will route this call properly.

Agent IDs should not match any of the extensions on the Cisco CallManager cluster. If you begin all agent IDs with the same number and they all have the same length, you could set up a generic wildcard string that matches all agent IDs so that you need only one entry in the DNP for agent-to-agent routing.

If your environment has multiple PIMs, then you must use an agent ID number plan to determine which PIM contains this agent. Agent IDs by themselves are not unique. Agent IDs are associated with a specific PIM and can be reused on other PIMs. By not repeating agent IDs across the enterprise and by setting up a consistent agent ID assignment plan (such as all PIM 1 agent IDs begin with a 1, all PIM 2 agent IDs begin with a 2, and so on), you can parse the CED field in the script editor to determine which PIM contains the agent. The parsing may be done via a series of "if" nodes in the script editor or via a "route select" node. The agent-to-agent node requires the PIM to be specified.

In the event that the target agent is not in a ready state, the agent-to-agent script editor node allows alternative routing for the call.

Transferring from an IVR to a Specific Agent

Many contact centers often wish to prompt the caller for something like an order number and then route the call to the agent who is working that order. This can be done on IPCC by having the router do a database lookup and then placing the agent ID in any of the peripheral variable fields and then using the agent-to-agent script editor node to route the call to that particular agent. The agent-to-agent node must be configured to look for the agent ID value in the peripheral variable field where the IVR placed the agent ID. If multiple PIMs exist, then the same configuration as discussed in the previous section would be required.

This type of scenario could also be used to prompt a caller for a specific agent ID and then route the caller to that agent ID.

Transfer Reporting

After a call transfer is completed, a call detail record for the original call leg will exist and a new call detail record will be opened for the new call leg. The two call records are associated with one another via a common call ID assigned by the ICM. The time during the consultation call leg, before the transfer is completed, is considered as talk time for the transferring agent.

For more details, refer to the *IPCC Reporting Guide*, available online at Cisco.com.

Combination or Multiple Transfers

During a consultative transfer, the consulted target agent may transfer the consultation call to another target agent. Then, when the transferring agent presses the transfer complete button, the original caller will be connected to the second consulted target agent.

After a call has been successfully transferred, it can be transferred again. Each call leg generates a call detail record in the ICM, and the talk time during that call leg is associated with the agent who received that call. All call detail records are associated with one another via a common call ID assigned by the ICM. This allows complete cradle-to-grave reporting for the call.

Transfers of Conferenced Calls

After a conference has been set up by an agent, transfer is no longer a valid operation, even if the conferenced party has released.

PSTN Transfers (Takeback N Transfer, or Transfer Connect)

Many PSTN service providers offer a network-based transfer service. These services are generally invoked by the customer premises equipment (CPE) outputting a series of DTMF tones. The PSTN is provisioned to detect these tones and perform some specific logic based upon the tones detected. A typical outpulse sequence might be something like *827500. This DTMF string could mean, "transfer this call to site 2 and use 7500 as the DNIS value when delivering the call to site 2." IPCC has the ability to invoke these types of transfers.

Call Admission Control

Quality of Service (QoS) is a necessary part of a Voice over IP (VoIP) environment. QoS has various mechanisms to give voice traffic priority over data traffic, but QoS alone is not enough to guarantee good voice quality. What is needed is a way to make sure that the bandwidth allocated on the WAN link is not exceeded. Call admission control is a methodology for ensuring voice quality by limiting the number of active calls allowed on the network at one time.

When voice is enabled as an application on a data network, a certain amount of bandwidth should be allocated for voice traffic. This total voice bandwidth must be able to support the voice call itself plus any call control traffic. For information on how to calculate the required bandwidth for voice traffic, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

For IPCC, the call center should be able to determine its BHCC within the WAN and use the information to determine the bandwidth that is needed for its calls. This bandwidth should be added to the data traffic and any other voice traffic that is on the network. The sum of all these applications should not exceed 75% of the available WAN bandwidth. The capacity of the WAN depends on the network infrastructure. For details, refer to the *Cisco AVVID Network Infrastructure Enterprise Quality of Service Design* guide, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/qos_register.html

Call admission control makes sure that the active calls do not exceed the voice bandwidth allocation.

When a voice call is made, the bandwidth needed for that call is subtracted from the available voice bandwidth pool. When a call disconnects, the bandwidth that was used for that call is returned to the voice bandwidth pool. If the voice bandwidth pool is exhausted, then the next call request will be rejected due to insufficient bandwidth. The entity that controls and manages this bandwidth pool is called a gatekeeper. It is the gatekeeper's job to make sure that voice calls stay within the bandwidth allotment.

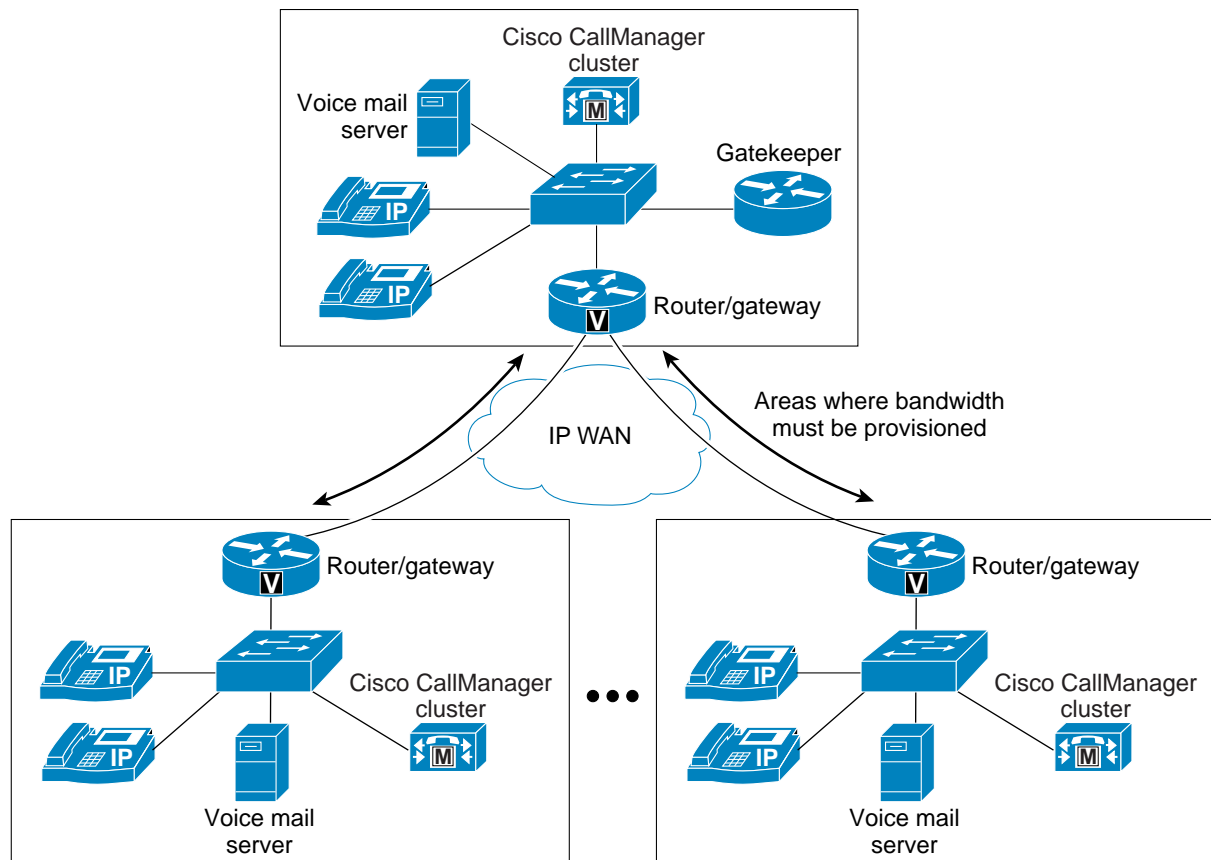
In a Cisco CallManager environment, there are two types of call admission control:

- [Gatekeeper Controlled, page 1-23](#)
- [Locations Controlled, page 1-24](#)

Gatekeeper Controlled

Gatekeeper control means that there is an independent entity acting as a gatekeeper. For distributed call processing deployments, the gatekeeper-controlled model is used. Before sending the call out the gateway or inter-cluster trunk, Cisco CallManager will ask the gatekeeper if there is enough bandwidth for the call to go through the WAN to another site. (See [Figure 1-6](#).)

Figure 1-6 Area of Concern for Distributed Call Processing Models



If the gatekeeper rejects the call, then Cisco CallManager can perform digit manipulation on the dialed digits and send this call transparently out the PSTN.

For IPCC, it is important to define this alternate route and digit manipulation within the dialing plan if the gatekeeper does not allow the call to go on the WAN. The reason this is important is that calls are sent to agents and IVRs via routing clients (CTI Desktop, IVR, or CTI Route Point), which are not able to hang up and redial the call. Therefore, the caller would receive busy tone and not be routed to its peripheral target (agent or IVR).

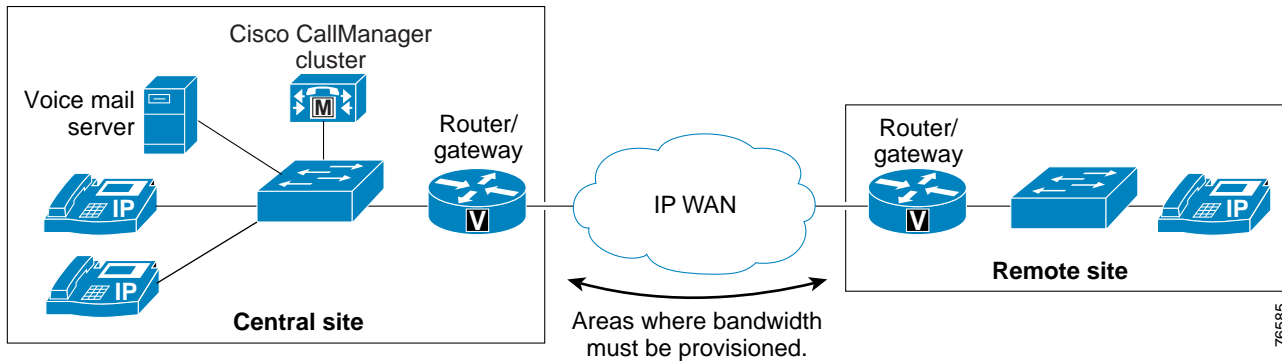
The ramification of having calls go out to the PSTN is that two ports are consumed because calls would have to come into and go out of the main or branch sites via voice gateway ports, which stay up if the call then gets transferred to another agent or IVR port at another site within the network.

The gatekeeper should be configured to allow enough bandwidth for call center traffic to go through. The total amount of bandwidth needed would depend on whether incoming traffic from the PSTN is routed through the WAN or if the WAN is used for inter-site transfers and conferences between agents.

Locations Controlled

For centralized call processing deployments, the locations-controlled model is used. In this model, Cisco CallManager (not the gatekeeper) decides if there is enough bandwidth available on the WAN to send the call. If there is not, then the call will fail. Transparent failover to the PSTN is not available with locations-based call admission control. (See [Figure 1-7](#).)

Figure 1-7 Area of Concern for Centralized Call Processing Models



For IPCC, this will result in the caller receiving busy tone because the call is routed by IVR or the CTI Desktop application, and there would not be a mechanism for the routing client to disconnect the call and then dial again.

Therefore, it is important to calculate the bandwidth allocation for each branch office properly. The number of simultaneous calls to each branch should be calculated. Inter-site transfer and conference situations as well as normal office traffic should also be taken into account. Ideally, agent phones should be allocated as one "location" within the location configuration of Cisco CallManager to make sure that traffic generated to and from office workers' phones do not interfere with the bandwidth allocated to call center traffic.



Deployment Models

There are numerous ways that IPCC can be deployed, but the deployments can generally be categorized into the following major types or models:

- [Single Site, page 2-2](#)
- [Multi-Site with Centralized Call Processing, page 2-4](#)
- [Multi-Site with Distributed Call Processing, page 2-9](#)

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

- Locations of IPCC servers and voice gateways
- Choice of inter-exchange carrier (IXC) or local exchange carrier (LEC) trunks
- Pre-routing availability
- IVR queuing platform
- Transfers
- Traditional ACD, PBX, and IVR integration
- Sizing
- Redundancy
- Placement of IPCC components (redundant ICM Central Controllers across a WAN)

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

A combination of these deployment models is also possible. For example, a multi-site 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 an IPCC deployment, with considerations on hybrid PBX/ACD deployments.

Sizing and redundancy are discussed in later chapters of this IPCC design guide.

For more information on the network infrastructure required to support an IPCC solution, refer to the *Cisco AVVID Network Infrastructure Enterprise Quality of Service Design* guide, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/qos_register.html

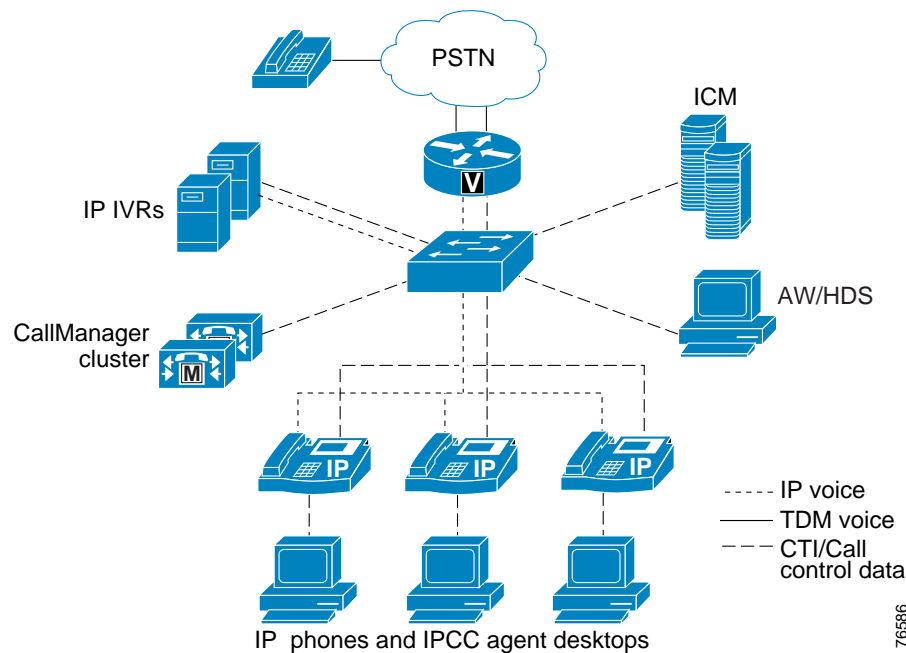
For more information on deployment models for IPCC and IP Telephony, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avid/ip_tel_register.html

Single Site

A single-site deployment refers to any scenario where all voice gateways, agents, desktops, IP Phones, and call processing servers (Cisco CallManager, ICM, and IP IVR) are located at the same site and have no WAN connectivity between any IPCC software modules. [Figure 2-1](#) illustrates this type of deployment.

Figure 2-1 Single-Site Deployment



[Figure 2-1](#) shows two IP IVRs, a two-node Cisco CallManager cluster, redundant ICM PROGGERS, an Administrative Workstation (AW) and Historical Data Server (HDS), and a direct connection to the PSTN from the voice gateways. The ICM PROGGER in this scenario is running the following major software processes:

- Router
- Logger
- Cisco CallManager Peripheral Interface Manager (PIM)
- Two IVR PIMs
- CTI Server
- CTI Object Server (CTI OS)

Within this model, many variations are possible. For example, the ICM Central Controller and Peripheral Gateways (PGs) could be split onto separate servers. For information on when to install the ICM Central Controller and PG on separate servers, refer to the chapter on [Sizing IPCC Components and Servers](#).

The ICM could also be deployed in a simplex fashion instead of redundantly. For information on the benefits and design for IPCC redundancy, refer to the chapter on [Design Considerations for High Availability](#).

The number of Cisco CallManager nodes and the hardware model used is not specified along with the number of IP IVRs. For information on determining the number and type of servers required, refer to the chapter on [Sizing IPCC Components and Servers](#).

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 IP Telephony design guides are available to assist in the design of these components. The chapter on [Sizing IP Contact Center Resources](#) discusses how to determine the number of gateway ports. The chapter on [Voice Gateway Considerations for IPCC](#) also discusses some specifics to contact center deployments.

Another variation in 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, page 2-13](#), and [Traditional IVR Integration, page 2-14](#).

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 or MGCP) to be used between the voice gateway and Cisco CallManager. The chapter on [Voice Gateway Considerations for IPCC](#) shows which combinations are supported.

The amount of digital signal processor (DSP) resources required for holds, consultative transfers, and conferencing is also not specified in this model. For information on sizing of these resources, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*.

The main requirement in 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 would not be required.

While WAN connectivity is not required in this deployment model, using Quality of Service (QoS) in your network infrastructure is still required. To ensure QoS is implemented properly, refer to the *Cisco AVVID Network Infrastructure Quality of Service Design* guide.

Queuing

In this deployment model, all initial and subsequent queuing is done on the IP IVR. If multiple IP IVRs are deployed, the ICM should be used to load-balance calls across those IP IVRs.

Transfers

In this deployment model (as well as in the multi-site centralized call processing model), both the transferring agent and target agent are on the same PIM. This also implies that both the routing client and the peripheral target are the same peripheral (or PIM). The transferring agent generates a transfer to a particular Dialed Number (for example, looking for any specialist in the specialist skill group). Assuming a match is found in the Dialed Number Plan (DNP) for the transfer request, the DNP type is allowed for the transferring agent, and the post-route option is set to **yes**, the Cisco CallManager PIM logic will generate a route request to the ICM router. The ICM router will match the dialed number to a call type and activate the appropriate routing script. The routing script looks for an available specialist.

If a target agent (specialist) is available to receive the transferred call, then the ICM router will return the appropriate label to the routing client (the Cisco CallManager PIM). In this scenario, the label is typically just the extension of the phone where the target agent is currently logged in. Upon receiving the route response (label), the Cisco CallManager PIM will then initiate the transfer by sending a JTAPI transfer request to the Cisco CallManager.

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 Cisco CallManager PIM. This is because the transferring agent and the target agent are both associated with the same PIM. In some of the more complex scenarios to be discussed in later sections, sometimes the routing client and peripheral target are not the same.

If a target agent is not available to receive the transferred call, then the ICM routing script is typically configured to transfer the call to an IVR so that queue treatment can be provided. In this scenario, the label is a dialed number that will instruct the Cisco CallManager to transfer the call to an IVR. Also in this scenario, the routing client and peripheral target are different. The routing client is the Cisco CallManager PIM, while the peripheral target is the specific IVR PIM to which the call is being transferred.

Multi-Site with Centralized Call Processing

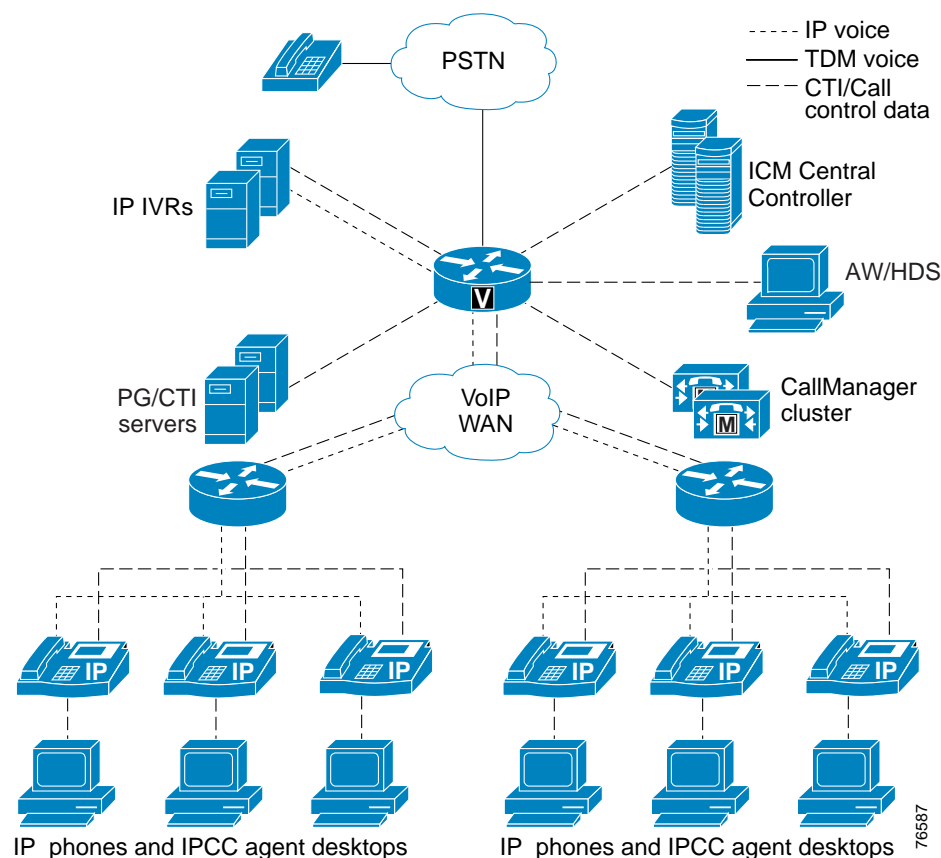
There are two variations of this model:

- [Centralized Voice Gateways, page 2-5](#)
- [Distributed Voice Gateways, page 2-7](#)

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 disperse, use of distributed voice gateways may be better. Figure 2-2 illustrates this model.

Figure 2-2 Multi-Site Deployment with Centralized Call Processing and Centralized Voice Gateways



Advantages of this model include the following:

- Only small data switch and router, IP Phones, and agent desktops are needed at remote sites where only a few agents exist, and 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.
- PSTN trunks are used more efficiently because the trunks for small remote sites are aggregated.
- IPCC Queue Points (IP IVR) are used more efficiently because all Queue Points are aggregated.
- No VoIP WAN bandwidth is used while calls are queuing (initial or subsequent). This provides "queuing at the edge" capability.

As with the single site deployment model, all the same variations exist. For example, multi-site deployments can run the ICM software all on the same server or on multiple servers. The ICM software can be installed as redundant or simplex. The number of Cisco CallManager and IP IVR servers is not specified by the deployment model, nor is the LAN/WAN infrastructure, voice gateways, or PSTN connectivity. For other variations, see [Single Site, page 2-2](#).

Considerations

The following considerations apply to the multi-site model with centralized call processing:

- VoIP WAN connectivity is required for RTP traffic to agent phones at remote sites.
- RTP traffic to agent phones at remote sites may 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 upon how the IP Telephony deployment is designed.
- Skinny Client Control Protocol (SCCP) call control traffic from IP Phones to the Cisco CallManager cluster flows over the WAN.
- CTI data to and from the IPCC Agent Desktop flows over the WAN. Adequate bandwidth and QoS provisioning is 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 1-800 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, and this must be managed via the Cisco CallManager dial plan. In most cases, local trunks are configured to dial out locally and for 911 emergency calls.
- Cisco CallManager locations-based call admission control failure will result in a routed call being disconnected (rerouting within Cisco CallManager is not currently possible). Therefore, it is important to provision adequate bandwidth to the remote sites. Also, an appropriately designed QoS WAN is critical.

Queuing

As in the single-site deployment, all call queuing is done on the IP IVR at a single central site. While calls are queuing, no RTP traffic flows over the WAN. If requeuing 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.

Transfers

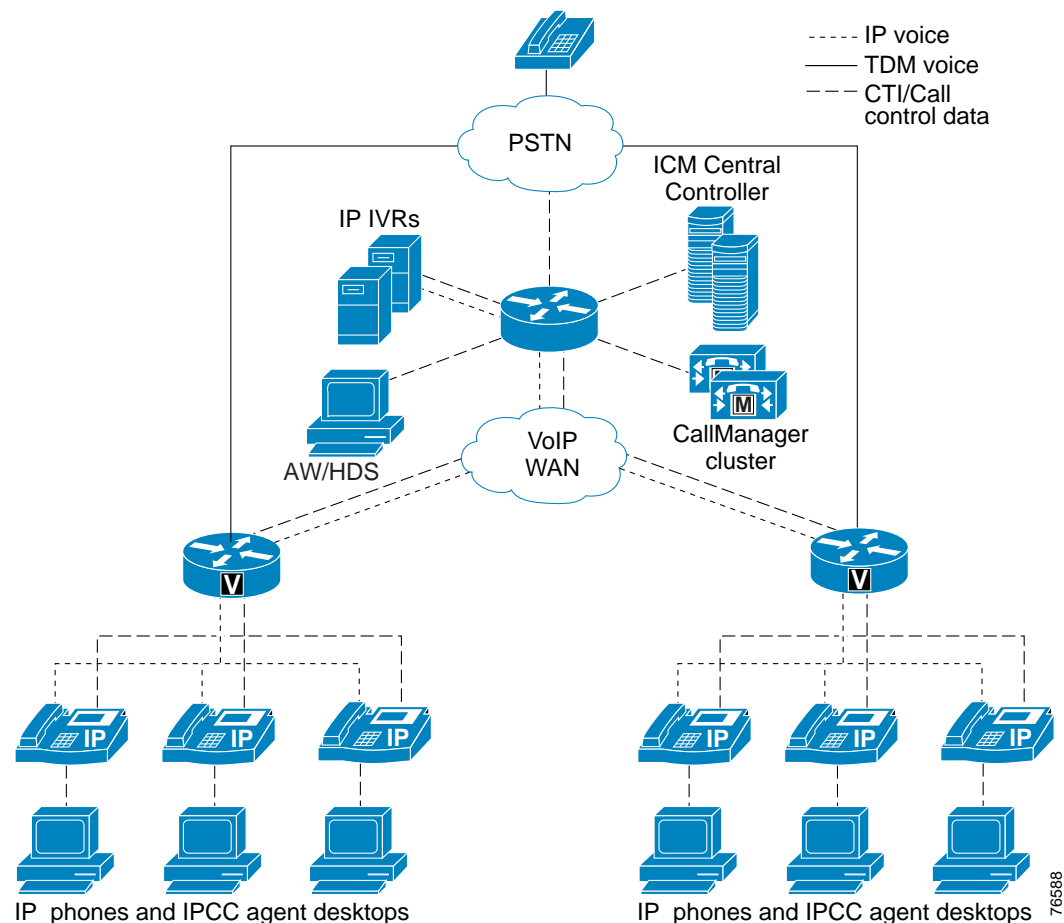
In this scenario the transferring agent and target agent are on the same Cisco CallManager cluster and Cisco CallManager PIM. Therefore, the same call and message flows will occur as in the single-site model, whether the transferring agent is on the same LAN as the target or on a different LAN. The only difference is that QoS must be enabled and that appropriate LAN/WAN routing must be established. For details on provisioning your WAN with QoS, refer to the *Cisco AVVID Network Infrastructure Quality of Service Design* guide.

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

Distributed Voice Gateways

A variation of the centralized call processing model is one with multiple voice gateway locations. This model variation may 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 2-3](#) illustrates this model.

Figure 2-3 Multi-Site Deployment with Centralized Call Processing and Distributed Voice Gateways



In this deployment model, 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:

- VoIP WAN bandwidth is reduced.
- Customer service levels for those callers arriving into that site might suffer due to longer queue times and handle times.
- Longer queue times can occur because, even though an agent at another site is available, the IPCC configuration may continue to queue for an agent at the local site only.
- Longer handle 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.

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.

An IPCC 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. Identify who is calling, why they are calling, where they are calling from, and how they are calling.

In multi-site deployments with distributed voice gateways, the ICM's pre-routing capability can also be used to load-balance calls dynamically across the multiple sites. A list of PSTN carriers that offer ICM pre-routing services can be found in the ICM product documentation available at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/>

In multi-site environments where the voice gateways have both local PSTN trunks and separate toll-free trunks delivering contact center calls, the 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 many calls in queue from locally originated calls and Site 2 has only a few calls in queue or maybe even a few agents currently available. In that scenario, you could have the ICM instruct the toll-free provider to route most or all of the toll-free calls to Site 2. This type of multi-site load balancing provided by the ICM is dynamic and automatically adjusts as call volumes at all sites change.

Just as in the two previous deployment models, much variation exists in number and type of ICM, Cisco CallManager, and IP IVR servers; LAN/WAN infrastructure; voice gateways; PSTN connectivity; and so forth.

Advantages of this deployment model are as follows:

- Only limited remote site systems management skills are needed because most servers, equipment, and system configurations are managed from a centralized location.
- The ICM pre-routing option can be used to load balance calls across sites, including sites with local PSTN trunks in addition to 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.
- Centralized IP IVRs provide efficiency of IP IVR ports when compared with smaller deployments of IP IVRs at each remote site.

Considerations of this deployment model are as follows:

- H.323 or MGCP signaling traffic between the voice gateways and the centralized Cisco CallManager servers will flow over the WAN. Proper QoS implementation on the WAN is critical, and signaling delays must be within tolerances listed in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at
http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html
- Queuing treatment is provided from centralized IP IVR servers, which can only generate G.711 traffic. Thus, transcoding to G.729 traffic to reduce WAN bandwidth to the remote sites is probably desirable. Sizing of the max number of simultaneous IP IVR queuing treatments to each remote site

is required to size your bandwidth for RTP and the number of transcoding resources needed. As call queuing treatment and average speed of answer (ASA) increase, it may be desirable to look at the distributed call processing deployment model to reduce WAN bandwidth and transcoding resources.

- The IP IVR, Cisco CallManager, and PG (for both Cisco CallManager and IVR) must be co-located. Current product releases do not support separating them across a WAN. The only IPCC communications that can be separated across a WAN are the following:
 - ICM Central Controller to ICM PG
 - ICM PG to IPCC Agent Desktops
 - Cisco CallManager to voice gateways
 - Cisco CallManager to IP Phones
- 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. Cisco CallManager locations-based call admission control failure will result in a routed call being disconnected (rerouting within Cisco CallManager 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.

Queuing

The IP IVR has the ability to generate only G.711 RTP traffic. In order to pass RTP traffic to the remotely located voice gateway using G.729, the G.711 RTP stream from the IP IVR must be transcoded at the central site.

Transfers

Intra-site or inter-site transfers using the VoIP WAN to send the RTP stream from one site to another will 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 PSTN transfer service. These services allow the IPCC voice gateways to output DTMF tones to instruct the PSTN to reroute (transfer) the call to another voice gateway location. In order for this to work, each site would have to be configured within the ICM as a separate peripheral. This is because the label is what indicates whether a transfer is intra-site or inter-site using Takeback N Transfer (TNT).

Multi-Site 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 Cisco CallManager cluster, IP IVRs, 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).

Regardless of how many sites are being deployed, there will still be only one logical ICM Central Controller. If the ICM Central Controller is deployed with redundancy, side A and B can be deployed side-by-side or geographically separated (remote redundancy). For details on remote redundancy, refer to the ICM product documentation available at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/>

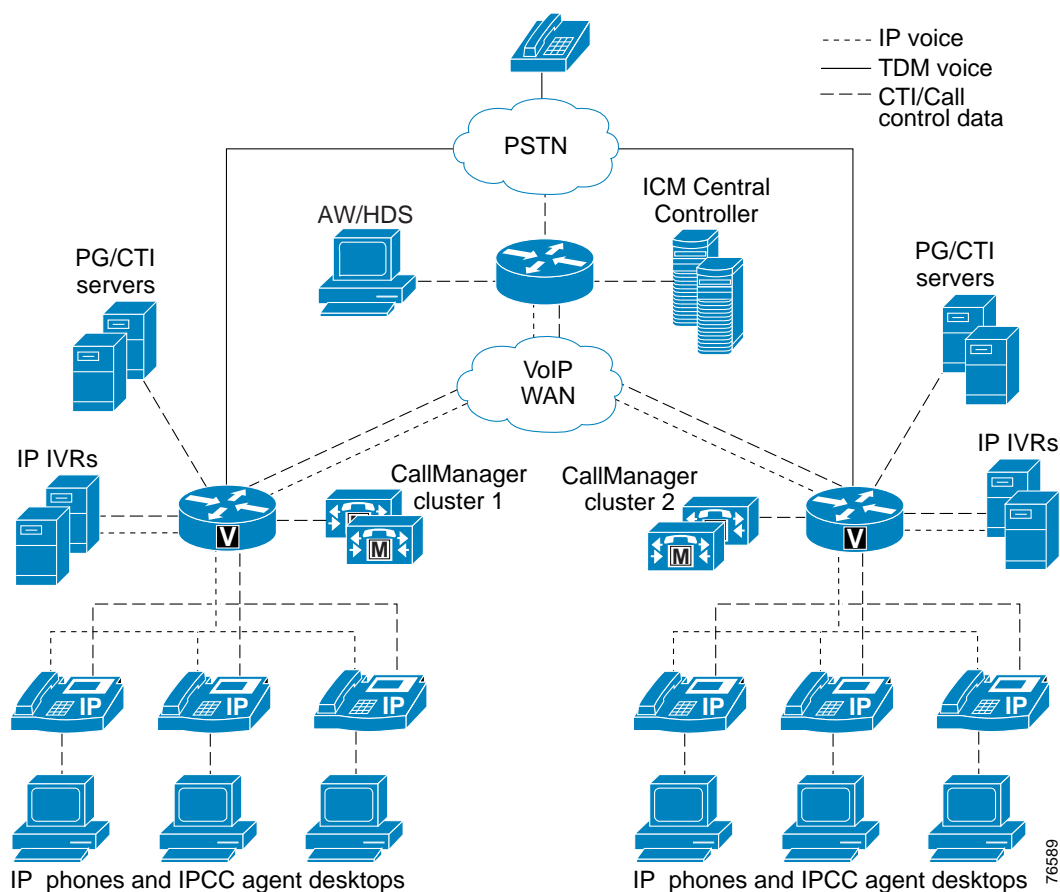
As with the centralized call processing model, there are two variations of the distributed call processing model:

- [Distributed Voice Gateways, page 2-10](#)
- [Centralized Voice Gateways, page 2-12](#)

Distributed Voice Gateways

This deployment model is a good choice if the company has multiple medium to large sites. Voice gateways with PSTN trunks terminate into each site. Just as in the centralized call processing model with distributed voice gateways, it may 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 recommended. [Figure 2-4](#) illustrates this model.

Figure 2-4 Multi-Site Deployment with Distributed Call Processing and Distributed Voice Gateways



As with the previous models, many variations are possible. The number and type of ICM Servers, Cisco CallManager servers, and IP IVR servers can vary. LAN/WAN infrastructure, voice gateways, PSTN trunks, redundancy, and so forth are also variable within this deployment model. In addition, the usage of a pre-routing PSTN Network Interface Controller (NIC) is also an option.

Advantages of this model are as follows:

- Each independent site can scale to support up to 1000 agents per Cisco CallManager cluster, and there is no software limit to the number of sites that can be combined by the ICM Central Controller to produce a single enterprise-wide contact center.
- All or most VoIP traffic can be contained within the LAN of each site, if desired. The QoS WAN shown in [Figure 2-4](#) would be required for voice calls to be transferred across sites. Usage 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.
- ICM pre-routing can be used to load-balance 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 ICM Central Controller provides centralized management for configuration of routing for all calls within the enterprise.
- The ICM Central Controller provides the capability to create a single enterprise-wide queue.
- The ICM Central Controller provides consolidated reporting for all sites.

Considerations of this model are as follow:

- The PG, Cisco CallManager cluster, and IP IVR must be co-located.
- The ICM Central Controller to PG communication link needs to be properly sized and provisioned for bandwidth and QoS. (For details, refer to the chapter on [Bandwidth Provisioning and QoS Considerations](#).)
- Gatekeeper-based Call Admission Control could be used to reroute calls between sites over the PSTN when WAN bandwidth is not available. However, no formal testing of this scenario has been done. It is best to ensure adequate WAN bandwidth exists between sites for the maximum amount of calling that may occur.
- If the communication link between the PG and the ICM CC is lost, then all contact center routing for calls at that site is lost. Therefore, it is important that a fault-tolerant WAN is implemented. Even when a fault-tolerant WAN is implemented, it is important to identify contingency plans for call treatment and routing when the ICM Central Controller to PG communication is lost. For example, in the event of a lost ICM Central Controller connection, the Cisco CallManager CTI Route Points could send the calls to IP IVR ports to provide basic announcement treatment or invoke a PSTN transfer to another site. Another alternative is for the Cisco CallManager cluster to route the call to another Cisco CallManager cluster that may have a PG with an active connection to the ICM Central Controller.
- 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 a logical hair-pinning and reducing the number of inter-cluster trunks by two).

Queuing

Initial call queuing is done on an IP IVR co-located with the voice gateways, so no transcoding is required. When a call is transferred and subsequent queuing is required, the queuing should be done on an 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, the call should be queued to an 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 IP Phone at Site 1. However, the two Cisco CallManager clusters would still logically see two calls in progress between the two clusters.

Transfers

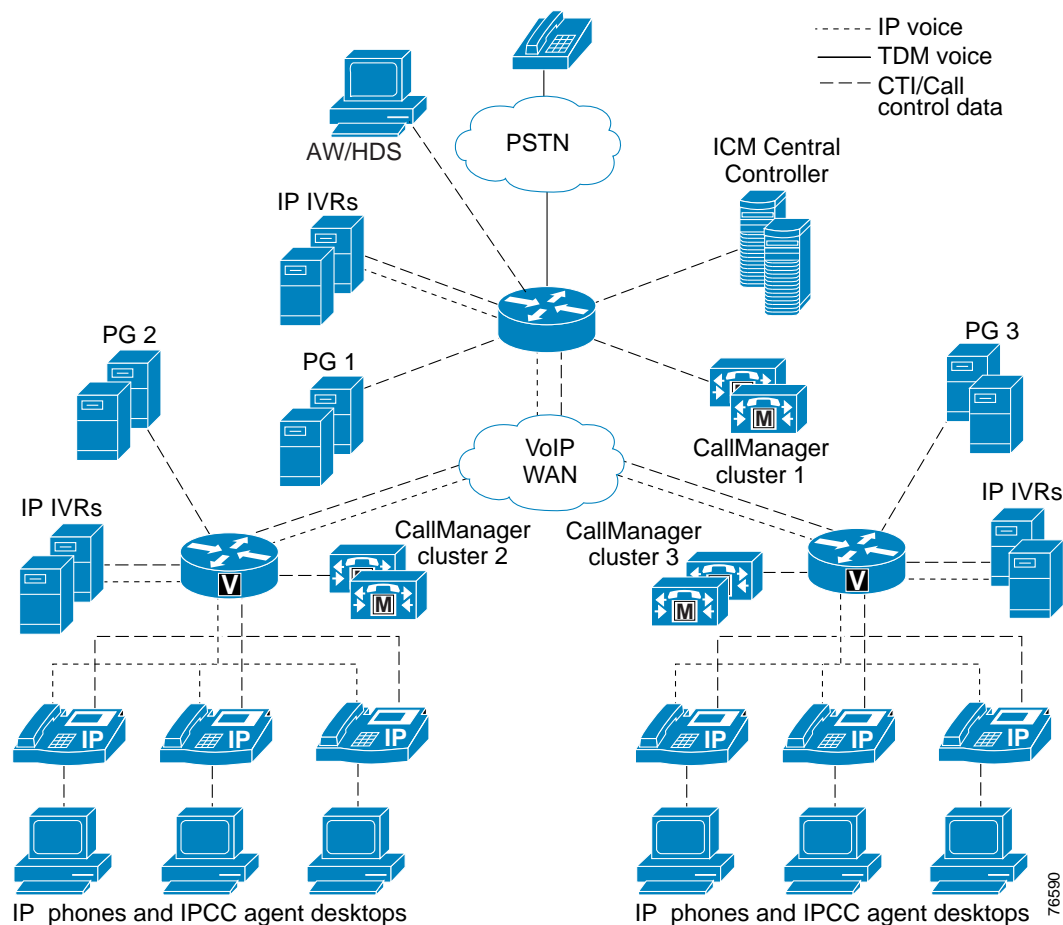
Transfers within a site function just like a single-site transfer. Transfers between Cisco CallManager clusters use either the VoIP WAN or a PSTN service.

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 IPCC voice gateways to outpulse DTMF tones to instruct the PSTN to reroute (transfer) the call to another voice gateway location. Another alternative is to have the Cisco CallManager 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.

Centralized Voice Gateways

This deployment model is really just an extension of the previous model. (See [Figure 2-5](#).)

Figure 2-5 Multi-Site Deployment with Distributed Call Processing and Centralized Voice Gateways



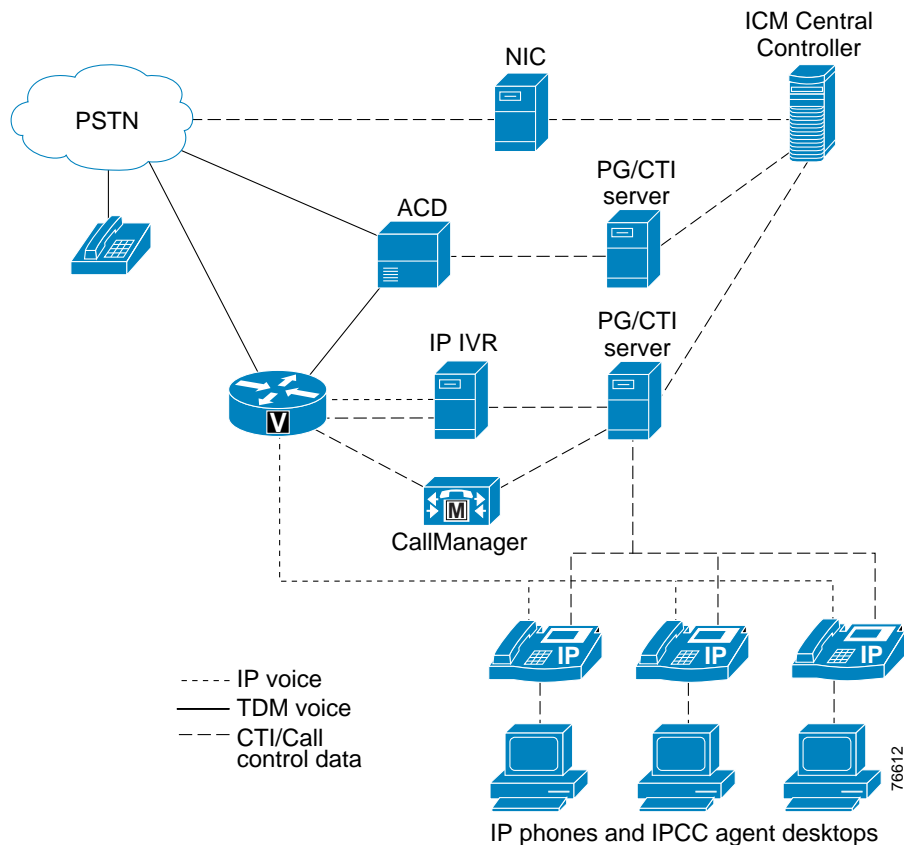
This model can often co-exist with a deployment where there are voice gateways at each remote site. The voice gateways at the remote sites would receive inbound calls made using a locally dialed number to that site. The voice gateways at the central site would receive inbound calls made using a generic toll-free number. Those calls would be distributed from the centralized site to the remote sites according to the ICM routing scripts. Calls arriving at the centralized site(s) are queued at the central site (edge) then, when agents become available, an inter-cluster call to that Cisco CallManager site is made.

