Design Guide for Cisco Unified Customer Voice Portal, Release 11.0(1)

First Published: 2015-05-08

Last Modified: 2016-04-06

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Preface

Change History

Change	Date	Link
Initial release of document for release 11.0(1)	Aug 27, 2015	
Added new topic for Cisco Integrated Services Router Generation 3 Gateway		Cisco Integrated Services Router Generation 3 Gateway, on page 7
Added new topic for CVP Whisper Announcement and Agent Greeting Configuration		CVP Whisper Announcement and Agent Greeting Configuration, on page 113
Added new section on Generate G729 Prompts for Unified CVP		Generate G729 Prompts for Unified CVP, on page 83
Added Cisco VVB content across the document wherever VoiceXML Gateway features are described.	Jan 25, 2016	VoiceXML Gateway, on page 6
Added a new topic on "Network-Based Recording" in the Features and Functions chapter	March 8, 2016	Network-Based Recording, on page 168

About this Guide

The *Design Guide for Cisco Unified Customer Voice Portal* provides the following information:

- Architecture of Unified Customer Voice Portal (CVP)
- Design considerations and guidelines for deploying enterprise network solutions that include Unified CVP
- Deployment characteristics and provisioning requirements of the Unified CVP network
- Types of managing, monitoring, and reporting functions that can be used with Unified CVP

Audience

This design guide is intended for the system architects, designers, engineers, and Cisco channel partners who want to apply best design practices for Unified CVP and are familiar with basic contact center terms and concepts along with the information presented in the *Cisco Unified CCE SRND* document. To view those terms and concepts, see http://www.cisco.com/en/US/products/sw/custcosw/ps1844/products_implementation_design_guides_list.html.

Related Documents

- *Configuration Guide for Cisco Unified Customer Voice Portal*
- *IInstallation and Upgrade Guide for Cisco Unified Customer Voice Portal*

Obtaining Documentation and Submitting a Service Request

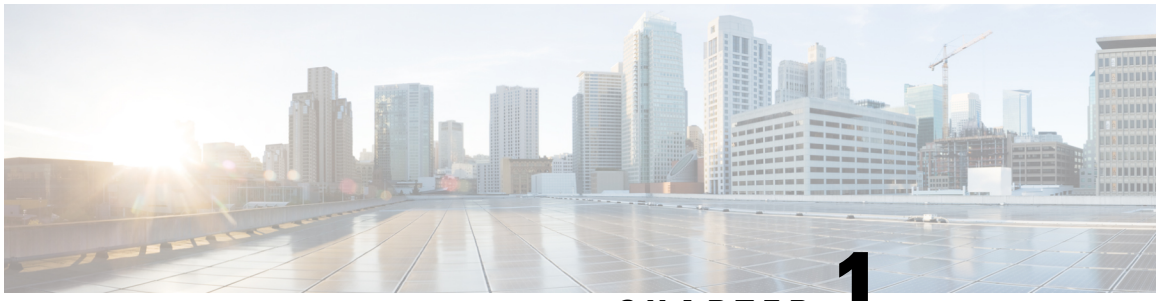
For information on obtaining documentation, using the Cisco Bug Search Tool (BST), submitting a service request, and gathering additional information, see What's New in Cisco Product Documentation.

To receive new and revised Cisco technical content directly to your desktop, you can subscribe to the . RSS feeds are a free service.

Documentation Feedback

Provide your comments about this document to: mailto:contactcenterproducts_docfeedback@cisco.com