Because this model is really just a combination of the previous models, Cisco recommends that you review the entire contents of this chapter before finalizing your design.

Traditional ACD Integration

For enterprises wishing to integrate traditional ACDs with their IPCC deployment, several options exist. For enterprises wishing to load-balance calls between a traditional ACD site and an IPCC site, a pre-routing NIC could be added. (See [Figure 2-6](#).) This requires that the ICM have a NIC that supports the PSTN service provider. In this scenario, the PSTN will query the ICM Central Controller (via the NIC) to determine which site is best, and the ICM response back to the PSTN will instruct the PSTN where (which site) to deliver the call. Any call data provided by the PSTN to the ICM will be passed to the agent desktop (traditional ACD or IPCC).

In order to transfer calls between the two sites (ACD site and IPCC 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 IPCC voice gateways. In that environment, any transferring 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 utilized.

Figure 2-6 Integrating a Traditional ACD with an IPCC Site

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 Cisco CallManager will generate a route request to the ICM 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. Determination of where calls should be routed, and if and when they should be transferred between sites, will depend upon the enterprise business environment, objectives, and cost components.

Traditional IVR Integration

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

Using PBX Transfer

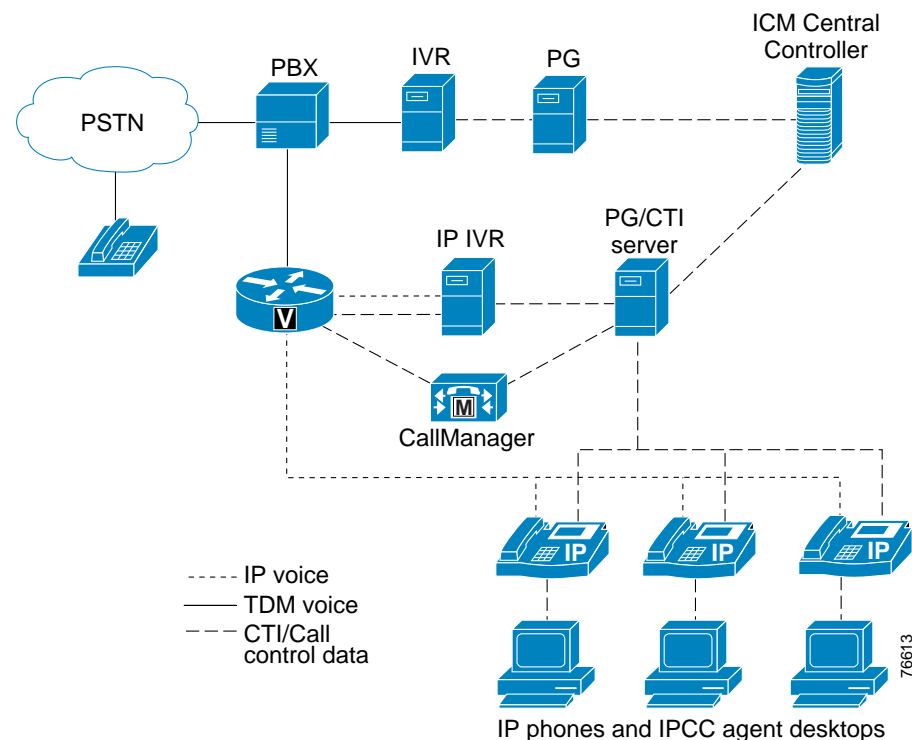
Many existing call centers have existing traditional IVR applications which they are not prepared to rewrite. In order to preserve these IVR applications, but yet integrate them into an IPCC environment, the IVR must have an interface to the ICM. (See [Figure 2-7](#).)

There are two versions of the IVR interface to the ICM. One is simply a post-routing interface, which just allows the IVR to send a post-route request with call data to the ICM. The ICM will return a route response instructing the IVR to transfer the call elsewhere. In this scenario, the traditional IVR will invoke a PBX transfer to release its port and transfer the call into the IPCC environment. Any call data passed from the IVR will be passed by the ICM to the agent desktop or IP IVR.

The other IVR interface to the ICM is the SCI interface. The SCI interface allows the IVR to receive queuing instructions from the ICM. In the PBX model, the SCI interface is not required.

Even if the IVR has the SCI interface, Cisco still recommends that you deploy IP IVR for all call queuing because this prevents any additional utilization of the traditional IVR ports. In addition, usage of the IP IVR for queuing provides a way to requeue calls on subsequent transfers or RONA treatment.

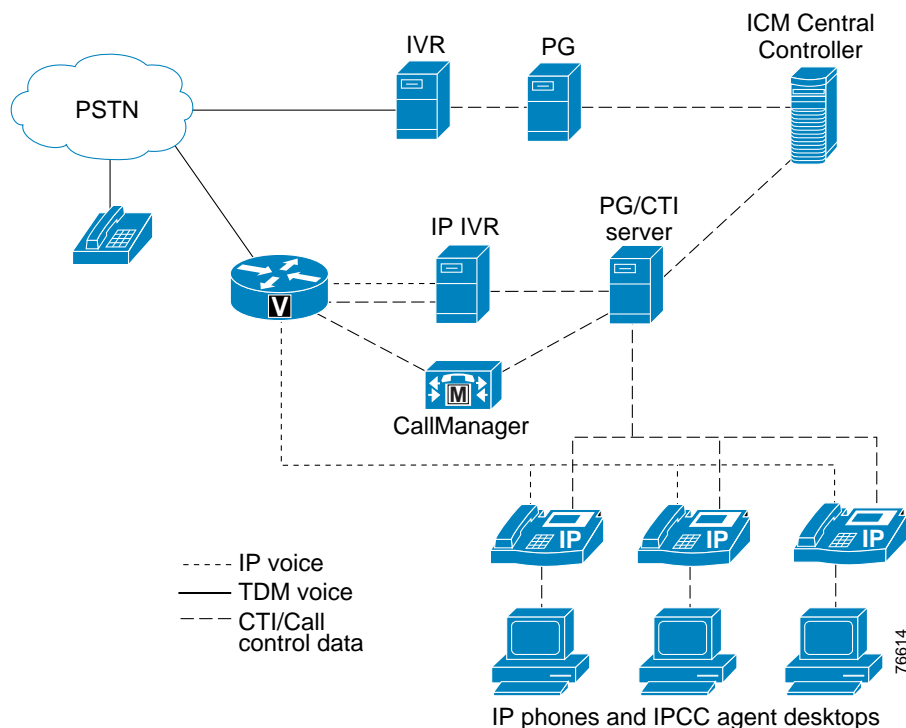
Figure 2-7 Traditional IVR Integration Using PBX Transfer



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 2-8](#).) Again, the IP IVR would be used for all queuing so that any additional occupancy of the traditional IVR ports is not required and also so that any double trunking in the IVR is avoided. Any call data collected by the traditional IVR application will be passed by the ICM to the agent desktop or IP IVR.

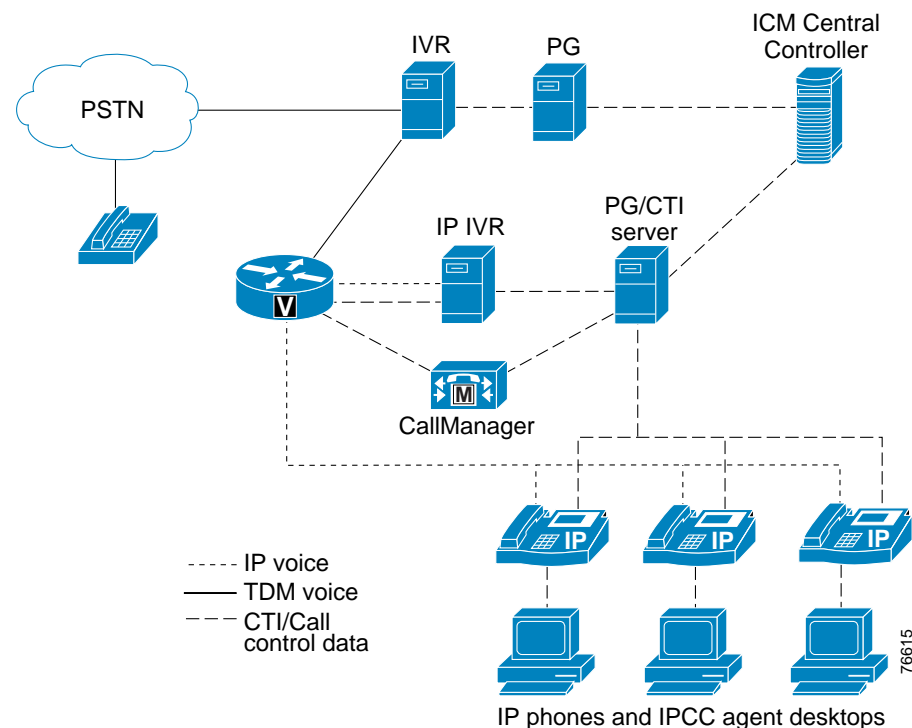
Figure 2-8 Traditional IVR Integration Using PSTN Transfer



Using IVR Double Trunking

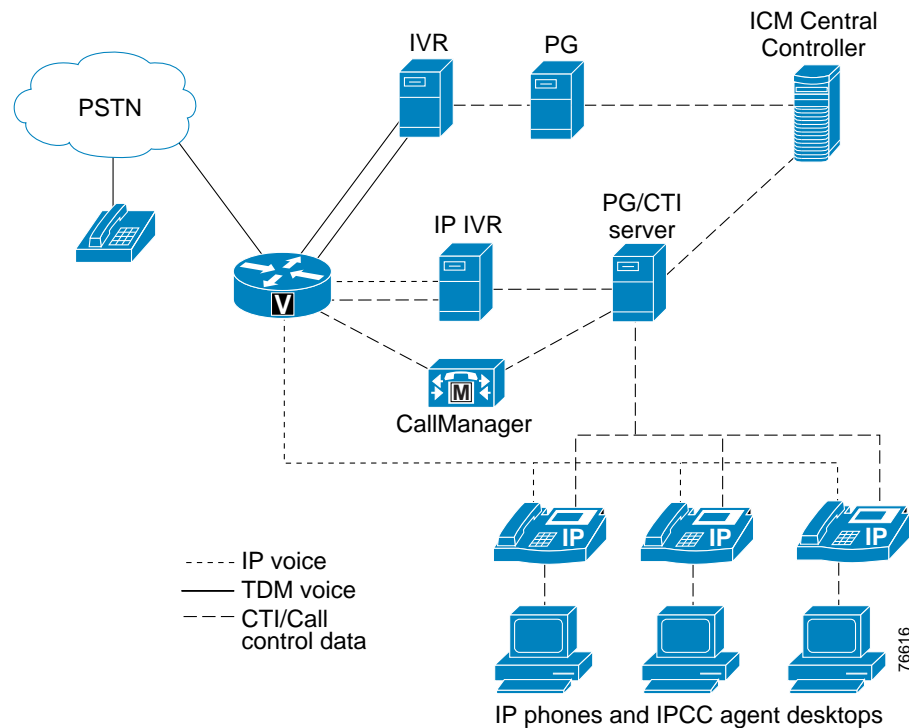
If your traditional IVR application has a very high success rate, where most callers are completely self-served in the traditional IVR and only a very small percentage of callers ever need to be transferred to an agent, then it may be acceptable to double-trunk the calls in the traditional IVR for that small percentage of calls. (See [Figure 2-9](#).) Unlike the previous model, if the traditional IVR has a SCI interface, 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 IP IVR, a second traditional IVR port would be used. 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 IP IVR to avoid any double trunking. If the traditional IVR does not have a SCI interface, then the IVR will just generate a post-route request to the ICM to determine where the call should be transferred. All queuing in that scenario would have to be done on the IP IVR.

Figure 2-9 Traditional IVR Integration Using IVR Double Trunking



Using Cisco CallManager Transfer and IVR Double Trunking

Slowly over time, it may be desirable to begin migrating the traditional IVR applications to the 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 2-10](#).) Calls arriving at the voice gateway from the PSTN would be routed by Cisco CallManager. Cisco CallManager could route specific DN's to the traditional IVR or let the ICM or IP-VR determine when to transfer calls to the traditional IVR. If calls in the traditional IVR need to be transferred to an IPCC agent, then a second IVR port, trunk, and voice gateway port would be used for the duration of the call. Care should be taken to ensure that transfer scenarios do not allow multiple loops to be created because voice quality could suffer.

Figure 2-10 Traditional IVR Integration Using Cisco CallManager Transfer and IVR Double Trunking

Hybrid IP Telephony and IPCC System

It is possible for a Cisco CallManager cluster to support IP Phones with both normal IP Telephony (office) extensions and IPCC (call center) extensions. For more information on this type of deployment, see [Hybrid IP Telephony and IPCC Cisco CallManager Clusters, page 1-15](#).



Voice Gateway Considerations for IPCC

After you decide which IPCC deployment model to use, the next step is to decide where the voice gateways should be positioned within the deployment model. Voice gateways are used as a bridge or interface between the Time Division Multiplexing (TDM) and IP networks.

For more information on how to choose a voice gateway, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html



Note

Cisco CallManager is an essential component of the IPCC solution. If you choose a voice gateway with a feature or capability that is not supported on Cisco CallManager, that feature or capability will not be available within the IPCC solution. The information in this chapter applies to Cisco CallManager Release 3.1 and later.

Within the IPCC environment, voice gateways connect the LAN to one of the following:

- Public Switched Telephone Network (PSTN)
- Private Branch Exchange (PBX) or automatic call distribution (ACD) system
- TDM on-premise interactive voice response (IVR) system
- Remote at-home agents

PSTN

All deployment models require this type of gateway unless the IPCC is behind a PBX.

PBX or ACD

This type of gateway is needed for a single-site deployment where the IPCC is behind a PBX. Some ACD systems are contained within, or work in conjunction with, the PBX (for example, Meridian, Definity, and Symposium), but some ACDs are standalone (such as Aspect or Rockwell).

TDM IVR

On-premise TDM IVRs can be connected via analog FXS or T1/E1 interfaces to Cisco gateways running Media Gateway Control Protocol (MGCP). The blind transfer operation (hookflash transfer) is currently supported only on the FXS interface. The FXS interface supports both dual-tone multifrequency (DTMF) and dial-pulse dialing.

Recommended Voice Gateways

Cisco recommends the following voice gateways for IPCC:

- IOS gateways:
 - Cisco 827-4V — Small office gateway that provides QoS.
 - Cisco 1751 — Small office voice gateway that provides QoS.
 - Cisco IAD2400 Series — Small office voice gateway that supports one T1 CAS/PRI line (24 channels) or eight FXO ports.
 - Cisco 2600, 3620, 3640, 3660, and 3700 — The gateway runs on a Voice Interface Card (VIC) module within a Cisco router.
 - Cisco VG200 — Standalone gateway.
 - Cisco MC3810-V3 — Entry level enterprise multiservice standalone gateway that provides improved performance over the Cisco 3810 gateway.
 - Cisco AS5400, AS5350, AS5850 — Gateway runs on the Network Access Service platform.
 - Cisco 7200 Multiservice Gateway — Contains voice gateway capability integrated into a router.
 - Cisco 7500 Series Router — Contains voice gateway capability integrated into a router.
- Integrated switch gateways:

For ease of deployment, integrated switch gateways (ISGs) are all-in-one hardware platforms that perform switching as well as gateway functions:

- Catalyst 4000 — Access gateway module that plugs into the Catalyst 4000 chassis and supports both analog and digital interfaces.
- Catalyst 4224 — Voice interface card (VIC) that plugs into the chassis and supports both analog and digital interfaces.
- Catalyst 6000 — Access gateways module that plugs into the chassis and supports both analog and digital interfaces.

For more information on a particular gateway, refer to the specific product documentation available online at Cisco.com.

Gateways That Are Not Recommended

Cisco recommends that you do *not* use the following voice gateways in an IPCC deployment:

- Cisco 1750

This gateway has been replaced by the Cisco 1751, which has better performance, value, and functionality.
- Cisco 3810

The performance of the Cisco 3810 gateway is not as good as the other H.323 gateways, so it is not recommended for use within an IPCC environment. However, the Cisco MC3810-V3 is recommended for IPCC because it has improved performance.
- Cisco VG248 Analog Gateway

This gateway supports analog phones with Cisco CallManager Release 3.1, but analog phones connected to this gateway are *not* supported as phones for IPCC agents.

- Cisco AS5800 and AS5300

The AS5300 and AS5800 are being discontinued because the new Cisco AS5000 Universal gateways (AS5400 and AS5850) are better able to meet the needs of Cisco voice and dial gateway customers. If your deployment already includes AS5300 gateways, they will support IPCC and do not have to be replaced, but AS5300 gateways are not recommended for new installations.

- Cisco Access DT-24+ and DE-30+

The Cisco Access DT-24+ and DE-30+ are no longer supported beginning with Cisco CallManager Release 3.1 and beyond.

- Cisco ATA 186 Analog Telephone Adapter

This gateway enables analog phones to interface with an IP telephony network. However, the analog phones connected to this gateway are not supported as phones for IPCC agents.

- Cisco ICS 7750 Integrated Communications System

This platform is a multi-service router processor with various cards for Cisco CallManager, voice gateways, router functions, voice mail, and contact center applications for up to 150 users. The contact center is a low-end ACD product called Integrated Contact Distribution (ICD). ICD is mutually exclusive with IPCC.

The ICS 7750 Analog Station Interface (ASI) card provides an FXS interface, but IPCC agents cannot use analog phones connected to this gateway.



Note

If a particular gateway is not listed as a supported IP telephony gateway in the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, do not use that gateway for IPCC deployments either. The IP Telephony design guide is available at http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Selection Criteria

You can apply the following selection criteria to refine your choice of voice gateways:

- [Port Density, page 3-3](#)
- [VoIP Interfaces, page 3-5](#)
- [TDM Interfaces, page 3-5](#)
- [Features, page 3-7](#)

Port Density

Port density is the number of available gateway ports per gateway. To determine the port density of a particular gateway, refer to the product documentation for that gateway, available online at Cisco.com.

The number of gateway ports needed for your deployment depends on the number of incoming trunks times the number of gateway ports that are consumed for each call. Incoming trunks are calculated based on busy hour call attempts (BHCA), average IVR treatment time (self-service or information-gathering applications), queue time, and talk time.

For more information on calculating trunk and gateway port requirements, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Once you have determined the total number of gateway ports needed per gateway deployment, the next step is to look at the port density of each of the gateways. Always allow for one to two years of growth by buying gateways with higher port density than is currently needed. Base your calculations on the normal office traffic that comes into each site. Pick some gateways that fulfill the port density requirements and then move on to the next selection criteria.

Number of Gateway Ports Needed per Call

Use call flows to determine how many gateway ports are needed per call. Once you have chosen an IPCC deployment model, call flows can help validate that choice. If the network is pure IP and does not include any TDM equipment, then only one voice gateway port is needed per call. However, this is not usually the case in a hybrid environment, especially if a lot of transfers and conferences take place across the IP and TDM networks. As much as possible, try to avoid double trunking at the gateways.

For help in determining the number of gateway ports needed for your particular IPCC deployment, consult your Cisco Systems Engineer (SE).

Media Resources

A media resource is a collection of digital signal processors (DSPs) that perform a single function such as transcoding or conferencing. The DSPs used for these functions are called *media resources* and are grouped together to form *media resource groups*. These media resources consume gateway port capacity. Therefore, to ascertain your complete gateway port capacity, you have to take these features into consideration. For more details, refer to the *Transcoding, Conferencing, and MTP Resources* chapter of the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Transcoding

Transcoding takes the output stream of one codec (for example, G.711) and converts it in real time to an input stream of another compression type (for example, G.729).

Because IP IVR supports only G.711 encoding, all IPCC sites that use IP IVR should use G.711 within their local network. However, for multi-site deployments, Cisco recommends that you use G.729 for WAN traffic. When a voice packet leaves the source site, the voice gateway must transcode the voice packet from G.711 to G.729; and when a voice packet arrives at the destination site, the destination voice gateway must transcode the voice packet from G.729 back to G.711.

For a list of the codecs supported for voice compression on the various gateway platforms, refer to the *Transcoding, Conferencing, and MTP Resources* chapter of the *Cisco IP Telephony Solution Reference Network Design Guide*.

A-Law to Mu-Law Conversion

A-law to mu-law conversions are needed for calls that originate outside of North America and terminate within North America, or vice versa. Therefore, if your IPCC installation includes sites in North America and sites outside of the North America, then the voice gateways must be configured to do the necessary conversions.

VoIP Interfaces

Cisco CallManager supports two types of voice-over-IP (VoIP) protocols for communicating with voice gateways: MGCP and H.323.

MGCP has better call survivability, but H.323 supports more types of TDM interfaces. All gateways that support H.323 protocol must run on Version 2 in an IPCC environment because Cisco CallManager supports H.323v2. All gateways that support MGCP protocol must support Version 0.1 with PRI backhaul via TCP because that is the type of MGCP supported by Cisco CallManager.

For a list of the protocols supported by each gateway platform, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html



Note

Beginning with Cisco CallManager Release 3.1, Skinny Client Control Protocol (SCCP) is no longer supported for communicating with voice gateways.

TDM Interfaces

This section describes the various digital and analog interfaces used to connect to TDM equipment. For more information on the analog and digital interfaces supported by each gateway platform, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

PRI Interfaces

Even though PRI is a standard, there are many different implementations of it. (See [Table 3-1](#).) Some of these interfaces are based on T1 PRI, and some are based on E1 PRI. If the gateway supports T1 PRI, then it should support all of the interfaces defined for T1 PRI; if the gateway supports E1 PRI, then it should support all PRI interfaces defined for E1. Not every Information Element (IE) is supported for these interfaces, but basic call setup and teardown, calling line ID (CLID), and called number are supported.

Table 3-1 PRI Interfaces

PRI Interface	Type	Standard	Vendor Support	Countries
4EES, 5EES	T1		AT&T Class 4 and 5 switches	North America
National ISDN (NI-2)	T1	Bellcore SR3887	AT&T National ISDN switch	North America
DMS100, DMS250	T1		Nortel Networks Class 5 switch	North America
NTT	E1		NTT	Japan

Table 3-1 PRI Interfaces (continued)

PRI Interface	Type	Standard	Vendor Support	Countries
TS014	E1		TS014	Australia
NET5	E1	ETSI 300, Euro-ISDN E-DSS1	NET5	Europe, New Zealand, and Asia

One side of the PRI connection is the user side and the other side is the network side. This distinction is used to arbitrate during a glare situation, when both sides want to grab the same B-channel at the same time. When this condition occurs, the user side acquiesces to the network side. When connected to the PSTN, the gateway is usually the user side. If the gateway is connected to a PBX, then the user side or network side depends on the PBX configuration. For more information on the PRI interfaces, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

QSIG Interface

QSIG was developed to rectify all the different types of PRI in order to achieve Multi-Vendor ISDN PBX-based private networks.

QSIG is supported between the Central Office (CO) and PBX and between PBXs. Cisco gateways can support the tunneling of QSIG messages between PBXs. This feature allows PBXs to communicate via the WAN transparently. For example, two legacy call centers could use this functionality to communicate with each other without going through the PSTN.

Basic call setup and teardown is supported when the voice gateway interfaces with a Central Office or PBX that is using QSIG protocol. The gateway converts the QSIG H.931 messages to H.225 messages supported for H.323. QSIG supplementary services will be supported in a future release. For more information on QSIG, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

SS7

SS7 is a signalling protocol used between Class 5 central office or Class 4 toll switches or Inter Exchange Carrier switches. If you want to connect a VoIP network to the PSTN via a SS7 interface, you can do it with the Cisco SC2200 Signaling Controller. The Cisco SC2200 speaks a proprietary Q.931+ protocol to the Cisco AS5XXX gateway but speaks SS7 to the PSTN. However, the bearer trunks from the Class 4 office are connected directly to the AS5XXX gateway.

The SS7 Interconnect for Cisco Any Service, Any Port (ASAP) provides for the ports on the AS5XXX gateway to be configured for either voice or data calls. This feature allows for a single SS7 interface to provide for data, voice, fax, or wireless data service on the network.

SS7 is used mostly in hosted deployments whose VoIP network consists of Cisco CallManagers.

BRI

Basic Rate Interface (BRI) consists of two bearer channels and one D-channel. This type of interface was designed for at-home usage, but can also be used for a small business office. IPCC uses BRI for the at-home agent solution. For more information on BRI, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Features

Besides basic call setup and teardown, Cisco voice gateways support the sending and receiving of CLID and called numbers, supplementary services, and networking capabilities, as described in the following sections.

Calling and Called Number Support

There are three ways to get calling and called number support:

- Analog (FXO and FXS)
- Digital (T1 and E1)
- ISDN (PRI, BRI, and QSIG)

Analog

The calling number, or calling line ID (CLID), is sent as tones within the media stream (in-band). The CLID is received via Frequency Shift Keying (FSK) tones between the first two ring cycles (on-hook CLID) to analog sets. Because Cisco CallManager does not support analog sets, this feature is not supported in an IPCC environment.

The called number, or direct inward dialing (DID) number, is sent via DTMF tones after the wink signal is received at the originating end. To receive the called number, special hardware is needed (VIC-2 DID) for the FXO interface in the Catalyst 2600 Series switches. The FXS interface does not support DID, but the E&M interfaces support DID.

For more information on DID and CLID, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

If the CLID does not appear on the agent desktop, then either it was not sent to the gateway or the gateway is not configured properly. A Display IE delivery flag should be configured on the gateway to enable CLID.

MGCP gateways do not support analog CLID on Cisco CallManager because the messaging to provide CLID on MGCP is not implemented. (PRI CLID is supported because PRI messaging is backhauled to Cisco CallManager.)

Digital

For digital interfaces, the calling and called numbers are sent as in-band digits.

For CLID on a T1 CAS interface, either FG-D or FG-B is required. If FG-B is used, then the automatic number identification (ANI) CLID delimiter feature is required to get CLID. See [ANI/DNIS Delimiter, page 3-8](#), for more information on this feature.

For CLID on an E1 CAS interface, R2 is required.

DID is supported on E&M signaling.

For more information on R2, FG-D, and FG-B, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

If the CLID does not appear on the agent desktop, then either it was not sent to the gateway or the gateway is not configured properly. A Display IE delivery flag should be configured on the gateway to enable CLID.

Not all of these interfaces have been tested directly, but if the gateway passes CLID to Cisco CallManager, then the CLID should be passed through the JTAPI link and sent to ICM via a Route Request to be used within the ICM script.

ISDN

For ISDN interfaces, calling and called numbers are sent out-of-band.

All ISDN interfaces (PRI, QSIG, and BRI) provide calling and called number support. For more information on PRI, BRI, and QSIG, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Cisco recommends that you use ISDN interfaces (PRI or QSIG) if the call center wants to receive CLID consistently. Use BRI for the at-home agent.

For more information on supported interfaces, consult your Cisco Systems Engineer (SE).

Calling Party Name

Calling party name is supported on all ISDN interfaces (PRI and QSIG) within Cisco CallManager. On PRI gateways, enable the Display IE delivery flag on the configuration page.

Calling party name is not currently passed through the JTAPI link to ICM.

ANI/DNIS Delimiter

This feature allows the Cisco AS5300 and AS5800 universal access server gateways to provide the ANI and dialed number identification service (DNIS) delimiter on incoming T1 CAS trunk lines. The digit collection logic in the call switching module (CSM) for incoming T1 CAS calls in dual tone multifrequency (DTMF) is modified to process the delimiters, the ANI digits, and the DNIS digits. This feature allows the Cisco AS5300 and AS5800 to support CLID on Feature Group-B (FGB).

For this feature to work, you have to define a CAS signaling class with the template to process ANI/DNIS delimiters. This creates a signaling class structure that can be referred to by its name.

This feature functions only in a T1 CAS configured for E&M-FGB (wink start).

For more information, refer to *ANI/DNIS Delimiter for CAS Calls on CT1*, available at

<http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121newft/121t/121t1/anidnis.htm>

Presentation Indicator

The Presentation Indicator is an Information Element within the ISDN protocol that designates whether the caller wants his or her CLID to be displayed on the terminating set.

Even if the Presentation Indicator is set to Restricted, Cisco CallManager still passes the CLID information to IPCC on the JTAPI link.

ISDN Non-Facility Associated Signaling (NFAS)

NFAS is an ISDN PRI feature that supports:

- Multiple ISDN PRI spans controlled by a single D-channel
- A primary and backup D-channel

The connecting switch must support one of the PRI interfaces that support NFAS (4ESS, DMS250, DMS100, or NI), and the voice gateway must also be able to support NFAS.

For more information on the NFAS feature, refer to *NFAS with D Channel Backup*, available at

http://www.cisco.com/univercd/cc/td/doc/product/software/ios113ed/113t/113t_3/nfas.htm

Supplementary Services

MGCP and H.323v2 protocols support supplementary services. (QSIG supplementary services will be supported in a future release.) All H.323 voice gateways must be able to support H.323v2 to take advantage of the supplementary services. Supplementary services supported by Cisco CallManager and the voice gateways include:

- Hold
- Restore
- Transfer
- Conference

Transfer support enables a call from a voice gateway to be transferred seamlessly to an IP IVR or IPCC agent either on the same Cisco CallManager cluster or across the WAN, using only one incoming gateway port because Cisco CallManager instructs the gateway to move the media stream to a different destination. For more information on how transfer operates, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Conferencing is supported for calls coming into gateways because Cisco CallManager instructs the gateway to redirect the media stream to the conference bridge.

If an IP phone puts a call from a gateway on hold, Cisco CallManager instructs the gateway to temporarily disable the transmission of the media stream to the IP phone. Tone on hold or music on hold can be provided by the gateway to the caller. When an IP phone restores the held call, Cisco CallManager instructs the gateway to remove the tone or music on hold and to start retransmitting the media stream to the IP phone.

QoS

Quality of Service (QoS) is required to give voice traffic priority over data traffic. Cisco IP Phones and Cisco IP Softphones enable the Class of Service (CoS) and Type of Service (ToS) bits in the Layer 2 and Layer 3 packets, respectively. All Cisco voice gateways use these bits to give voice packets priority over data packets in various QoS prioritization schemes. For more information, refer to the *Network Infrastructure Requirements* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Ringback on Transfer

If an agent answers an incoming call from a gateway and then decides to transfer the call to another skill group, the caller will hear a ringback tone when the call rings on the other agent's phone. For H.323 gateways, Cisco IOS Release 12.2.2T or later is required. Ringback on transfer is also supported if a TDM IVR or legacy ACD agent initiates the transfer.

DTMF Relay

DTMF relay is used to send dialed digits during an active call. These digits are responses to prompts from an IVR device.

All of the gateways support conversion of in-band DTMF (from the TDM network) to out-of-band DTMF (to the IP network), and vice versa. When a person uses the dial pad on an IP phone, an out-of-band DTMF signal is sent to Cisco CallManager. If the IP phone is connected to a H.323 gateway, Cisco CallManager converts the out-of-band DTMF signal to an H.245v3 message and forwards it to the H.323 gateway. If the IP phone is connected to an MGCP gateway, Cisco CallManager sends the original out-of-band DTMF message to the gateway. The gateway then converts the out-of-band DTMF to in-band DTMF and sends it out the speech path within the TDM network.

In an IPCC environment that uses Network IVR with outpulse transfer, DTMF relay is required so that the IPCC agent desktop can let the Network IVR know to disconnect this leg of the call and either keep the call at the IVR pending further instructions from ICM or transfer the call to another agent. This type of protocol is known as Take Back and Transfer (TNT) and is supported on AT&T Toll Free Transfer Connect Service as well as Telera IVR.

In ICM 4.6, a DTMF label (prefixed by DTMF*) must be defined in ICM. When the agent initiates the transfer operation via the dialed number plan, the DTMF label is returned to the Cisco CallManager Peripheral Gateway (PG). This causes out-of-band DTMF digits to be sent to the gateway. The gateway then converts the DTMF digits to in-band DTMF and sends them out the speech path to the TDM IVR.

DTMF relay within the Real-Time Transport Protocol (RTP) stream (RFC 2833) is not supported by IPCC. Even though some gateways such as the Cisco 2600 Series do support RFC 2833, RFC 2833 requires H.323 v4 but Cisco CallManager supports only H.323 v1.

Hookflash Transfer

A hookflash is a brief on-hook condition during a call. The on-hook indication is not long enough to be interpreted as a signal to disconnect the call. Users can create a hookflash indication by quickly depressing and releasing the hookswitch on their phone.

PBXs and telephone switches are frequently programmed to intercept hookflash indications and use them as a way to allow a user to invoke supplemental services. For example, your local service provider might allow you to enter a hookflash as a means of switching between calls if you subscribe to a

call-waiting service. In a traditional telephony network, a hookflash results in a voltage change on the telephone line. Because there is no equivalent of this voltage change in an IP network, the ITU H.245 standard defines a message representing a hookflash. To send a hookflash indication using this message, an H.323 endpoint sends an H.245 user input indication message containing an H.245-signal or H.245-alpha structure with a value of "!". This value represents a hookflash indication.

To connect a TDM IVR directly to the voice gateway, it is desirable for the TDM IVR to send a hookflash to the voice gateway and have the voice gateway send an appropriate H.323 or MGCP transfer message to Cisco CallManager so that Cisco CallManager will transfer the media stream from the TDM IVR port on the gateway to the IPCC agent.

Hookflash relay is currently supported on an MGCP gateway for a FXS interface. Only blind transfer (complete transfer before the other party answers) is supported, and not consultative transfer (complete transfer after the other party answers).

Although, H.323 gateways support sending of the H.245 hookflash message to Cisco CallManager, Cisco CallManager currently ignores this message.

In Cisco CallManager Release 3.1, MGCP supports T1 CAS, but it does not support hookflash transfer from a T1 CAS interface.

Tone on Hold

Tone on hold gives the caller a comfort tone while the agent puts the call on hold to confer with another agent or supervisor. This feature is supported for calls coming from MGCP gateways only.

Cisco CallManager Release 3.1 introduced music on hold, which should be used in lieu of tone on hold for calls coming from H.323 gateways.

Music on Hold

Music on hold (MOH) can be provided to callers connected to Cisco IOS H.323 and MGCP gateways. A separate MOH server contains the media resources needed to provide the music on hold. For IPCC, Cisco does not recommend combining the MOH server with Cisco CallManager because of capacity reasons. For each MOH server, you can define up to 50 music channels that reference a music file and one channel that references live audio. Music on hold sources can be defined globally, per device group, per device, and per line on a device. The device can be an IP hard phone or a media termination point (MTP) phone. The held device determines which MOH media resource to use, but the device initiating the hold action determines the audio source to use.

There are two different transport media for MOH: unicast and multicast. Unicast uses a separate source stream for each user or gateway port. Multicast allows multiple users to use the same audio source stream. Although multicast is more efficient, it requires more configuration of the network devices (such as routers) to understand the multicast address. Therefore, Cisco recommends that you use unicast as a transport medium.

Music is streamed to the caller across the gateway when an IPCC agent activates the hold button at the desktop or during the consultation time of the transfer or conference operation. If an agent transfers or conferences with another agent and then reconnects to the caller via the Alternate button, then the other agent will hear the same music on hold source that the caller heard. If the agent activates emergency or supervisor assist buttons, then the caller will hear music on hold if it is defined for the agent's phone.

Cisco CallManager Redundancy

If the primary Cisco CallManager does not respond, the voice gateways have the ability to fail-over to a backup Cisco CallManager. Conversely, when the primary Cisco CallManager is back on-line, the voice gateways can re-home to the primary Cisco CallManager.

MGCP gateways require manual configuration of the primary and backup Cisco CallManagers. Beginning with Cisco CallManager Release 3.2, you can generate automatic configuration files using the latest version of Cisco IOS software. Polling messages are used so that the gateways know when the primary Cisco CallManager is not responding. MGCP gateways have the additional capability to fail-back after a configurable amount of time or only when all connected sessions have been released.

The primary and backup Cisco CallManagers must be defined manually within the Cisco IOS H.323 gateways.

For more details, refer to the *Choosing a Cisco IP Telephony Gateway* chapter in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Call Preservation

Call preservation means that the call will stay up even if Cisco CallManager goes down. Call preservation depends on the survivability of the endpoints within the call.

Survivable endpoints include:

- Analog ports and digital interfaces on MGCP gateways
- IP phones
- Media termination point (MTP) phones

Non-survivable endpoints include:

- IP IVR
- H.323 gateways

Table 3-2 shows what happens to both survivable and non-survivable endpoints if Cisco CallManager fails to respond.

Table 3-2 Call Preservation Scenarios

Called Party	Calling Party:	
	Survivable Endpoint	Non-Survivable Endpoint (Cisco CallManager fails)
Survivable endpoint	Active call stays up ¹	Call torn down
Non-survivable endpoint (Cisco CallManager stays up)	Active call stays up	Call torn down

1. Signaling to Cisco CallManager is lost for this call. Therefore, subsequent actions that would necessitate a message being sent to Cisco CallManager on behalf of this call are not supported (for example, transfer, conference, hold, or DTMF relay). Active calls do not include calls on hold or in progress (ringing but not answered). Conference call survivability depends on the conference bridge type and on whether the Cisco CallManager controlling the conference bridge fails.

ICM Reporting

The ICM WebView and Monitor ICM reports do not list gateway ports because IPCC does not have visibility into specific gateway ports used on the JTAPI link. However, the CDR Analysis and Reporting (CAR) tool – formerly known as the Administrative Reporting Tool (ART) – in Cisco CallManager can extract call detail records (CDRs) that you can coordinate with the ICM database through Professional Services Organization (PSO) customization.

There is no real-time reporting provided for all-trunk-busy conditions on gateway ports.

An outbound call is considered external (TALKING OUT) if the call rings on a device that is connected through a voice gateway. All other outbound calls are counted as internal (TALKING OTHER).

Network Management

The Cisco CallManager Serviceability Real-Time Monitor (RTM) tool can be used to monitor the status of MGCP gateway channels.

The CiscoWorks2000 Voice Health Monitor (VHM) provides "dashboard" views that demonstrate the real-time health of gateways within a Cisco CallManager VoIP environment.



Design Considerations for High Availability

This chapter covers several possible IPCC failover scenarios and explains design considerations for providing high availability of system functions and features in each of those scenarios. This chapter contains the following sections:

- [Terminology and Conventions, page 4-1](#)
- [Designing for High Availability, page 4-2](#)
- [Data Network Design Considerations, page 4-5](#)
- [Cisco CallManager and CTI Manager Design Considerations, page 4-7](#)
- [IP IVR \(CRS\) Design Considerations, page 4-10](#)
- [Peripheral Gateway Design Considerations, page 4-12](#)
- [Understanding Failure Recovery, page 4-19](#)
- [CTI OS Considerations, page 4-24](#)
- [Other Considerations, page 4-24](#)

Terminology and Conventions

This chapter uses the following terminology and conventions:

- **Intelligent Contact Management (ICM) Central Controller** — a server running only the ICM Router and Logger services.
- **ICM Progger** — a server running all the ICM services except the Administrative Workstation (AW) and Historical Data Server (HDS).
- **Cisco CallManager** — the Cisco CallManager service on the Cisco CallManager server.
- **Cisco CallManager cluster** — a collection of Cisco CallManagers with at least one server as the database publisher and one as a subscriber.
- **Primary Computer Telephony Integration (CTI) Manager** — the CTI Manager service running on a Cisco CallManager server. This is the CTI Manager where the JTAPI user logs in for ICM Peripheral Gateway (PG) side A.
- **Secondary CTI Manager** — the CTI Manager service running on a Cisco CallManager server. This is the CTI Manager where the JTAPI user logs in for Cisco CallManager PG side B.
- **Cisco IP Interactive Voice Response (IP IVR)** — part of the Cisco Customer Response Solutions (CRS) platform, an open, multimedia platform for building customer response applications.

Cisco recommends using only duplex (redundant) Cisco CallManager configurations for all IPCC deployments that require high availability. This chapter assumes that the IPCC failover feature is a critical requirement for all deployments, and therefore it presents only deployments that use a redundant (duplex) Cisco CallManager configuration, with each Cisco CallManager cluster having at least one publisher and one subscriber.

The CTI Manager and Cisco CallManager are separate services running on the same Cisco CallManager server, just as the Trivial File Transfer Protocol (TFTP) service does. Throughout this chapter, all references to CTI Manager and Cisco CallManager refer to each individual service only and not to the entire Cisco CallManager server, unless otherwise specified.