CONTENTS



CHAPTER 1

Unified Customer Voice Portal Overview

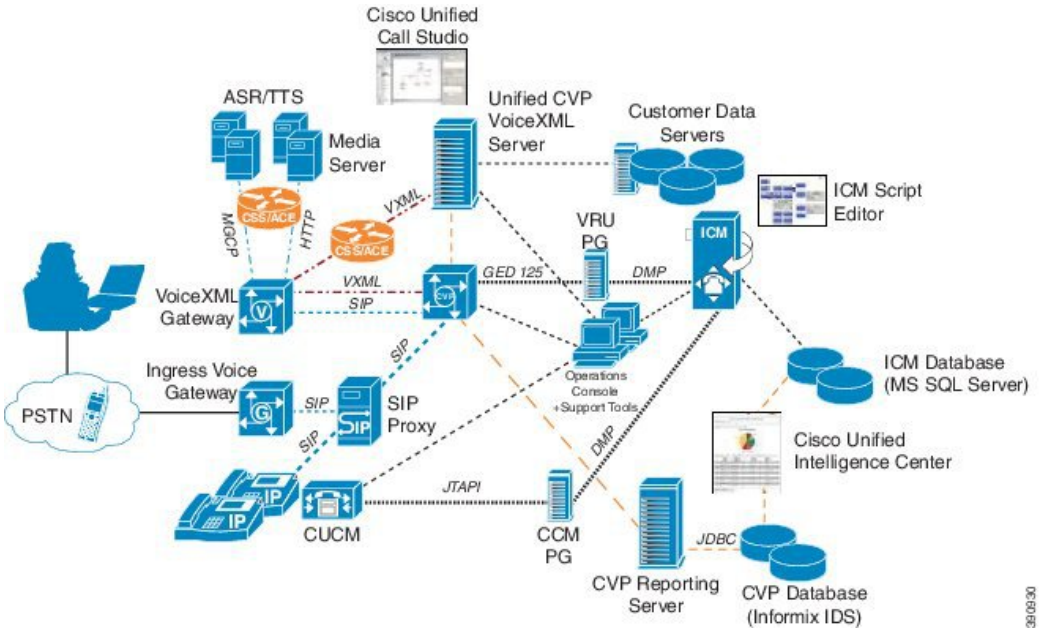
Overview

The Unified Customer Voice Portal (Unified CVP) is a web-based platform that provides carrier-class Interactive Voice Response (IVR) and IP switching services over VoIP networks. The platforms work together so that you can create and deploy IVR applications. The IVR applications include voice and video interaction and traditional numeric input to provide intelligent, personalized self-service over the phone.

Unified CVP is based on VoiceXML (VXML). VXML provides a flexible application development and deployment environment for creating IVR applications. IVR applications control audio input and output, presentation logic, call flow, telephony connections, and error handling. The Unified CVP suite of components enables the VXML to receive and report IVR events and interface with customer database components.

This chapter provides a high-level view of the Unified CVP product as it relates to the VXML and call flow model. The following figure illustrates the Unified CVP and components in a network. The other chapters in this guide provide more details about Unified CVP, including system design considerations such as call flow models and implementation factors.

Figure 1: Unified CVP and Its Components in a Network



Unified CVP Product Components

This section describes the Cisco Unified Customer Voice Portal (CVP) product components.

Note Call Server, VXML Server, and Media Server are combined as one component, which is known as CVP Server. Installing CVP Server installs all three components.

Call Server

The Call Server component provides the following independent services, which all run on the same Windows server:

- **SIP service**: This service communicates with the Unified CVP solution components such as the SIP Proxy Server, Ingress Gateway, Unified CM SIP trunks, and SIP phones. The SIP service implements a Back-to-Back User Agent (B2BUA). This B2BUA accepts SIP invites from ingress voice gateways and typically directs those calls to an available Voice XML gateway port. After completing call setup, the Unified CVP B2BUA acts as an active intermediary for any subsequent call control. While the Unified CVP SIP signaling is routed through this service, this service does not touch the RTP traffic. Integrated into this B2BUA is the ability to interact with the Cisco Unified ICM through the ICM Service. This integration provides the ability for the SIP Service to query the Unified ICM for routing instruction and service control. This integration also allows Unified ICM to begin subsequent call control to do things such as transfers.
- **ICM service**: This service is responsible for all communication between Unified CVP components and Unified ICM. It sends and receives messages on behalf of the SIP Service and the IVR Service.

- **IVR service**: This service creates the Voice XML pages that implement the Unified CVP Micro applications based on RunExternalScript instructions received from Unified ICM. The IVR Service functions as the VRU leg (in Unified ICM Enterprise parlance). Calls must be transferred to it from the SIP Service in order to execute Micro applications. The Voice XML pages created by this module are sent to the Voice XML gateway to be executed.

Call Server can be deployed as part of the Enterprise Windows Domain.