Also, the Cisco Agent Desktop differs in many ways from the CTI Object Server (CTI OS), which are not covered in detail in this chapter. Currently Cisco Agent Desktop release 4.4 provides a redundant server for failover, but this capability was not supported in the previous Cisco Agent Desktop release 4.2. Within the scope of this chapter, all statements regarding agent desktops refer to CTI OS only, unless specified otherwise.

Designing for High Availability

IPCC is a distributed solution that uses numerous hardware and software components, and it is important to design each deployment in such a way that a failure will impact the fewest resources in the call center. The type and number of resources impacted will depend on how stringent your requirements are and which design characteristics you choose for the various IPCC components, including the network infrastructure. A good IPCC design will be tolerant of most failures (defined later in this section), but not all failures can be made transparent.

IPCC is a sophisticated solution designed for mission-critical call centers. The success of any IPCC deployment requires a team with experience in data and voice internetworking, system administration, and IPCC application configuration.

Before implementing IPCC, 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 IPCC sites.

In summary, plan ahead and follow all the design guidelines and recommendations presented in this guide and in the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avid/ip_tel_register.html

For assistance in planning and designing your IPCC solution, consult your Cisco Systems Engineer (SE).

Figure 4-1 shows a high-level design for a fault-tolerant IPCC single-site deployment.

Public network

T1 lines

T1 lines

Voice gateway 1

Voice gateway 2

Gatekeepers

Firewall

Corporate LAN

IP IVR group

IP IVR 1

IP IVR 2

CTI OS A

CTI OS B

CTI server A

CTI server B

ICM A

ICM B

ICM central controllers

CM PG A

CM PG B

VRU PG A

VRU PG B

AW A

AW B

AW client

Call control, CTI data, IP messaging

TDM voice lines

Ethernet lines

76600

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

- Routed and answered by an available IPCC agent
- Sent to an available IP IVR (CRS) port
- Answered by the Cisco CallManager AutoAttendant
- Prompted by an IP IVR (CRS) announcement that the call center is currently experiencing technical difficulties, and to call back later

The components in Figure 4-1 can be rearranged to form two connected IPCC sites, as illustrated in Figure 4-2.

Figure 4-2 IPCC Single-Site Redundancy

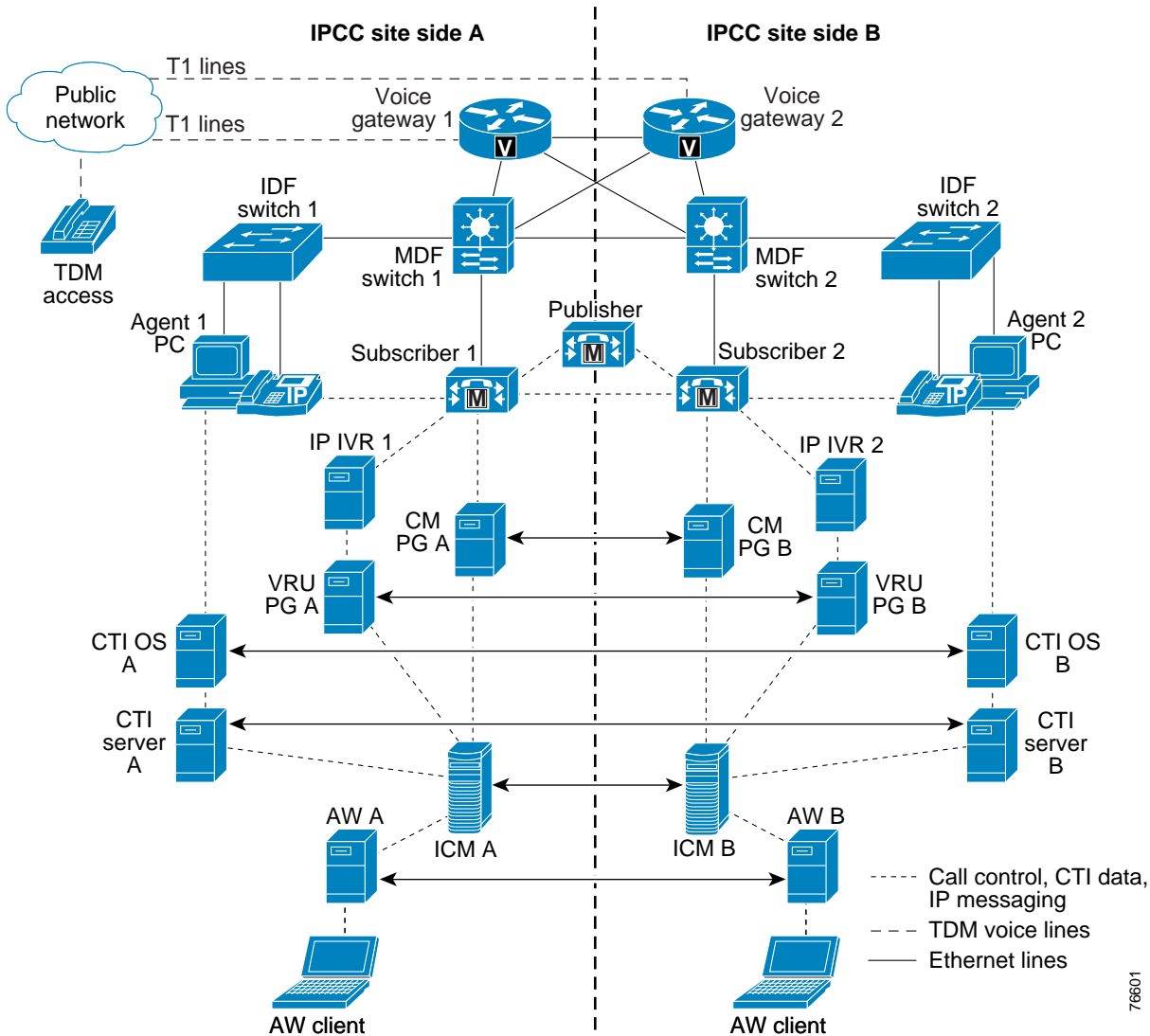


Figure 4-2 emphasizes the redundancy of the single site design in Figure 4-1. Side A and Side B are basically mirrors of each other. In fact, one of the main IPCC features to enhance high availability is its simple mechanism for converting a site from non-redundant to redundant. To implement IPCC high availability, all you need to do is to duplicate the first side and cross-connect all the corresponding parts.

The following sections use Figure 4-1 as the model design to discuss issues and features that you should consider when designing IPCC 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.

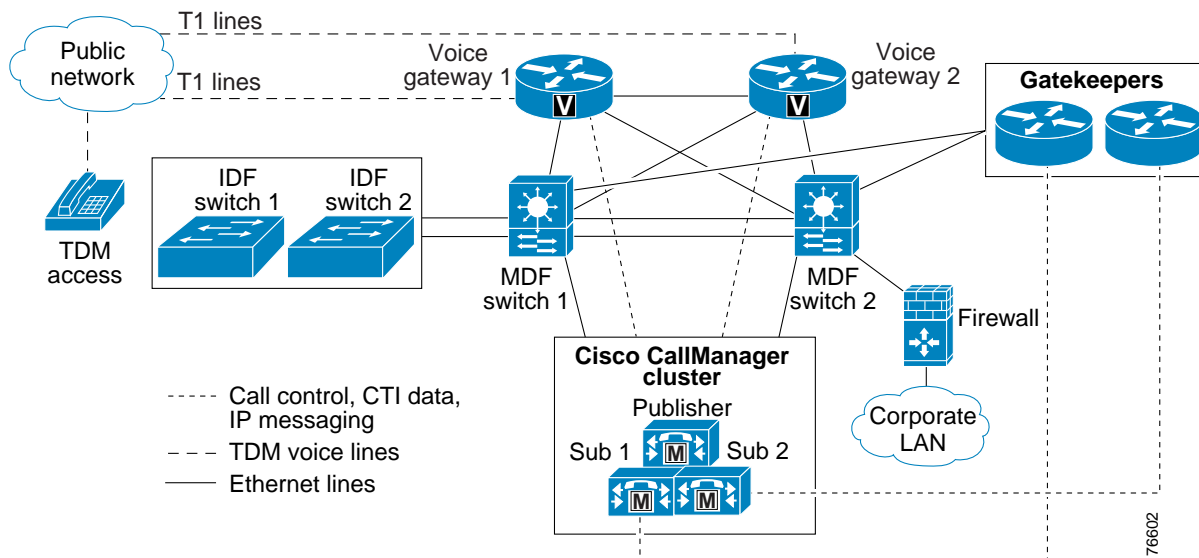
Data Network Design Considerations

The IPCC design shown in [Figure 4-3](#) starts from a time division multiplexing (TDM) call access point and ends where the call reaches an IPCC agent. The bottom of the network infrastructure in the design supports the IPCC environment for data and voice traffic. The network, including the PSTN, is the foundation for the IPCC solution. If the network is poorly design to handle failures, then everything in the call center is prone to failure because all the servers and network devices depend on the network for communication. Therefore, the data and voice networks must be a primary part of your solution design and the first stage for all IPCC implementations. For more information, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

In addition, the choice of voice gateways for a deployment is critical because some protocols offer more call resiliency than others. (For more information, refer to the chapter on [Voice Gateway Considerations for IPCC](#).) This chapter focuses on how the voice gateways should be configured for high availability with the Cisco CallManager cluster(s).

Figure 4-3 High Availability in a Network with Two Voice Gateways and One Cisco CallManager Cluster



Using multiple voice gateways avoids the problem of a single gateway failure causing a blockage of all calls. In a configuration with two voice gateways and one Cisco CallManager cluster, each gateway should register with a different primary Cisco CallManager to spread the workload across the Cisco CallManagers in the cluster. Each gateway should use the other Cisco CallManager as a backup in case its primary Cisco CallManager fails. Refer to the *Cisco IP Telephony Solution Reference Network Design Guide* for details on setting up Cisco CallManagers redundancy groups for backup.

When calculating the number of trunks from the PSTN, be sure enough trunks are available to handle the maximum busy hour call attempts (BHCA) when one or more voice gateways fail. During the design phase, first decide how many simultaneous voice gateway failures are acceptable for the site. Based upon this requirement, the number of voice gateways used, and the distribution of trunks across those voice gateways, you can determine the number of trunks required. The more you distribute the trunks over multiple voice gateways, the fewer trunks you will need. However, using more voice gateways will increase the cost of that component of the solution, so you should compare the annual operating cost of the trunks (paid to the PSTN provider) to the one-time fixed cost of the voice gateways.

For example, assume the call 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 in this case, then each voice gateway should be provisioned with four T1 lines (total of eight). If three voice gateways are deployed, then two T1 lines per voice gateway (total of six) would be enough to achieve the same level of availability. If five voice gateways are deployed, then one T1 per voice gateway (total of five) would be enough to achieve the same level of availability. Thus, you can reduce the number of T1 lines required by adding more voice gateways.

The operational cost savings of fewer T1 lines may be greater than the one-time capital cost of additional voice gateways. In addition to the recurring operational costs of the T1 lines, you should also factor in the 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. Therefore, it is a worthwhile design practice to perform this cost comparison.

After you have determined the number of trunks needed, the PSTN provider has to configure them in such a way 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 re-routing 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 the carrier level signaling on the digital circuit has dropped. Loss of carrier level signaling causes the PSTN to busy out all trunks on that digital circuit, thus 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.

Because the voice gateways register with a primary Cisco CallManager, an increase in the amount of traffic on a given voice gateway will result in more traffic being handled by its primary Cisco CallManager. Therefore, when sizing the Cisco CallManager servers, plan for the possible failure of a voice gateway and calculate the maximum number of trunks that may be in use on the remaining voice gateways registered with each CallManager server.

With standalone voice gateways, it is possible that the voice gateway itself is operational but that its communication paths to the Cisco CallManager servers are severed (for example, a failed Ethernet connection). If this occurs in the case of a H.323 gateway, 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.

When the voice gateway interface to the switch fails, the voice gateway automatically busies out all its trunks. This prevents new calls from being routed to this voice gateway from the PSTN. Calls in progress do not survive because the Real-Time Transport Protocol (RTP) stream connection no longer exists. Parties at both ends of the line receive silence and, after a configurable timeout, calls are cleared. You can set the Transmission Control Protocol (TCP) timeout parameter in the voice gateway, and you can also set a default timeout in Cisco CallManager. The calls are cleared by whichever timeout expires first. When the voice gateway interface to the switch recovers, the trunks are automatically idled and the PSTN should begin routing calls to this voice gateway again (assuming the PSTN has not permanently busied out those trunks).

Cisco CallManager and CTI Manager Design Considerations

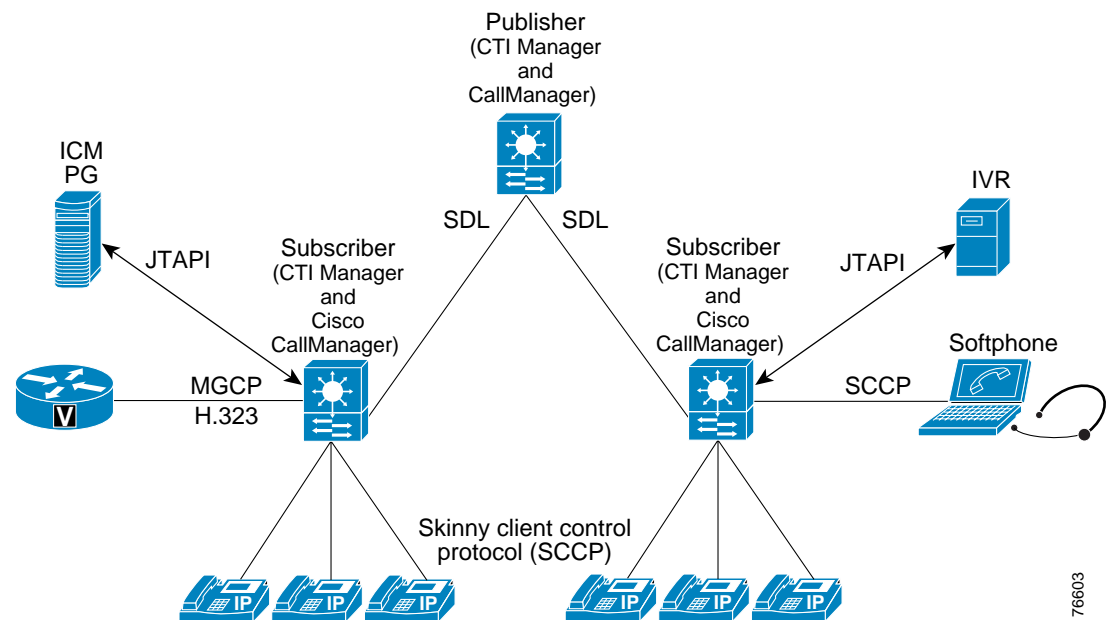
Cisco CallManager Release 3.1 introduced the CTI Manager, a service that acts as an application broker and abstracts the physical binding of the application to a particular Cisco CallManager server to handle all its CTI resources. (Refer to the *Cisco IP Telephony Solution Reference Network Design Guide* for further details about the architecture of the CTI Manager.) The CTI Manager and Cisco CallManager are two separate services running on a Cisco CallManager server. Some other services running on a Cisco CallManager 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 Cisco CallManager cluster. The CTI Manager uses the Cisco JTAPI link to communicate with the applications. It acts like a JTAPI messaging router. Cisco CallManager Release 3.1 modified the JTAPI client library to connect to the CTI Manager instead of connecting to Cisco CallManager, as in previous releases. In addition, Cisco CallManager Release 3.1 modified the JTAPI client library to be aware of all the CTI Managers in its Cisco CallManager cluster (via the Cisco CallManager service, which is explained later in this section.) The CTI Manager uses the same Signal Distribution Layer (SDL) signaling mechanism that the Cisco CallManagers 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 Cisco CallManager is to register and monitor all the IP telephony devices. It basically acts as a switch for all the IP telephony resources and devices in the system, while the CTI Manager 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 Cisco CallManager service are the IP phones, CTI ports, and CTI route points.

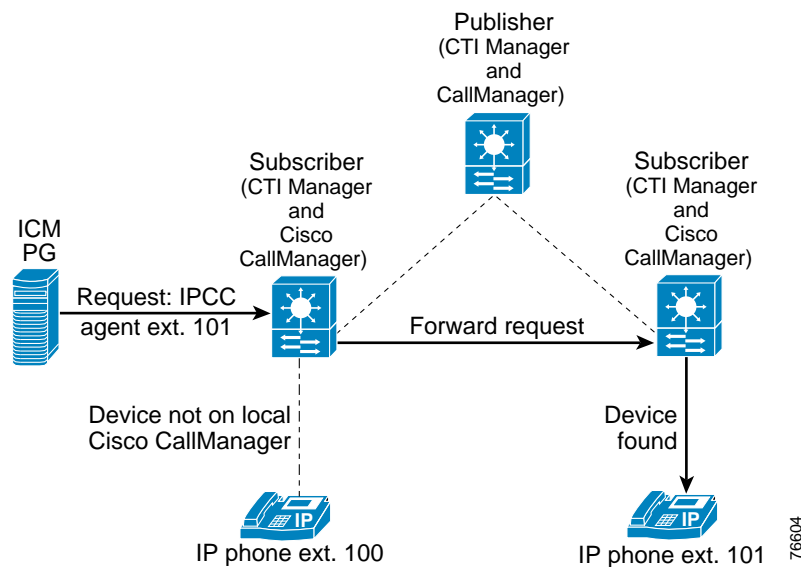
Figure 4-4 illustrates some of the functions of Cisco CallManager and the CTI Manager.

Figure 4-4 Functions of Cisco CallManager and the CTI Manager



The servers in a Cisco CallManager cluster communicate with each other using the Signal Distribution Layer (SDL) service. SDL signaling is used only by the Cisco CallManager service to talk to the other Cisco CallManager services to make sure everything is in sync within the Cisco CallManager cluster. The CTI Managers in the cluster are completely independent and do not establish a direct connection with each other. CTI Managers only route the external CTI application requests to the appropriate devices serviced by the Cisco CallManager service. If the device is not resident on its local Cisco CallManager, then the Cisco CallManager service forwards the application request to the appropriate Cisco CallManager in the cluster. Figure 4-5 shows the flow of a device request to another Cisco CallManager in the cluster.

Figure 4-5 CTI Manager Device Request to a Remote Cisco CallManager



It is important to load-balance devices and CTI applications evenly across all the nodes in the Cisco CallManager cluster.

The external CTI applications use a JTAPI user account on the CTI Manager to establish connection and assume control of the Cisco CallManager devices registered to this JTAPI user. In addition, given that the CTI Managers are independent from each other, any CTI application can connect to any CTI Manager 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 VRU PG allows the administrator to input two CTI Managers, primary and secondary, in its JTAPI subsystem. The Cisco CallManager Peripheral Gateway (PG) handles failover for the CTI Manager by using its two sides, sides A and B, which both log into the same JTAPI user upon initialization of the two CTI Managers. However, only one Cisco CallManager PG side allows the JTAPI user to register and monitor the user devices to conserve system resources. The other Cisco CallManager and VRU PG side stays in hot-standby mode, waiting to be activated immediately upon failure of the active side.

The CTI applications can use the same JTAPI user multiple times to log into separate CTI Managers. This feature allows you to load-balance the CTI application connections across the cluster, and it adds an extra layer of failover and redundancy at the CTI application level by allowing multiple connections to separate CTI Managers while using the same JTAPI user to maintain control. However, keep in mind that every time a JTAPI connection is established with a CTI Manager (JTAPI user logs into a

CTI Manager), the server CPU and memory usage will increase because the CTI application registers and monitors events on all the devices associated with the JTAPI user. Therefore, make sure to allocate the CTI application devices so that they are local to the CTI Manager where the application is connected. (See Figure 4-6.)

Figure 4-6 CTI Application Device Registration

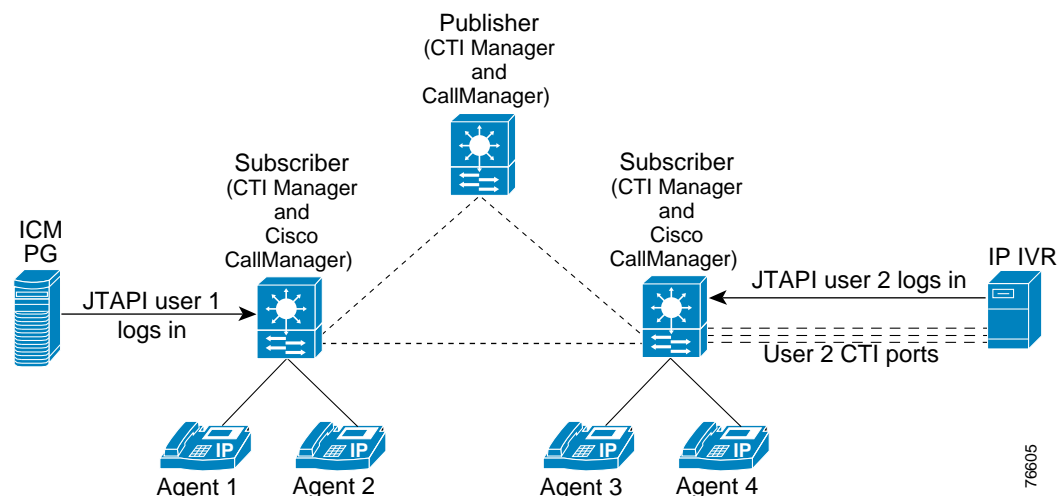


Figure 4-6 shows two external CTI applications using the CTI Manager, the Cisco CallManager PG, and the IP IVR (CRS). The Cisco CallManager PG logs into the CTI Manager using the JTAPI account User 1, while IP IVR (CRS) uses User 2. Each subscriber has two phones to load-balance the calls, and each server has one JTAPI connection to load-balance the CTI applications.

To avoid overloading the available resources, it is best to load-balance devices and CTI applications evenly across all the nodes in the Cisco CallManager cluster.

Cisco CallManager and CTI Manager design should be the second design stage, right after the network design stage, and deployment should occur in this same order. The reason is that the IP telephony 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 Cisco CallManager cluster is the heart of the IPCC system, and any server failure in a cluster will take down two services (CTI and Cisco CallManager), thereby adding extra load to the remaining servers in the cluster.

Distribute Cisco CallManager devices (phones, CTI ports, and CTI route points) evenly across all Cisco CallManagers. Also be sure that all servers can handle the load for the worst-case scenarios, where they are the only remaining server in their cluster. For more information on how to load-balance the Cisco CallManager clusters, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

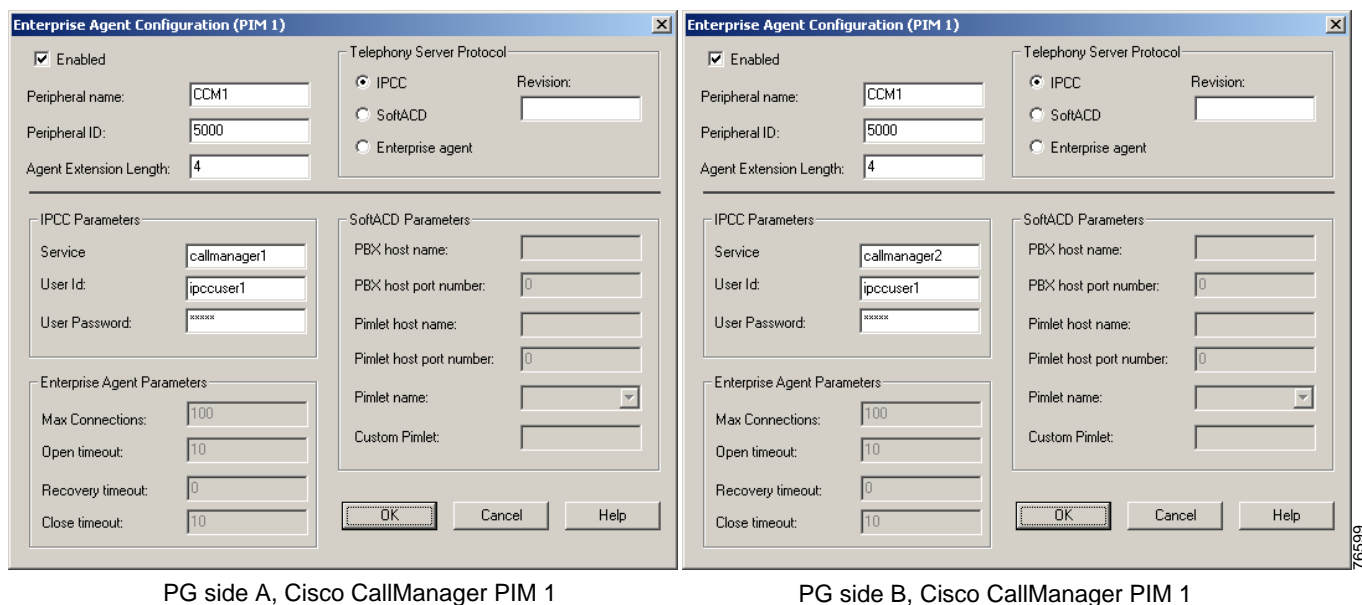
http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Configuring ICM for CTI Manager Redundancy

To enable Cisco CallManager support for CTI Manager failover in a duplex Cisco CallManager model, perform the following steps:

- Step 1** Create a Cisco CallManager redundancy group, and add a publisher and any subscriber to the group.
- Step 2** Designate two CTI Managers to be used for each side of the duplex Peripheral Gateway (PG).
- Step 3** Assign one of the CTI Managers to be the JTAPI service of the Cisco CallManager PG side A. (See [Figure 4-7.](#))
- Step 4** Assign the remaining CTI Manager to be the JTAPI service of the Cisco CallManager PG side B. (See [Figure 4-7.](#))

Figure 4-7 Assigning CTI Managers for PG Sides A and B



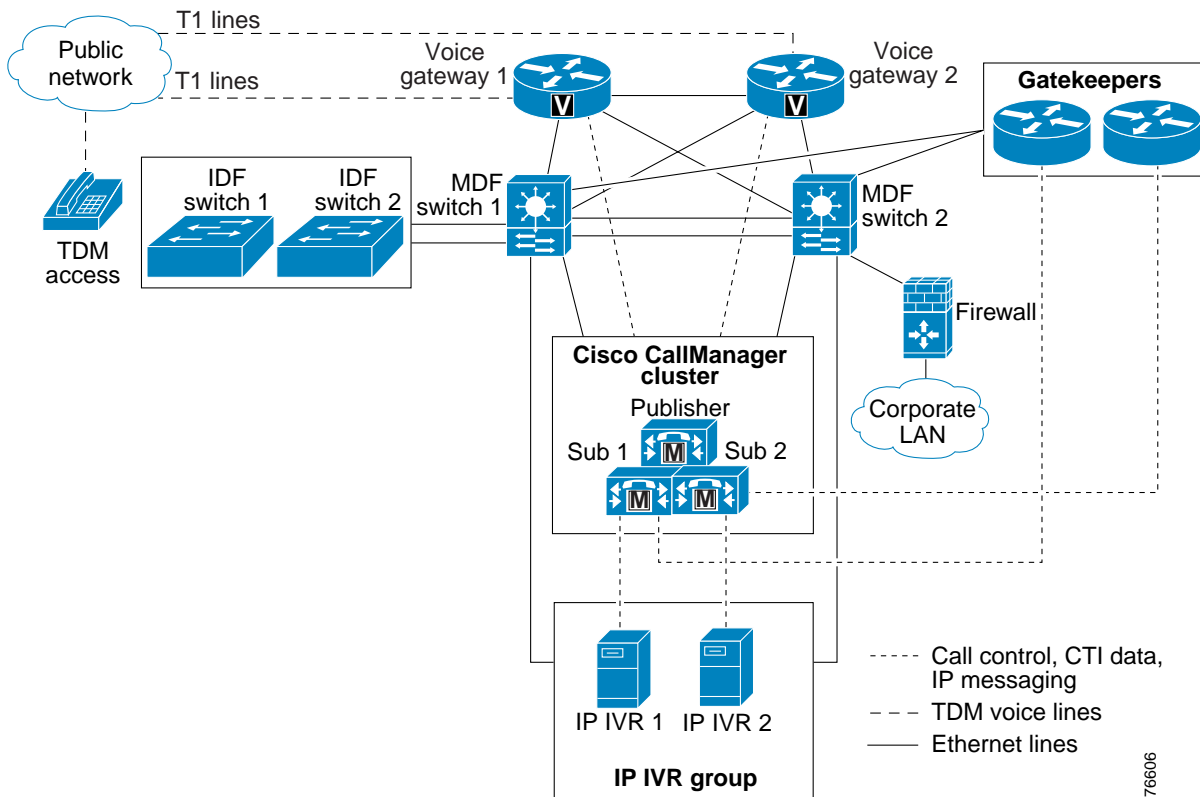
IP IVR (CRS) Design Considerations

Beginning with IP IVR (CRS) release 2.2 and Cisco CallManager release 3.1, the JTAPI subsystem in IP IVR can establish connections with two CTI Managers. This feature allows IPCC designs to add IP IVR redundancy at the CTI Manager level in addition to using the ICM script to check for the availability of IP IVR before sending a call to it. Load balancing is highly recommended to ensure that all IP IVRs are used in the most efficient way.

[Figure 4-8](#) shows two IP IVR (CRS) servers configured for redundancy within one Cisco CallManager cluster. The IP IVR group should be configured so that each server is connected to a different CTI Manager for load balancing and high availability. Using the redundancy feature of the JTAPI

subsystem in the IP IVR server, you can implement redundancy by adding the IP addresses or host names of two Cisco CallManagers from the cluster. Then, if one of the Cisco CallManagers fails, the IP IVR associated with that particular Cisco CallManager will failover to the second Cisco CallManager.

Figure 4-8 High Availability with Two IP IVR Servers and One Cisco CallManager Cluster



You can increase IP IVR (CRS) availability by using one of the following optional methods:

- Call-forward-busy and call-forward-on-error features in Cisco CallManager. This method is more complicated, and Cisco recommends it only for special cases where a few critical CTI route points and CTI ports absolutely must have high availability down to the call processing level in Cisco CallManager. For more information on this method, see [IP IVR \(CRS\) High Availability Using Cisco CallManager, page 4-12](#).
- ICM script features to check the availability of an IP IVR prior sending a call to it. For more information on this method, see [IP IVR \(CRS\) High Availability Using ICM, page 4-12](#).



Note

Do not confuse the IP IVR (CRS) subsystems with services. IP IVR uses only one service, the Cisco Application Engine service. The IP IVR subsystems are connections to external applications such as the CTI Manager and ICM.

IP IVR (CRS) High Availability Using Cisco CallManager

You can implement IP IVR (CRS) port high availability by using any of the following call forward features in Cisco CallManager:

- **Forward Busy** — forwards calls to another port or route point when Cisco CallManager detects that the port is busy. This feature can be used to forward calls to another CTI port when an IP IVR CTI port is busy due to an IP IVR application problem, such as running out of available CTI ports.
- **Forward No Answer** — forwards calls to another port or route point when Cisco CallManager detects a port has not picked up a call within the timeout period set in Cisco CallManager. This feature can be used to forward calls to another CTI port when an IP IVR CTI port is not answering due to an IP IVR application problem.
- **Forward on Failure** — forwards calls to another port or route point when Cisco CallManager detects a port failure caused by an application error. This feature can be used to forward calls to another CTI port when an IP IVR CTI port is busy due to a Cisco CallManager application error.



Note

When using the call forwarding features to implement IP IVR port high availability, avoid creating a loop in the event that all the IP IVR servers are unavailable. Basically, do not establish a path back to the first CTI port that initiated the call forwarding.

IP IVR (CRS) High Availability Using ICM

You can implement IP IVR (CRS) high availability through ICM scripts. You can prevent calls from queuing to an inactive IP IVR by using the ICM scripts to check the IP IVR status before sending the calls to it. For example, you can program an ICM script to check if the 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). This method can be modified to load-balance ports across multiple IP IVRs, and it is easily scalable to virtually any number of IP IVRs.



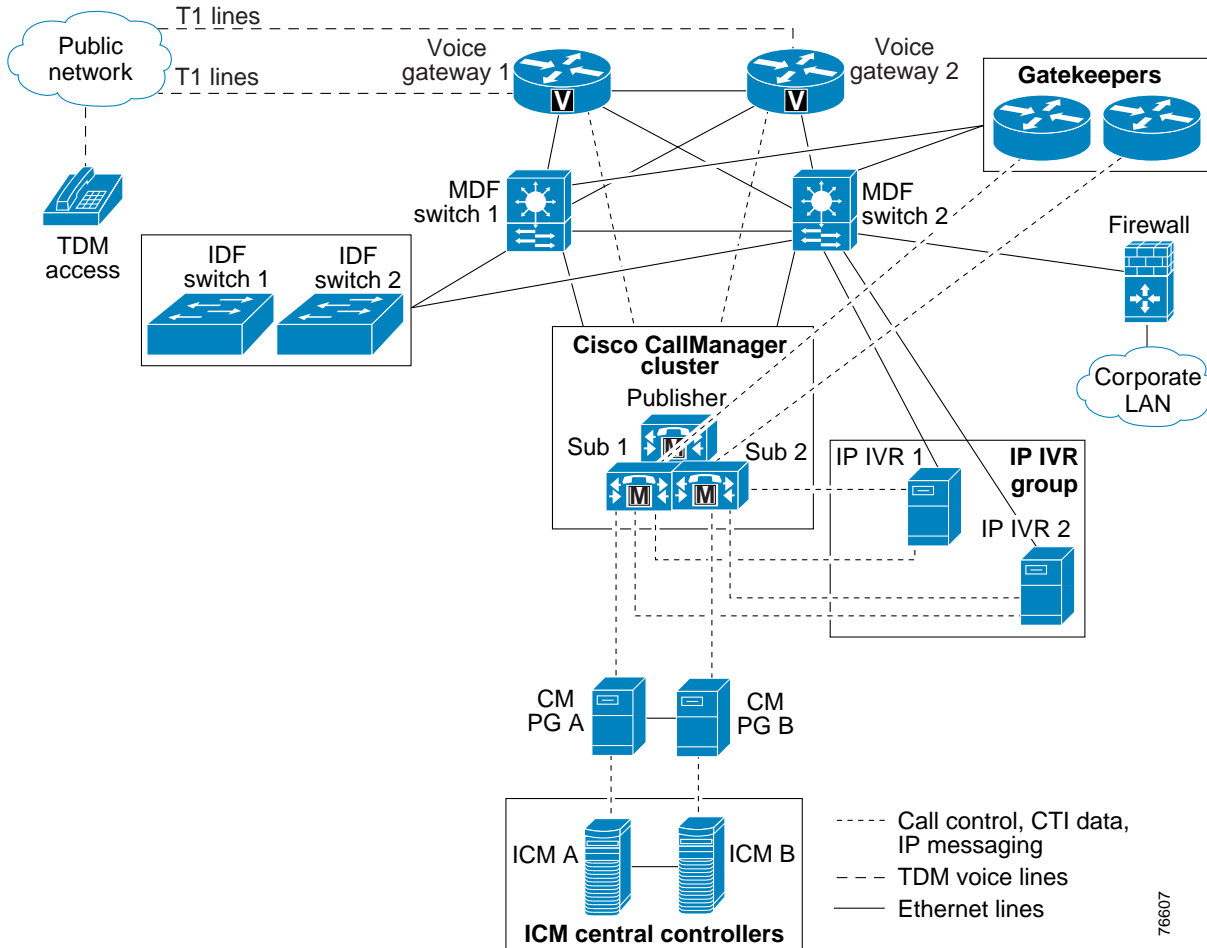
Note

All calls at the IP IVR are dropped if either the IP IVR server or the IP IVR PG fails.

Peripheral Gateway Design Considerations

Starting with ICM Release 4.6.1, the Peripheral Gateway (PG) incorporates the high-availability features in Cisco CallManager to work with the existing ICM high-availability architecture. In addition, one of the major enhancements in ICM Release 4.6.1 is that a PG supports one Cisco CallManager Peripheral Interface Manager (PIM) per Cisco CallManager cluster. This means that a Cisco CallManager PIM can control phones anywhere in the cluster.

Only duplex Cisco CallManager PG implementations take full advantage of the CTI Manager high-availability features. Cisco recommends that all the ICM services be duplex, each with a side A and side B. The minimum requirement for ICM high-availability support for CTI Manager and IP IVR (CRS) is a duplex (redundant) Cisco CallManager PG environment with one Cisco CallManager cluster with at least two servers. Therefore, the minimum configuration for a Cisco CallManager cluster in this case is one publisher and one subscriber. (See [Figure 4-9](#).)

Figure 4-9 ICM High Availability with One Cisco CallManager Cluster

Redundant ICM servers can be co-located or distributed. In both cases, they are connected through a private LAN. If the servers are co-located, you can provide the private LAN by inserting a second NIC card in each server (sides A and B) and connecting them with a crossover cable. If the servers are distributed, you can provide the private LAN by inserting a second NIC card in each server (sides A and B) and connecting them with a dedicated T1 line.

Upon initialization of the Cisco CallManager PG service, JTAPI detects the Cisco CallManager version. It then restarts itself to re-read the registry setting and load the correct library. This process causes a slight delay when the Cisco CallManager PG initializes for the first time after a successful installation and configuration.

In a duplex ICM environment, both JTAPI services from both Cisco CallManager PG sides log into the CTI Manager upon initialization. Cisco CallManager PG side A logs into the primary CTI Manager, while PG side B logs into the secondary CTI Manager. However, only the active side of the Cisco CallManager PG registers monitors for phones and CTI route points. The standby side logs into the secondary CTI Manager only to initialize the interface and prime it for a failover. The registration and initialization services of the Cisco CallManager devices take a significant amount of time, and having the CTI Manager primed significantly decreases the time for failover.

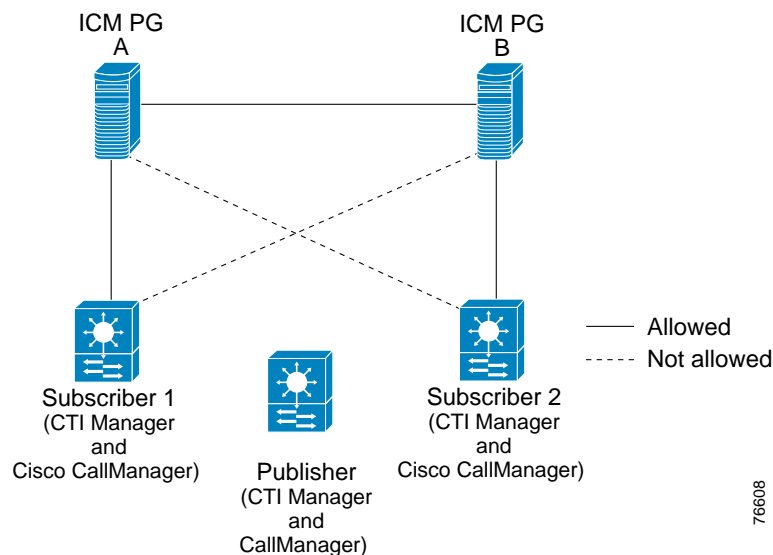
There is no need to configure the CTI Manager port in the PG. The CTI Manager port is stored in the JTAPI gateway, which is a component of the PG in charge of establishing the JTAPI link to a CTI Manager. Finally, failover between the ICM Router and Logger services on ICM side A and ICM side B is completely transparent to the other ICM services.

Cisco CallManager Failure Scenarios

A duplex ICM model contains no single points of failure. However, there are scenarios where a combination of several factors, mostly failures, can prevent IPCC from accepting new incoming calls. Also, if a component of the IPCC solution does not itself support redundancy and failover, existing calls on that component will be dropped. The following ICM failure scenarios have the most impact on high availability:

- Cisco CallManager Peripheral Interface Managers (PIMs) cannot activate if either of the following failure scenarios occurs (see [Figure 4-10](#)):
 - PIM side A and secondary CTI Manager both fail.
 - PIM side B and primary CTI Manager both fail.

Figure 4-10 Cisco CallManager PGs Cannot Cross-Connect to Backup CTI Managers



ICM Failover Scenarios

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

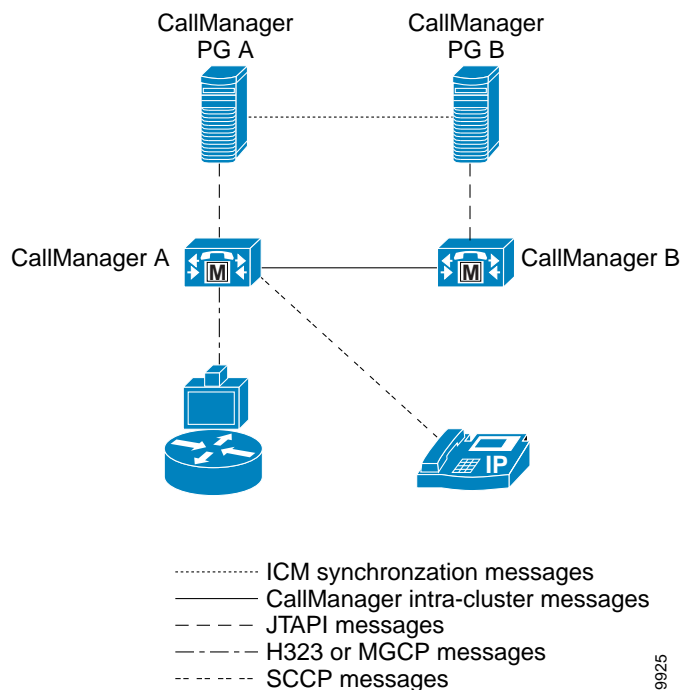
- [Scenario 1 - Cisco CallManager and CTI Manager Fail, page 4-15](#)
- [Scenario 2 - Cisco CallManager PG Side A Fails, page 4-16](#)
- [Scenario 3 - Only Cisco CallManager Fails, page 4-16](#)
- [Scenario 4 - Only CTI Manager Fails, page 4-17](#)

Scenario 1 - Cisco CallManager and CTI Manager Fail

Figure 4-11 shows a complete system failure or loss of network connectivity on Cisco CallManager A. The CTI Manager and Cisco CallManager services are both active on the same server, and Cisco CallManager A is the primary CTI Manager in this case. The following conditions apply to this scenario:

- All phones and gateways are registered with Cisco CallManager A.
- All phones and gateways are configured to re-home to Cisco CallManager B (that is, B is the backup).
- Cisco CallManagers A and B are each running a separate instance of CTI Manager.
- When the entire Cisco CallManager A system fails, all phones and gateways re-home to Cisco CallManager B.
- PG side A detects a failure and induces a failover to PG side B.
- PG side B becomes active and registers all Dialed Numbers and phones; call processing continues.
- After an agent disconnects from all calls, that agent's desktop functionality is restored to the same state prior to failover.
- When Cisco CallManager A recovers, all phones and gateways re-home to it.
- PG side B remains active, using the CTI Manager on Cisco CallManager B.

Figure 4-11 Scenario 1 - Cisco CallManager and CTI Manager Fail

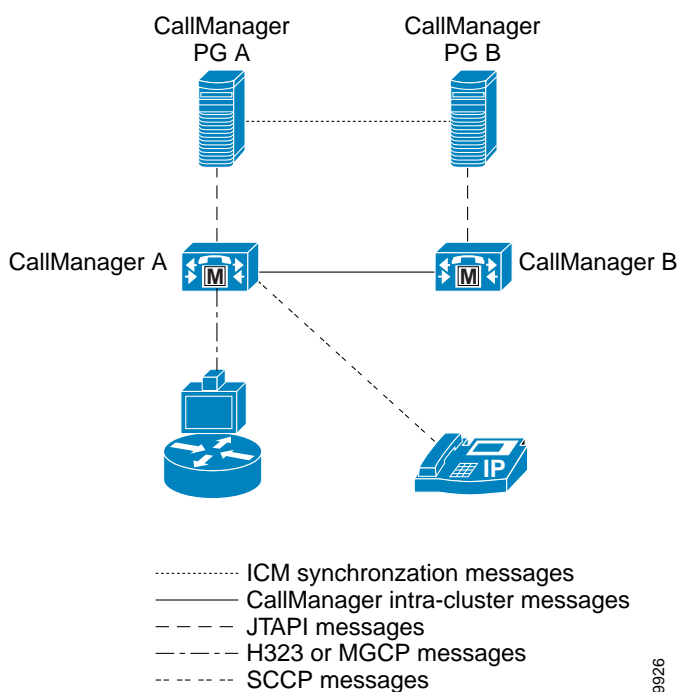


Scenario 2 - Cisco CallManager PG Side A Fails

Figure 4-12 shows a failure on PG side A and a failover to PG side B. All CTI Manager and Cisco CallManager services continue running normally. The following conditions apply to this scenario:

- All phones and gateways are registered with Cisco CallManager A.
- All phones and gateways are configured to re-home to Cisco CallManager B (that is, B is the backup).
- Cisco CallManagers A and B are each running a separate instance of CTI Manager.
- When PG side A fails, PG side B immediately becomes active.
- PG side B registers all Dialed Numbers and phones; call processing continues.
- After an agent disconnects from all calls, that agent's desktop functionality is restored to the same state prior to failover.
- When PG side A recovers, PG side B remains active and uses the CTI Manager on Cisco CallManager B.

Figure 4-12 Scenario 2 - Cisco CallManager PG Side A Fails



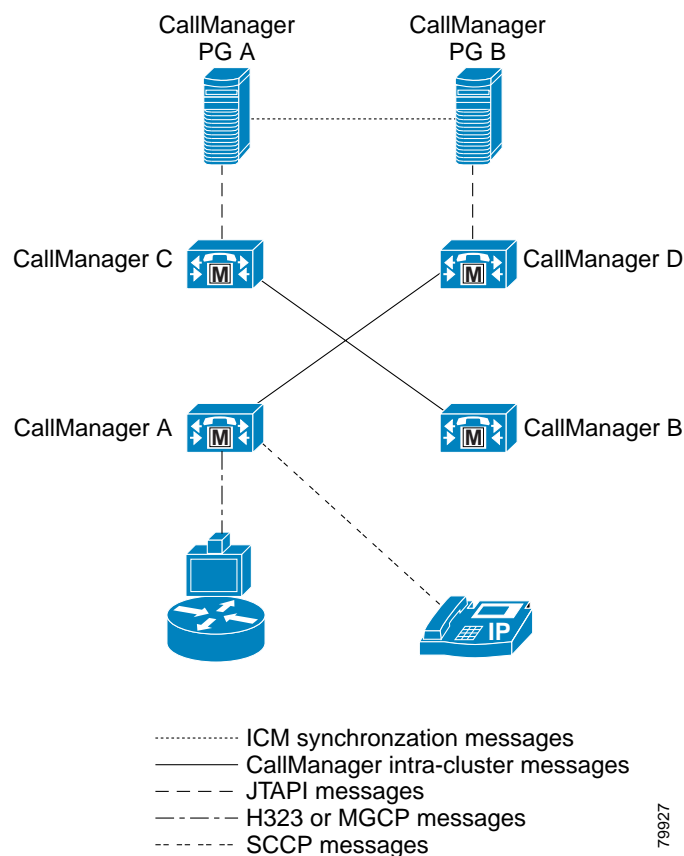
Scenario 3 - Only Cisco CallManager Fails

Figure 4-13 shows a failure on Cisco CallManager A. The CTI Manager services are running on Cisco CallManager C and D, and Cisco CallManager C is acting as the primary CTI Manager. However, all phones and gateways are registered with Cisco CallManager A. During this failure, Cisco CallManager is not affected because the PG communicates with the CTI Manager service, not the Cisco CallManager service. All phones re-home individually to the standby Cisco CallManager B if they are not in a call. If a phone is in a call, it re-homes to Cisco CallManager B after it disconnects from the call.

The following conditions apply to this scenario:

- All phones and gateways are registered with Cisco CallManager A.
- All phones and gateways are configured to re-home to Cisco CallManager B (that is, B is the backup).
- Cisco CallManagers C and D are each running a separate instance of CTI Manager.
- When Cisco CallManager A fails, phones and gateways re-home to Cisco CallManager B.
- PG side A does not see any failure, but agent phones and desktop controls become disabled.
- Call processing continues.
- After an agent disconnects from all calls, that agent's desktop functionality is restored to the same state prior to failover.
- When Cisco CallManager A recovers, phones and gateways re-home to it.
- Call processing continues.

Figure 4-13 Scenario 3 - Only Cisco CallManager Fails



Scenario 4 - Only CTI Manager Fails

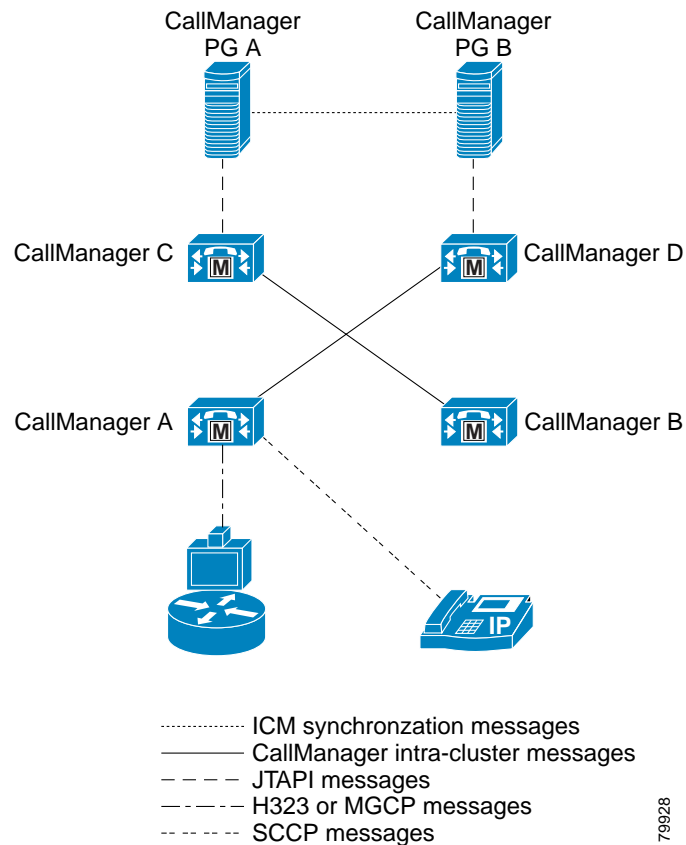
Figure 4-14 shows a CTI Manager service failure on Cisco CallManager C. The CTI Manager services are running on Cisco CallManager C and D, and Cisco CallManager C is the primary CTI Manager. However, all phones and gateways are registered with Cisco CallManager A. During this failure, both

the CTI Manager and the PG failover to their secondary sides. Because the JTAPI service on PG side B is already logged into the secondary (now primary) CTI Manager, the device registration and initialization time is significantly shorter than if the JTAPI service on PG side B had to log into the CTI Manager.

The following conditions apply to this scenario:

- All phones and gateways are registered with Cisco CallManager A.
- All phones and gateways are configured to re-home to Cisco CallManager B (that is, B is the backup).
- Cisco CallManagers C and D are each running a separate instance of CTI Manager.
- When Cisco CallManager C fails, PG side A detects a failure of the CTI Manager on that server and induces a failover to PG side B.
- PG side B registers all Dialed Numbers and phones with Cisco CallManager D, and call processing continues.
- After an agent disconnects from all calls, that agent's desktop functionality is restored to the same state prior to failover.
- When Cisco CallManager C recovers, PG side B continues to be active and uses the CTI Manager on Cisco CallManager D.

Figure 4-14 Only CTI Manager Fails



Understanding Failure Recovery

This section analyzes the failover recovery of each individual part (products and sub-components inside each product) of the IPCC solution.

Cisco CallManager Service

In larger deployments, it is possible that the Cisco CallManager where agent phones are registered will not be running the CTI Manager service that communicates with the Cisco CallManager. When an active Cisco CallManager service fails, all the devices registered to it are reported out of service by the CTI Manager service. Cisco CallManager reporting shows the call as terminated when the Cisco CallManager failure occurred because, from a Cisco CallManager reporting perspective, any calls in progress are terminated and the agents are logged out so that future calls are not routed to them. IP phones of agents not on calls at the time of failure will quickly register with the backup Cisco CallManager. The IP phone of an agent on a call at the time of failure will not register with the backup Cisco CallManager until after the agent completes the current call. If MGCP gateways are used, then the calls in progress survive, but further call control functions (hold, retrieve, transfer, conference, and so on) are not possible.

When the active Cisco CallManager fails, the agent desktops show the agents as being logged out, their IP phones display a message stating that the phone has gone offline, and all the IP phone soft keys are grayed out until the phones failover to the backup Cisco CallManager. To continue receiving calls, the agents must wait for their phones to re-register with a backup Cisco CallManager to have their desktop functionality restored by the CTI server to the state prior to the Cisco CallManager service failure. Upon recovery of the primary Cisco CallManager, the agent phones re-register with their original service because all the Cisco CallManager devices are forced to register with their home Cisco CallManager.

In summary, the Cisco CallManager service is separate from the CTI Manager service, which talks to the Cisco CallManager PG via JTAPI. The Cisco CallManager service is responsible for registering the IP phones, and its failure does not affect the Cisco CallManager PGs. From a Cisco CallManager perspective, the PG does not go offline because the Cisco CallManager server running CTI Manager remains operational. Therefore, the PG does not need to failover.

IP IVR (CRS)

When a CTI Manager fails, the IP IVR (CRS) JTAPI subsystem shuts down and restarts by trying to connect to the secondary CTI Manager, if a secondary is specified. In addition, all voice calls at this IP IVR are dropped. If there is an available secondary CTI Manager, it logs into this CTI Manager again and re-registers all the CTI ports associated with the IP IVR JTAPI user. After all the Cisco CallManager devices are successfully registered with the IP IVR JTAPI user, the server resumes its Voice Response Unit (VRU) functions and handles calls.

ICM

The ICM is a collection of services and processes within these services. 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 IPCC solution, including another ICM service.

As stated previously, all redundant ICM services discussed in this chapter must be co-located and connected through a private LAN. You can provide the private LAN by installing a second network interface card (NIC) in each server (sides A and B) and connecting them with a crossover cable. By doing this, you can eliminate all external network equipment failures.

Cisco CallManager PG and CTI Manager Service

When the active CTI Manager or PG fails, the JTAPI detects an `OUT_OF_SERVICE` event and induces a failover to the standby PG. Since the standby PG is logged into the standby CTI Manager already, it registers monitors for the phones with logged-in agents and configured Dialed Numbers and CTI Route Points. This initialization service takes place at a rate of about 5 devices per second. The agent desktops show them as being logged out, and a message displays stating that their routing client or peripheral (Cisco CallManager) has gone offline. (This warning can be turned on or off, depending on the administrator's preference.) All agents lose their desktop functionality until the failure recovery is complete. The agents can recognize this event because the agent state display on their desktop will show *logged out*, and the login button will be the only button available. Any existing calls handled by the agent should remain alive without any impact to the caller.

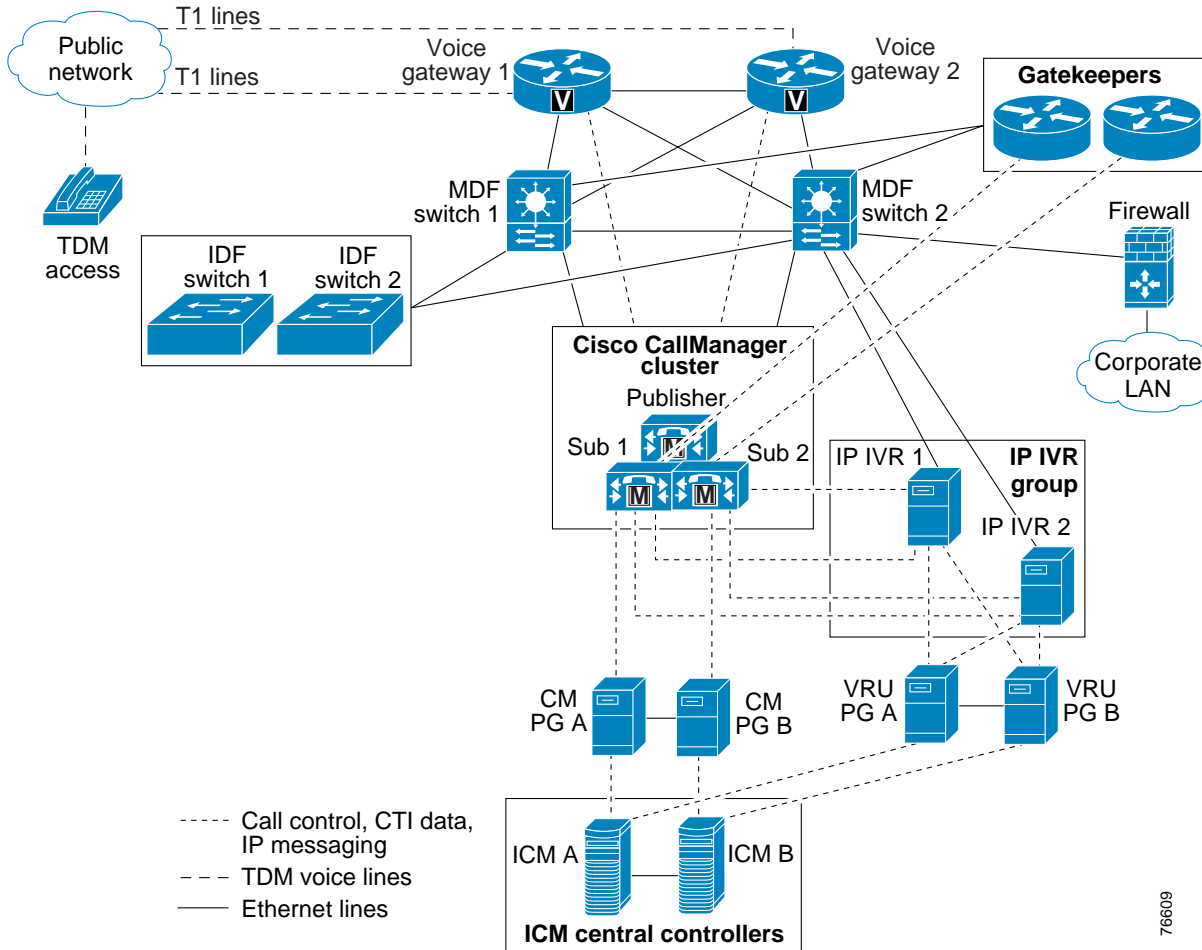
**Note**

Agents should not push any buttons during desktop failover because this can confuse the CTI server when it completes its failover and restores the agent states.

Once the CTI Manager or PG completes its failover, the agents can return to their previous call state (talking, ready, not ready, and so forth). At this point, the agents should also be able to release, transfer, or conference calls if they were on a call at the time of the failure. All the call data that had been collected and stored via a call data update message is retained on the agent desktops, recovered, and matched with call context information saved on the PG. However, all agents without active calls are reset to the default Not Ready state. In addition, the Longest Available Agent (LAA) algorithm resets the timers for all the agents to zero.

ICM Voice Response Unit PG

When a Voice Response Unit (VRU) PG fails, all the calls currently in queue on that IP IVR (CRS) are dropped. However, the Service Control Interface (SCI) link of the failed VRU PG automatically connects to the backup VRU PG so that all new calls can be handled properly. Upon recovery of the failed VRU PG, the currently running VRU PG continues to operate as the active VRU PG. Therefore, having a backup VRU PG adds significant value because it allows an IP IVR to continue to function as an active IP IVR. Without VRU PG redundancy, a VRU PG failure would block use of that IP IVR even though the IP IVR was working properly. (See [Figure 4-15](#).)

Figure 4-15 Redundant ICM VRU PGs with Two IP IVR Servers

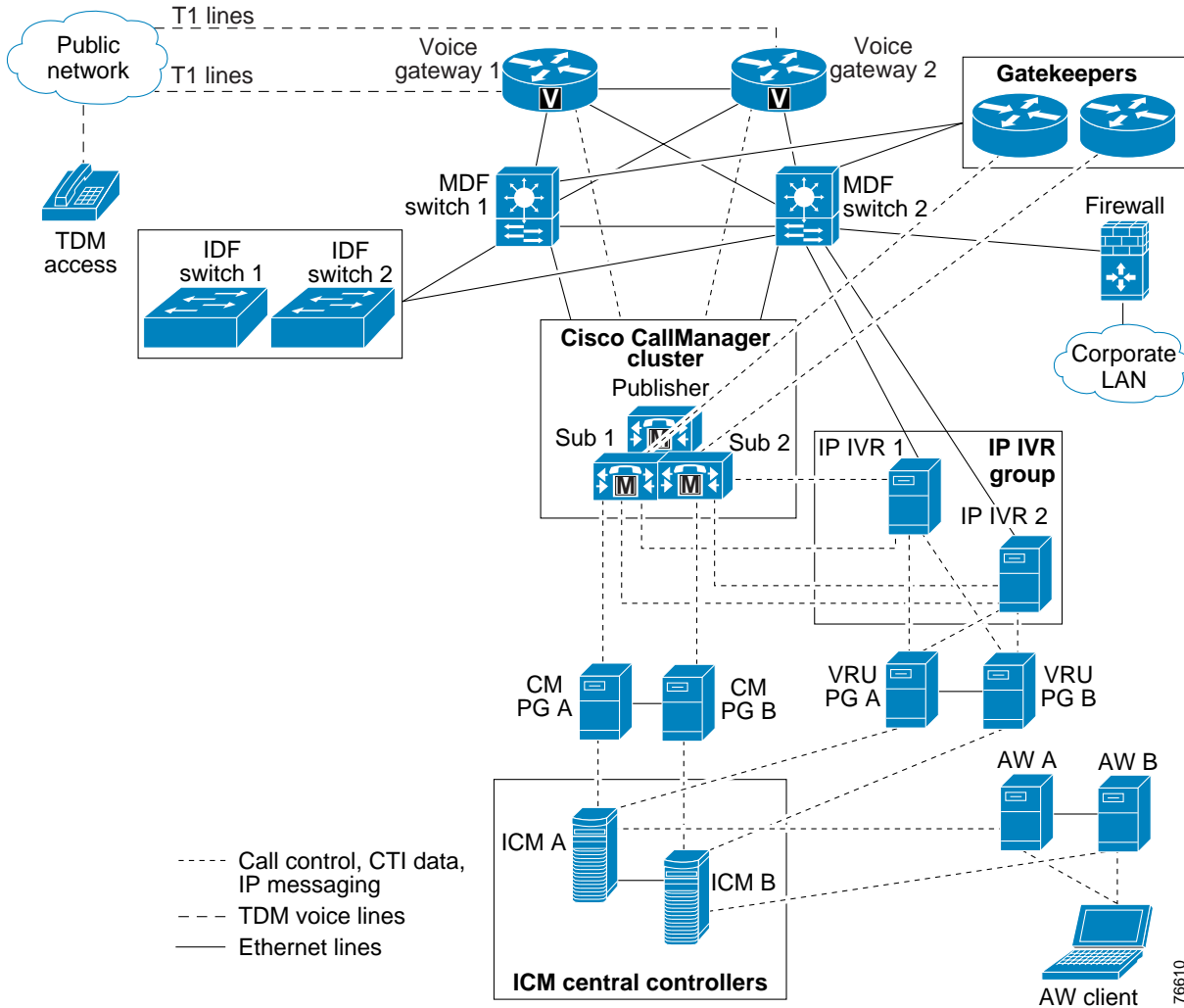
ICM Router and Logger

A ICM Router or Logger failover does not affect other ICM services. The failover and recovery are transparent even to the agent desktops because the ICM Central Controller enables the ICM services to synchronize with each other constantly through the Historical Data Server (HDS) database and to maintain a mirror image of each other in any failure event. When the active side loses connectivity to the side in hot-standby mode, the hot-standby side goes active after missing three heartbeats.

Real-Time Distributor

The Real-Time Distributor supports primary and secondary servers, similar to the rest of the ICM servers, which have sides A and B. Failover and redundancy also work the same way as with the ICM servers. (See [Figure 4-16](#).)

Figure 4-16 Redundant ICM Distributors and AW Servers



The active and standby Distributors send keepalive messages to each other. They do not synchronize their transactions with each other because each Distributor synchronizes its transactions with the active Logger. If the active Distributor fails, the standby Distributor re-synchronizes its tables with the active Logger upon initialization.

The client Administrative Workstations (AWs) first try to connect to the primary Distributor at the ICM site. If the primary Distributor is not active, then the AWs try to connect to the secondary Distributor. If the connected Distributor fails, the AWs stay in a busy state until the client AW connects to the standby Distributor. If no Distributor is available, then the client AW times out and displays an error message to the user stating that there are no real-time distributors available.

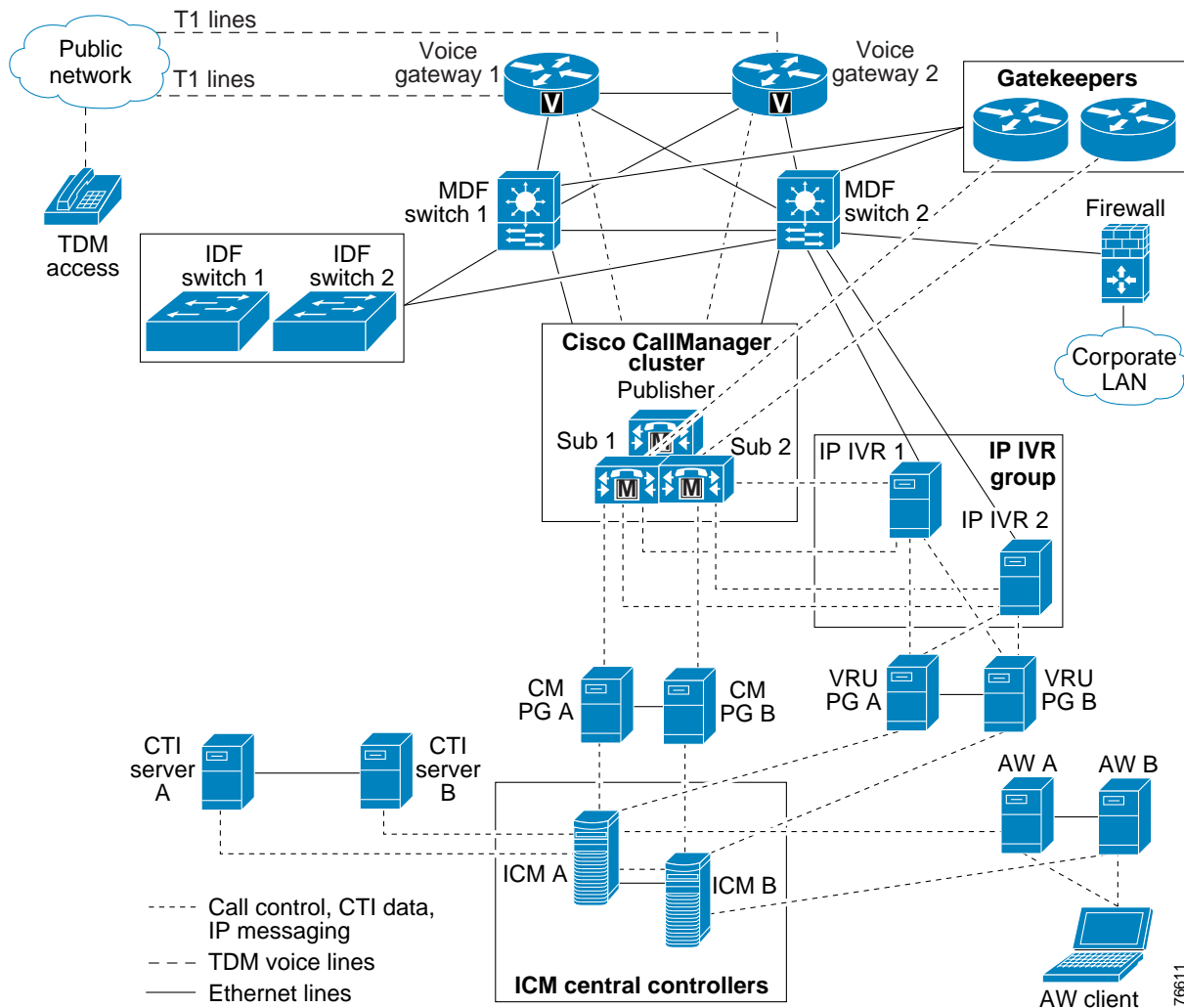
Note that, if the primary Distributor comes back online after a failure, all client AWs connected to the secondary will be disconnected. These client AWs will need to close and relaunch to work.

CTI Server

When the active CTI Server fails, it takes 50 seconds for the standby CTI server to detect a failure and take over as the active CTI server. (30 seconds for the standby CTI server to detect a failure; plus 15 seconds for service activation, agent registration, and agent state synchronization and recovery; plus 5 seconds to compensate for network and system latency.) During this period, the agents will have no desktop functionality, and a message will display stating that the CTI server has gone offline. Once agents are logged in again, any active calls are recovered and matched with call context data saved on the PG. The agents are notified of which calls they currently have in progress, and they can then control those calls from their desktops. However, any agents without active calls are reset to the default Not Ready state. In addition, the LAA algorithm resets the timers of all the agents to zero. (See [Figure 4-17](#).)

If the CTI server fails, the CTI driver notifies the CTI Object Server (CTI OS) and Cisco Agent Desktop server connected to it. The CTI driver then attempts to reconnect to the CTI OS and Cisco Agent Desktop servers. If it fails to reconnect, it then attempts to connect with the backup CTI server. The CTI driver attempts to connect with the original server and the backup server alternately every 50 seconds until it succeeds.

Figure 4-17 Redundant CTI Servers with No Cisco Agent Desktop Server Installed



CTI OS Considerations

The CTI Object Server (CTI OS) consists of two services, the CTI OS service and the CTI driver. If either of these two fails, then the active CTI OS fails over to its peer server. Therefore, it is important to keep both of these services active at all times.

Other Considerations

An IPCC failover can affect other parts of the solution. Although IPCC may stay up and running, some data could be lost during its failover, or other products that depend on IPCC to function properly might not be able to handle an IPCC failover. This section examines what happens to other critical areas in the IPCC solution during and after failover.

Reporting

The IPCC reporting feature uses half-hour intervals to build its history. Therefore, at the end of each half-hour interval, all the real-time reporting statistics are recorded by the Logger and subsequently stored on the Historical Data Server (HDS). The first half-hour interval starts with the activation of the ICM Central Controller.

When failover of an ICM service occurs, the reporting for that particular service is affected for the half-hour interval when the failover occurred. The new reporting statistics are not valid again until the ICM service has recovered and the next half-hour interval has begun. If two ICM services such as the Router and the Logger both fail, the next half-hour interval begins when the Router and Logger both are active again.

When agents log out, all their reporting statistics stop. The next time the agents log in, their statistics start from zero. Although an ICM failover does not force the agents to log out, it does reset their agent statistics when the ICM failover is complete, but their agent desktop functionality is restored back to its pre-failover state.

For further information, refer to the *Cisco IP Contact Center Reporting Guide*, available at

<http://www.cisco.com/warp/public/78/ipccreporting.html>



Sizing IP Contact Center Resources

Central to designing an IP Contact Center (or any call center) is the proper sizing of its resources. This chapter discusses the tools and methodologies needed to determine the required number of call center agents based on customer requirements such as call volume and service level desired, the number of IP IVR ports required for various call scenarios (for example, call treatment, 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 sources such as PBXs and IVRs.

The methodologies and tools presented in this chapter are based on traffic engineering principles using the Erlang models applied to the various resource deployments in an IPCC solution. Examples are provided for an IPCC deployment to illustrate how resources can be impacted under various call scenarios such as call treatment in the IP IVR and agent wrap-up time. These tools and methodologies are intended as building blocks for sizing call center resources and for any telephony applications in general.

Sizing Call Center Resources

The focus of this chapter is on sizing the main three resources in a call center:

- Agents
- Gateway ports (PSTN trunks)
- IP IVR ports

It is very important to be familiar with the common call center terminology and to use it consistently because many of the terms directly affect the sizing and design of an IPCC solution. The following terms and definitions apply throughout this document.

Busy Hour

The busy hour is the one-hour period of the day when the most traffic occurs. The busy hour varies with different days, weeks, and months. Common practice is to design for the average busy hour (the average of the 10 most busy hours in one year). This is not always the case, however, when staffing is required to accommodate a marketing campaign or a seasonal busy hour such as an annual holiday peak. In a call center, maximum agent staffing 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 or participate in offline activities such as training or coaching. In most cases it is not practical to add or remove trunks or IVR ports daily, so they are sized for the peak periods.

Busy Hour Call Attempts (BHCA)

BHCA is the total number of calls that are attempted or received during the peak traffic hour in the call center. For the sake of simplicity, we assume that all calls offered to the voice gateway are received and serviced by the call center resources (agents and IP IVR ports). Calls normally originate from the PSTN, although calls to a call center can also be generated internally (for example, by a Help Desk application) or transferred from other call agents such as another Cisco CallManager, PBX, or TDM IVR.

Servers

Servers are resources that handle traffic loads or calls. There are many types of servers in a call center – for example, PSTN trunks and gateway ports, agents, voice mail ports, and IVR ports.

Talk Time

This is the amount of time an agent spends talking to a caller, including the time an agent places a caller on hold and time spent during consultative conferences.

Wrap-Up Time

After the call is terminated (caller finishes talking to an agent and hangs up), the wrap-up time is the time it takes an agent to finish tasks associated with "wrapping up" the call, such as updating a database, recording notes from the call, or any other activity associated with the call, until the agent becomes available to answer another call.

Average Handle Times (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 "handle time" such as call treatment time, talk time queuing time, and so forth. After-call or wrap-up time is considered by some as another type of AHT, but it should be clearly differentiated from the other types for proper sizing of resources. (See [Sizing Call Center Resources \(Agent Wrap-up Time Example\)](#), page 5-11.)

Erlang

Erlang is a measurement of traffic load during the busy hour. The Erlang is based on having 3600 seconds (60 minutes, or one 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 calls lasts.) For example, if the call center receives 30 six-minute calls in the busy hour, then it received 180 call minutes, or 3 Erlangs. If the call center receives 100 calls averaging 36 seconds each in the busy hour, then it received 3600 call seconds, or 1 Erlang.

Use the following formula to calculate the Erlang value:

$$\text{Traffic in Erlangs} = (\text{Number of calls in the busy hour}) * \text{AHT} / 3600$$

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

This is the traffic load during the busy hour, and it is represented as the product of the BHCA and the AHT normalized to one hour:

$$\text{BHT} = (\text{BHCA} * \text{AHT seconds}) / 3600, \text{ or}$$

$$\text{BHT} = (\text{BHCA} * \text{AHT minutes}) / 60$$

For example, if the call center receives 600 calls in the busy hour, averaging 2 minutes each, then busy hour traffic load is:

$$(600 * 2) / 60 = 20 \text{ Erlangs}$$

Grade Of Service (% Blockage)

This is the probability that a resource or server is busy during the busy hour. All resources may be occupied when a user places a call. In that case the call is lost or blocked. Grade of service 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 percent 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; this is a typical figure to use for PSTN trunks, but different applications may 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 their call is 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 is a standard term used 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" call center is 90% of all calls answered in less than 10 seconds (some calls will be delayed in a queue). A "support" oriented call center may have a different service level goal, such as 80% of all calls answered within 30 seconds in the busy hour. A customer's service level goal drives the number of agents they will need, the percent of calls that will be queued, the average time calls will spend in queue, and the number of trunks and IP IVR ports they will need.

Queuing

When all agents are busy with other callers, subsequent callers must be placed in a queue until an agent becomes available. The percentage of calls queued and the average time spent in a queue is determined by the service level desired and by agent staffing. Cisco's IPCC solution uses an IP IVR to place callers in a queue and to play announcements. An IVR can also be used to handle all calls initially (call treatment or information gathering) and for self-service applications, where the caller is serviced without needing to talk to an agent (such as for obtaining an account balance). Each of these scenarios requires a different number of IP IVR ports to handle the different applications because each has a different average handle time and possibly a different call load. The number of trunks or gateway ports will also differ accordingly. (See examples later in this chapter.)

Design Tools - Erlang Calculators

There are many traffic models that 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, or delayed)
- Call arrival patterns (random, smooth, or peaked)

For purposes of this document, there are mainly two traffic models that are commonly used in sizing call center resources: Erlang-B and Erlang-C. There are many resources on the web that go into detailed explanations of the various models if one is interested (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 IP IVR ports do I need?

Before you can answer these basic questions, you need to have the following minimum set of information that are used as input to these calculators:

- The busy hour call attempts (BHCA) or traffic load presented to each of the resources you want to size
- Average handle time (AHT) for each of the resources, based on the different call scenarios (call treatment, queuing, agent talk time, agent wrap-up time, and so on)
- Service level (percentage of calls that are answered within x seconds)
- Grade of service, or percent blockage, desired for PSTN trunks and IP IVR ports

This remaining sections of this chapter help explain the differences between the Erlang-B and Erlang-C traffic models in simple terms, list which model to use for sizing the specific call center resource (agents, gateway ports, and IP IVR ports), and show examples of how to use them for different call scenarios. The tools used here are obtained from various web sites that provide them 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 suites their needs. The input required for any of the tools and the methodology used is the same regardless of the tool itself.

Erlang-C

The Erlang-C model is used when sizing agents in call centers that queue calls before presenting them to agents. This model assumes:

- Calls are presented randomly to the servers (agents).
- A percentage of callers finding all agents busy will be queued but not blocked.

The input parameters required for this model are:

- Number of calls in the busy hour (BHCA) to be answered by agents
- Average handle time (AHT), the average talk time and wrap-up time
- 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 is 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 for sizing PSTN trunks, gateway ports, or IP IVR ports. It assumes the following:

- Calls are presented randomly.
- A percentage of calls are lost or blocked, 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 — 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 IP IVR or gateway ports

Sizing Call Center Resources (Basic Example)

The call center example in this section is the basis for all examples to follow in this chapter. We start with a basic call flow, where all incoming calls are presented to the voice gateway from the PSTN (no internal calls are generated to agents). Calls are routed directly to an agent, if available; otherwise calls are queued until an agent becomes available. After demonstrating how to calculate the three resources in this basic example, we 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.

The following parameters apply to this example:

- Total BHCA received into the center via a voice gateway = 14,400
- Average Handle Time (agent talk time) = 3 minutes and 16 seconds (total 196 seconds)
- No agent wrap-up time (For an example that adds agent wrap-up time, see [Sizing Call Center Resources \(Agent Wrap-up Time Example\)](#), page 5-11.)
- No call treatment (or information gathering) is implemented initially; all calls are routed to available agents or are queued until an agent becomes available
- Desired service level of 80% of calls answered within 10 seconds
- Desired grade of service for the PSTN ports on the voice gateway = 1%
- IP IVR ports are used for queuing and are sized with a low percentage of blockage to allow calls answered by the gateway to be queued (See [Sizing IP IVR Ports](#), page 5-8, for more information on this point.)

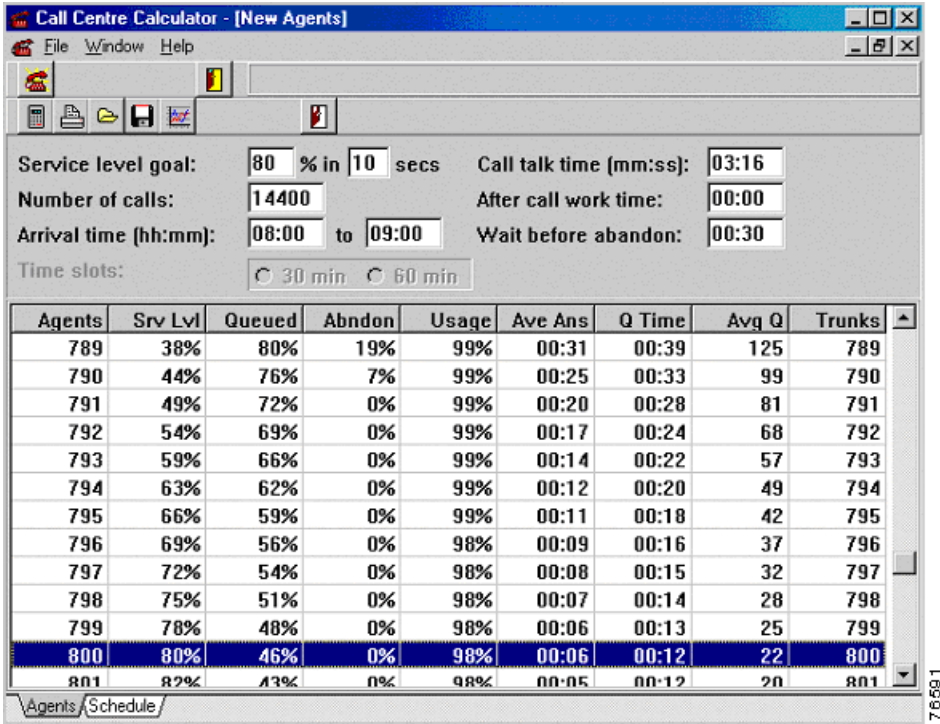
Sizing Agents

The number of agents is determined by using the Erlang-C model because calls are queued to this resource (agents). The call center calculator used in this case is a free version of one of the many Erlang-C calculators on the web (see [Figure 5-1](#)), and it is available at

<http://www.erlang.co.uk/ccr.htm>

The input required for this example is the 14,400 BHCA, service level of 80% of calls answered within 10 seconds, and average handle time (talk time) of 3 minutes and 16 seconds. Entering these parameters in the top portion of the calculator and then pressing the “calc” button on the top left corner will generate the rows of data representing the number of agents needed to handle the call load at various service levels. (The optimal row of output is highlighted automatically in this calculator.) For this example, 800 agents are needed. With this calculator, be sure to adjust the “arrival time” input to one hour, representing the busy hour; start and ending time does not matter as long as the interval is one hour.

Figure 5-1 Using an Erlang-C Calculator to Size Agents



The calculator also lists the percent of calls queued (46%) in the "Queued" column and the average queue time (12 seconds) in the "Q Time" column. We will use these values later to size the number of IP IVR queuing ports needed for the calls that are queued. (We will use an Erlang-B calculator for that calculation.)

From this Erlang-C calculator we are able to determine the following:

- Agents needed = 800
- Percent of calls queued = 46%
- Average time calls spend in a queue = 12 seconds

Number of trunks is ignored here, but an Erlang-B calculator will be used to size trunks.

Sizing PSTN Trunks (Gateway Ports)

There are two traffic loads that impact the required number of trunks:

- Busy Hour Traffic (BHT) — the load presented by the incoming traffic, with an average holding time equal to the average agent talk time for all calls
- The load presented by the subset of calls that queue when no agents are available (For details, see [Sizing Trunks for Queued Calls, page 5-7.](#))

Sizing Trunks for Busy Hour Traffic

Using the same example as in the previous section, where the call center receives 14,400 incoming calls in the busy hour, averaging 3 minutes and 16 seconds (total 196 seconds) each, the Busy Hour Traffic (BHT) load is:

$$(14,400 \text{ calls}) * (196 \text{ seconds}) / 3600 = 784 \text{ Erlangs}$$

The grade of service used for sizing trunks could be different than the one used for sizing IP IVR ports and may vary depending on the type of call scenarios. For example, call center managers need to specify the percent of blockage for the PSTN trunks carrying traffic into the center. The percent blockage could be as low as 1% of the calls getting busy, 2%, or even as high as 5% in some cases (but this would not be typical). This determination is driven by the nature and the requirements of the call center. In a sales environment, the call center wants to capture all the revenue it can, so the percent blockage needs to be low (probably 1%). In a non-call center environment, you might use 3% blockage because callers can try to call again if they hear a busy tone the first time.