For hardware details, see the latest version of the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* (formerly called the *Bill of Materials*), available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

VXML Server

The VXML Server executes advanced IVR applications by exchanging VoiceXML pages with the VoiceXML gateway's built-in voice browser. Like almost all other Unified CVP product components, it runs within a Java 2 Enterprise Edition (J2EE) application server environment such as Tomcat. You can add either custom-built or standard J2EE components to interact with back-end hosts and services. The VXML Server applications are written using Cisco Unified Call Studio and are deployed to the VXML Server for execution. The applications are invoked on an as-needed basis by a special Micro application which must be executed from within the Unified CCE routing script.

The VXML Server can also be deployed in a standalone configuration that does not include any Unified ICM components. Applications are invoked as a direct result of calls arriving in the VoiceXML gateway, and a single post application transfer is allowed.

The VXML Server can execute on Windows Server 2012 R2 Standard Edition. For hardware requirements and details, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal*(formerly called the *Bill of Materials*), available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html

For more information about the VXML Server and its latest features, see the *User Guide for Cisco Unified CVP VXML Server and Cisco Unified Call Studio*.

Media Server

The Media Server component is a simple web server which provides prerecorded audio files, external VoiceXML documents, or external ASR grammars to the VoiceXML gateway. Some of these files can be stored in the local file system on the gateways. However, in practice, most installations use a centralized media server to simplify distribution of prerecorded customer prompt updates. Media Server functionality can also include a caching engine. The gateways themselves, however, can also do prompt caching when configured for caching. Typical Media Servers used are Microsoft IIS and Apache, both of which are HTTP-based.

Note

- The Media Server component in Unified CVP is installed by default, along with Unified CVP Call Server and Unified CVP VXML Server.
- Unified CVP does not support the use of Tomcat as a Media Server.

Media Servers can be deployed as a simplex operation, as a redundant pair, or with supported load balancers in a farm. The VoiceXML Gateway caches.wav files it retrieves from the Media Server. In most deployments, the Media Server encounters low traffic from Unified CVP.

For the most current information on Media Servers, see the latest version of *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* (formerly called the *Bill of Materials*), available at

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html

Call Studio

Cisco Unified Call Studio is the service creation environment (script editor) for Unified CVP VXML Server applications. It is based on the open source Eclipse framework, which provides an advanced drag-and-drop graphical editing feature. Call Studio also provides options to insert vendor-supplied and custom-developed plug-ins that enable applications to interact with other services in the network. Call Studio basically is an offline tool. The only interaction with the Unified CVP VXML Server is to deliver compiled applications and plugged-in components for execution.

The Call Studio license is associated with the MAC address of the machine on which it is running. You typically designate one or more data center servers for that purpose. Cisco Unified Call Studio cannot run on machines.

For additional hardware details, see the latest version of the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* (formerly called the *Bill of Materials*), available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html

Reporting Server

The Reporting Server is a Windows server that hosts an IBM Informix Dynamic Server (IDS) database management system. The Reporting Server provides consolidated historical reporting for a distributed self-service deployment. The database schema is specified by the Unified CVP product, but the schema is fully published so that customers can develop custom reports based on it. The Reporting Server receives reporting data from the IVR Service, the SIP Service (if used), and the VXML Server. The Reporting Server depends on the Call Server to receive call records.

For standalone VXML Server deployments, one Call Server is needed per Reporting Server. The Reporting Server must be local to the Call Servers and VXML Servers. Deploying the Reporting Server at a remote location across the WAN is not supported. Use multiple Reporting Servers and place them at each site when Call Servers and VXML Servers exist at multiple locations.

The Reporting Server does not perform database administrative and maintenance activities such as backups or purging. However, Unified CVP provides access to such maintenance tasks through the Operations Console Server.

Operations Console

The Unified CVP Operations Console is a Windows server that provides a console for the browser-based administration and configuration for all Unified CVP components. It offers shortcuts to the administration and configuration interfaces of other Unified CVP solution components. The Operations Console is a required component in all Unified CVP deployments. The Operations Console must be run on a separate server from other Unified CVP devices. The Operations Console is, in effect, a dashboard from which an entire Unified CVP deployment can be managed.

The Operations Console must be configured with a map of the deployed solution network. It can then collect and maintain configuration information from each deployed component. You can use standard tools to back up both the network map and the configuration information that are stored locally on the server. The Operations Console provides options to display and modify the network map and stored configuration data and to distribute such modifications to the affected solution components.