Assuming a desired grade of service of 0.01, or 1% blockage, we can calculate the PSTN trunk ports required for a voice gateway to carry this traffic. For this example, we used one of the free Erlang-B calculators on the web (see [Figure 5-2](#)), available at

<http://mmc.et.tudelft.nl/~frits/Erlang.htm>

Figure 5-2 Using an Erlang-B Calculator to Size Trunks

Block probability: 0.01 calculate B

traffic (in erlang): 784.0 calculate A

Number of lines: 812.2 calculate N

(c) Frits Schoute (1878-1929) A.K. Erlang 76692

Based on these calculations, we find that 812 trunks are needed.

These trunks are required just to carry the incoming traffic load based on the agent talk time alone. This number does not include the additional trunks required for those calls that spend time in a queue waiting for an agent to become available. A call center typically needs more trunks for the calls that are queued because the trunk is held longer for these calls than for calls that are answered immediately.

Sizing Trunks for Queued Calls

The percent of calls queued and the average hold time in queue can be obtained from the output of the Erlang-C calculator shown in [Figure 5-1](#):

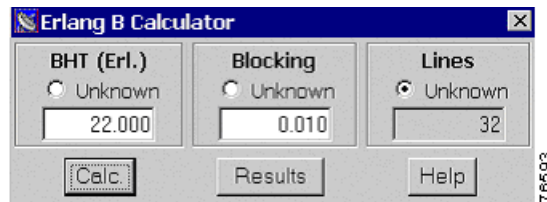
- Percent of calls queued = 46%
- Number of calls queued in the busy hour = 14,400 * 46% = 6,624 calls
- Average time calls spend in a queue = 12 seconds

From this input data, we first calculate the BHT in Erlangs:

$$\text{BHT} = (6,624 \text{ calls}) * (12 \text{ seconds}) / 3600 = 22 \text{ Erlangs of traffic}$$

Assuming the same desired grade of service of 0.01, or 1% blockage, we can calculate the additional PSTN trunk ports required to carry the queued traffic. Using another free Erlang-B calculator from the web (see [Figure 5-3](#)), we determine that we need 32 additional trunks to carry the queued traffic load.

Figure 5-3 Using an Erlang-B Calculator to Size Trunks for Queued Traffic



Calculating Total PSTN Trunks Required

The total number of trunks required to handle the traffic load for this call scenario (talk time and queue time for the respective call loads) is the sum of both results:

$$\text{Total trunks required} = 812 + 32 = 844 \text{ trunks}$$

This does not include trunks that may be required for call scenarios that require all calls to be treated before they are presented to available agents. For more information on call treatment calculations, see [Sizing Call Center Resources \(Front-End IP IVR for Call Treatment Example\)](#), page 5-9.

Sizing IP IVR Ports

In the preceding example, the IP IVR is used as a queue manager to queue calls when no agents are available. Following the same methodology used in that example, we first need to calculate the BHT presented to the IP IVR.

Again from the output of the Erlang -C calculator, we know the total number of calls queued and the average queue time for these calls:

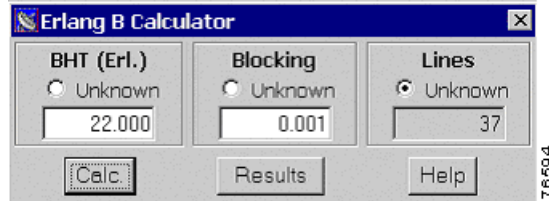
- Percent of calls queued = 46%
- Number of calls queued in the busy hour = $14,400 * 46\% = 6,624$ calls
- Average time calls spend in a queue = 12 seconds

From this input data we can calculate the BHT in Erlangs:

$$\text{BHT} = (6,624 \text{ calls}) * (12 \text{ seconds}) / 3600 = 22 \text{ Erlangs of traffic}$$

This is similar to calculating the additional trunks required for queued calls, but the difference is in the assumption we use for the grade of service to calculate the required IP IVR ports. In this case, however, once a caller gets through the gateway and no agents are available, we do not want that call to be blocked or lost (due to not having enough IP IVR ports). Instead, we want the call to find a port where it can queue. When sizing IP IVR ports for queuing, you may want to use the least possible blockage percent, such as 0.001. This would allow only one call blocked out of 1,000 instead of one out of 100, as was the case for sizing the PSTN trunks for incoming and queued calls.

Assuming the desired grade of service of 0.001, or 0.1% blockage, we can calculate the number of IP IVR ports required for queuing. Using any Erlang-B calculator (see [Figure 5-4](#)), we determine that we need 37 IP IVR ports to carry the queued traffic load.

Figure 5-4 Using an Erlang-B Calculator to Size IP IVR Ports for Queued Traffic

Sizing Call Center Resources (Front-End IP IVR for Call Treatment Example)

The example presented in this section builds upon the basic example from the previous section. First, all incoming calls are presented to the voice gateway from the PSTN (no internal calls are generated to agents), then calls are routed to the IP IVR for call treatment (such as initial greeting or gathering account information) 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 having all calls presented to the IP IVR first for initial call treatment is that the PSTN trunks are held longer for the period of call treatment. More IP IVR ports are also required to carry this additional load, beyond the number of ports required for queued calls.

Initial call treatment does not impact the number of required agents because the traffic load presented to the agents (number of calls, talk time, and service level) has not changed.

The following input parameters for this example are used to calculate the additional trunk ports and IP-IVR ports:

- Total BHCA received into the center via a voice gateway = 14,400.
- Average Handle Time (agent talk time) = 3 minutes and 16 seconds (total 196 seconds).
- No agent wrap-up time (For an example that adds agent wrap-up time, see [Sizing Call Center Resources \(Agent Wrap-up Time Example\)](#), page 5-11.)
- Call treatment is implemented, and all calls are treated before being routed to available agents. (Some calls are queued if no agent is available.)
- Desired service level of 80% of calls answered within 10 seconds.
- Desired grade of service for the PSTN ports on the voice gateway = 1%.
- IP IVR ports are used for call treatment and queuing and are sized with 0.001 blockage.

Sizing Additional PSTN Trunks for Initial IP IVR Call Treatment

More trunks are required to carry the additional traffic load for all 14,400 calls based on the average call treatment time of 30 seconds.

Using the same methodology as in the previous sections, we calculate the additional BHT for call treatment presented to the gateway from the PSTN, as follows:

$$\text{BHT} = (14,400 \text{ calls}) * (30 \text{ seconds}) / 3600 = 120 \text{ Erlangs of call treatment traffic}$$

Assuming the same desired grade of service of 0.01, or 1% blockage, we can calculate the additional PSTN trunk ports required to carry the call treatment traffic. Using any Erlang-B calculator (see [Figure 5-5](#)), we can determine that 138 additional trunks are needed.

Figure 5-5 Using an Erlang-B Calculator to Size Trunks for Call Treatment Traffic

The total number of trunks required to handle the traffic load for this call scenario (call treatment, talk time, and queue time for the respective call loads) is the sum of all three results:

$$\text{Total trunks required} = 138 + 812 + 32 = 982 \text{ trunks}$$

When sizing trunks, consider all call scenarios that may have a different load (BHCA) and a different average handle time presented to the gateway, then calculate the required trunks or resources for each situation and add the results. This total may include call flows such as self-service applications, where callers are routed to an IPI IVR and presented with a menu of choices to access information from various databases (such as bank accounts). At the end of the transaction, the caller hangs up and does not have to speak to an agent, so in this case only trunk and IP IVR resources are required.

Sizing Additional IP IVR Ports for Call Treatment

In this example, the IP IVR is used as a queue manager to queue calls when no agents are available and to provide initial call treatment. Following the same methodology used in the previous example, we first find the BHT presented to the IP IVR for call treatment, which is the same traffic load presented to the gateway from the PSTN:

$$\text{BHT} = (14,400 \text{ calls}) * (30 \text{ seconds}) / 3600 = 120 \text{ Erlangs of call treatment traffic}$$

This is similar to calculating the IP IVR ports required for queued calls.

Assuming a desired grade of service of 0.001, or 0.1% blockage, we can calculate the number of IP IVR ports required for call treatment. Using any Erlang-B calculator (see [Figure 5-6](#)), we determine that we need 151 IP IVR ports to carry the call treatment traffic load.

Figure 5-6 Using an Erlang-B Calculator to Size IP IVR Ports for Call Treatment Traffic

The total number of IP IVR ports required to handle the traffic load for this call scenario (call treatment and queue time for the respective call loads) is the sum of both results:

$$\text{Total IP IVR ports required} = 151 + 37 = 188 \text{ ports}$$

Remember that the number of calls being treated (normally, all calls received into the call center) and the average holding time are different than for calls that may need to queue. Normally, only a subset of all calls are queued, with a different average queue time.

Again, sizing IP IVR may include call flows such as self-service applications, where callers are routed to an IP IVR and presented with a menu of choices to access information from various databases (such as bank accounts). At the end of the transaction, the caller hangs up and does not require to speak to an agent, so in this case only trunk and IP IVR resources are required.

Sizing Call Center Resources (Agent Wrap-up Time Example)

Using the same data as in the basic example (see [Sizing Call Center Resources \(Basic Example\)](#), page 5-5) and assuming the agents spend an average of 45 seconds wrap-up time after each call, we can then use the Erlang-C calculator to size the required agents to handle the same traffic load (see [Figure 5-7](#)).

Figure 5-7 Using an Erlang-C Calculator to Size Agents with Call Wrap-up Time

Agents	Srv Lvl	Queued	Abndon	Usage	Ave Ans	Q Time	Avg Q	Trunks
972	49%	72%	1%	99%	00:22	00:30	86	972
973	53%	68%	0%	99%	00:18	00:27	73	973
974	57%	66%	0%	99%	00:16	00:24	63	974
975	60%	63%	0%	99%	00:14	00:22	55	975
976	64%	60%	0%	99%	00:12	00:20	48	976
977	67%	57%	0%	99%	00:11	00:19	42	977
978	69%	55%	0%	99%	00:09	00:17	38	978
979	72%	52%	0%	98%	00:08	00:16	33	979
980	74%	50%	0%	98%	00:07	00:15	30	980
981	77%	47%	0%	98%	00:07	00:14	27	981
982	79%	45%	0%	98%	00:06	00:13	24	982
983	80%	43%	0%	98%	00:05	00:13	22	983

Agent wrap-up time (after caller hangs up) does not impact trunk or IP IVR resources and should not be included in the average handle time when sizing trunks and IP IVR ports.

Every time you have a different call scenario where you need to size agents, make sure you use the output of that scenario to calculate trunks and IP IVR ports needed for queuing. In this case, the percent of calls queued is 43% and average queue time is 13 seconds.

Call Center Design Considerations

Observe the following design considerations when sizing call center resources:

- When sizing agents, IP IVR ports, and PSTN trunks, it is better to over-provision than to under-provision. The cost of trimming excess capacity is much less than lost revenue, bad service, or legal risks. In addition, some governmental agencies are required to meet minimum service levels, and out-sourced call centers may have to meet specific service level agreements.
- Consider the seasonal busy hour versus daily busy hour and the cost of additional resources versus incremental seasonal business. Retail business call centers need to staff agents according to the seasonal demands, such as holiday seasons. Every business has a different call load throughout the day or the week, and agents must be staffed accordingly (different shifts or staffing levels).
- If the call center receives different incoming call loads on different trunk groups, you must size the required trunks for each trunk group type.
- Consider marketing campaigns where call loads peak for a certain period of time. In addition, agent absenteeism may cause service levels to fluctuate, requiring additional IP IVR queuing ports and longer average queue times.
- There are many criteria to consider when deciding to use the same or separate IP IVR resources for queuing and self-service applications: service level tolerance (waiting in queue), cost, excess capacity, different busy hours for different applications, different business owners, different maintenance windows, and so forth.
- Allow for growth and unforeseen fluctuations. Reserve about 20% of capacity to accommodate changes from assumptions or averages (reality will be different than assumptions).



Sizing IPCC Components and Servers

Proper sizing of your Cisco IP Contact Center (IPCC) solution is important for optimum system performance and scalability. Sizing considerations include how many agents the solution can support, the maximum busy hour call attempts (BHCA), and other variables that affect the number and type of servers required to support the deployment. Regardless of the deployment model chosen, IPCC 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 a single-site IPCC deployment. Design considerations for a multi-site deployment are similar and are mainly a matter of scaling the design upward. The design considerations, best practices, and performance characteristics presented in this chapter are derived from testing and, in some cases (noted where appropriate), extrapolated test data. This information is intended to enable you to size and provision IPCC solutions appropriately.

Sizing Considerations for IPCC

This section discusses sizing considerations for the following IPCC components:

- [Core IPCC Components, page 6-1](#)
- [CTI Components, page 6-9](#)
- [Additional Sizing Variables, page 6-4](#)

Core IPCC Components

When sizing IPCC deployments, IP Telephony components are typically the most critical factor in capacity planning. Good design, including multiple Cisco CallManagers, can be configured to support significant call loads. For additional information on capacity and sizing of IP Telephony components, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*.

Additionally, because of varying agent and skill group capacities, proper sizing of the CTI OS and Cisco Agent Desktop servers should be considered directly after the IP Telephony components.

Finally, the remaining ICM components, while able to scale extremely well, are affected by specific sizing variables that also have a dramatic impact on processor usage.

These factors, discussed in this section, must be considered and included in the planning of any deployment.

The information presented in [Table 6-1](#) does not apply equally to all implementations of IPCC. The data is based on testing in particular scenarios, and it serves only as a guide, along with the sizing variables information in this chapter. As always, you should be conservative when sizing and should plan for growth. In addition, this section does not show the Administrative Workstation (AW) in a co-resident configuration with other components because Cisco no longer recommends a co-resident AW configuration.

**Note**

Sizing considerations are based upon performance test data. Major ICM software processes were run on individual servers to measure their specific CPU and memory usage. Reasonable extrapolations were used to derive capacities for co-resident software processes and multiple CPU servers (dual and quad CPUs). This information is meant as a guide for determining when ICM software processes can be co-resident within a single server and when certain processes need their own dedicated server. [Table 6-1](#) assumes that the deployment scenario includes *two* fully redundant servers that are synchronized. While a non-redundant deployment might theoretically deliver higher capacity, no testing has been done to validate this theory. Therefore, you can and should refer to [Table 6-1](#) for sizing information on either redundant and non-redundant deployments.

Table 6-1 Minimum Hardware Configurations for IPCC Core Components

Configuration	Minimum Hardware Requirement	Minimum RAM	Number of CPUs on each Server	Maximum BHCA	Maximum Agents
Peripheral Gateway, Router, and Logger (PROGGER) ¹	1.13 GHz Pentium III	1 GB	1	7,000	200 (max. 5 skill groups/agent)
		1 GB	2	10,400	300 (max. 5 skill groups/agent)
		2 GB	4	14,500	415 (max. 5 skill groups/agent)
Router and Logger (ROGGER) ²	1.13 GHz Pentium III	1 GB	1	87,000	2,480
		1 GB	2	100,000	2,850
		2 GB	4	150,000	4,280
“Hybrid 1” Peripheral Gateway, CTI Server ³	1.13 GHz Pentium III	1 GB	1	22,000	630
		1 GB	2	33,000	940
		2 GB	4	55,000	1,570
Cisco CallManager Peripheral Gateways ⁴	1.13 GHz Pentium III	1 GB	1	35,000	850
		1 GB	2	45,000	1,200
		2 GB	4	65,000	1,800
IP IVR Peripheral Gateways ⁵	1.13 GHz Pentium III	1 GB	1	80,000	2,280
		1 GB	2	120,000	3,400
		2 GB	4	180,000	3,400
ICM Router ⁶	1.13 GHz Pentium III	2 GB	1	110,000	3,100
	1.26 GHz Pentium III	2 GB	2	140,000	4,000
	900 MHz Xeon	4 GB	4	195,000	5,500
ICM Logger	733 MHz Pentium III	2 GB	1	100,000	2,850
	1.26 GHz Pentium III	2 GB	2	150,000	3,570
	900 MHz Xeon	2 GB	4	250,000	5,000

Table 6-1 Minimum Hardware Configurations for IPCC Core Components (continued)

Configuration	Minimum Hardware Requirement	Minimum RAM	Number of CPUs on each Server	Maximum BHCA	Maximum Agents
CTI OS Server ⁷	1.13 GHz Pentium III	1 GB	1	24,000	500 (max. 5 skill groups/agent)
Administrative Workstation (AW) Historical Data Server (HDS)	1.13 GHz Pentium III	1 GB	1	80,000	2,280
	1.26 GHz Pentium III	2GB	2	100,000	2,850
	900 MHz Xeon	4GB	4	125,000	3,500

1. PROGGER is a configuration in which all of the ICM software and processes except the AW/HDS are on a single server, and it is subject to the following restrictions: single site only, maximum of six Peripheral Interface Managers (PIMs); Cisco Agent Desktop requires a separate server; CTI OS maximum of five skill groups; Historical Data Server (HDS) must be on a separate AW (no reports run on PROGGER); Logger database is limited to 14 days. This configuration requires dual (and separate) private network NICs in addition to a visible network NIC.
2. ROGGER is a configuration in which the Router and Logger share a server, and the Peripheral Gateway (PG) and AW processes are off-loaded to other servers.
3. Hybrid Peripheral Gateway (PG) running on a single server with Cisco CallManager PG or PIM, IP IVR PG or PIM, and CTI Server.
4. Peripheral Gateway for Cisco CallManager PG and CTI Server (no Voice Response Unit PG, CTI OS, or Cisco Agent Desktop).
5. Peripheral Gateway for IP IVR.
6. Router is a single-threaded process that does not have significant benefits on a multi-processor platform, but some benefit is realized because the OS can still take advantage of the extra processor.
7. Adding multiple skill groups can significantly reduce the number of agents supported by a CTI OS server. Also see CTI OS scaling in [Table 6-4](#).

The effects of the sizing considerations in [Table 6-1](#) must be applied to each deployment to gauge server sizing requirements accurately. The following considerations and guidelines apply to the information in [Table 6-1](#):

- Formal and critical call center deployments are encouraged to use dual CPU configurations, especially on the PROGGER.
- When BHCA exceed the capacity of a PG, more PGs can be added, potentially affecting the ICM Router.
- You must observe both the maximum BHCA and the maximum agent count for each specific configuration.
- It is possible to have different server configurations for the ROGGER and the PGs, as long as you observe the BHCA and agent maximums for each separate component.
- These maximums assume a normal amount of CTI traffic for each given configuration. Extraordinary CTI traffic (from very large IVRs, for example) will cause the BHCA and agent maximums to be too high.
- These numbers are based on an average of nine “Run Voice Response Unit (VRU) scripts,” running consecutively in the ICM script, per IVR call. If a deployment has a more complex ICM/IVR script than this, it will also decrease the maximum BHCA shown in [Table 6-1](#).
- These numbers are based on approximately 45% of the calls being queued to the IVR. For the PROGGER, the percentage of calls queued was 23%.

Additional Sizing Variables

Many variables in the IPCC 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 IPCC components. In addition, [Table 6-2](#) summarizes the sizing variables and their effects.

Busy Hour Call Attempts (BHCA)

The number of calls attempted during a busy hour, while not the only variable for sizing, is by far the most important metric. You should design your IPCC solution to achieve 100% completion of all attempted calls. That is, Busy Hour Call Completions (BHCC) should equal BHCA. As BHCA increase, there is a linear or exponential increase in the load on all IPCC components, most notably on Cisco CallManager, IP IVR, and the Cisco CallManager PG.

Example effects (most affected components):

- Cisco CallManager PG CPU usage increased from 5% at 1,250 BHCA to 20% at 5,000 BHCA.
- Cisco CallManager CPU usage increased from 9% at 1,250 BHCA to 30% at 5,000 BHCA.

Agents

The number of agents in a contact center is limited by many factors, including the capacity of Cisco CallManager. Currently, this limit is 250 agents per Cisco CallManager server or 1,000 agents per cluster. In addition, the Logger CPU usage increases linearly as agents are added.

Example effect (most affected component):

- Cisco CallManager PG CPU usage increased from 18% with 200 agents to 22% with 800 agents.

Skill Groups

The number of skill groups has significant effects on the CTI OS Server, the Cisco CallManager PG, and the ICM Router. Cisco recommends that you limit the number of skill groups per agent to less than 15, when possible, and that you periodically remove unused skill groups so that they do not affect system performance. You can also manage the affects on the CTI OS server by adjusting the frequency of statistical updates.

Example effects (most affected components):

- CTI OS users must follow the sizing guidelines in [Table 6-4](#).
- Cisco CallManager PG CPU usage increased from 11% with 6 skill groups to 26% with 10 skill groups. CPU usage doubled for every additional 4 skill groups per agent. Similar results were observed for the ICM Router.

Queuing

IP IVR places calls in a queue and plays announcements until an agent answers the call. For sizing purposes, it is important to know whether the IVR will handle all calls initially 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 significantly affects the performance of the ICM Router and Voice Response Unit (VRU) PG.

Example effect (most affected component):

- Logger CPU usage increased from 10% with no calls queued to 20% with 45% of calls queued.

ICM Script Complexity

As the complexity and/or number of ICM scripts increase, the processor and memory overhead on the ICM Router and VRU PG will increase significantly. The delay time between replaying RunVRU scripts also has an impact.

Example effects (most affected components):

- ICM Router CPU usage increased from 5% with 1 Run VRU script node to 10% with 9 Run VRU script nodes, when run consecutively without delay times between “Run VRU” nodes.
- VRU PG CPU usage increased from 9% with 1 Run VRU script node to 24% with 9 Run VRU script nodes, when run consecutively without delay times between “Run VRU” nodes.

Reporting

Real-time reporting can have a significant effect on the Logger’s processing due to SQL writes. This is why a separate server is recommended for an AW and/or Historical Data Server (HDS) to off-load reporting overhead from the Logger.

Conferences Setup

Contact center agents sometimes conference in other agents or a supervisor. This call setup typically requires the use of transcoding resources if agents are using dissimilar codecs (such as in multi-site deployments over a WAN). Transcoding resources have a device weight in Cisco CallManager that you must also consider. To size an IPCC solution based on the number of conferencing resources needed, refer to the device weights in the *Cisco IP Telephony Solution Reference Network Design Guide*.

SoftPhone

The media-terminated IP softphone that is integrated with the agent desktop uses Skinny Client Control Protocol (SCCP) and has a device weight of 1.

IVR Script Complexity

As IVR script complexity increases with features such as database queries, the load placed upon the IVR server also increases. There is no good rule of thumb or benchmark to characterize the IP IVR performance when used for complex scripting, complex database queries, or transaction-based usage. Cisco recommends that you test complex IVR configurations in a lab or pilot deployment to determine how they affect the processor.

IP IVR Self-Service Applications

In deployments where the IP IVR is also used for self-service applications, the self-service applications are in addition to the IPCC load and must be factored into the sizing requirements.

Cisco CallManager Database

Very large updates to the Cisco CallManager database during busy hours negatively affect call processing. Therefore, Cisco recommends that you do not perform large database updates (involving more than 1,000 records) during busy hours. Instead, perform them either off-line, during a maintenance window, or during non-peak hours.

Third-Party Database and Cisco Resource Manager Connectivity

Carefully examine connectivity of any IPCC solution component to an external device and/or software to determine the overall affect on the solution. Cisco IPCC 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, Cisco recommends that you engage a firm with the appropriate experience and certifications to help you design your IPCC solution.

Table 6-2 Sizing Variables for IPCC Components

IPCC Component	Sizing Variables	How Variables Affect Overall Solution
Router	Skill groups Script complexity	Mean CPU usage doubles with every 4 skill groups per agent (based on 600 agents and 20,000 BHCA). Mean CPU usage doubles when running 9 “Run VRU Script” nodes instead of just one.
Logger	Agents Queuing CPU spikes Reporting Data retention period Event tables	CPU usage increases incrementally (1% for every 200 agents) with every agent added. Mean CPU usage doubles with every 12 skill groups per agent (based on 600 agents and 20,000 BHCA). Mean CPU usage doubles when going from 0 to 45% calls queued. Real-time reporting can have a significant effect on the Logger’s processing due to Sequel writes. If no HDS is used, real-time reporting also affects CPU usage. (AW HDS should be on a separate server.)
Cisco CallManager PG	Skill groups Multi-site Translation routing	Mean CPU usage doubles with every 4 skill groups per agent (based on 800 agents and 20,000 BHCA).
VRU PG	Script complexity	Mean CPU usage increases by 200% when running 9 “Run VRU Script” nodes instead of just one.
CTI OS	Agents Skill groups	CTI OS supports up to 450 agents per server with 5 skill groups, but fewer as skill groups increase.
PROGGER <ul style="list-style-type: none"> Router Logger CCM PG PIM VRU PG PIM CTI Server CTI OS Server 	Agents Skill groups Script complexity Queuing CPU spikes Skill groups	Single site only (redundant or non-redundant), on the same LAN. 6 PIMs (mixed). CTI OS only (Cisco Agent Desktop requires a separate server). 100 agents AW with HDS must be on a separate server. No reports run on PROGGER. Maximum of 5 skill groups per agent. Logger database limited to 14 days (HDS can grow). Requires 3 separate Network Interface Cards (NICs), 2 private and 1 visible network
ROGGER <ul style="list-style-type: none"> ICM Router Logger 	Agents Skill groups Script complexity Queuing CPU spikes	Same as Router plus Logger.

Table 6-2 Sizing Variables for IPCC Components (continued)

IPCC Component	Sizing Variables	How Variables Affect Overall Solution
Hybrid PG	Skill groups	100 agents with ICM 4.6.1; 200 agents with ICM 4.6.2.
• CCM PG PIM	Script complexity	Maximum of 5 skill groups per agent.
• VRU PG PIM	Maximum agents (CTI OS)	CTI OS only (Cisco Agent Desktop requires a separate server).
• CTI Server		
• CTI OS Server		

Peripheral Gateway Sizing Recommendations

An ICM Peripheral Gateway (PG) translates messages coming from the Cisco CallManager servers, the IP IVR, or other third-party automatic call distributors (ACDs) or voice response units (VRUs), into messages that are then sent to and understood by the ICM. It also translates ICM messages so that they can be sent to and understood by the peripheral devices.

The Peripheral Interface Manager (PIM) is the software process that runs on the PG and performs the message translation and control. Every peripheral device that is part of the IPCC solution has to be connected to a PG and PIM.

When sizing PGs and PIMs, consider the following questions:

- Can PGs be co-resident with other ICM components? If so, what is the BHCA limit before the PGs must be moved to separate servers?
- How many PIM types can run on the same PG?

[Table 6-3](#) lists PG and PIM sizing recommendations based on ICM Software Release 4.6.1.

Table 6-3 PG and PIM Sizing Recommendations

Sizing Variable	Recommendation, Based on ICM Software Release 4.6.2
Maximum number of PGs per ICM	40
Protocol used between the ICM and PG	IP, Computer Supported Telecommunications Applications (CSTA)
Can PGs be remote from ICM?	Yes
Can PGs be remote from Cisco CallManager or IP IVR?	No
PIM types	Cisco CallManager, IVR, and ACD ¹
Maximum PG types per server platform	2
Maximum number of PIMs per PG	32 ²
Maximum number of PIM types per PG (CTI server may be added)	2 + CTI Server
Maximum number of IVRs controlled by one Cisco CallManager	4
Maximum number of CTI servers per PG	1

Table 6-3 PG and PIM Sizing Recommendations (continued)

Sizing Variable	Recommendation, Based on ICM Software Release 4.6.2
Single PG server capacity ³ : <ul style="list-style-type: none"> • Cisco CallManager PIMs • IP IVR PIMs • CTI Server 	<ul style="list-style-type: none"> • 2 • 2 • 1
PG co-resident with Cisco CallManager on Media Convergence Server (MCS)?	No

1. One PIM is required per IP IVR, and one PIM is required per Cisco CallManager cluster.
2. The theoretical maximum is 32 PIMs on one PG. Actual number of IVR PIMs is determined by the size of the IPCC deployment (agents, IVR ports, BHCA, and so forth). Under most circumstances, 5 PIMs per PG is a reasonable limit.
3. Assume all listed items are running on the same server. The configuration can be duplexed for redundancy. Small IPCC configurations of less than 10,000 BHCA could exceed this capacity.

Other ICM Applications

Although the following products have not been tested, they can be used in an IPCC solution:

- **WebView**

A dedicated WebView server should be able to handle 50 simultaneous WebView clients. Additional performance profiling for WebView will be done in future testing. For more information, refer to *Cisco ICM Software Pre-installation Planning: Network and Site Requirements*, available online at <http://www.cisco.com/univercd/cc/td/doc/product/icm/icm46/plan/plannets.pdf>

- **Historical Data Server (HDS)**

Cisco recommends that you install the HDS on its own server. Additional sizing considerations depend on the frequency and type of reporting required, and these factors vary widely by deployment. Cisco recommends that you determine the reporting requirements prior to ICM configuration and, if possible, prior to sizing the HDS(s).

CTI Components

The main CTI components are:

- [CTI OS, page 6-9](#)
- [Cisco Agent Desktop and Cisco Supervisor Desktop, page 6-9](#)

CTI OS

The CTI Object Server (OS) capacity decreases as skill groups are added because the processor usage increases significantly as the server queries for agent and skill group statistics every ten seconds. (See [Table 6-4](#).) You may use multiple servers for additional scaling. Testing for dual and quad processor servers has not yet been completed.

Table 6-4 CTI OS Scaling (Dedicated Single-Processor Server)

Skill Groups	Maximum Agents
5	500
10	300

Cisco Agent Desktop and Cisco Supervisor Desktop

For an architectural overview of Cisco Agent Desktop and Cisco Supervisor Desktop, refer to the chapter on [IPCC Agent Desktop and Supervisor Desktop](#).



Note

The guidelines in this section for sizing Cisco Agent Desktop (CAD) deployments are based on testing completed by Spanlink Communications and are based on a combination of real and simulated agents. All sizing is based on the current Bill of Materials for ICM PGs and CAD.

Server capacities for the Cisco Agent Desktop (CAD) CTI Option vary based on the total agents and on whether or not Silent Monitoring and Recording are required. While both CAD 4.2.1 and 4.4 support these features, CAD 4.4 scales far better and overcomes the need for multiple instances of a set of CAD servers for deployments having more agents than the capacity of one Silent Monitor application server.

Component Sizing

This section presents sizing guidelines for the following Cisco Agent Desktop (CAD) components:

- [CAD Base Server, page 6-9](#)
- [CAD Silent Monitor Server, page 6-10](#)
- [RASCAL Application Server, page 6-11](#)

CAD Base Server

The CAD Base Servers consist of a set of application servers that run as Windows NT services. The server applications include the Lightweight Directory Access Protocol (LDAP) Primary, Enterprise application server, Call/Chat application server, and Synchronization application server. The Enterprise application interfaces with the standard CTI Server typically running on a PG. In addition, there are application servers that may be placed on the same or separate computers as the CAD Base Servers. These additional applications include Silent Monitor, RASCAL, IP Phone Agent, and Tomcat.

A set of CAD Base Servers plus the additional application servers corresponds to a logical call center (LCC). The maximum number of agents that can be supported by a single LCC is as follows:

CAD 4.2.1 ¹	CAD 4.4
400	800

1. Applying CAD Hot Fix 4 increases the capacity of CAD 4.2.1 to 800 agents within an LCC when silent monitoring is not used.

Note that 800 agents corresponds to approximately 15,000 BHCA, based on the assumption of 20 calls per agent per hour and on the other call volume assumptions discussed in the section on the [RASCAL Application Server](#), page 6-11.

CAD Silent Monitor Server

The CAD Silent Monitor Server simultaneously monitors the RTP streams spanned on the local switch. It uses a SPAN Port to sniff each RTP stream that might potentially be monitored by a supervisor. All streams are being monitored all the time. The maximum number of simultaneously monitored sessions (that is, the total number of supervisors actively monitoring agents, plus the number of concurrent recording sessions) is 30 for both CAD versions.

The CAD Silent Monitor Server must be at the *same physical location* as the agents to be monitored. A physical location is defined as a set of interconnected switches with no intervening routers. Therefore, deployments with agents at multiple sites must provision for a Silent Monitor server at each site. Additionally, the Silent Monitor server is currently certified for uni-processors, an important design consideration when considering the co-resident options with CAD 4.4.



Note

In CAD 4.4, there may be multiple Silent Monitor servers that are all part of a single LCC and appear as a single instance for administrative and management purposes. This is a significant enhancement over CAD 4.2.1, which requires a separate instance of CAD (that is, the LCC) for each Silent Monitor server.

The Silent Monitor server is sized based on a combination of active calls and total streams monitored by the server. For example, if the percentage of time that agents are actually talking is low, more calls can be monitored; if agents are on the phone most of the time, fewer calls can be monitored.

[Table 6-5](#) and [Table 6-6](#) define the capacity in terms of the number of agents and the percent of time that agents are talking. It is important to note that the percent of time that agents are talking is the maximum possible peak time. The total number of sessions that can be monitored and/or recorded simultaneously is fixed at 30.

Table 6-5 CAD Silent Monitor Server 4.4 Sizing

Maximum Number of Agents	Peak % Time Agents Are On-Call (Talking)	Maximum Number of Simultaneous Calls	Maximum Simultaneous Sessions Monitored
300	100	300	30
333	90	300	30
375	80	300	30

Table 6-6 CAD Silent Monitor Server 4.2.1 Sizing

Maximum Number of Agents	Peak % Time Agents Are On-Call (Talking)	Maximum Number of Simultaneous Calls	Maximum Simultaneous Sessions Monitored
128	100	128	30
142	90	128	30
160	80	128	30

RASCAL Application Server

The RASCAL Application Server stores the recorded conversations and makes them available to the Supervisor Log Viewer application. Additionally, the server stores call and agent logs and assembles the agent and team statistics for Supervisor Desktop. The RASCAL application server does not have to be located with the agents.

The speech encoding algorithm directly affects the capacity. RASCAL can support a total of 16 simultaneously recorded G.711-encoded calls versus a total of four G.729-encoded calls. The RASCAL server recording is primarily I/O-bound for G.711 and primarily CPU-bound for G.729. Recording quality might be affected if the maximum number of recordings is exceeded (for example, there may be speech break-up).

Table 6-7 summarizes the raw RASCAL performance capacity.

Table 6-7 Effect of Compression Type on RASCAL Server Capacity

CODEC	Maximum Simultaneous Recordings
G.711	16
G.729	4

To use this performance data in the context of a contact center, a number of assumptions must be made about the call volume. The following example is appropriate for many contact centers, but you should make specific assumptions and calculations expressly for any customer planning to do a significant amount of recording. The following assumptions about the call volume and contact center apply to this example:

- 10-hour contact center working day
- 20 calls per agent per hour
- Average handle time (AHT) of 3 minutes per call (2 minutes on the call and 1 minute wrap-up)
- 0.5% of calls (or 1 call per agent per day) are recorded
- An agent-to-supervisor ratio of 10:1

Agent resources are calculated using an Erlang C computation based on the assumption of 180 second average handle time and a service level of 90% of calls answered in 10 seconds.

Recording resources are calculated using the above assumptions and an Erlang B computation assuming 0.001 blockage. In the case of recording, there is no actual blockage; instead, an additional recording resource is used. The following formula can be used to calculate recording Erlangs in the busy hour based on percentage of calls monitored:

$$\text{BHCC calls per hour} * (\% \text{ Recorded}) * (2 \text{ minutes per call}) * (1 \text{ hour}) / (60 \text{ minutes}), \text{ or} \\ \text{BHCC} * (\% \text{ Recorded}) / 30$$

For example, in a contact center with 1000 calls per hour and a 0.5% call recording rate, the recording Erlang is

$$(1000 * 0.005) / 30 = 0.1667$$

Using an Erlang B calculator with 0.001 blocking, this contact center would require no more than 3 simultaneous recording resources.

[Table 6-8](#) summarizes the number of agents and recording resources required for BHCC values from 1000 to 20,000.

Table 6-8 Agents and Recording Resources Required for Various Call Volumes

BHCC	Number of Agents	0.5% Calls Recorded	Number of Supervisors
1000	59	3	6
2000	112	4	11
4000	215	5	22
6000	317	6	32
8000	419	7	42
10,000	520	8	52
12,000	621	8	62
14,000	722	9	72
16,000	822	10	82
18,000	923	10	92
20,000	1023	11	102

Co-Residency Options for CAD

This section lists the server co-residency options for Cisco Agent Desktop (CAD).

Both CAD 4.2.1 and CAD 4.4 support a single computer configuration, with PG, Router, Logger, and CAD co-resident on the same server. This configuration can support a maximum of 50 agents.

[Table 6-9](#) lists recommendations for installing CAD 4.2.1 co-resident on a PG. For CAD 4.2.1 to be co-resident on a duplexed (redundant) PG, CAD 4.2.1 Hot Fix 4 should be installed.

Table 6-9 CAD 4.2.1 Configuration Recommendations

Number of Agents	CAD Base Servers Plus RASCAL Server	Monitor Server
Up to 64	Co-resident on PG	Co-resident on PG
65 to 128 ¹	Co-resident on PG	Dedicated server
Up to 300	Co-resident on PG	No monitoring
301 to 800	Dedicated server	No monitoring

1. The number of agents may be up to 160 if the peak percent of the time agents are talking does not exceed 80%. See [Table 6-6](#).

New in CAD 4.4 is the ability to have multiple monitor servers with a single set of CAD Base Servers. The CAD Base Servers can be co-resident on a PG, including one that is duplexed with a private LAN. [Table 6-10](#) summarizes the co-residency options.

Table 6-10 CAD 4.4 Configuration Recommendations

Number of Agents	CAD Base Servers Plus RASCAL Server	Monitor Server
Up to 100	Co-resident on PG	Co-resident on PG
101 to 300	Co-resident on PG	Dedicated server
301 to 800	Dedicated server	Dedicated server

Summary

Proper sizing of IPCC 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 should uncover critical sizing variables, and these considerations should be applied to the final design and hardware selection.

Correct sizing and design can ensure stable deployments up to 1,000 agents and 30,000 BHCA. For smaller deployments, cost savings can be achieved with careful planning and co-resident ICM components (for example, PROGGER, ROGGER, and Hybrid PG).

Additionally, designers should pay careful attention to the CTI components, which can often be the limiting factor in a deployment. The number of skill groups, while often hard to determine in the pre-sales phase, is a critical factor in the number of agents a single CTI OS server can support, especially when the process is co-resident on a PG or PROGGER. 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.



IPCC Agent Desktop and Supervisor Desktop

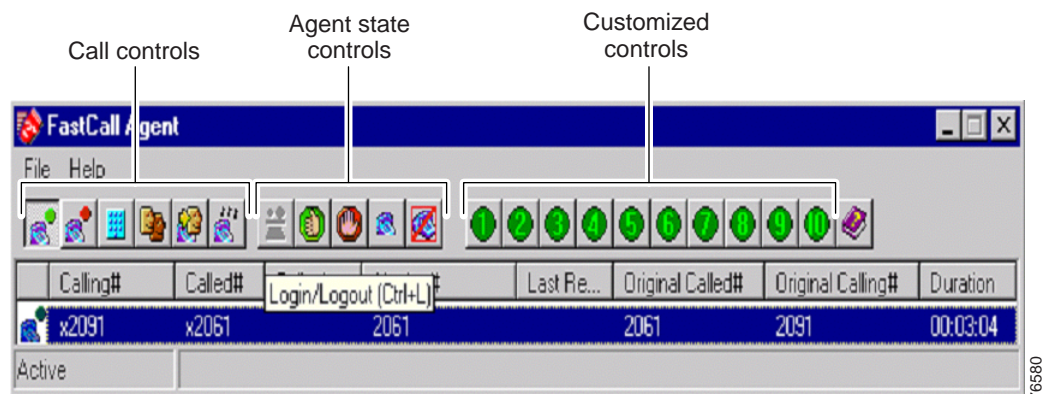
A required component of an IPCC deployment is the IPCC agent desktop (see [Figure 7-1](#)). From the agent desktop, the agent performs agent state control (login, logout, ready, not ready, and wrap-up) and call control (answer, release, hold, retrieve, make call, transfer, and conference).



Note

While it is possible for call control to be performed from the IP Phone, Cisco recommends that all call control (except answer and release) be done from the agent desktop application.