The Operations Console can display two views of configuration parameters for managed components. The runtime view shows the status of all configuration parameters as the managed components use them. The configured or offline view shows the status of all configuration parameters that are stored in the Operations Server database and deployed to the device when you run the Save and Deploy option.

The Operations Console allows configuration parameters to be updated or preconfigured even when the target component is not online or running. If the target server (without its services) is online, you can apply the configured settings to that server. These settings become active when that server's services also come online, only then are reflected in the runtime view.

The Operations Console Server is not a redundant component. As such, you cannot duplicate the Operations Console Server within a deployment. It backs up the configuration database regularly, or whenever changes are made.

Additional Components

The following components are not part of the Unified CVP software, but you can use them with the CVP components for a complete solution:

- Cisco Ingress Voice Gateway
- Cisco VoiceXML Gateway
- Cisco Egress Gateway
- Video Endpoints
- Cisco Unified Communications Manager
- Cisco Unified Intelligence Center
- Cisco Unified Contact Center
- SIP Proxy Server
- DNS Server
- Load Balancers
- Cisco Unified Border Element

- Video Media Server
- Automatic Speech Recognition Server/Text-to-Speech Server
- Network Monitor

Cisco Ingress Voice Gateway

The Cisco Ingress Voice Gateway is the point at which an incoming call enters the Unified CVP system. It terminates time division multiplexing (TDM) calls on one side and implements VoIP on the other side. It serves as a pivot point for extension of calls from the TDM environment to VoIP endpoints. Hence, WAN bandwidth is conserved because no hairpinning of the media stream occurs. The Cisco Ingress Voice Gateway also provides for call switching capabilities at the command of other Unified CVP solution components.

Unified CVP Ingress Voice Gateways supports Session Initiation Protocol (SIP). Media Gateway Control Protocol (MGCP) voice gateways are supported if they are registered with Cisco Unified Communications Manager.

For the list of supported gateways, see ISO Gateway Choices, on page 85. For approved gateway and software combinations, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

VoiceXML Gateway

In this document, the term "VoiceXML Gateway" can either refer to Cisco Virtualized Voice Browser (Cisco VVB) or IOS VoiceXML Gateway that can be deployed on an Ingress Voice Gateway. VoiceXML Gateways are often deployed in farms for Centralized deployment models. VoiceXML Gateway interprets VoiceXML pages either from the Unified CVP Server IVR Service or the VXML Server.

Audio prompts retrieved from a third-party media server can be cached in VoiceXML Gateway to reduce WAN bandwidth and prevent poor voice quality. The VoiceXML document provides a pointer to the location of the audio file to be played or it provides the address of a text-to-speech (TTS) Server to stream the audio to the user. The VoiceXML Gateway interacts with automatic speech recognition (ASR) and TTS Servers through Media Resource Control Protocol (MRCP).

Cisco Virtualized Voice Browser

You can deploy Cisco VVB on a separate virtual machine. This model is suitable for both standalone, and comprehensive deployments. Cisco VVB communicates with ASR/TTS using MRCP.

For more information on Cisco VVB support, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Cisco IOS VoiceXML Gateway

You can deploy Cisco IOS VoiceXML Gateway on the same router as you deploy Unified CVP Ingress Voice Gateway. This model is suitable for deployments with small branch offices. The Cisco IOS VoiceXML Gateway can also run on a separate router platform. This model is suitable for deployments with large or multiple voice gateways, where only a small percentage of the traffic is for Unified CVP. Using this model, an organization can share public switched telephone network (PSTN) trunks between office users and contact center agents, and route calls based on the dialed number. VoiceXML Gateway can store audio files on flash memory or on a third-party media server.

For the supported Cisco IOS VoiceXML Gateways, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Unless a Cisco IOS VoiceXML Gateway is combined with an Ingress Voice Gateway, the Cisco IOS VoiceXML Gateway does not require TDM hardware. It interacts with VoIP on one side and HTTP (carrying VXML or .wav files) and MRCP (carrying ASR and TTS traffic) on the other side. As with Ingress Voice Gateways, Cisco IOS VoiceXML Gateways are often deployed in farms for Centralized deployment models, or one in each office in Branch deployments.