Figure 7-1 IPCC Agent Desktop

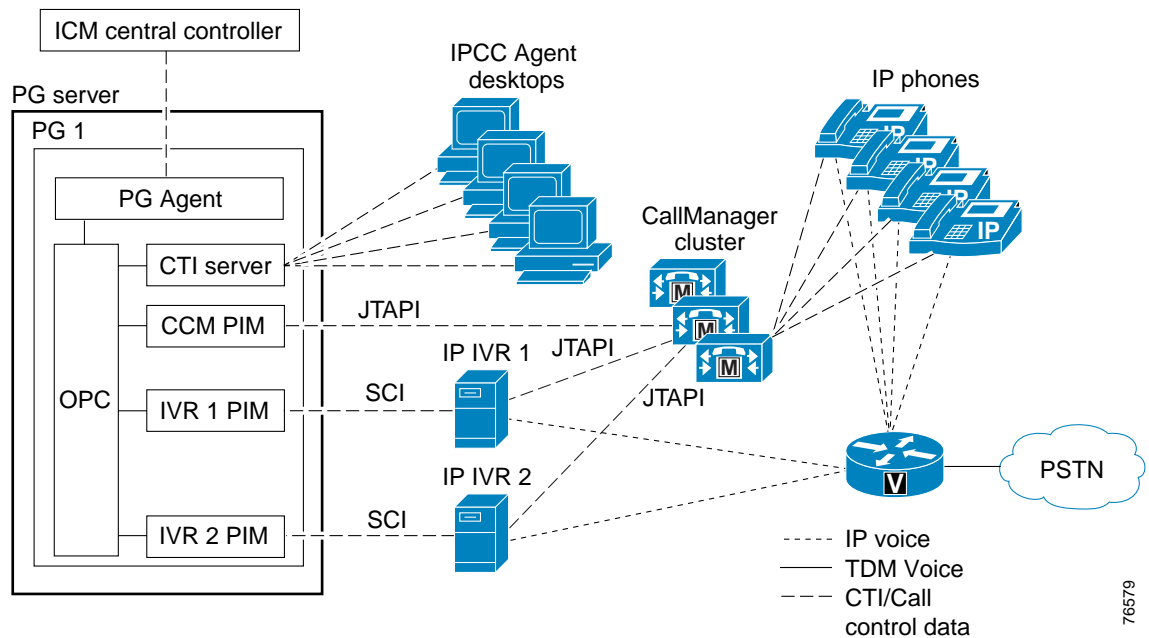


Within the Cisco Intelligent Contact Management (ICM) configuration, an IPCC agent desktop is not statically associated with any specific agent or IP Phone extension. Agents and IP Phone extensions (device targets) must be configured within the ICM configuration, and both are associated with a specific Cisco CallManager cluster. When logging in from their agent desktop, the agent is presented with a dialog box prompting them for their agent ID, password (optional, depending upon agent configuration in the ICM), and the IPCC phone extension to be used for this login session. It is at login time that the agent ID, IP Phone extension (device target), and agent desktop IP address are all dynamically associated. The association is released upon agent logout. This allows an agent to hot-desk from one agent desktop to another. This also provides for laptop roaming so that an agent can take their laptop to any IP Phone and log in from that IP Phone (assuming the IP Phone has been configured in the ICM and Cisco CallManager to be used in an IPCC deployment). Agents can also log in to other IP Phones using the extension mobility feature.

All communication from the agent desktop passes through the CTI Server process (see [Figure 7-2](#)). The CTI Server process can run on the same PG server as the Cisco CallManager PIM process (typical scenario) or on a separate server. If the CTI Server runs on its own server, then that server is sometimes

called a CTI Gateway (CG) as opposed to a Peripheral Gateway (PG). The hardware and third-party software requirements for a CG and PG are the same. Server sizing is discussed in the chapter on [Sizing IPCC Components and Servers](#).

Figure 7-2 Agent Desktop Communication with CTI Server



For each Cisco CallManager PIM (and Cisco CallManager cluster), there is one CTI Server. The CTI Server and the Cisco CallManager PIM communicate with each other via the Open Peripheral Controller (OPC) process. All communications from the CTI Server are passed on by OPC to the Cisco CallManager PIM process and then typically onto either the ICM Central Controller or the Cisco CallManager. All agent state change requests flow from the agent desktop through the CTI Server to the Cisco CallManager PIM to the ICM Central Controller as the ICM Central Controller is monitoring 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. Most call control (answer, release, hold, retrieve, make call, and so on) will flow from the agent desktop through the CTI Server to the Cisco CallManager PIM and then to the Cisco CallManager. The Cisco CallManager then performs the requested call or device control. It is the role of the Cisco CallManager PIM to keep the IPCC agent desktop and the IP Phone in sync with one another.

Types of IPCC Agent Desktops

There are two types of IPCC agent and supervisor desktops available:

- Cisco Agent Desktop, a packaged agent desktop solution
- CTI Object Server (CTI OS) Toolkit, for agent desktops that need to be customized or integrated with other applications on the desktop or with customer databases such as a Customer Relationship Management (CRM) application

In addition to an agent desktop, a supervisor desktop is available with each option.

Cisco Agent Desktop is a packaged agent and supervisor desktop application. It has a system administration interface that allows configuration of the desktop and workflow automation. Desktop configuration includes defining what buttons are visible; specifying call, voice, and data processing functions for buttons; and specifying what telephony data will appear on the desktop. The workflow automation enables data processing actions to be scheduled based on telephony events (for example, popping data into third-party applications on answer and sending email on dropped events). The workflow 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.

While CTI OS is a toolkit, it does provide a pre-built, operational agent and supervisor desktop executable. Source code for these executables is provided with the toolkit to allow for easy customization.

One major feature distinction between the two desktop solutions is that Cisco Agent Desktop offers a silent monitoring and recording capability while CTI OS does not. Silent monitoring and recording will be made available in a future CTI OS toolkit release.

Cisco Agent Desktop, Supervisor Desktop, and CTI OS can co-exist with the same Cisco CallManager cluster and Cisco CallManager PIM, but the configuration of agents, supervisors, and skill groups must be kept separate. Cisco Supervisor Desktop cannot be used to monitor a CTI OS agent desktop, nor can a CTI OS supervisor monitor a Cisco Agent Desktop agent.

The following sections cover these two desktop options separately. Both rely upon communication with the CTI Server as described in the previous section.

CTI Object Server (CTI OS) Toolkit

Cisco CTI Object Server (CTI OS) is a high-performance, scalable, fault tolerant server-based solution for deploying CTI applications. It is Cisco's latest version of the CTI implementation. CTI OS serves as a single point of integration for third-party applications, including Customer Relationship Management (CRM) systems, data mining, and workflow solutions. Configuration and behavior information is managed at the server, simplifying customization, updates, and maintenance. Servers can be accessed and managed remotely. Thin-client and browser-based applications that do not require Cisco software on the desktop can be developed and deployed with CTI OS.

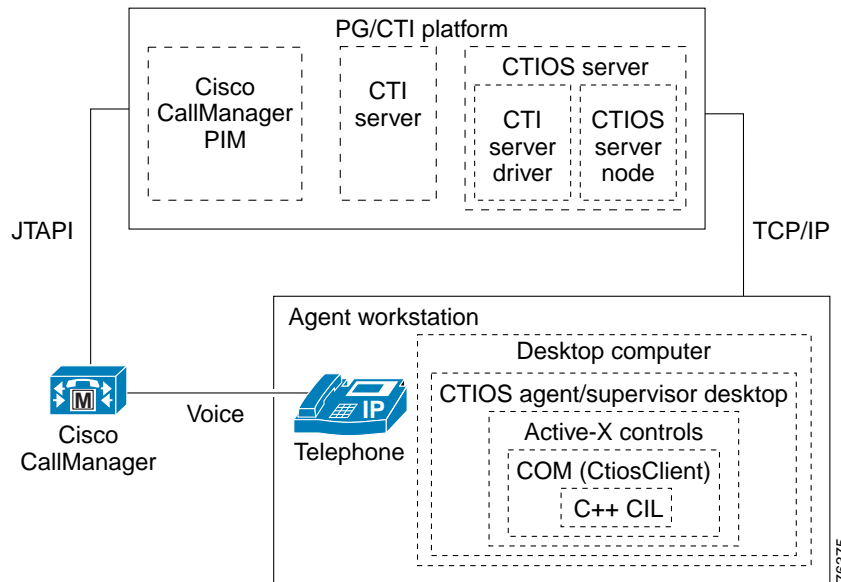
CTI OS incorporates the following major components:

- CTI OS Toolkit
- Client Interface Library
- CTI OS Agent Phone
- CTI OS Supervisor Phone

Architecturally, CTI OS Server is positioned between the CTI OS agent desktop and the CTI Server. CTI OS Server provides a mechanism to maintain agent and call state information so that the agent desktop can be stateless. This architecture provides the necessary support to develop a browser-based agent desktop if desired.

The CTI OS system consists of three major components (see [Figure 7-3](#)):

- CTI OS Server
- CTI OS Agent Desktop
- CTI OS Supervisor Desktop (only on Cisco IPCC for now)

Figure 7-3 CTI OS Basic Architecture

CTI OS Server connects to CTI Server via TCP/IP.

CTI OS typically runs on the same server as the CTI Server and Cisco CallManager PIM processes. As an IPCC site gets larger, the first process to split off of the PG/CTI Server is the CTI OS server process. Multiple CTI OS server processes can connect to a CTI Server. The maximum number of simultaneous agent logins for a CTI OS server is 500. Server sizing for CTI OS is covered in the chapter on [Sizing IPCC Components and Servers](#).

CTI OS is typically installed in duplex mode, with two CTI OS servers running in parallel for redundancy. CTI OS desktop application will randomly connect to either server and automatically fail over to the other server if the connection to the original CTI OS server fails. CTI OS can also run in simplex mode with all clients connecting to one server, but Cisco does not recommend this configuration.

Cisco Agent Desktop and Cisco Supervisor Desktop

Throughout this section, the usage of Cisco Agent Desktop refers to capabilities of both Cisco Agent Desktop and Cisco Supervisor Desktop, except where specifically noted. The Cisco Supervisor Desktop integrates with Cisco Agent Desktop and allows supervisory functions such as barge-in, intercept, and silent monitoring.

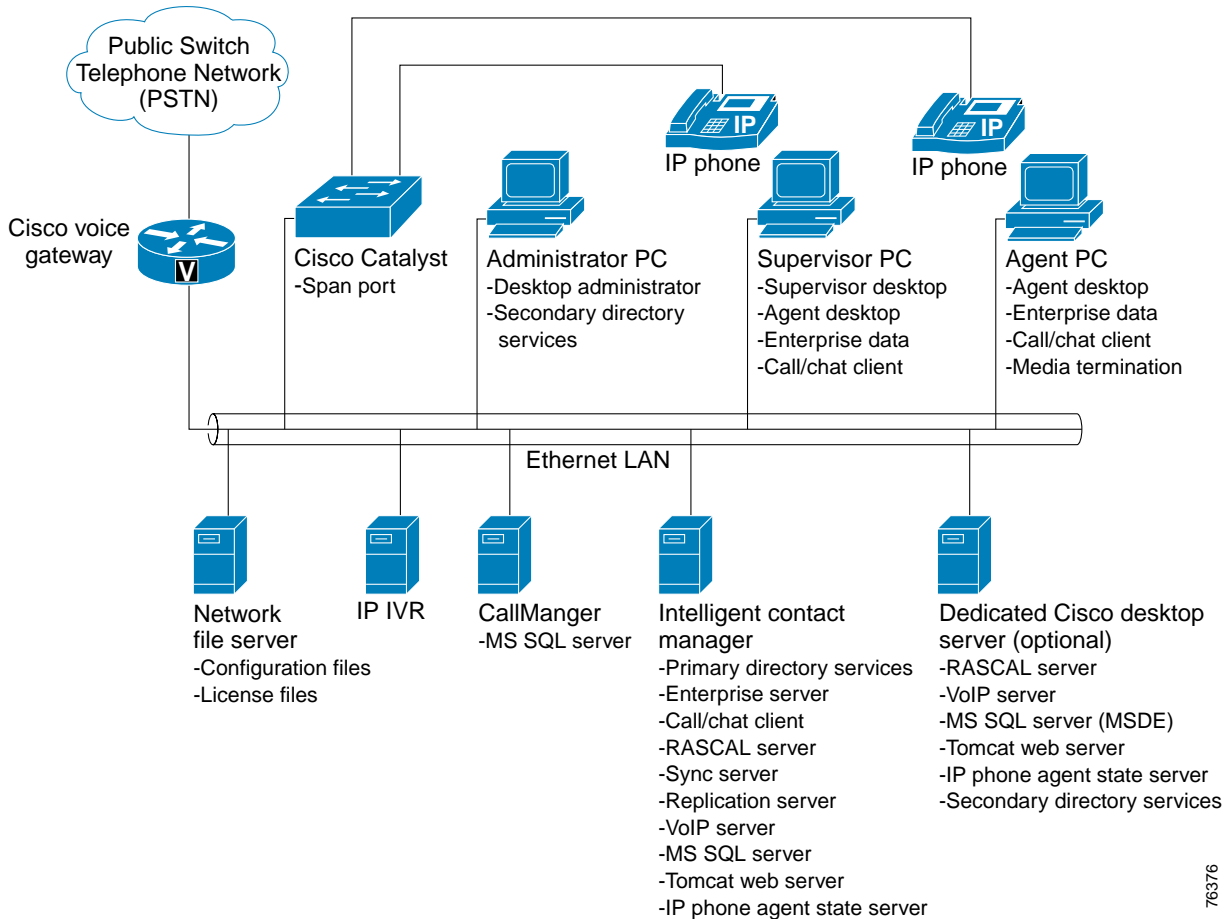
Cisco Agent Desktop and Supervisor Desktop (formerly Turnkey CTI) Product Suite v4.2 is a client-server application providing packaged CTI functionality for Cisco ICM and CTI. Cisco Agent Desktop includes a set of base server applications as well as a VoIP Monitor server application.

One of the base Cisco Agent Desktop servers, the Enterprise server, is a monitor-only CTI application that provides value-added services to the agent desktop. Similarly, the other Cisco Agent Desktop servers provide value-added features such as recording and chatting. The agent desktop receives its CTI feed from the ICM CTI server.

The Cisco Agent Desktop servers may be co-resident on the Peripheral Gateway (PG). As the number of agents increases, the Cisco Agent Desktop servers may require a dedicated server. For more information on server requirements, refer to the chapter on [Sizing IPCC Components and Servers](#).

Figure 7-4 illustrates the system components.

Figure 7-4 Cisco Desktop Product Suite for IPCC



Cisco IPCC supports multiple sites. The Cisco Agent Desktop 4.2 product can be installed at each site, but each site, for the most part, operates independently of the others. Administration must be done separately at each site, and agents at one site are not visible to a supervisor at another site.

Cisco Agent Desktop 4.4 lifts these restrictions and makes all sites appear as one logical call center.

One exception to site independence is Expanded Call Context (ECC) variables, which are able to follow a call between sites, depending on how the call is routed.

For more information, refer to the Cisco Agent Desktop and Supervisor Desktop product documentation, available at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/icm46/ipcc/>

Agent Desktop

The Agent Desktop software provides the core component of the Agent Desktop application: the softphone, workflow automation, login, call and agent event logging, and agent real-time statistics.

With Agent Desktop, you have the option of using either a Cisco IP Phone (model 7940 or 7960) or media termination (softphone). The media termination softphone allows the agent to make, answer, transfer, and conference calls. If your version of Agent Desktop includes media termination, agents do not need a physical IP phone; they can use the Agent Desktop softphone by itself.

Cisco Supervisor Desktop

Cisco Supervisor Desktop provides a graphical view of the agent team being managed by the supervisor. An expandable navigation tree control, similar to Windows Explorer, is used to navigate to and manage the team's resources.

Cisco Supervisor Desktop requires an instance of Cisco Agent Desktop running co-resident on the supervisor's PC. This instance of Agent Desktop is the same as the instance of Agent Desktop on the agent PCs.

The Supervisor Desktop installation includes installation of both the Supervisor Desktop software and the instance of Agent Desktop software. During the Supervisor Desktop installation process, you are prompted to choose the option of using either a hard IP phone (Cisco IP Phone 7940 or 7960) or media termination (softphone). The instance of Agent Desktop allows the supervisor to take calls and enables barge-in, intercept, and retrieval of skill group statistics.

For more information on the Supervisor Desktop, refer to the *IPCC Supervisor Desktop User Guide*, available at

<http://www.cisco.com/univercd/cc/td/doc/product/icm/icm46/ipcc/>

Cisco Supervisor Desktop Considerations and Guidelines

Observe the following guidelines when configuring Supervisor Desktop:

Sound Cards

With some older sound cards, you cannot run other audio-capable applications such as Windows Media Player while running Supervisor Desktop. This is because the card and/or drivers do not support simultaneous use of these applications. Most new sound cards do not have this limitation. For sound card recommendations, refer to the *Cisco Agent Desktop Service Information*, available online at Cisco.com.

Remote Supervisor Desktop

Security and restrictions may be an issue in some out-source situations where agents handle calls for multiple out-source clients. In these cases, the out-source client user of Remote Supervisor Desktop can monitor calls that the agent handles for the user's organization as well as calls that the agent handles for other organizations. Cisco Supervisor Desktop version 4.4 limits supervisors to the teams to which they are assigned.

The configuration must satisfy the following conditions for Remote Supervisor Desktop monitoring to succeed:

- The Remote Supervisor Desktop on the remote supervisor PC must have a connection over which packets can travel via TCP/IP services from the PC to the Call/Chat, VoIP Monitor and Recording, and Statistics servers. In many situations firewall issues will require careful consideration.
- The connection must support VoIP Quality of Service (QoS) for a single conversation.

The configuration requirements are more demanding if the user also wants to use the Supervisor Log Viewer to play back recorded calls. The Remote Supervisor Desktop software on the PC must have access to the Recording and Statistics server and to the .wav files at the pathname set during installation for the Agent Desktop and Supervisor Desktop configuration files.

Instead of giving an out-source contact center access to the recorded conversations, an alternative approach is to email the .wav file as an attachment to the remote client representative. The remote client representative can then listen to the .wav file in an appropriate media player.

Silent Monitor

Two supervisors can simultaneously silent-monitor the same agent that is configured in their agent team(s). However, only one supervisor can intercept or barge-in to the call, and the monitoring session automatically ends for the supervisor that intercepts or barges in. Once a supervisor has intercepted or barged in, the intercept and barge-in features are disabled for the other supervisor for the remainder of that particular monitoring session.

Required Switched Port Analyzer (SPAN) Port Configuration

Voice monitoring and recording capabilities are not built into IPCC. The Voice over IP (VoIP) Monitor server accomplishes these functions by "sniffing" voice packets sent to and from IP phones.

Because network switches do not normally deliver packets to Ethernet ports other than the destination port (in this case, an IP phone), the switch must be configured to perform this function. To accomplish this, you must configure the Ethernet port for the VoIP Monitor server to monitor the Ethernet ports for all agent IP phones. If the voice packets going to and from an agent's IP phone are not sent to the VoIP Monitor server's port for any reason, that conversation will not be available to the supervisor. For additional information, consult your Cisco Systems Engineer (SE).

Having the VoIP Monitor server monitor a port that all voice traffic goes through (for instance, the Ethernet port to which a gateway to the PSTN is connected) is not sufficient. It must monitor the Ethernet ports to which the IP phones are directly connected. The reason for this is that the server identifies packets by the IP phone's media access control (MAC) address. The packet's MAC address changes as the packet moves around the network. There must not be a router between the IP phone and the port the server is monitoring.

The port-monitoring feature on Cisco Catalyst switches is called Switched Port Analyzer (SPAN). For detailed information on SPAN, see *Configuring the Catalyst Switched Port Analyzer (SPAN) Feature*, available online at

<http://www.cisco.com/warp/public/473/41.html>

Cisco CallManager Interfaces

Cisco Agent Desktop and Supervisor Desktop depend on the following Cisco CallManager interfaces:

- SQL Server database on the publisher Cisco CallManager and one subscriber Cisco CallManager for failover
The VoIP Monitor server uses the extension of the agent phone to retrieve the phone's MAC address from the SQL server database on the publisher Cisco CallManager.
- Cisco CallManager Services administration for services on the Cisco IP Phones 7940 and 7960
The IP Phone Agent service is one of the services for the Cisco IP Phones 7940 and 7960.

Intelligent Contact Management (ICM) Interfaces

Cisco Agent Desktop and Supervisor Desktop depend on the following ICM components:

- CTI server
The CTI server receives the call and agent state change commands from Agent Desktop and the IP Phone Agent server and supplies event information to Agent Desktop, the IP Phone Agent server, and the Enterprise server. In addition, it provides skill group statistic information to Agent Desktop.
- SQL server database on the Logger
The Synchronization server periodically retrieves agent, team, and supervisor information from the SQL server database on the Logger to update the corresponding information in the Directory Services server LDAP directory.

Packet Sniffing and Network Configuration

The Voice over IP (VoIP) Monitor server process sniffs voice packets from a VLAN segment using the SPAN port of a Cisco Catalyst switch. With Cisco Agent Desktop version 4.2.1, there can be only one VoIP Monitor server per logical call center, but Cisco Agent Desktop version 4.4 supports multiple VoIP Monitor servers per logical call center. Multiple logical call centers can be configured per Cisco CallManager cluster, and this requires multiple Cisco CallManager PIMs per Cisco CallManager cluster. As the number of simultaneous calls on the VLAN segment increases, so does the workload of this server process. For very large sites, more than one VoIP Monitor server process will likely be required. Server sizing for this process is covered in the chapter on [Sizing IPCC Components and Servers](#).

The VoIP Monitor server supports G.711 mu-law and a-law and G.729 codecs. Conversations using any codec other than G.711 or G.729 will not be available for monitoring. The codec that an IP phone uses is configurable in Cisco CallManager.

In order to monitor voice conversations on the network, the VoIP Monitor server must be connected to a port on the data switch that has been configured for port monitoring. In addition, this port must receive all voice traffic for the call center.

Cisco recommends the configurations depicted in the section on [Verified Network Configurations, page 7-10](#). Other configurations may be possible based on the rules for SPAN port monitoring. For additional information, consult your Cisco Systems Engineer (SE).

The monitor server is independent of H.323 and Standard Interface Protocol (SIP). Both of these protocols use Real-Time Transport Protocol (RTP) to transport voice. The monitor server looks specifically for RTP version 2 packets.

**Note**

The RTP packets must be carried over User Datagram Protocol (UDP), IPv4, and Ethernet II.

Because a network switch does not normally deliver packets to Ethernet ports other than the destination (an IP phone in this case), the switch must be configured to do so. You must configure the Ethernet port of the monitor server to monitor the Ethernet ports for all of the agent IP phones. If the voice packets to and from an agent's IP phone are not sent to the monitor server's port for any reason, that conversation will not be available to the supervisor.

When a request is made to monitor an agent, the monitor server looks up the MAC address of the agent's IP phone in the Cisco CallManager database. The monitor server then looks for packets to and from this MAC address; if it is an RTP packet, it is forwarded to the Supervisor.

It is not sufficient for the monitor server to monitor a port that all voice traffic goes through, such as the Ethernet port to which a gateway to the PSTN is connected. The monitor server must monitor the Ethernet ports to which the IP phones are directly connected. This is because MAC addresses change as packets pass through OSI Layer 3 devices such as routers.

The monitor server sniffs packets on a single network interface card (NIC) and, therefore, a single Ethernet port. You must configure this port to monitor the Ethernet ports of all agent IP phones. This configuration does not necessarily require that the monitor server and all agent IP phones be connected to the same network switch. The connections depend on the monitoring capabilities of the network switch.

Catalyst Switch Capabilities

Cisco Catalyst switches use Switched Port Analyzer (SPAN) to monitor ports. Some of the capabilities and restrictions of Catalyst switches are as follows:

- Catalyst 2900XL and 3500XL Switches
 - A monitor port cannot be in a Fast EtherChannel or Gigabit EtherChannel port group.
 - A monitor port cannot be enabled for port security.
 - A monitor port cannot be connected to multiple virtual LANs (VLANs).
 - A monitor port must be a member of the same VLAN as the port monitored. VLAN membership changes are disallowed on monitor ports and ports being monitored.
 - A monitor port cannot be a dynamic-access port or a trunk port. However, a static-access port can monitor a VLAN on a trunk, a multi-VLAN, or a dynamic-access port. The VLAN monitored is the one associated with the static-access port.
 - Port monitoring does not work if both the monitor and monitored ports are protected ports.
- Catalyst 4000, 5000, and 6000 Series Switches
 - You can monitor ports belonging to multiple VLANs on these switches.
 - The Catalyst 6000 with CatOS Release 5.3 or higher has a feature called Remote SPAN (RSPAN), which allows you to monitor ports spread throughout a switched network. With RSPAN on a Catalyst 6000, the monitor server and IP phones can be on separate switches.

Verified Network Configurations

Figure 7-5 through Figure 7-11 are verified network configurations using Cisco Catalyst switches.

Figure 7-5 Catalyst 3500 with Single VLAN

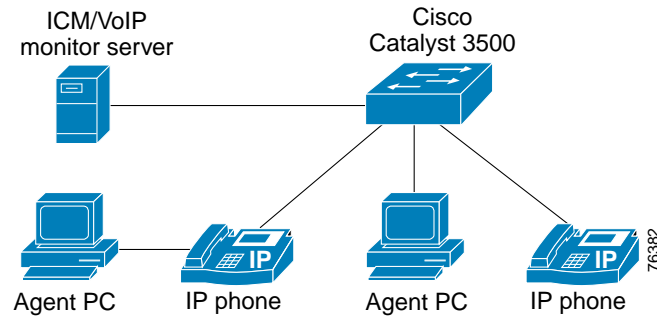


Figure 7-6 Catalyst 3500 with Separate Voice and Data VLANs

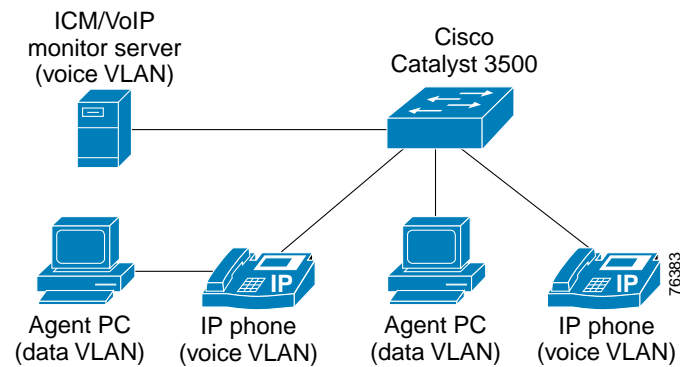


Figure 7-7 Catalyst 6000 with Single VLAN

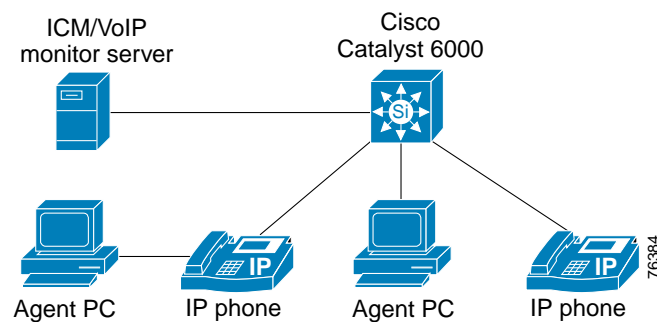


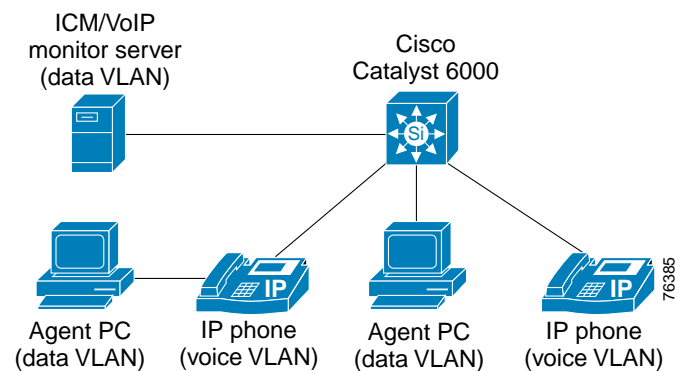
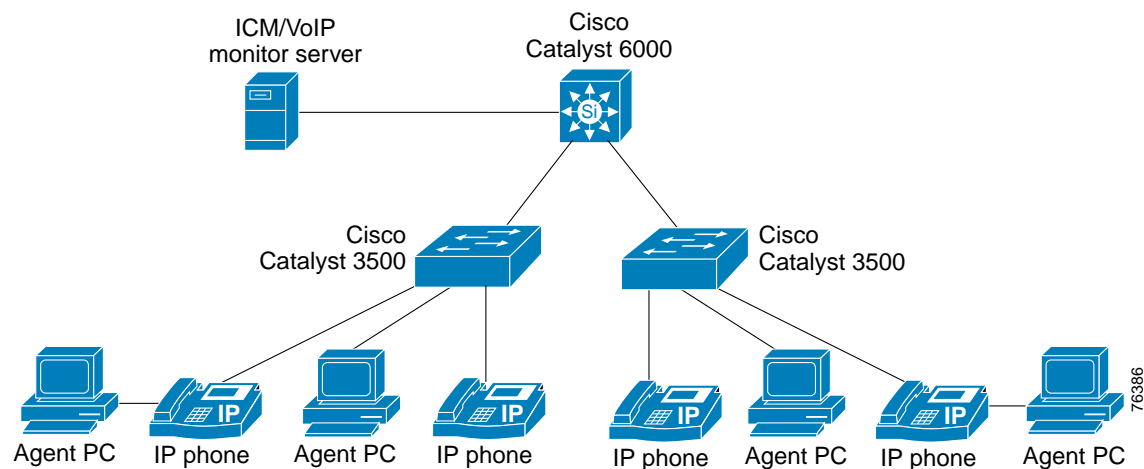
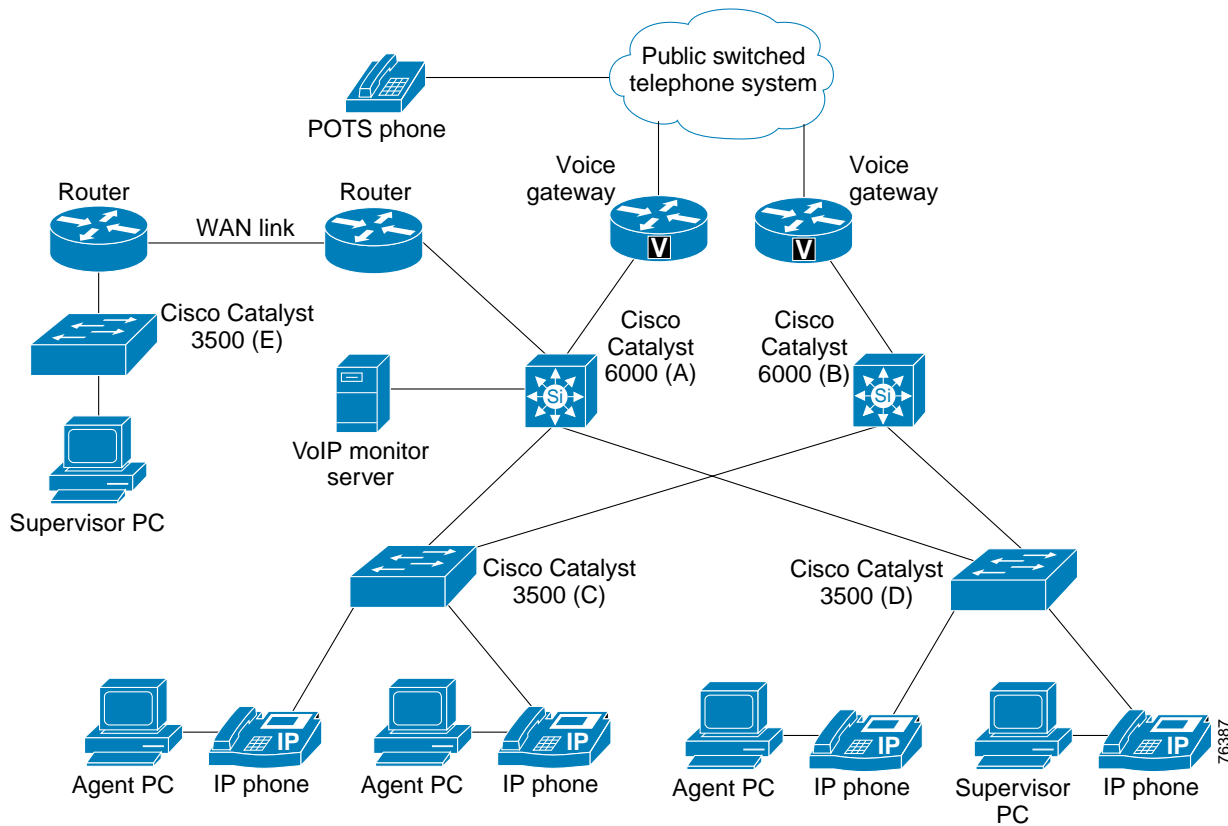
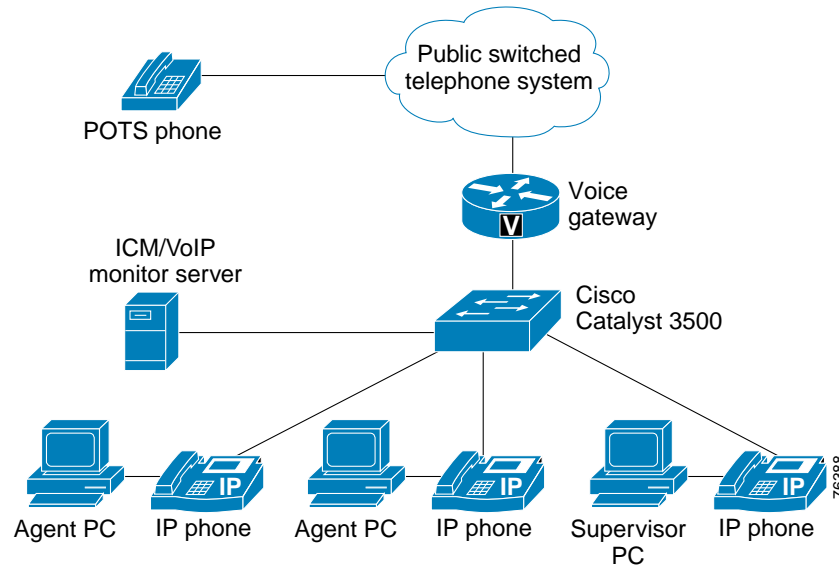
Figure 7-8 Catalyst 6000 with Separate Voice and Data VLANs**Figure 7-9 System with One Catalyst 6000 and Two Catalyst 3500s**

Figure 7-10 High-Availability Network**Notes on the High-Availability Network (Figure 7-10)**

This network configuration is modeled after a highly available, single-site IPCC deployment, and it has the following characteristics:

- The VoIP Monitor server is connected to a data port on a core switch (switch A in Figure 7-10). That port is configured using SPAN to monitor traffic on the ports going to the closet switches (switches C and D).
- Because there is no router between the core and closet switches, the VoIP Monitor server is able to identify correctly RTP streams to and from agent phones.
- Conversations between agents on the same closet switch (C, for example) are visible to the VoIP Monitor server even though traffic would not normally be sent to the core switch A.
- While the Cisco IPCC servers support fault tolerance with redundant servers, the VoIP Monitor currently does not. If the VoIP Monitor server or its connected core switch is down, the monitoring feature will not be available in the Supervisor.
- If any agents are connected to switch E, they are not available for monitoring. This is due to MAC addresses changing as packets are sent to the routers.
- A supervisor on switch E is able to monitor agents on switches C and D. The supervisor's PC or IP phone does not have to be monitored by the VoIP Monitor server. The only requirement is that a route exists with sufficient bandwidth from the VoIP Monitor server to the supervisor's PC.

Figure 7-11 Simple Network**Notes on the Simple Network (Figure 7-11)**

The high-availability network model illustrated in Figure 7-10 is expensive and might not be practical in many cases. The simple network is an alternative configuration, and it has the following characteristics:

- The simple network is not fault tolerant to network hardware failures.
- The Catalyst 3500 switch has a variety of restrictions on the monitor ports and the ports being monitored. For details, refer to *Configuring the Catalyst Switched Port Analyzer (SPAN) Feature*, available at www.cisco.com/warp/public/473/41.html
- A key requirement is that the monitor port must be a member of the same VLAN as the port being monitored. In this configuration, the VoIP Monitor server must be on the same VLAN as the agent IP phones.



Bandwidth Provisioning and QoS Considerations

This chapter presents an overview of the various IPCC components, their roles, and the network traffic flows between remote components over the WAN. It includes guidelines for estimating network link bandwidth between remote Peripheral Gateways (PGs) and the ICM Central Controller, along with some examples. This chapter also presents recommendations on how to apply proper QoS to the traffic flows over the WAN, with an example of how to configure the flows. For a detailed description of the IPCC architecture and various component interworking, refer to the IPCC and Intelligent Contact Management (ICM) documentation available online at Cisco.com.

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, IP IVR Q-points (ports), and IP Phones.
- 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 are, for instance, those used to set up, maintain, tear down, or redirect a call. In case of IPCC, control traffic includes Device Management Protocol (DMP) and External Message Transport (EMT) messages required to route voice calls to agents and other media termination resources, such as IP IVR ports, and the transmission of real-time status information.
- Data traffic

Data traffic could 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. IPCC priority data includes data associated with non-real-time system states, such as events involved in reporting and configuration updates.

This chapter focuses on the types of data flows and bandwidth generated between a remote Peripheral Gateway (PG) and the ICM Central Controller. Guidelines and examples are presented to help estimate the bandwidth required between the PG and ICM Central Controller and to illustrate how to provision QoS for these data flows.

For bandwidth estimates for the voice RTP stream generated by the calls to IPCC agents and the associated call control traffic generated by the various protocols, refer to the *Cisco IP Telephony Solution Reference Network Design Guide*, available at

http://www.cisco.com/warp/public/779/largeent/netpro/avvid/iptel_register.html

Data traffic consisting of various HTTP, email, and other non-IPCC mission critical traffic will vary by the specific integration and deployment model used, and this type of traffic is not addressed in this chapter.

IPCC Network Components Overview

The IP Contact Center (IPCC) is a solution that combines Cisco IP Telephony products and Intelligent Contact Management (ICM) software to create an IP-based customer contact solution.

IPCC consists of three main Cisco products:

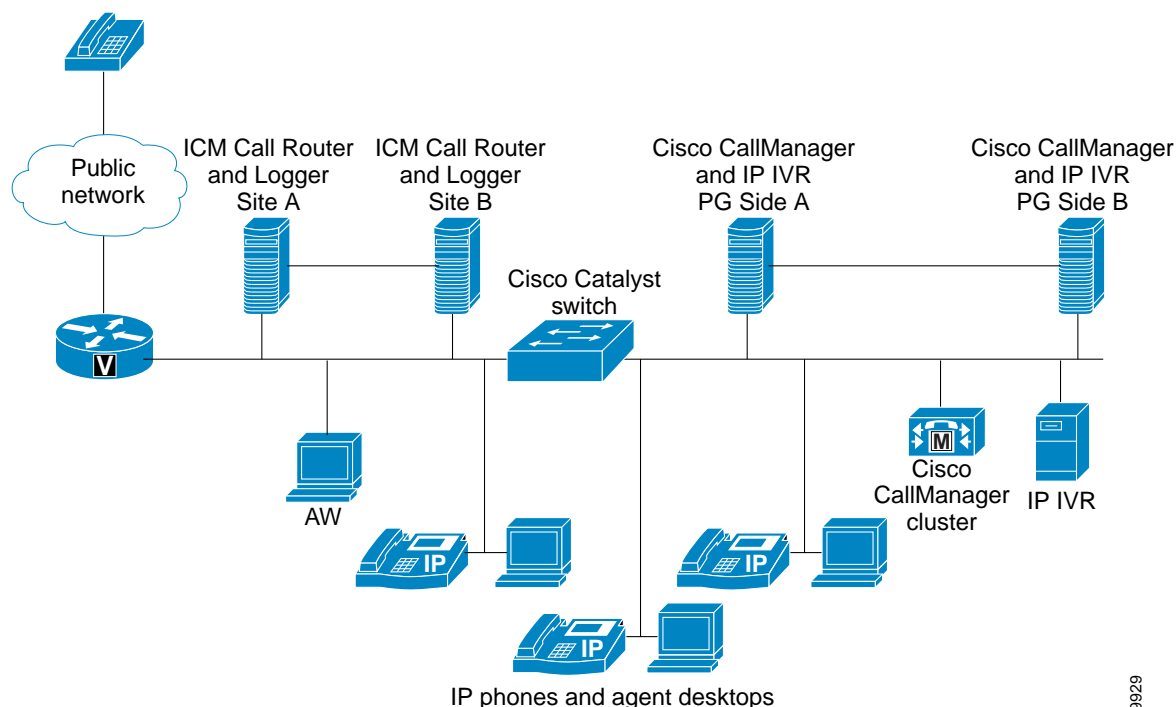
- Cisco CallManager — Provides call processing and traditional telephony features in an IP telephony environment.
- Cisco IP IVR — An application server that provides traditional interactive voice response (IVR) features such as call treatment and queuing.
- Cisco Intelligent Contact Management (ICM) — The platform that provides Cisco CallManager and IP IVR. The major components in the ICM are the Central Controller (Router and Logger), Peripheral Gateway (PG), CTI Object Server (OS), Agent Desktop, Historical Data Server (HDS), and Administrative Workstation (AW). A combination of these processes can run on one physical server or on separate servers, depending on capacity required, configuration requirements, and server types (as described in the chapter on [Sizing IPCC Components and Servers](#)).

Single-Site Characteristics

For a single contact center, all the IPCC components reside in the same physical site with no WAN (Layer 3) connectivity, and all connections run at 10/100 Mbps Ethernet speed. (See [Figure 8-1](#).) A single-site IPCC has the following characteristics:

- All the IPCC components use 10/100 Ethernet Cards connected to ports on the Catalyst data switch.
- The data switch and IPCC servers (nodes) must be "hard-set" to select the port speed and duplex setting.
- Port speed across the solution depends upon the customer-provided equipment. Cisco does not recommend mixing Cisco CallManagers at 10Mbps and PGs at 100Mbps.
- Duplexed pairs of systems (PGs and ICM Central Controllers) require a "private" connection between them, typically provided by a simple crossover cable and set to 100 Mbps, Full Duplex.

Figure 8-1 Typical Single-Site IPCC



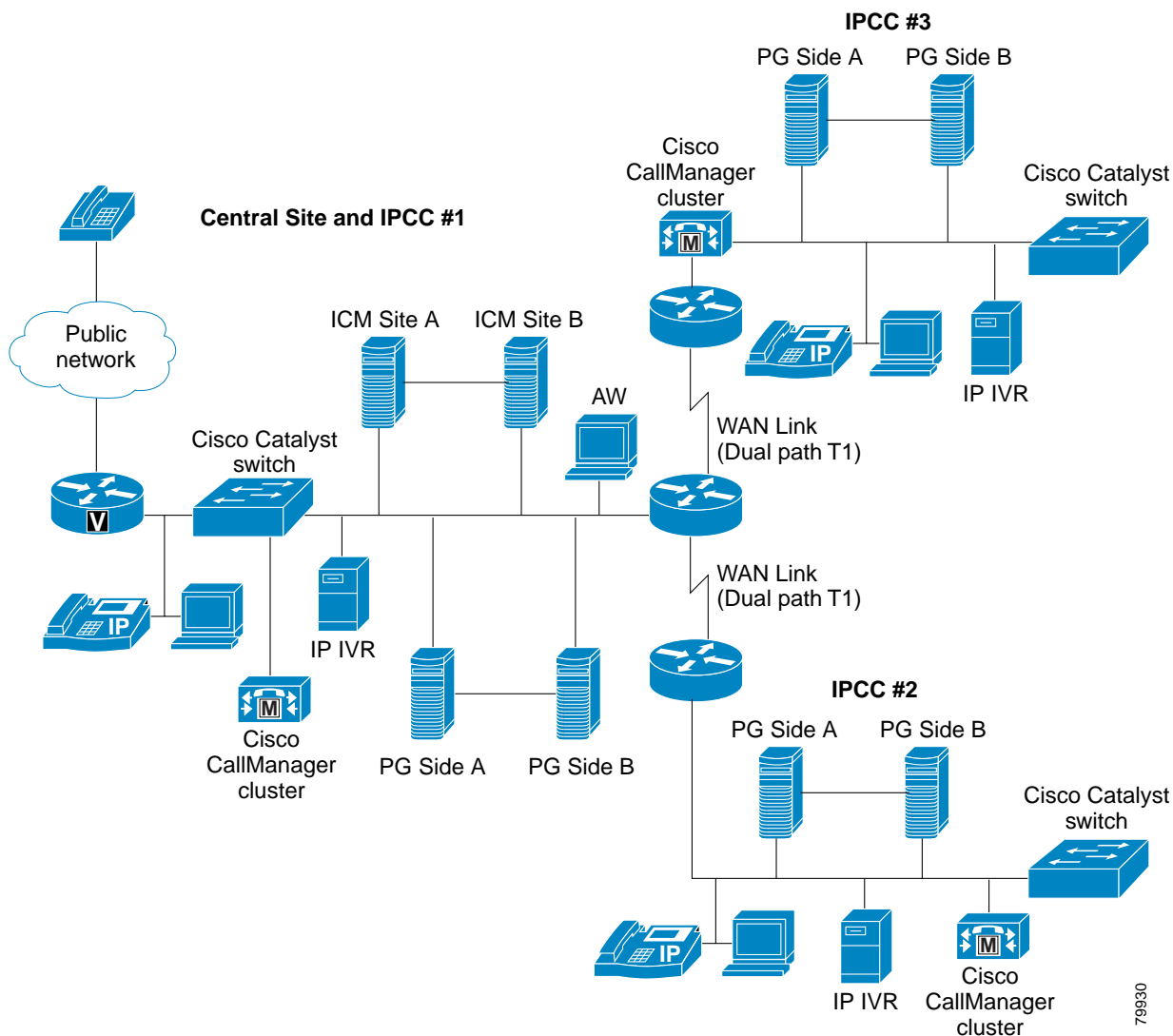
79929

Multi-Site Characteristics

For multi-site IP Contact Centers, the ICM Central Controller components are located at a single, central site, and the contact centers with remote PGs are connected by WAN links. (See [Figure 8-2](#).) A multi-site IPCC has the following characteristics:

- All of the single-site Ethernet port speed and duplex considerations apply to the multi-site configuration as well, but on a site-by-site basis.
- Contact center sites are connected to the Central Controller site via a dual-path WAN connection; with required bandwidth, QoS prioritization, and latency.

Figure 8-2 Typical Multi-Site IPCC



The IPCC components are designed to be distributed, fault tolerant network applications that rely heavily on a network infrastructure with sufficient performance to meet real-time data transfer requirements. The IPCC network is characterized by high bandwidth, low latency, and a prioritization scheme favoring specific User Datagram Protocol (UDP) and Transmission Control Protocol (TCP)

traffic. This design is necessary to ensure both the fault-tolerant message synchronization, in the case of duplexed ICM nodes (Router, Logger, Peripheral Gateway, CTI Server, and AW), as well as the delivery of time-sensitive system status data.

Expeditious delivery of ICM data to the PG is necessary for accurate and timely call routing as well as for skill-group queue real-time reporting data. To meet the stringent bandwidth and latency requirements, data flows between the PG and the ICM Central Controller are identified and prioritized according to type and function, as discussed later in this chapter. The recommended approach is to mark the traffic flow at the edge of the network, based on the guidelines given in this chapter. PG configuration of bandwidth should be set to the LAN bandwidth (large) so that fragmentation is not attempted by the application but is left to the network instead.

In duplexed IPCC configurations, the ICM is deployed with two Central Controller servers. One Central Controller is typically referred to as Side A and the other as Side B. The two sides exchange message synchronization data within a fraction of a second. For this synchronization to work, you must configure a path between the two sides so that heartbeats ("I'm alive" messages) can be sent symmetrically in both directions. The heartbeats allow rapid detection of a circuit (or other) failure. Similarly, the two Central Controller sides communicate with the PGs on the network, and a similar heartbeat mechanism is used, although the timing constraints are less stringent.

Network Segmentation

The ICM Central Controller uses multiple network segments to perform specific jobs in the system as well as to maintain traffic isolation and minimize network impact. There are three types of network segments:

- [Private Network, page 8-5](#)
- [Visible \(Public\) Network, page 8-7](#)
- [Signaling Access Network \(PSTN Interfaces\), page 8-8](#)

Private Network

The private network is a logical and/or physical network that allows specific nodes to communicate with each other using the highest possible priority. This network carries the data that is necessary to maintain and restore synchronization between the systems.

The ICM private network is used to interconnect duplex systems such as the Call Routers, Loggers and Database Servers, and Peripheral Gateways to exchange messaging between the systems as part of the fault-tolerant architecture. In co-located systems, the private network can be as simple as a crossover cable between servers or, if desired for manageability, as independent VLANs on a common switch with a visible network.

Private Network Functions

The ICM uses the private network for two major functions:

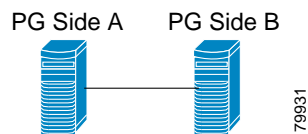
- Failure detection between redundant pairs of systems or networks. On this segment, the private network generates keep-alive heartbeats at 100 ms intervals so that it can detect network or system problems after 5 consecutive missed heartbeats (500 ms). Without this segmentation, the entire network would have to be able to maintain a high level of service for the system to function.
- Fast transmission of system requests and data updates between the two redundant systems to ensure that both sides are always "in sync" for making routing decisions. Route requests as well as state data and Logger database requests are passed across the private network to ensure delivery.

Private Network Characteristics Between Peripheral Gateways

The private network between Side A and Side B of the Peripheral Gateway has the following characteristics (see [Figure 8-3](#)):

- Used to isolate traffic between the PGs from the rest of the system.
- Used to send real-time updates (route requests and data network state) between sides of the PGs.
- Used to maintain memory image, or state, between the two PGs.
- Typically deployed as a red crossover cable, and no LAN hardware is used. 100 Mbps, full duplex is acceptable. This configuration requires an additional network interface card (NIC) in each PG.
- PGs use heartbeats at 100 ms intervals.
- Can be geographically distributed with dedicated bandwidth equivalent to a T1 link. However, this configuration is not recommended over a converged network at this time because no data profiling has been conducted yet to estimate required bandwidth or level of QoS prioritization.

Figure 8-3 Private Network Between Peripheral Gateways

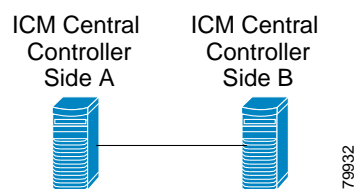


Private Network Characteristics Between Central Controllers

The private network between Side A and Side B of the ICM Central Controller has the following characteristics (see [Figure 8-4](#)):

- Used to isolate traffic between the Routers and Loggers (Rogger) from the rest of the system.
- Used to send real-time updates (Route requests, Logger database updates, and so on) between sides of the system.
- Used to maintain memory image, or state, between the two call Routers.
- For co-located systems running on co-resident (Rogger) platforms, the private network could be red crossover cable because there are only two machines on this network. This configuration requires an additional network interface card (NIC) in each server.
- If the Routers and Loggers were deployed on individual platforms, you would have to install network hardware to support the private VLAN segment locally if the systems were co-located.
- Can be geographically distributed with dedicated bandwidth equivalent to a T1 link. However, this configuration is not recommended over a converged network at this time because no data profiling has been conducted yet to estimate required bandwidth or level of QoS prioritization.

Figure 8-4 Private Network Between Central Controllers



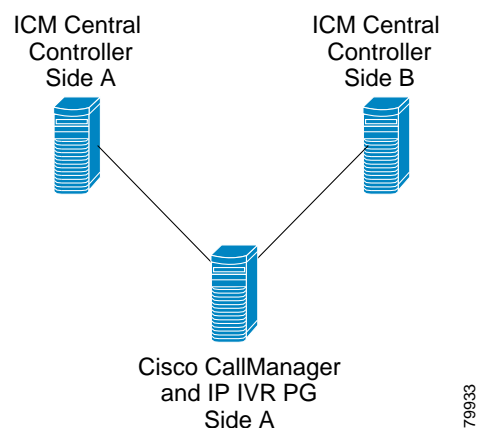
Visible (Public) Network

The visible network is the primary network for the contact center, and it connects the ICM site with the PG site. (See [Figure 8-5](#).) This network allows the Central Controller to communicate to PGs and remote distributor AW machines/HDS (historical data server), and from distributor AW/HDS/Webview server to client AWs and client browsers. It carries traffic between each side of the synchronized system and remote systems. The visible network is also used by the fault tolerance software as an alternate network to distinguish between node failures and network failures.

The visible network has the following characteristics:

- Can be either a dedicated ICM network or part of the shared (converged) corporate network with appropriate bandwidth sizing and QoS prioritization (discussed later in this chapter).
- Used to send skill group real-time statistics and call routing notification from the ICM to the PG.
- Uses heartbeats at 400-ms intervals to monitor the health of the network and systems, taking up to 2 seconds for failure detection.

Figure 8-5 Visible Network



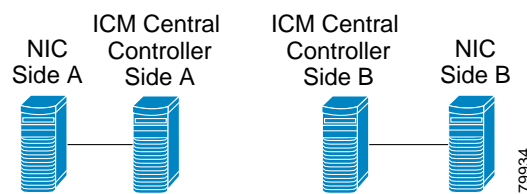
Signaling Access Network (PSTN Interfaces)

The signaling access network connects the IPCC system to a carrier network. (See [Figure 8-6](#).) It has the following characteristics:

- Used to isolate network carrier traffic from the rest of the system.
- Deployed at the central site(s) with a red crossover cable (if there is only the ICM Call Router and one network interface card) or with network hardware for multi-system connections such as MCI Gateway.
- Can be set up as 100 Mbps, full duplex.

Figure 8-6 *Signaling Access Network*

Signal Access Network – ICM CC & NIC



IPCC Network Bandwidth and QoS Overview

The IPCC network design must fully guarantee the specific traffic latency, bandwidth, and prioritization requirements of all the various IPCC components. Cisco Quality of Service (QoS) mechanisms designed to support latency-sensitive applications are required when IPCC deployments extend beyond the local LAN into campus or remote WAN sites.

Traffic Flows from PG to ICM Central Controller

The active PG continuously updates the ICM Central Controller with the state of the agents, calls, and so on, for each of the peripherals connected to it (for example, Cisco CallManagers and IP IVRs). This type of traffic is referred to as real-time traffic, and the amount of traffic varies by peripheral device, call load, and configuration.

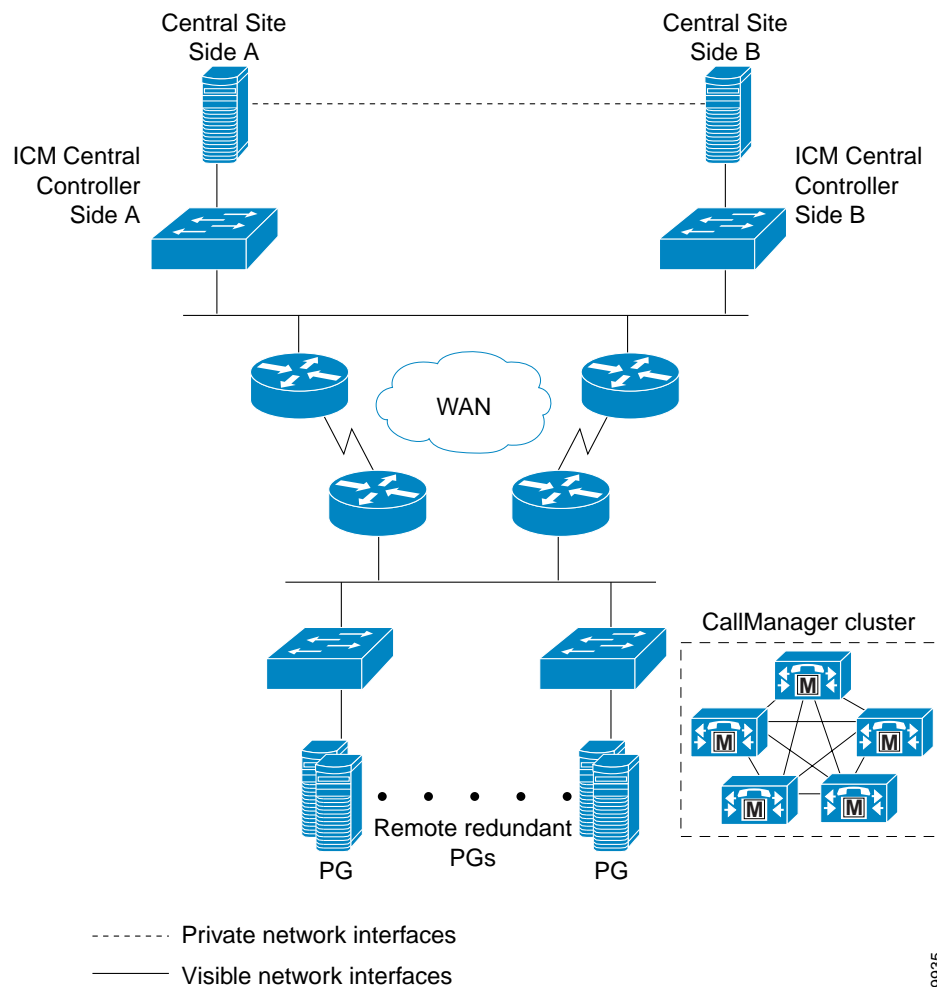
The PGs also send historical data to the Central Controller every five minutes and every half hour on the half hour. The amount of this data varies, depending on how many agents, skill groups per agent, and services are configured. These historical data updates are low priority and are sent on the normal priority address to the call Router. The historical data is buffered in the PG memory until it can be transmitted, and it does not impact the real-time data sent for call arrival and agent state changes. The historical data must complete its journey to each side of the Central Controller within the half hour so that the system can be ready for the next half hour's data. In a redundant ICM Central Controller, the primary side that receives this data will replicate the data and pass it to the other half of the Central Controller.

When a PG starts up, the central site supplies its configuration data so that it can know what agents, ports, and so on, it needs to monitor. In duplex mode where a failed redundant PG is coming back up, this traffic can be a significant network bandwidth transient and needs to be factored into network bandwidth calculations when sizing the various network traffic for each segment or circuit.

Traffic flows from PG to ICM Central Controller can be classified into four distinct flows, as illustrated in Figure 8-7:

- **High Priority traffic** — Includes Device Management Protocol (DMP) TCP traffic for events involved in, or synchronized with, actual voice call routing to agents and agent states updates. Examples would be routing requests and responses for post routing and routing label or release messages for translation routing.
- **Heartbeat traffic** — Is also high priority traffic that includes User Datagram Protocol (UDP) messages consisting of heartbeats transmitted at 400 ms intervals between all active components, PGs, and ICM Central Controllers.
- **Medium Priority traffic** — Includes DMP traffic for events for peripheral state and configuration table updates such as agent state changes.
- **Low priority traffic** — Includes DMP traffic for events involved in system configuration and reporting functions. Examples would be call status and reporting statistics updates, call close notifications and PG configuration updates.

Figure 8-7 Traffic Flows from PG to ICM Central Controller



79935

The Central Controller is configured with two IP addresses on the visible network interface, and the data flows are mapped to these addresses as follows:

- High priority IP address traffic association:
 - DMP high priority traffic
 - DMP medium priority traffic
 - UDP heartbeats
- Normal priority IP address traffic association:
 - DMP low priority traffic

Latency and QoS Requirements for PG to ICM Central Controller

Latency and QoS requirements for the various traffic types mentioned in the preceding section are critical in a campus deployment due to buffer management. (Refer to the *Cisco IP Telephony Solution Reference Network Design Guide* for QoS requirements in the campus.) Latency, bandwidth, and QoS are also critical for multi-site deployments where any one of the components (such as agents or peripherals) resides in a remote site over a WAN, regardless of whether the call processing is centralized or distributed.

The following latency requirements for pre- and post-translation routing must be met to ensure proper operation of the IPCC solution:

- Pre-routing
 - Ensure accuracy of the real-time agent and queue state, and use a 200 ms one-way "delay budget" between the PG and the ICM Central Controller.
- Post-translation routing
 - Post-translation routing has the additional requirement to coordinate call and data delivery for the CTI data being passed by the data network. The voice network may be able to deliver the voice call faster than the data network can deliver the data associated with the call. Use a 100 ms one-way "delay budget" between the PG and ICM central controller to ensure that the voice calls and CTI data information arrive together and successfully complete the CTI-enabled translation route.

Given the preceding latency requirements, the recommendations for classification of the ICM traffic control data are as follows (see [Table 8-1](#)):

- ICM control data associated with real-time system state and call routing should be mapped to the "call signaling" queue
 - DMP high and medium priority traffic classified as AF31
 - Heartbeats classified as AF31
- ICM control data associated with non-real-time system state should be mapped to the "priority data" queue
 - DMP low priority traffic classified as AF21

These recommendations have been validated in lab testing under heavy loads of converged traffic over a WAN using Frame Relay.

Table 8-1 Recommended Traffic Classifications

Layer 2	Layer 3 Classification			Application
CoS	IP Precedence	Pre-Hop Behavior (PHB)	DSCP	
7	7	—	56-63	Reserved
6	6	—	48-55	Reserved
5	5	EF	46	Voice Bearer
4	4	AF41	34	Video Conferencing
3	3	AF31	26	Call Signaling (DMP high and medium priority, and UDP heartbeats)
2	2	AF2y	18, 20, 22	High Priority Data (DMP low priority)
1	1	AF1y	10, 14, 16	Medium Priority Data
0	0	BE	0	Best Effort Data

QoS Classification Implementation for PG to ICM Central Controller

ICM currently does not perform DSCP marking automatically (but this feature is planned for a future release). DSCP marking can be done by setting the IP Precedence bits and enabling the QoS features in the IP routers and switches that come in contact with this packet, or by using hard-coded Access Control Lists (ACLs) at each IP router and switch across the network.

Using either method, implement the following classifications:

- AF31 for all packets with a source or destination IP address matching the high priority IP address of the ICM Central Controller and that are either:
 - TCP packets in the port range 40,000 to 49,999
 - UDP packets in the port range of 39,000 to 39,999
- AF21 for all packets with a source or destination IP address matching the normal priority IP address of the ICM Central Controller and that are
 - TCP packets in the port range 40,000 to 49,999

The following example configurations for an ICM QoS implementation are based on [Figure 8-7](#), and they use the following address scheme:

- ICM Central Controller, Side A
 - High Priority IP address = 10.81.201.51
 - Low Priority IP address = 10.82.201.50
- ICM Central Controller, Side B
 - High Priority IP address = 10.81.201.51
 - Low Priority IP address = 10.82.201.50

Example 8-1 QoS Configuration for ICM Central Controller Side A

```

! Enable QoS
ipcc-6500-side-a> (enable) set qos enable
! Create an access list to mark all IPCC traffic from the ICM Central Controller with the
appropriate DSCP
ipcc-6500-side-a> (enable) set qos acl ip ipcc dscp 26 tcp host 10.81.201.51 range 40000
49999 any
ipcc-6500-side-a> (enable) set qos acl ip ipcc dscp 26 udp host 10.81.201.51 range 39000
39999 any
ipcc-6500-side-a> (enable) set qos acl ip ipcc dscp 18 tcp host 10.81.201.50 range 40000
49999 any
! Activate the ACE in Hardware
ipcc-6500-side-a> (enable) commit qos acl ipcc

! Apply this QoS policy to the required ICM Central Controller visible network server port
ipcc-6500-side-a> (enable) set port qos 3/6 trust untrusted
ipcc-6500-side-a> (enable) set port qos 3/6 port-based
ipcc-6500-side-a> (enable) set qos acl map ipcc 3/6

```

Example 8-2 QoS Configuration for PG

```

! Enable QoS
ipcc-6500-CC> (enable) set qos enable

! Create an access list to mark all IPCC traffic to the ICM Central Controller with the
appropriate DSCP
ipcc-6500-CC> (enable) set qos acl ip ipcc dscp 26 tcp any host 10.81.201.51 range 40000
49999
ipcc-6500-CC> (enable) set qos acl ip ipcc dscp 26 udp any host 10.81.201.51 range 39000
39999
ipcc-6500-CC> (enable) set qos acl ip ipcc dscp 18 tcp any host 10.81.201.50 range 40000
49999
ipcc-6500-CC> (enable) set qos acl ip ipcc dscp 26 tcp any host 10.81.231.51 range 40000
49999
ipcc-6500-CC> (enable) set qos acl ip ipcc dscp 26 udp any host 10.81.231.51 range 39000
39999
ipcc-6500-CC> (enable) set qos acl ip ipcc dscp 18 tcp any host 10.81.231.50 range 40000
49999

! Activate the ACE in Hardware
ipcc-6500-CC> (enable) commit qos acl ipcc

! Apply this QoS policy to the required ICM PG visible network server port
ipcc-6500-CC> (enable) set port qos 5/2 trust untrusted
ipcc-6500-CC> (enable) set port qos 5/2 port-based
ipcc-6500-CC> (enable) set qos acl map ipcc 5/2

```

Example 8-3 QoS Configuration for Router

```

! Define the DSCP markings of interest
class-map match-any mission-critical
  match ip dscp af21
class-map match-any control
  match ip dscp af31
class-map match-any voice
  match ip dscp ef
!

```

```

! Define the policy map queue sizes based on calculated IPCC and voice traffic loads
!
policy-map ipcc-policy
  class voice
    priority 578 10000
  class control
    bandwidth 319
    queue-limit 128
  class mission-critical
    bandwidth 140
  class class-default
!
. . .
! Apply the IPCC map class on the link from the Data Center to the Call Center
!
interface Hssi3/0
  mtu 1500
  no ip address
  encapsulation frame-relay
  load-interval 30
  tx-ring-limit 4
  serial restart-delay 0
  frame-relay traffic-shaping
!
interface Hssi3/0.1 point-to-point
  bandwidth 1400
  ip address 10.80.1.193 255.255.255.252
  frame-relay interface-dlci 101
  class ipcc-ts
!
. . .
!
! Associate the IPCC policy map with the map class
map-class frame-relay ipcc-ts
  frame-relay cir 1400000
  frame-relay mincir 1400000
  no frame-relay adaptive-shaping
  service-policy output ipcc-policy

```

For more information on QoS configuration, refer to the *Cisco AVVID Network Infrastructure Enterprise Quality of Service Design* guide, available online at Cisco.com

Traffic Flows from ICM to ICM Central Controller

The Central Controllers have access to both the private and visible networks for the ICM. The activity on these two networks differs greatly.

The visible network is used only for "Test Other Side" messages in the event of a failure of the private network. No private traffic is ever redirected over this link. Bandwidth required for "Test Other Side" messages is minimal.

On the private network, all state, update, and route requests are sent over the link between the Central Controllers, creating a steady stream of data and typically requiring a full T1 in the case where both sides of the ICM are geographically distributed. Redundant network links are also recommended in this case.

The specific data that is carried by the private network between the Central Controllers includes:

- Heartbeats that are generated every 100 ms.
- Route Requests and Responses generated from both the carriers and the Cisco CallManager and IP IVR peripherals.

- PG updates. The active data stream updates only one side of the Central Controller directly from the PG, and the call routers have to update each other.
- Logger database updates. In the event that a Logger has been off-line, it gets updated from the active Logger over the private network.
- State transfers. All of the ICM configuration and scripting is resident in the call router memory, along with all the associated data in the rolling five minutes. This data is periodically sent across the private network from the enabled synchronizer to the other side.

Latency and QoS Requirements for ICM to ICM Central Controller

The majority of the communications that flow over the private network are time-sensitive. The heartbeats are generated every 100 ms to maintain very tight latency.

The end-to-end latency or delay budget is 200 ms, but large configurations run a risk of exceeding the state transfer time and causing the system to shut down to simplex mode because it believes the private network has failed. The optimal system performance would be to have the delay between 100 ms and 200 ms, and the lower the better.

Missing the latency on this link puts the system at risk of potentially running the system in simplex mode rather than the fully redundant duplex mode because the system cannot validate and support the remote connection.

On the private network, the ICM also recognizes certain types of data as being high priority, or needing to get across the network as fast as possible, and some data as normal (that is, not critical).

Examples of high-priority data include:

- Heartbeat packets (sent as UDP packets in the port range of 39,000 to 39,999)
- Route request and response packets
- State transfer

Examples of normal-priority data include:

- Logger database updates
- Five-minute and half-hour updates

The ICM does not set any of the IP Precedence bits in the packets to identify their priority to the network. It is up to the edge network hardware to mark the packets or treat the packets differently based upon:

- Source or destination IP address matching the high-priority IP address of the call routers
- UDP packets in the port range of 39,000 to 39,999

Packet marking can be done by setting the IP Precedence bits and enabling the QoS features in the IP routers and switches that come in contact with the packets, or by using hard-coded Access Control Lists (ACLs) at each IP router and switch across the network.

Using either method, implement the following classifications:

- AF31 for all packets with a source or destination IP address matching the high priority IP address of the ICM Central Controller and that are either:
 - TCP packets in the port range 40,000 to 49,999
 - UDP packets in the port range of 39,000 to 39,999
- AF21 for all packets with a source or destination IP address matching the normal priority IP address of the ICM Central Controller and that are
 - TCP packets in the port range 40,000 to 49,999

Bandwidth and QoS Requirements for CTI Server to IPCC Desktop

The CTI Server acts as an interface between the ICM system and client CTI applications running on the agent desktop. The CTI Server software runs on a PG hardware platform. It typically shares the same PG hardware platform with the Cisco CallManager PIMs (and sometimes the IP IVR PIMs). Cisco strongly recommends that you run the CTI Server process on the same PG platform as the Cisco CallManager PIM in all cases, except in very large deployments when the CTI traffic is extensive (then the CTI Server can run on a separate platform from the Cisco CallManager PIM). See the chapter on [Sizing IPCC Components and Servers](#) for more details on when to bundle or to split the ICM components on separate servers.

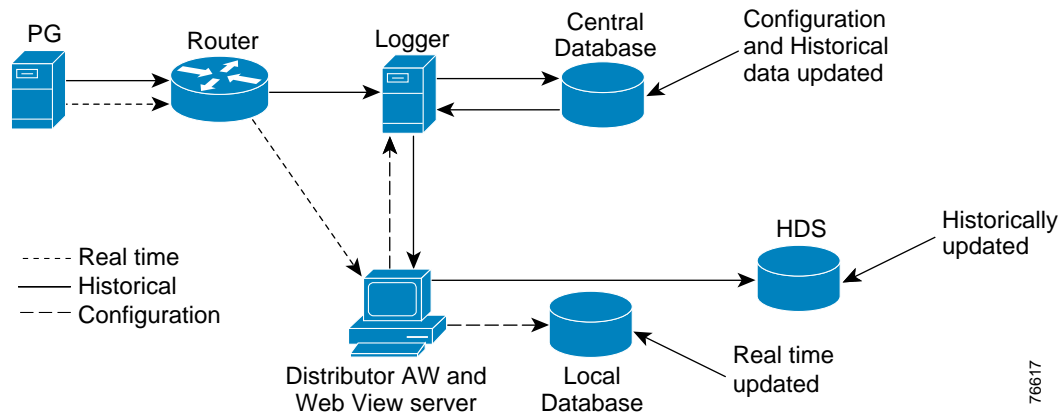
The CTI Server reports call events and agent work state changes as they occur through each stage of the call flow, from the moment a call arrives at an answering resource (Cisco CallManager) until the caller hangs up. In a desktop application environment, call event information is delivered to the targeted agent desktop at the same time the call is delivered. In this environment, applications may also share information with each other (for example, extended call variables such as ANI, caller entered digits, agent transfers, and conferences).

The amount of data associated with a call depends upon the CTI solution and how much data is collected. For example, if all 10 of the standard 40-byte fields are used, they add only 400 bytes (or 3.2 kbps) per event to any of the existing links. If all of the Extended Call Variables (ECV) are used, they add 2,000 bytes per second per event.

In general, when deploying agents at remote sites over the WAN, it is critical to account for CTI data sent to the agent desktop as well as all other types of traffic (for example, voice and call control) traversing the WAN. CTI Server traffic flow to and from the desktop will vary based on the solution implemented (see [Network Bandwidth Provisioning](#), page 8-16). Control traffic such as agent state changes should be marked as AF31.

Administrative Workstation (AW) Traffic Flows

Often, Administrative Workstations (AWs) are co-located with the HDS on the same server PGs and share the same physical WAN and LAN circuits as the ICM Central Controller. When AWs are remote over a WAN, network activity for the AW must also be factored into network bandwidth calculations. The traffic from the AW/HDS to the Central Controller represents system configuration changes, such as when a script is saved. Similarly, traffic from the Central Controller is sent to the AW (real-time feeds) for monitoring purposes. (See [Figure 8-8](#).) This traffic depends on configuration and operational activity and must be considered when sizing network bandwidth. Refer to the ICM product documentation, available online at Cisco.com, for additional details regarding AW configurations. Also see [Table 8-19](#) for examples of traffic volumes.

Figure 8-8 Current Database Architecture

76617

Network Bandwidth Provisioning

The amount of traffic that is sent between the ICM Central Controller and PGs depends on the call load and many configuration parameters such as number of agents, number of peripherals, average number of skill groups per agent, average number of call variables, percentage of calls queued, and average number of "RUN VRU" script nodes executed by a call. Transient boundary conditions (for example, startup configuration load) and specific configuration sizes also factor into the equation.

Cisco Systems Engineers (SEs) have access to tools to help customers estimate the volume of traffic sent between the Cisco CallManager or IP IVR PG and the ICM Central Controller. These tools show the bandwidth required for all traffic flows (TCP and UDP) and the priority level required for QoS.

These tools are based on the following sources:

- Extensive testing in the Cisco Enterprise Solutions Engineering (ESE) labs, where many bandwidth and QoS profiling tests were conducted with different dynamic loads (calls per second) and varying configurations.
- Development methodologies for calculating network bandwidth derived from information documented in various IPCC and ICM specifications identifying the various types of network messages and the number and size of those messages associated with various IPCC and ICM communications protocols and call flows.

Properly provisioning the network bandwidth is essential for any Cisco AVVID network design. Bandwidth provisioning should include all types of traffic, from RTP voice stream traffic and voice call control traffic to customer application traffic, including email, web activity, and Customer Relations Manager (CRM).

Check with your Cisco Partner or SE to help you estimate bandwidth for your specific configuration.

Bandwidth Sizing Examples for CTI Server to Agent Desktop

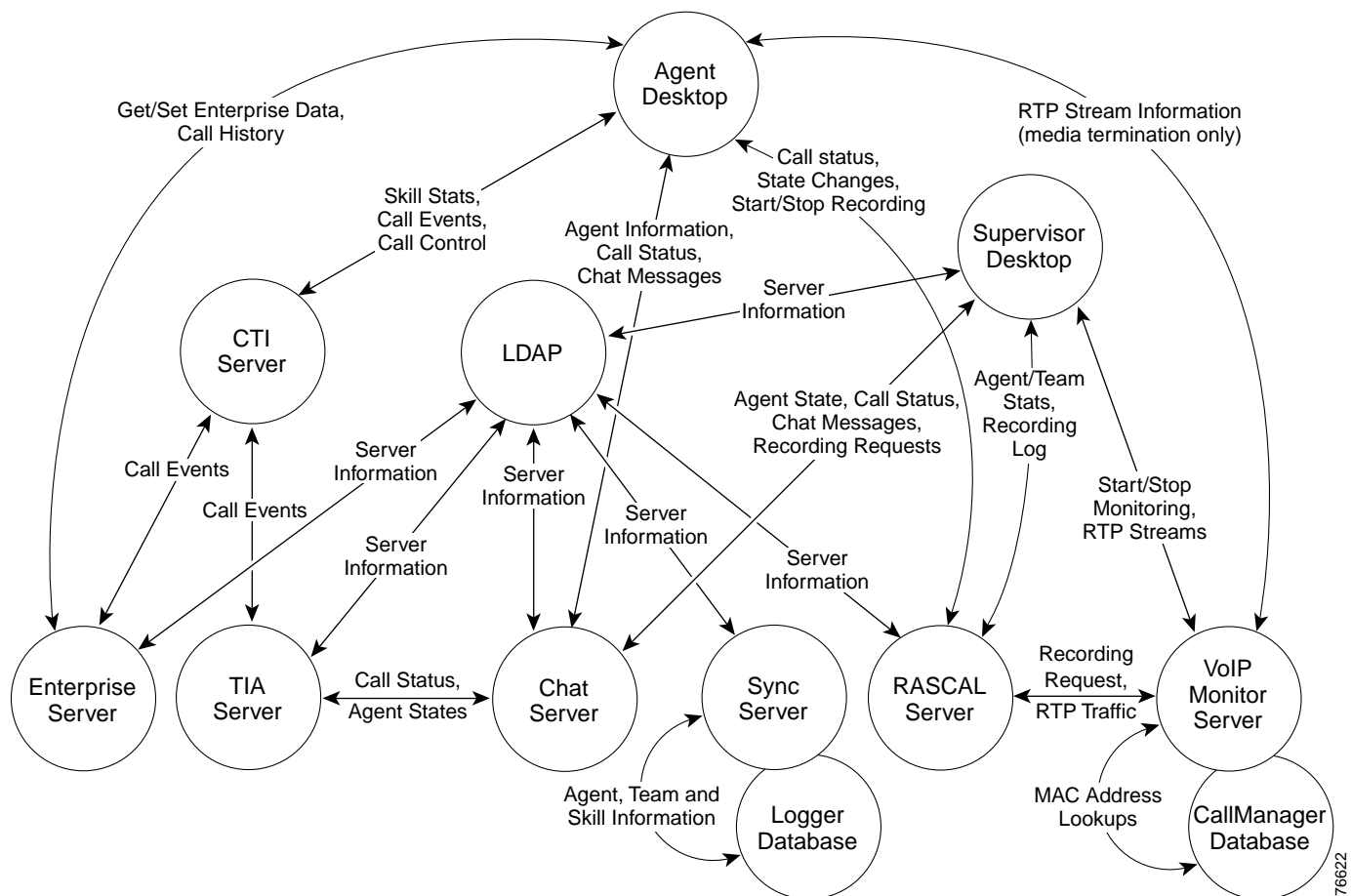
The amount of traffic sent between the CTI Server and the CTI Desktop depends on the call load and many configuration parameters. (The CTI Server can be on the same PG platform as Cisco CallManager or its own server platform.) This traffic can vary if Extended Call Variables (ECV) are used.

For remote agents over a WAN, the skill group statistics updates (sent at a default 10-second interval) are the most significant sizing criteria for network capacity. The network bandwidth requirements increase linearly as a function of agent skill group membership.

Each customer configuration is different, and Cisco advises designing networks with enough bandwidth to account for all types of traffic traversing the various segments, based on specific customer requirements. It is important to monitor and measure network bandwidth and to adjust for actual traffic requirements early in the deployment cycle.

Figure 8-9 illustrates the communication flows between the various IPCC servers and processes.

Figure 8-9 Inter-Process Communications



76622

Cisco Agent Desktop Call Control Bandwidth Usage

All of the call scenarios in this section are based on use of the Cisco Agent Desktop softphone option to perform the agent functions. The bandwidth usage values represent the total number of bytes sent for the scenario listed. The values include the bandwidth for call control (hang-up, conference, make call, and so on) as well as any CTI events that are returned from the CTI Server. Use of the softphone option represents the worst-case scenario; if a hardware phone is used for call control, the bandwidth numbers will be less. By default, all communication between the Cisco Agent Desktop (CAD) and the CTI Server is done on server port 42027, which is configurable via the ICM software.

Skill Statistics Refresh

This value is the amount of bandwidth between the agent desktop and the servers for skill group statistics. The statistics are refreshed every 5 seconds. Servers not listed in [Table 8-2](#) and [Table 8-3](#) (and those listed with NA) are not used for skill statistics refresh and do not add to the bandwidth for that scenario. The total number of skill groups in the system does not affect the amount of bandwidth between the Cisco Agent Desktop (CAD) and the CTI Server. Only the number of skills per agent has an effect because the CAD receives events only for skill groups for which the agent is a member.

Table 8-2 Skill Statistics Refresh Bandwidth with One Skill per Agent

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	NA	NA
Rascal Server	NA	NA
CTI Server	650	130

Table 8-3 Skill Statistics Refresh Bandwidth with Five Skills per Agent

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	NA	NA
Rascal Server	NA	NA
CTI Server	3250	650

The number of bytes is proportional to number of skills per agent. For example, if there are 25 remote agents with 5 skill groups per agent, the number of bits per second sent from the CTI Server to the remote agents' desktops (traffic on the WAN link) will be:

$25 * 3250 = 81,250$ bytes every 5 seconds, or an average of 16,250 bytes per second

Converting this value to bits per second gives a bandwidth of:

$16250 * 8 = 130,000$ bps, or 130 kbps

Agent State Change

This value represents the bandwidth used to change agent state from ready to not-ready.

Table 8-4 Agent State Change Bandwidth with One Skill per Agent

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	60	150
Rascal Server	170	350
CTI Server	340	270

Table 8-5 Agent State Change Bandwidth with Five Skills per Agent

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	60	150
Rascal Server	170	350
CTI Server	700	270

If there are 25 remote agents with 5 skills each, and each agent changes state one time, the total number of bytes would be $25 * 700 = 17,500$ bytes.

Make Call

This value represents the bandwidth used when the agent makes a call using call control, and the other party answers.

Table 8-6 Bandwidth to Make a Call Using Call Control

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	450	650
Rascal Server	250	460
CTI Server	1280	600

Answer Incoming Call

This value represents the bandwidth used by an agent to answer an incoming call using call control.

Table 8-7 Bandwidth to Answer an Incoming Call Using Call Control

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	180	810
Rascal Server	340	690
Enterprise Server	2050	960
CTI Server	1230	340

Hang Up

This value represents the bandwidth used when an agent hangs up a call using call control.

Table 8-8 Bandwidth to Hang Up a Call Using Call Control

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	120	350
Rascal Server	160	620
CTI Server	1140	470

Transfer

This value represents the bandwidth used when an agent transfers an existing call, assuming that the other party answers and the transfer is completed.

Table 8-9 Bandwidth to Transfer a Call

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	720	1400
Rascal Server	480	1530
CTI Server	1990	660

Conference

This value represents the bandwidth used when an agent completes a conference of an existing call using call control, assuming the other party answers.

Table 8-10 Bandwidth to Conference a Call Using Call Control

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	800	1650
Rascal Server	320	960
CTI Server	1690	1010

Hold

This value represents the bandwidth used when an agent puts an existing call on hold using call control.

Table 8-11 Bandwidth to Put a Call on Hold Using Call Control

Server	To CAD (bytes)	From CAD (bytes)
Chat Server	120	350
Rascal Server	80	220
CTI Server	340	190

For example, assume there are 25 remote agents and each agent answers 20 calls in the busy hour. Also assume every call is only answered and hung up (no transfer, conference, or hold). In this case, the total number of bytes would be the sum of the value in [Table 8-7](#) (answer incoming call) and the value in [Table 8-8](#) (hang up) multiplied by the total number of calls answered by the 25 agents, as follows:

$$(1230 + 1140) \text{ bytes} * (25 * 20 \text{ calls}) * (1 \text{ hour} / 3600 \text{ seconds}) * (8 \text{ bits per byte}) = 2633 \text{ bits per second, or 2.63 kbps}$$

Thus, the traffic from the CTI Server to all 25 desktops (WAN link) would be 2.63 kbps.

Cisco Agent Desktop Bandwidth for an Inbound Call

For a typical inbound call, an agent will answer the call, have an agent state change, and hang up. On a full-duplex connection, these actions require 5890 bytes of bandwidth per agent per call (adding bytes from each of the tables above for answering the call, agent state change, and hanging up).

Table 8-12 Bandwidth for an Inbound Call

Server	To CAD (bytes)	From CAD (bytes)	Total
Chat Server	360	1310	1670
Rascal Server	670	1660	2330
Enterprise Server	2050	960	3010
CTI Server	2810	1080	3890
Total	5890	5010	10900

For a site with 100 agents, each handling 30 calls in the busy hour (for a total of 3000 calls in the busy hour, or 0.83 calls per second), the bandwidth is:

$$5980 * 0.83 * (8 \text{ bits per byte}) = 40 \text{ kbps}$$

The access to LDAP is not included in this calculation because both Cisco Agent Desktop and Cisco Supervisor Desktop read their profiles only once at start up and cache it. Note that the values are not based on calls in progress but on calls attempted and completed. In other words you cannot calculate bandwidth based upon the number of calls that are in progress; you must calculate it based upon the number of calls completed because the amount of bandwidth used is per call and does not depend upon on how long the call lasts. A one-minute call and a 10-minute call will typically generate the same amount of bandwidth (excluding voice traffic). This example does not take into account additional traffic generated if calls are transferred, held, or conferenced while in progress before they are terminated (hung up). Traffic generated by such actions would be additional, as shown in [Table 8-9](#) through [Table 8-11](#).