Cisco Egress Gateway

The Cisco Egress Voice Gateway is used only when calls need to be extended to TDM networks or equipment, such as a PSTN or a TDM automatic call distributor (ACD). While the Real-time Transport Protocol (RTP) stream runs between the ingress and egress voice gateway ports, the signaling stream logically goes through the Unified CVP Server and Cisco Unified Intelligent Contact Management (Unified ICM). This allows subsequent call control (such as transfers).

Cisco Integrated Services Router Generation 3 Gateway

Cisco's Generation 3 Integrated Services Router(ISR G3) 4xxx series gateway is certified with Cisco Unified Customer Voice Portal (CVP) version 10.0(1) or later. You can deploy the ISR G3 gateways with Unified CVP in a Cisco Unified Contact Center Enterprise (CCE) solution or in a standalone Unified CVP self-service deployment. For more information about 4xxx series gateway variants, see *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* available at, http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-technical-reference-list.html.

Note

- ISR G3 gateways do not have a built-in Voice-XML (VXML) browser. Therefore, deploying the ISR G3 ingress gateway with Unified CVP requires the use of a separate ISR G2 gateway to provide the VXML browser.
- Support of the ISR G3 gateways with Unified CCE Outbound Option with Call Progress Analysis requires Unified CCE version 10.5(2) or later.

Video Endpoints

Unified CVP Basic Video Service supports the following features:

- Cisco Unified Video Advantage
- Cisco TelePresence
- Video in Queue (VIQ)

Cisco Unified Communications Manager

Cisco Unified Communications Manager (Unified CM) is the main call processing component of a Cisco Unified Communications system. It manages and switches VoIP calls among IP phones. The Unified CM

combines with Cisco Unified Intelligent Contact Management (Unified ICM) to form Cisco Unified Contact Center Enterprise (Unified CCE). Unified CVP interacts primarily with Unified CM as a means for sending PSTN-originated calls to Unified CCE agents.

The following common scenarios require calls to Unified CVP to originate from Unified CM endpoints:

- An office worker (not an agent) on an IP phone dials an internal help desk number.
- An agent initiates a consultative transfer that gets routed to a Unified CVP queue point.
- A Cisco SCCP Dialer port transfers a live call to a Unified CVP port for an IVR campaign.

A single Unified CM can originate and receive calls from SIP. PSTN calls that arrived on MGCP Voice Gateways registered with Unified CM can be routed or transferred to Unified CVP only through SIP (and not through Cisco Unified Border Element).

Unified CM is an optional component in the Unified CVP solution. Its use in the solution depends on the type of call center being deployed. TDM-based call centers using ACDs, for example, typically do not use Unified CM (except when they are migrated to Cisco Unified CCE), nor do strictly self-service applications that use the Unified CVP Standalone self-service deployment model. Unified CM is used as part of the Cisco Unified CCE solution, in which call center agents are part of an IP solution that uses Cisco IP Phones, or when migrated from TDM ACDs.

Only specific versions of Unified CM are compatible with Unified CVP solutions. Unified CVP is supported with SIP only, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Cisco Unified Intelligence Center

Cisco Unified Intelligence Center (Unified Intelligence Center) is a web-based reporting application that provides real-time and historical reporting in an easy-to-use, wizard-based application for Unified CCE and Unified CVP. It allows Contact Center supervisors and business users to report on the details of every contact across all channels in the Contact Center from a single interface. Unified Intelligence Center allows you to extend the boundaries of traditional contact center reporting to an information portal where data can be easily integrated and shared throughout the organization.

Administrators can use Unified Intelligence Center to control access to features, reports, and data by granting privileges only to authorized individual users or groups of users. For example, you can assign each supervisor to a group of agents, skills, and call types that are the most relevant to them so that each report provides focused, actionable insights into data that is appropriate to their role.

Several features in this product allow you to extend the Unified Intelligence Center platform beyond traditional Contact Center reporting and into an enterprise-wide information portal. You can use data from nontraditional sources to improve business efficiency and effectiveness.

Cisco Unified Contact Center

Use either Cisco Unified Contact Center Enterprise (Unified CCE) or Cisco Unified Intelligent Contact Management (Unified ICM) when advanced call control, such as IP switching and transfers to agents, is required in Unified CVP. Unified ICM provides call center agent-management capabilities and call scripting capabilities. Variable storage capability and database access through the Unified CCE or Unified ICM application gateways are also powerful tools. A Unified CVP application can take advantage of these capabilities because Unified CVP applications can be called from within a Unified CCE or Unified ICM script in a non-standalone Unified CVP deployment model.

The Unified CVP Call Server maintains a GED-125 Service Control Interface connection to Unified CCE or Unified ICM. GED-125 is a third-party control protocol in which a single socket connection is used to control many telephone calls. For Unified CCE or Unified ICM, Call Server is a voice response unit (VRU) connected to Unified CCE or Unified ICM, as all other GED-125 VRUs are connected. Unified CVP is a VRU peripheral to Unified CCE or Unified ICM.

Note You can use the hosted versions of Unified CCE and Unified ICM for advanced call control.

SIP Proxy Server

SIP Proxy Server routes SIP messages among SIP endpoints. SIP Proxy Server is required for Unified CVP high-availability architecture for call switching. SIP Proxy Server is designed to support multiple SIP endpoints of various types and to implement load balancing and failover among these endpoints. Deployment of a SIP Proxy in the solution enables a more centralized configuration of the dial plan routing configuration.

You can configure a SIP Proxy with multiple static routes to do load balancing and failover with outbound calls. The static routes can point to an IP address or a DNS.

Domain Name System (DNS) Service Record (SRV) is also supported. However, it is not qualified for use on the Cisco Unified SIP Proxy (CUSP) Server. It is qualified for the devices that must reach the CUSP Server, such as Unified CVP, Ingress Voice Gateway, and Unified Communications Manager.

You can deploy Unified CVP without a SIP Proxy Server, depending on the design and complexity of the solution. In such cases, some of the functions that a SIP Proxy Server provides are provided by the Unified CVP Server SIP service.

Following are the benefits of using a SIP Proxy Server:

- You can use priority and weight routing with the routes for load balancing and failover.
- If a SIP Proxy Server exists in your SIP network, then Unified CVP acts as an additional SIP endpoint. The Unified CVP fits incrementally into the existing SIP network.

If you do not use a SIP Proxy Server, then the Ingress Voice Gateways and Unified CMs must point directly to Unified CVP. In such a deployment, perform the following tasks:

- Perform load balancing using DNS SRV lookups from gateway to DNS Server; balance SIP calls using this procedure.
- Perform load balancing of calls outbound from Unified CVP (outbound call leg) using DNS SRV lookups.
- Perform failover of SIP rejections (code 503 only), if you configure SRV records with ordered priorities.

DNS Server

You can install Domain Name System (DNS) Server anywhere in the network. Its purpose is to resolve hostnames to IP addresses. Unified CVP can make both Type A record lookups and SRV Type record lookups. If a DNS Server is slow to respond, is unavailable, or is across the WAN, the performance of Unified CVP is affected.

The DNS Server is used during SIP interactions in the following situations:

- When a call arrives at an Ingress Voice Gateway, the dial peer can use DNS to alternate calls between the two SIP Proxy Servers. The SIP Proxy Servers can also use DNS to distribute incoming calls among multiple SIP Services. If SIP Proxy Servers are not being used, then the Ingress Voice Gateway can use DNS directly to distribute inbound calls among multiple SIP Services.
- Unified CCE directs the SIP service to transfer calls to the VRU leg and can use DNS to alternate such requests between two SIP Proxy Servers. If SIP Proxy Servers are not being used, the SIP Service can use DNS directly to distribute VRU legs among multiple VoiceXML Gateways.
- When transferring a call to an agent using a SIP Proxy Server, the SIP Proxy cannot use DNS SRV for outbound calls; it must be configured with multiple static routes to do load balancing and failover. The static routes can point to an IP address or a regular DNS A host record. If SIP Proxy Servers are not being used, then the SIP Service can use DNS to locate the IP address of the target agent.

The use of the DNS Server for SIP routing is optional in Unified CVP. You do not need to have a dedicated DNS Server, as the existing DNS server handles the additional load of Unified CVP. For every call destined for Unified CVP that comes into the network, there are approximately three to four DNS lookups. You can determine the number of DNS queries per second by determining the number of calls per second for the solution, and multiplying that number by 4.