Bandwidth for Silent Monitor to Supervisor

The amount of traffic between the VoIP Monitor server and the monitoring supervisor is the same as that required for an IP phone, minus any control protocols.

The VoIP Monitor server sends two Real-Time Transport Protocol (RTP) streams to the supervisor application (one for the stream to the agent's phone and one for the stream from the agent's phone). The bandwidth for these streams depends on the codec being used. The VoIP Monitor server supports G.711 and G.729 with and without silence suppression. For G.711 the bit rate is 64 kbps, so the two streams require 128 kbps per monitoring supervisor. The bit rate for G.729 is 8 kbps, which results in 16 kbps per monitoring supervisor. This accounts for only the RTP data, but we also need to account for UDP, IP, and Link headers by using a VoIP bandwidth calculator similar to the one at

<http://www.packetizer.com/iptel/bandcalc.html>

Using this calculator on our example reveals that a single G.711 64-kbps stream actually uses 87.2 kbps and a G.729 8-kbps stream uses 31.2 kbps. If silence suppression is used, the bandwidth may be reduced significantly, depending on the conversation background noise and other factors. (See [Table 8-13](#).)

Table 8-13 Bandwidth per Monitoring Supervisor

Codec	Average kbps per monitoring supervisor	Maximum kbps per monitoring supervisor
G.711	174.4	174.4
G.711 with silence suppression	61	174.4
G.729	62.4	62.4
G.729 with silence suppression	21.8	62.4

In [Table 8-13](#), the maximum kbps is the maximum instantaneous bandwidth. When silence suppression is used on a physical channel that has fixed capacity, you must carefully consider this metric because, when a voice signal is present, the maximum bandwidth is required (and average bandwidth is not really useful). For example, based on the calculated average bandwidth, it might seem as though a 64-kbps G.711 stream should be able to transmit at 160 samples per packet using RTP silence suppression and using only 61 kbps of bandwidth. However, when there is a voice signal present for both sides of the conversation, 174.4 kbps are needed, and you must provision for this maximum bandwidth.

Bandwidth for VoIP Monitor to RASCAL

The RASCAL server is used to record agent conversations. The bandwidth requirements between the RASCAL server and the VoIP Monitor server are shown in [Table 8-14](#).

Table 8-14 Bandwidth per Recorded Call

Codec	Average kbps per recorded call	Maximum kbps per recorded call
G711	174.4	174.4
G711 with silence suppression	61	174.4
G729	62.4	62.4
G729 with silence suppression	21.8	62.4

Bandwidth for Cisco Supervisor Desktop to Servers

There will be additional traffic from the Cisco Supervisor Desktop application to the Chat Server and the CTI Server. The amount of traffic is based upon the number of agents in the supervisor's team and what they are doing with the interface.

For each agent on the supervisor's team, there will be 1000 bytes per call sent between the Cisco Supervisor Desktop (CSD) and the Chat Server (800 bytes from Chat Server to CSD and 200 bytes from CSD to Chat Server). For example, if there are 10 agents on the supervisor's team and each one takes 20 calls an hour, the traffic will be:

$$10 \text{ agents} * (20 \text{ calls per agent}) * 1000 \text{ bytes} * (8 \text{ bits per byte}) / 3600 = 444 \text{ bps} = 0.44 \text{ kbps}$$

There will be additional traffic sent if the supervisor is looking at agent, team, or skill statistics, as follows:

- Agent statistics

An additional 750 bytes (400 bytes from CSD to RASCAL server and 350 bytes from RASCAL server to CSD). This is a one-time transfer that occurs when the supervisor clicks on the agent node. Agent statistics are not automatically refreshed.

- Team statistics

An additional 750 bytes per agent on the team (400 bytes from CSD to RASCAL server and 350 bytes from RASCAL server to CSD). This is a one-time transfer that occurs when the agent clicks on the team node. Team statistics are not automatically refreshed. For example, if there are 10 agents on the team, there will be 7500 bytes sent when the supervisor clicks on the team node.

- Skill statistics (Skill statistics are not available to remote supervisors)

An additional 850 bytes per skill are sent from the CSD to the CTI Server (130 bytes from CSD to CTI Server and 720 bytes from CTI Server to CSD). An additional 1420 bytes per skill are sent from the CSD to the Chat Server (730 bytes from CSD to Chat Server and 690 bytes from CTI Server to Chat Server). In Cisco Supervisor Desktop version 4.2.1, the supervisor must manually refresh this data by clicking on the node; however, in version 4.4 this data will automatically refresh at an interval that is configurable by the Administrative Workstation (AW). These statistics will be refreshed only while the supervisor is on the Skill statistics node. The default for the refresh interval is 5 seconds. If there are 5 skills refreshed at 5-second intervals, the bandwidth for skill statistics while the supervisor is on the node is:

$$(5 \text{ skills} * (850 + 1420)) / (5 \text{ seconds refresh rate}) = 2.27 \text{ kbps} * (8 \text{ bits per byte}) = 18 \text{ kbps}$$

This number can be reduced by using a longer refresh interval (for example, 10 seconds instead of 5 seconds).

Module Interactions

Table 8-15 summarizes the communication between the various CAD servers and IPCC components. Clients are listed in the top row, and servers are listed in the left column. The Xs represent client-to-server communication. For example, RASCAL is a client to the VoIP Monitor, but the VoIP Monitor is not a client to RASCAL.

Table 8-15 Module Interactions

Server	Client							
	VoIP Monitor	Enterprise	Call Chat	RASCAL	TAI	Sync Server	CAD	CSD
CTI Server		X			X		X	
VoIP Monitor				X			X	X
Enterprise							X	
Call Chat					X		X	X
RASCAL							X	X
Telecaster Agent Interface (TAI)								
Sync Server								
Cisco Agent Desktop (CAD)			X					
Cisco Supervisor Desktop (CSD)	X		X					
Primary LDAP	X	X	X	X	X	X	X	X
Secondary LDAP ¹								
Cisco CallManager database server	X					X		
ICM Logger database server						X		

1. If the primary LDAP server fails, all clients will talk to the secondary instead.

The following notes apply to Table 8-15:

- The Telecaster Agent Interface (TAI) Server is a client to the Chat Server. The TAI Server sends call information to the Chat Server, similar to the way the Agent Desktop does.
- The TAI Server is a client to the CTI Server. The TAI Server sends IP Phone agent call events and agent state changes to the CTI Server.
- The Enterprise Server is a client to the CTI Server, and it sends Agent Desktop call events.
- The RASCAL Server is a client to the VoIP Monitor Server, and it sends recording requests and RTP streams.
- The VoIP Monitor Server is a client to the CallManager SQL Server database, and it sends MAC address lookups.
- The Sync Server is a client to the ICM Logger SQL Server database, and it sends agent, team, and skill lookups.
- All of the CAD servers are clients to the Directory Services Server (LDAP). Server location (IP address and port) is written on startup. The Sync Server writes agent, team, and skill information.

Server Placement Recommendations

For a deployment with centralized call processing and remote agents, [Table 8-16](#) lists the recommended server placements to minimize bandwidth to the desktop.

Table 8-16 Recommended Server Placements

Software	Site or Location	Reason
CTI Server	Central	Communication with PG.
VoIP Monitor	Remote (near agents)	Span of agent traffic to server. (This is a requirement, not a recommendation, for silent monitoring and recording.)
Enterprise	Central (near CTI Server)	CTI traffic to CTI Server outweighs traffic to agents.
Call Chat	Central (with Enterprise Server)	Install requires Call Chat Server to be co-located with Enterprise Server. Enterprise Server CTI bandwidth outweighs Call Chat to agent bandwidth.
RASCAL	Remote (near agents and VoIP Monitor)	Needs to be near agents for statistics update, and near VoIP Monitor for recordings.
TAI and IP Phone agent	Central (near CTI Server)	CTI traffic to CTI Server outweighs traffic to agents.
Sync Server	Central (near Logger database)	Should be close to ICM Logger database.
Primary Directory Services (LDAP)	Central (with Sync Server)	Installed with Sync Server.
Secondary Directory Services (LDAP)	Central (near primary LDAP)	Replication with primary.
Desktop Supervisor	With the agents	Close to the VoIP Monitor Server.

For multiple remote locations, each remote location requires a VoIP Monitor Server. In Cisco Agent Desktop version 4.4, multiple VoIP Monitor Servers are supported in a single Logical Contact Center (LCC). The RASCAL Server may be moved to the central location if the WAN connections are able to handle the traffic. If not, each site should have its own separate installation of the Cisco Agent Desktop software (its own LCC).

QoS Considerations

When considering which traffic flows are mission-critical and need to be put into a priority queue for quality of service (QoS), use the following list to rank importance:

1. Customer experience
2. Agent experience
3. Supervisor experience
4. Administrator experience

The customer's experience is the most important, but the agent is second because the agent is in direct contact with the customer.

Using this ranking for the server-to-server flows listed in [Table 8-15](#), the traffic from Enterprise Server to CTI Server (call events) is the most critical. Based on the server placement recommendation, both are located in the central location, but QoS must still be applied. This traffic should be classified as AF31, similar to voice call control or signaling traffic. The traffic from the CAD (desktop) to and from the CTI servers (call events and call control) should also be prioritized and classified as AF31.

For IP Phone agents, the TAI-to-CTI Server communication is also important because it affects how quickly agents can change their state. This traffic should also be classified as AF31.

The traffic from the CAD (desktop) to and from the Chat Server (agent information and call status) is less critical and should be classified as AF21 or AF11.

Cisco Agent Desktop Software Component Port Usage

[Table 8-17](#) lists port usage information for the various Cisco Agent Desktop software modules.

Table 8-17 Software Module Port Usage

Module	TCP/UDP	Port Number
Chat Server	TCP	59000
VoIP Monitor Server	TCP	59002
RASCAL Server	TCP	59003
Enterprise Server	TCP	59004
TAI Server	TCP	59010
Synchronization Server	TCP	59011
Chat Server DLL (Call/Chat)	TCP	59020
Chat Server DLL (Supervisor)	TCP	59021
Chat Server DLL (TAI Server)	TCP	59022 ¹
Chat Server DLL (Reserved 1)	TCP	59023
Chat Server DLL (Reserved 2)	TCP	59024
Chat Server (VPN thread)	TCP	59025 ²
VoIP Monitor Server (VPN Thread)	TCP	59026 ²

1. Actual port used will be the same as the TAI Server (that is, 59010).

2. Added in Cisco Agent Desktop version 4.4.1 to support VPN desktops.

By default, all communication between the CAD (desktop) and the CTI Server is done on server port 42027.

Bandwidth Sizing Examples for PG to ICM Central Controller

[Table 8-18](#) shows bandwidth examples for specific configurations. Different configurations will have different bandwidth requirements. Cisco Systems Engineers (SEs) have access to tools that can help customers estimate bandwidth based on their specific configuration. These planning tools will be included in a future IPCC release.

**Note**

Check with your Cisco partner or SE to help you estimate bandwidth for your specific configuration.

Table 8-18 Network Bandwidth Sizing Values for Traffic Between PG and ICM Central Controller

IPCC Example Configurations	Calculations are based on 30 calls per agent in the busy hour, 5 skill groups per agent, and 1000 bytes (out of 2000) for all Extended Call Variables (ECV). 70% of calls are queued. There are 4 IP IVR PIMs and 3 "RUN VRU" script nodes.											
Number of agents	25	50	75	100	200	300	400	500	600	800	1000	1200
Calls per second (cps)	.2	.42	.62	.83	1.66	2.5	3.33	4.2	5	6.66	8.33	10
Bandwidth for AF31 (high and medium priority) traffic between Cisco CallManager PG and ICM	34	40	46	52	77	102	126	151	175	224	274	323
Bandwidth for AF21 (low priority) traffic between Cisco CallManager PG and ICM	2	4	6	7	14	21	27	34	41	54	67	81
Total bandwidth (in kbps) for link between Cisco CallManager PG and ICM	36	44	52	59	91	123	153	185	216	278	341	404
Bandwidth for AF31 (high and medium priority) traffic between IP IVR PG and ICM	71	85	99	110	163	216	268	323	374	479	584	690
Bandwidth for AF21 (low priority) traffic between IP IVR PG and ICM	1	2	2	3	6	8	11	14	16	22	27	32
Total bandwidth (in kbps) for link between IP IVR PG and ICM	72	87	101	113	169	224	279	337	390	501	611	722

ICM Inter-Server Network Traffic Volume

As mentioned previously, traffic and bandwidth measurements were also taken between other IPCC components such as ICM Central Controllers and PG redundant sides. However, no profiling was conducted nor calculators developed for such traffic flows, but these enhancements are planned for a future release. The measurements presented here merely give an indication of the relative volume of traffic between the various ICM components.

The data in [Table 8-19](#) is specific to the following test configuration (different configurations will have different bandwidth requirements):

- 500 agents defined and active
- 14 skill groups per agent
- The following reporting options were set (changed from the system defaults):
 - Trunk Group Reporting enabled
 - 5-minute Historical Data Server (HDS) with skill-group and trunk-group reporting enabled
 - Real Time Data with trunk-group reporting enabled
 - IVR PG Queue Reporting enabled

**Note**

The terms *active* and *standby* are used to denote which Call Router or PG has active connections on the visible (public) network (PG side A and ICM side B can be active at the same time.) *Sync High* denotes high-priority synchronization traffic to be marked as AF31, along with External Message Transport (EMT) heartbeats. *Sync Low* denotes low-priority traffic to be marked as AF21.

Table 8-19 Network Traffic Volume Between ICM Servers

Flow Direction	Flow Type	Average Bits per Second at 0 BHCA	Average Bits per Second at 5,000 BHCA	Average Bits per Second at 10,000 BHCA	Average Bits per Second at 15,000 BHCA	Average Bits per Second at 20,000 BHCA	Average Bits per Second at 25,000 BHCA	Average Bits per Second at 30,000 BHCA
PG to PG, Private Network								
Active IVR PG to Standby IVR PG	Sync Low	52	44	52	52	53	52	54
Standby IVR-PG to Active IVR-PG	Sync Low	52	54	49	49	50	48	51
Active IVR PG to Standby IVR PG	Sync High	3,625	133,392	262,768	390,691	522,101	621,989	774,239
Standby IVR PG to Active IVR PG	Sync High	3,625	10,545	19,435	28,179	37,121	43,956	54,232
Active IVR PG to Standby IVR PG	External Message Transport (EMT) Heartbeat	5,110	5,110	5,110	5,110	5,110	5,110	5,110
Standby IVR PG to Active IVR PG	EMT Heartbeat	5,110	5,110	5,110	5,110	5,110	5,110	5,110
Active Cisco CallManager PG to Standby Cisco CallManager PG	Sync Low	53	150	73	83	97	108	140
Standby Cisco CallManager PG to Active Cisco CallManager PG	Sync Low	53	137	57	60	65	69	82

Table 8-19 Network Traffic Volume Between ICM Servers (continued)

Flow Direction	Flow Type	Average Bits per Second at 0 BHCA	Average Bits per Second at 5,000 BHCA	Average Bits per Second at 10,000 BHCA	Average Bits per Second at 15,000 BHCA	Average Bits per Second at 20,000 BHCA	Average Bits per Second at 25,000 BHCA	Average Bits per Second at 30,000 BHCA
Active Cisco CallManager PG to Standby Cisco CallManager PG	Sync High	2,821	139,645	302,586	468,919	642,235	782,551	1,006,501
Standby Cisco CallManager PG to Active Cisco CallManager PG	Sync High	2,821	16,401	32,281	48,293	64,802	77,730	99,421
Active Cisco CallManager PG to Standby Cisco CallManager PG	EMT Heartbeat	5,110	5,110	5,110	5,110	5,109	5,108	5,107
Standby Cisco CallManager PG to Active Cisco CallManager PG	EMT Heartbeat	5,110	5,110	5,110	5,110	5,110	5,110	5,110
Router to Router, Private Network								
Active Router to Standby Router	Sync Low Priority	5,918	12,902	29,174	37,107	45,376	46,458	58,343
Standby Router to Active Router	Sync Low Priority	348	887	1,260	1,503	1,738	1,813	2,931
Active Router to Standby Router	Sync Medium Priority	582	25,011	43,392	60,863	78,555	92,392	11,6871
Standby Router to Active Router	Sync Medium Priority	185	2,030	2,534	2,943	3,352	3,663	4,201
Active Router to Standby Router	Sync High Priority	5,385	18,650	26,586	34,244	42,109	48,019	59,539
Standby Router to Active Router	Sync High Priority	4,341	15,509	19,821	23,728	27,750	30,580	36,814

Table 8-19 Network Traffic Volume Between ICM Servers (continued)

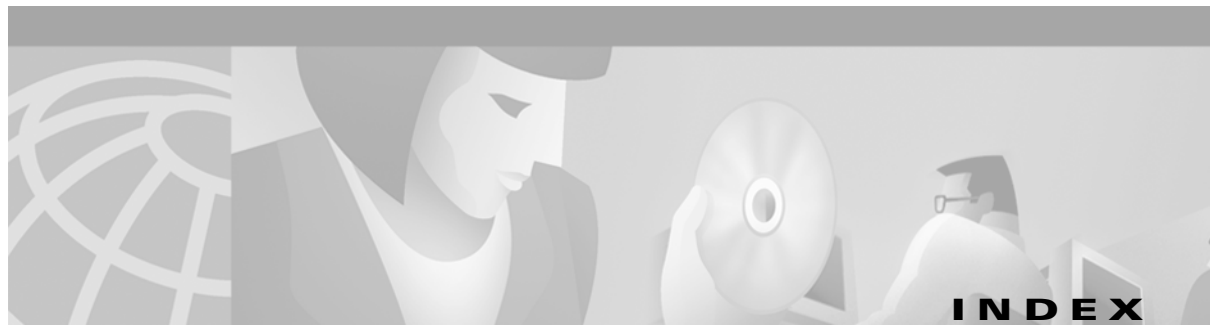
Flow Direction	Flow Type	Average Bits per Second at 0 BHCA	Average Bits per Second at 5,000 BHCA	Average Bits per Second at 10,000 BHCA	Average Bits per Second at 15,000 BHCA	Average Bits per Second at 20,000 BHCA	Average Bits per Second at 25,000 BHCA	Average Bits per Second at 30,000 BHCA
Active Router to Standby Router	EMT Heartbeat	5,114	5,114	5,114	5,114	5,114	5,114	5,114
Standby Router to Active Router	EMT Heartbeat	5,114	5,114	5,114	5,114	5,115	5,114	5,114
Active Logger to Standby Logger	Phone Home Customer Support Forwarding Service (CSFS)	245	245	244	246	237	245	244
Standby Logger to Active Logger	Phone Home CSFS	238	238	236	238	229	238	236
Active Logger to Standby Logger	EMT Heartbeat	171	170	170	170	170	171	170
Standby Logger to Active Logger	EMT Heartbeat	171	170	170	171	170	171	171
Active Logger to Standby Logger	EMT Heartbeat #2	170	170	171	170	170	170	170
Standby Logger to Active Logger	EMT Heartbeat #2	170	170	170	170	171	170	171
Router to PG, Visible Network								
Active Call Router to Active IVR PG	Device Management Protocol (DMP) Low Priority	19	366	400	491	569	629	741
Active IVR PG to Active Call Router	DMP Low Priority	127	3,329	6,161	8,944	11,733	13,920	17,177
Active Call Router to Active IVR PG	DMP Medium Priority	181	1,236	1,280	1,269	1,245	1,227	1,169
Active IVR PG to Active Call Router	DMP Medium Priority	499	4,121	4,990	5,487	6,118	6,436	6,900

Table 8-19 Network Traffic Volume Between ICM Servers (continued)

Flow Direction	Flow Type	Average Bits per Second at 0 BHCA	Average Bits per Second at 5,000 BHCA	Average Bits per Second at 10,000 BHCA	Average Bits per Second at 15,000 BHCA	Average Bits per Second at 20,000 BHCA	Average Bits per Second at 25,000 BHCA	Average Bits per Second at 30,000 BHCA
Active Call Router to Active IVR PG	DMP High Priority	85	6,977	13,744	20,578	27,516	33,081	41,021
Active IVR PG to Active Call Router	DMP High Priority	94	4,429	8,362	12,298	16,168	19,276	23,652
Active Call Router to Active IVR PG	EMT Heartbeat	1,305	1,305	1,305	1,305	1,305	1,306	1,305
Active IVR PG to Active Call Router	EMT Heartbeat	1,305	1,306	1,305	1,305	1,305	1,305	1,305
Active Call Router to Active Cisco CallManager PG	DMP Low Priority	141	439	871	1,139	1,403	1,504	1,867
Active Cisco CallManager PG to Active Call Router	DMP Low Priority	5,658	9,645	23,179	28,437	34,181	33,240	41,077
Active Call Router to Active Cisco CallManager PG	DMP Medium Priority	27	874	1,311	1,758	2,182	2,532	3,075
Active Cisco CallManager PG to Active Call Router	DMP Medium Priority	43	20,511	38,233	55,606	72,900	86,987	11,1042
Active Call Router to Active Cisco CallManager PG	DMP High Priority	57	6,318	12,513	18,675	24,873	29,840	38,786
Active Cisco CallManager PG to Active Call Router	DMP High Priority	72	4,201	7,907	11,561	15,152	18,072	23,675

Table 8-19 Network Traffic Volume Between ICM Servers (continued)

Flow Direction	Flow Type	Average Bits per Second at 0 BHCA	Average Bits per Second at 5,000 BHCA	Average Bits per Second at 10,000 BHCA	Average Bits per Second at 15,000 BHCA	Average Bits per Second at 20,000 BHCA	Average Bits per Second at 25,000 BHCA	Average Bits per Second at 30,000 BHCA
Active Call Router to Active Cisco CallManager PG	EMT Heartbeat	1,305	1,306	1,305	1,305	1,305	1,305	1,305
Active Cisco CallManager PG to Active Call Router	EMT Heartbeat	1,305	1,306	1,305	1,305	1,305	1,305	1,305
Router and Logger to Administrative Workstation (AW) and Historical Data Server (HDS), Visible Network								
Active Router to AW/HDS	Real Time Feed	541	22,851	40,791	55,969	73,971	87,102	103,941
AW/HDS to Active Router	Real Time Feed	170	561	843	1,095	1,401	1,617	1,876
Active Router to AW/HDS	EMT Heartbeat	522	521	522	522	522	522	522
AW/HDS to Active Router	EMT Heartbeat	523	522	523	522	523	523	522
Active Logger to AW/HDS	Logger Replication	7,395	20,925	34,944	43,249	60,159	62,953	80,925
AW/HDS to Active Logger	Logger Replication	1,606	1,783	2,221	2,457	2,792	2,759	3,265
Active Logger to AW/HDS	EMT Heartbeat	174	174	174	173	174	174	174
AW/HDS to Active Logger	EMT Heartbeat	174	173	175	174	175	173	173
Active Logger to AW/HDS	EMT Heartbeat	174	174	174	173	174	174	174
AW/HDS to Active Logger	EMT Heartbeat	174	176	175	173	175	173	173



A

ACD integration [2-13](#)
Administrative Workstation (AW) [1-6, 8-15](#)
admission control [1-22](#)
Agent Desk Settings [1-11](#)
Agent Desktop
 bandwidth requirements [8-15, 8-17](#)
 Base Server [6-9](#)
 call control [8-18](#)
 Cisco Agent Desktop [7-4](#)
 CTI OS Toolkit [7-3](#)
 described [1-6, 7-1](#)
 details [7-4](#)
 enhanced option [7-6](#)
 port usage [8-26](#)
 QoS requirements [8-15](#)
 settings [1-11](#)
 Silent Monitor Server [6-10](#)
 sizing [6-9](#)
 standard option [7-6](#)
 types [7-2](#)
agents
 general [1-11](#)
 login [1-12](#)
 settings [1-11](#)
 sizing [5-4, 5-5, 6-4](#)
 state changes [8-19](#)
 transfers between [1-20](#)
 wrap-up time [5-2, 5-11](#)
agent-to-agent transfers [1-20](#)
AHT [5-2](#)
a-law [3-4](#)

alternate [1-20](#)
ANI [3-8](#)
architectural overview [1-1](#)
audience for this document [x](#)
automatic call distribution (*see* ACD)
automatic number identification (ANI) [3-8](#)
availability of functions and features [4-1](#)
average handle time (AHT) [5-2](#)
AW [1-6, 8-15](#)

B

bandwidth
 Agent Desktop [8-17](#)
 agent state changes [8-19](#)
 answering a call [8-19](#)
 call control [8-18](#)
 conference call [8-20](#)
 hanging up a call [8-20](#)
 hold [8-20](#)
 inbound calls [8-21](#)
 making a call [8-19](#)
 module interactions [8-24](#)
 provisioning [8-1](#)
 RASCAL server [8-22](#)
 server placement [8-25](#)
 sizing examples [8-17, 8-26](#)
 skill statistics refresh [8-18](#)
 supervisor [8-22](#)
 Supervisor Desktop [8-23](#)
 transferring a call [8-20](#)
 VoIP Monitor [8-22](#)
Base Server [6-9](#)

BHCA 5-2, 6-4
 BHT 5-2, 5-7
 blind transfer 1-18
 blockage 5-3
 blocked calls 5-3
 BRI interface 3-7
 busy hour 5-1
 Busy Hour Call Attempts (BHCA) 5-2, 6-4
 busy hour traffic (BHT) 5-2, 5-7

C

CAD (*see* Agent Desktop)
 calculators for Erlang values 5-3
 call admission control 1-22
 call control traffic 8-1, 8-18
 calling line ID (CLID) 3-7
 calling party name 3-8
 CallManager (*see* Cisco CallManager)
 call processing
 centralized 2-4
 distributed 2-9
 calls
 blocked 5-3
 flow 1-3
 gateway ports needed 3-4
 inbound 8-21
 preservation 3-12
 queuing
 defined 5-3
 on IP IVR 2-3, 6-4
 scenarios 1-16
 sizing trunks 5-7
 with translation routing 1-13
 recording 6-11
 routing 1-13
 transferring 1-16
 treatment of 5-9
 wrap-up time 5-2, 5-11

 Catalyst switches 7-9
 centralized call processing 2-4
 Cisco.com xii
 Cisco Agent Desktop (CAD, *see* Agent Desktop)
 Cisco CallManager
 database 6-5
 described 1-1
 failover 4-14, 4-19
 high availability 4-7
 interfaces 7-8
 recovery 4-19
 redundancy 3-12, 4-7
 with IP IVR 4-12
 Cisco Supervisor Desktop (*see* Supervisor Desktop)
 classification of traffic 8-10, 8-11, 8-14
 Class of Service (CoS) 8-10
 CLID 3-7
 clients for routing 1-10
 client-to-server relationships 8-24
 combination transfers 1-21
 computer telephony integration (*see* CTI)
 conferences
 sizing 6-5, 8-20
 transfers of 1-21
 Configuration Manager 1-6
 configurations, verified 7-10
 consultative transfer 1-19
 conversion of a-law to mu-law 3-4
 co-resident servers 6-12
 CoS 8-10
 CTI
 components 6-9
 Manager 4-7, 4-10, 4-20
 Object Server (*see* CTI OS)
 Server 1-4, 4-23

CTI OS

- architecture [7-3](#)
- failover [4-24](#)
- sizing [6-9](#)
- Toolkit [7-3](#)

D

- database [6-5](#)
- deployment models [2-1](#)
- design tools [5-3](#)
- device targets [1-10](#)
- dialed number (DN) [1-12](#)
- dialed number identification service (DNIS) [3-8](#)
- Dialed Number Plan (DNP) [1-17](#)
- dial plans [1-17](#)
- Differentiated Services Code Point (DSCP) [8-10](#)
- digital signal processor (DSP) [3-4](#)
- directory number (DN) [1-12](#)
- distributed call processing [2-9](#)
- DN [1-12](#)
- DNIS [3-8](#)
- DNP [1-17](#)
- documentation
 - CD-ROM [xi](#)
 - feedback [xii](#)
 - obtaining [xi](#)
 - ordering [xi](#)
- double trunking [2-17](#)
- DSCP [8-10](#)
- DSP [3-4](#)
- DTMF relay [3-10](#)

E

- Erlang calculations [5-2, 5-3, 5-4](#)
- extensions for IPCC and IP Telephony on same phone [1-15, 2-18](#)

F

- failover
 - Cisco CallManager [4-14](#)
 - CTI OS [4-24](#)
 - design considerations [4-1](#)
 - ICM [4-14](#)
 - recovery [4-19](#)
 - scenarios [4-14](#)
- features of voice gateways [3-7](#)
- feedback [xii](#)
- flow of calls and messages [1-3, 8-8, 8-13](#)

G

- gatekeeper [1-23](#)
- gateways
 - BRI interface [3-7](#)
 - centralized [2-5, 2-12](#)
 - considerations for IPCC [3-1](#)
 - distributed [2-7, 2-10](#)
 - failover [3-12](#)
 - features [3-7](#)
 - monitoring [3-13](#)
 - not recommended [3-2](#)
 - ports
 - density [3-3](#)
 - sizing [5-6, 5-9](#)
 - PRI interfaces [3-5](#)
 - QoS [3-10](#)
 - QSIG interface [3-6](#)
 - recommended [3-2](#)
 - selecting [3-1](#)
 - selection criteria [3-3](#)
 - SS7 interface [3-6](#)
 - supplemental services [3-9](#)
 - TDM interfaces [3-5](#)
 - VoIP interfaces [3-5](#)

grade of service [5-3](#)
 groups of media resources [3-4](#)

H

HDS [4-24, 6-8](#)
 heartbeat traffic [8-8, 8-13](#)
 high availability [4-1, 7-12](#)
 Historical Data Server (HDS) [4-24, 6-8](#)
 hold
 bandwidth usage [8-20](#)
 music [3-11](#)
 tone [3-11](#)
 hookflash transfer [3-10](#)
 hybrid IP Telephony and IPCC system [1-15, 2-18](#)

I

ICM

Central Controller [1-4](#)
 components [1-4](#)
 described [1-3](#)
 failover recovery [4-20](#)
 failover scenarios [4-14](#)
 interfaces [7-8](#)
 inter-server traffic volume [8-27](#)
 IP IVR redundancy [4-12](#)
 reporting [3-13](#)
 routing clients [1-10](#)
 scripts [6-5](#)
 software modules [1-4](#)
 inbound calls [8-21](#)
 integration
 with ACD [2-13](#)
 with IVR [2-14](#)
 Intelligent Contact Management software (*see* ICM)
 interactions between software modules [8-24](#)
 Interactive Voice Response (*see* IVR)

interfaces

BRI [3-7](#)
 Cisco CallManager [7-8](#)
 ICM [7-8](#)
 PRI [3-5](#)
 PSTN [8-8](#)
 QSIG [3-6](#)
 SCCP [3-5](#)
 SCI [1-2](#)
 SS7 [3-6](#)
 TDM [3-5](#)
 VoIP [3-5](#)

inter-server network traffic [8-27](#)

IPCC

agent desktop [7-2](#)
 architecture [1-1](#)
 call flows [1-3](#)
 components [1-6, 6-1](#)
 extensions [1-15, 2-18](#)
 message flows [1-3](#)
 network components [8-2](#)
 overview [1-1](#)
 supervisor desktop [7-2](#)

IP IVR

described [1-2](#)
 failover recovery [4-19](#)
 high availability [4-10](#)
 ports [5-8, 5-10](#)
 redundancy [4-10](#)
 self-service applications [6-5](#)
 with Cisco CallManager [4-12](#)
 with ICM [4-12](#)

IP Precedence [8-10](#)

IP Telephony extensions [1-15, 2-18](#)

IVR

integration [2-14](#)
 scripts [6-5](#)
see also IP IVR
 transfers to agents [1-21](#)

J

JTAPI [1-7](#)

L

labels [1-10](#)

latency [8-10, 8-14](#)

Layer 2 traffic [8-10](#)

Layer 3 traffic [8-10](#)

level of service [5-3](#)

locations for call admission control [1-24](#)

Logger [1-4, 4-21, 4-24, 6-1](#)

login [1-12](#)

M

managing the network [3-13](#)

media resources [3-4](#)

message flows [1-3](#)

models for deployments [2-1](#)

module interactions [8-24](#)

monitoring gateways status [3-13](#)

mu-law [3-4](#)

multiple transfers [1-21](#)

multi-site deployment

 centralized call processing [2-4](#)

 characteristics [8-4](#)

 distributed call processing [2-9](#)

music on hold [3-11](#)

N

network

 bandwidth [8-8, 8-16](#)

 components [8-2](#)

 configuration [7-8, 7-10](#)

 high availability [7-12](#)

inter-server traffic [8-27](#)

management [3-13](#)

packet sniffing [7-8](#)

private [8-5](#)

public [8-7](#)

QoS [8-8](#)

segmentation [8-5](#)

signaling access [8-8](#)

simple design [7-13](#)

visible [8-7](#)

NFAS [3-9](#)

no answer [1-15](#)

Non-Facility Associated Signaling (NFAS) [3-9](#)

non-ICM transfers [1-20](#)

O

organization of this document [x](#)

overview of IPCC architecture [1-1](#)

P

packet sniffing [7-8](#)

PBX transfers [2-15](#)

percent blockage [5-3](#)

Peripheral Gateway (*see* PG)

Peripheral Interface Manager (*see* PIM)

PG

 design considerations [4-12](#)

 for Cisco CallManager [4-20](#)

 for Voice Response Unit [4-20](#)

 sizing [6-7](#)

PHB [8-10](#)

PIM [1-4, 4-14, 6-7](#)

placement of servers [8-25](#)

ports

- configuring [7-7](#)
 - density of [3-3](#)
 - IP IVR [5-8, 5-10](#)
 - server port usage [8-26](#)
 - sizing [5-6, 5-9](#)
 - Switched Port Analyzer (SPAN) [7-7](#)
- post route [1-17](#)
- post-translation routing [8-10](#)
- precedence of IP traffic [8-10](#)
- preface [ix](#)
- pre-hop behavior (PHB) [8-10](#)
- pre-routing [8-10](#)
- presentation indicator [3-9](#)
- preservation of calls [3-12](#)
- PRI interfaces [3-5](#)
- private network [8-5](#)
- PROGGER [6-1](#)
- provisioning (*see* sizing)
- PSTN
- interfaces [8-8](#)
 - transfers [1-21, 2-16](#)
 - trunks [5-8](#)
- public network [8-7](#)
- purpose of this document [ix](#)

Q

QoS

- configuration examples [8-11](#)
 - design considerations [8-1, 8-25](#)
 - latency [8-10, 8-14](#)
 - on voice gateways [3-10](#)
 - requirements [8-10, 8-14](#)
- QSIG interface [3-6](#)
- Quality of Service (*see* QoS)
- queuing of calls
- defined [5-3](#)
 - on IP IVR [2-3, 6-4](#)

- scenarios [1-16](#)
- sizing trunks [5-7](#)
- with translation routing [1-13](#)

R

- RASCAL application server [6-11, 8-22](#)
- Real-Time Distributor [4-21](#)
- Real-Time Monitor (RTM) [3-13](#)
- recommended
- server placement [8-25](#)
 - software releases [ix](#)
 - voice gateways [3-2](#)
- reconnect [1-19](#)
- recording calls [6-11](#)
- recovery from failover [4-19](#)
- redundancy [3-12, 4-1, 4-10](#)
- related documentation [xi](#)
- releases, recommended software [ix](#)
- Remote Supervisor [7-6](#)
- reports
- Historical Data Server [4-24](#)
 - ICM [3-13](#)
 - Logger [4-24](#)
 - sizing [6-5](#)
 - transfers [1-21](#)
- reroute on no answer (RONA) [1-10, 1-15](#)
- resource sizing [5-1](#)
- ringback on transfers [3-10](#)
- ROGGER [6-1](#)
- RONA [1-10, 1-15](#)
- Router [1-4, 4-21, 6-1](#)
- route request [1-18](#)

routing

- calls [1-13](#)
- clients [1-10](#)
- post-translation [8-10](#)
- request [1-18](#)
- scripts [1-12, 1-13](#)
- translation [1-13](#)

RTM [3-13](#)

S

SCCP [2-6, 3-5, 6-5](#)

SCI [1-2](#)

scope of this document [ix](#)

Script Editor [1-6](#)

scripts [1-12, 6-5](#)

segmentation of the network [8-5](#)

selection criteria for voice gateways [3-3](#)

self-service applications [6-5](#)

servers

- co-resident [6-12](#)
- placement [8-25](#)
- port usage [8-26](#)
- sizing [5-2, 6-1](#)

Service Control Interface (SCI) [1-2](#)

service grade [5-3](#)

service level [5-3](#)

settings for agent desktop [1-11](#)

signaling access network [8-8](#)

silent monitor [7-7](#)

Silent Monitor Server [6-10, 8-22](#)

simple network [7-13](#)

single site

- characteristics [8-3](#)
- deployment model [2-2](#)

single-step transfer [1-18](#)

SIP [7-8](#)

sizing

- Agent Desktop [6-9](#)
- agents [5-4, 5-5, 6-4](#)
- bandwidth [8-1, 8-17, 8-26](#)
- basic example [5-5](#)
- BHCA [6-4](#)
- busy hour traffic [5-7](#)
- call center resources [5-1](#)
- call treatment example [5-9](#)
- conferences [6-5](#)
- CTI components [6-9](#)
- CTI OS [6-9](#)
- database [6-5](#)
- gateway ports [5-6, 5-9](#)
- IPCC components [6-1](#)
- IPCC resources [5-1](#)
- IP IVR ports [5-8, 5-10](#)
- PG [6-7](#)
- PIM [6-7](#)
- ports [5-6, 5-9](#)
- PSTN trunks [5-4, 5-6, 5-8, 5-9](#)
- queued calls [5-7, 6-4](#)
- reports [6-5](#)
- scripts [6-5](#)
- self-service applications [6-5](#)
- servers [6-1](#)
- skill groups [6-4](#)
- softphone [6-5](#)
- Supervisor Desktop [6-9](#)
- variables [6-4](#)
- wrap-up time example [5-11](#)

skill groups [1-11, 6-4](#)

skill statistics refresh [8-18](#)

Skinny Client Control Protocol (SCCP) [2-6, 3-5, 6-5](#)

sniffing packets [7-8](#)

softphone [1-6, 6-5](#)

software releases, recommended [ix](#)

SPAN [7-7, 7-9](#)

SS7 interface [3-6](#)

Standard Interface Protocol (SIP) [7-8](#)

state changes [8-19](#)

state control [1-12](#)

Supervisor Desktop

bandwidth [8-23](#)

Cisco Supervisor Desktop [7-4](#)

CTI OS Toolkit [7-3](#)

described [7-1](#)

details [7-4](#)

guidelines [7-6](#)

Remote Supervisor [7-6](#)

requirements [7-6](#)

silent monitor [7-7](#)

sizing [6-9](#)

types [7-2](#)

supplemental services on voice gateways [3-9](#)

switch capabilities [7-9](#)

Switched Port Analyzer (SPAN) [7-7, 7-9](#)

T

TAC [xii, xiii](#)

Takeback N Transfer (TNT) [1-21](#)

talk time [5-2](#)

target devices [1-10](#)

TDM interfaces [3-5](#)

Technical Assistance Center (TAC) [xii, xiii](#)

time-division multiplexing (*see* TDM)

TNT [1-21](#)

tone on hold [3-11](#)

toolkit for CTI OS [7-3](#)

tools for designing an IPCC solution [5-3](#)

traffic

classification [8-10, 8-11, 8-14](#)

design considerations [8-1](#)

flows [8-8, 8-13](#)

volume [8-27](#)

transcoding [3-4](#)

transfer connect [1-21](#)

transfers

agent-to-agent [1-20](#)

alternate [1-20](#)

bandwidth usage [8-20](#)

blind [1-18](#)

conferenced calls [1-21](#)

consultative [1-19](#)

described [1-16](#)

hookflash [3-10](#)

IVR to agent [1-21](#)

multiple [1-21](#)

non-ICM [1-20](#)

reconnect [1-19](#)

reporting [1-21](#)

ringback [3-10](#)

single-site deployments [2-3](#)

single step [1-18](#)

using Cisco CallManager [2-17](#)

using PBX [2-15](#)

using PSTN [1-21, 2-16](#)

Translation Route to VRU [1-13](#)

translation routing [1-13](#)

treatment of calls [5-9](#)

trunks

double trunking [2-17](#)

sizing [5-4, 5-6, 5-9](#)

types of dial plans [1-17](#)

V

variables for sizing IPCC components [6-4](#)

verified network configurations [7-10](#)

VHM [3-13](#)

visible network [8-7](#)

voice gateways

- BRI interface [3-7](#)

- centralized [2-5, 2-12](#)

- considerations for IPCC [3-1](#)

- distributed [2-7, 2-10](#)

- failover [3-12](#)

- features [3-7](#)

- monitoring [3-13](#)

- not recommended [3-2](#)

- ports

 - density [3-3](#)

 - sizing [5-6, 5-9](#)

- PRI interfaces [3-5](#)

- QoS [3-10](#)

- QSIG interface [3-6](#)

- recommended [3-2](#)

- selecting [3-1](#)

- selection criteria [3-3](#)

- SS7 interface [3-6](#)

- supplemental services [3-9](#)

- TDM interfaces [3-5](#)

- VoIP interfaces [3-5](#)

Voice Health Monitor (VHM) [3-13](#)

Voice Response Unit (VRU) [1-13, 4-20](#)

VoIP interfaces [3-5](#)

VoIP Monitor [7-7, 7-8, 8-22](#)

volume of traffic [8-27](#)

VRU [1-13, 4-20](#)

W

WebView [6-8](#)

World Wide Web [xi](#)

wrap-up time [5-2, 5-11](#)