DNS lookups are needed for DNS SRV queries, not necessarily for A record queries, which can also be configured locally in the system file. You can use Unified CVP Server Groups alternately to avoid DNS SRV lookups.

Load Balancers

Application Control Engine

As a load-balancing device, ACE determines which server in a set of load-balanced servers receives the client request for service. Load balancing helps fulfill the client request in the shortest amount of time without overloading either the server or the server farm as a whole.

See *Cisco ACE 4700 Series Appliance Server Load-Balancing Configuration Guide* (http://www.cisco.com/en/US/docs/app_ntwk_services/data_center_app_services/ace_appliances/vA3_1_0/configuration/slb/guide/slbgd.html) to learn more about load-balancing with ACE.

To migrate from CSS to ACE, use the CSS2ACE Converter tool. See http://wwwin.cisco.com/dss/adbu/dcas/adoptions/cssmigration/ for more information.

To configure Unified CVP for ACE. See the *Configuration Guide for Cisco Unified Customer Voice Portal* available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/tsd_products_support_series_home.html.

You must have an ACE license to use ACE under load conditions. The minimum licensing requirements for ACE are:

- 1-Gbps throughput license (ACE-AP-01-LIC)
- A non-default SSL feature license, if you intend to use ACE for SSL
- Application Acceleration License (ACE-AP-OPT-LIC-K9), which allows more than 50 concurrent connections on ACE

See the *ACE product documentation* and *ACE release notes* for more licensing information.

Note There are two features for the VXML Server that assist with load balancing:

- Limiting load balancer involvement
- Enhanced HTTP probes for load balancers

See the configuration options ip_redirect and license_depletion_probe_error in the *User Guide for Cisco Unified CVP VXML Server and Cisco Unified Call Studio*, available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_user_guide_list.html.

Third-Party Load Balancers

Note The Unified CVP solution components recommend the use of load balancers to provide load distribution and high availability for HTTP, HTTPS, and MRCP traffic. Load Balancers provides the capability to load balance the rendering of the VXML pages between VoiceXML Gateways and VXML Server and also fetching of the media files for IVR scripts execution from media servers.

The Unified CVP now supports any third-party load balancer possessing the following features:

- Both SSL offloading and SSL pass-through has to be supported.
- Load balancer high availability.
- Session stickiness should not be mandatory.
- Persistence is cookie-insert.
- Distribution algo is round-robin.

The interoperability notes and the known caveats for most commonly used third-party load balancers, such as the Big-IP F5 and the Citrix NetScalar 1000v, can be referred from the following locations:

- For BIG-IP F5 refer to:

  http://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise/interoperability-portal/bigip.pdf

- For Citrix NetScalar 1000v refer to:

  http://www.cisco.com/c/en/us/solutions/enterprise/interoperability-portal/voice_portal.html

Cisco Unified Border Element

Cisco Unified Border Element is a Cisco IOS Session Border Controller (SBC) that interconnects independent Voice over IP (VoIP) and video over IP enterprise networks for data, voice, and video transport. SBCs are an industry class of devices that are critical components for scaling networks from VoIP islands within a single customer network to an end-to-end IP community, both inside the enterprise and to communicate beyond the enterprise across service provider networks or to other enterprises and small and medium businesses (SMBs).

Cisco Unified Border Element is an integrated Cisco IOS software application that runs on Cisco Integrated Services Routers (ISRs), Cisco universal gateways, and ASR routers. For details on supported versions, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Design Considerations

Please observe the following restrictions when deploying CUBE with SIP Trunks:

- Before 15.2(1)T IOS release, a dial-peer was required to pass the Refer-to header URI through CUBE. Starting from 15.2(1)T release onwards refer-to-passing command can be used without the need for a dial-peer.
- CUBE must be configured in media pass through mode in the Unified CVP deployment. Media flow around mode cannot be used because it is not supported or validated. Only media pass through mode, the default mode on the dial-peer, is supported for CUBE.
- CUBE does not currently support passing the Refer-To header URI designation from CVP when a REFER call flow is initiated. It rewrites the destination address based on the dial peer configuration. Therefore the dial plan must be configured on CVP and CUBE. The note below explains the behavior.
- REFER passthrough cannot be used in conjunction with Survivability. The script does not let REFER messages be relayed to a SIP service provider regardless of other CUBE configuration.
- REFER consume cannot be used in conjunction with Survivability and Router Requery. Survivability always accepts the REFER, even if the transfer does not complete. Unified CCE deems the transfer successful and does not attempt to requery.
- Survivability cannot be used when service provider Alternate Destination Routing (ADR) is used because the script does not let error messages (ring-no-answer or busy) reach the service provider. Manipulation in the script does not let error messages (ring-no-answer or busy) reach the service provider. Manipulation in the Remote-Party-ID header is required instead.
- If GTD is present on the incoming call or if Unified CCE sets a value for the UUI variable, Unified CVP will send a BYE immediately after outpulsing digits in a DTMF transfer. If a delay is required between the digits then comma should be used at the end of the label.
- If GTD is not present on the incoming call, Unified CCE does not set a value for the UUI variable and the service provider does not disconnect a call after receiving digits in a DTMF transfer. Unified CVP will send a BYE request after the SIP.ExternalTransferWait timer expires. Previous versions of Unified CVP did not disconnect the call.
- Survivability is required when Courtesy Callback is used.

Video Media Server

Video Media Server is required for uploading, storing, and playing back of video prompts. Cisco MediaSense is a Video Media Server that provides network-based multimedia capture, streaming, and recording. Cisco MediaSense records conversations on the network rather than on a device. This process simplifies the architecture, lowers costs, provides optimum scalability, and facilitates use by analytics applications from Cisco technology partners.

Automatic Speech Recognition Server and Text-to-Speech Server

Speech recognition servers can provide Automatic Speech Recognition (ASR), Text-to-Speech (TTS), and DTMF recognition services for a VoiceXML Gateway. Communication between the speech recognition servers and the VoiceXML Gateway uses Media Resource Control Protocol (MRCP).

The following table describes support for the MRCP protocol on IOS VXML Gateway and Cisco VVB:

Application	Supports MRCP v1	Supports MRCP v2
Micro-Apps with speech server	IOS VXML and VVB	VVB
Studio applications with speech server	IOS VXML and VVB	OS VXML and VVB

For capacity and redundancy reasons, an Application Control Engine (ACE) is used to mediate between a VoiceXML Gateway and a farm of ASR or TTS servers. If ACE is not used, then a VoiceXML Gateway can support a maximum of two ASR or TTS Servers.

Cisco does not sell or support any ASR or TTS software or servers. However, Cisco tests Unified CVP with Nuance products. A certification process is currently being developed to allow additional vendors to qualify their products against Unified CVP VoiceXML, and the World Wide Web Consortium (W3C) provides a rich feature set to support the ASR grammars. It is easy to implement and support inline grammars, by which the set of acceptable customer responses is passed to the VoiceXML Gateway. Another form is external grammars, where Unified ICM passes a pointer to an external grammar source. The VXML Server adds this pointer to the VoiceXML document that it sends to the VoiceXML Gateway, which then loads the grammar and uses it to check ASR input from the caller. In this case, the customer creates the grammar file. A third type of grammar is the built-in grammar. For a complete explanation of grammar formats, see the W3C website at http://www.w3.org/TR/speech-grammar/.

The text for TTS is passed directly from the VXML Server to the VoiceXML gateway. This action is referred to as inline TTS in this document.

The actual speech recognition and speech synthesis are performed by a separate server that communicates with the VoiceXML Gateway through MRCP. Currently, Nuance is the only tested ASR and TTS engine. The ASR and TTS engine also supports (with limitations) voice recognition and synthesis for multiple languages.

For information on Nuance, see http://www.nuance.com.

Nuance is a third-party product, which the customer or partner must purchase directly from the vendor. The customer also receives technical support directly from the vendor. However, that does not mean that the vendor's latest software version can be used. Unified CVP is tested with specific versions of each vendor's product. Cisco Technical Assistance Center (TAC) does not support Unified CVP customers who use different ASR and TTS versions apart from those which have been tested. For details on the supported ASR and TTS products, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Network Monitor

You can use Simple Network Management Protocol (SNMP) station to monitor the solution deployment status. See the *Operations Console User Guide for Cisco Unified Customer Voice Portal*, available at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-user-guide-list.html.

Call Flows

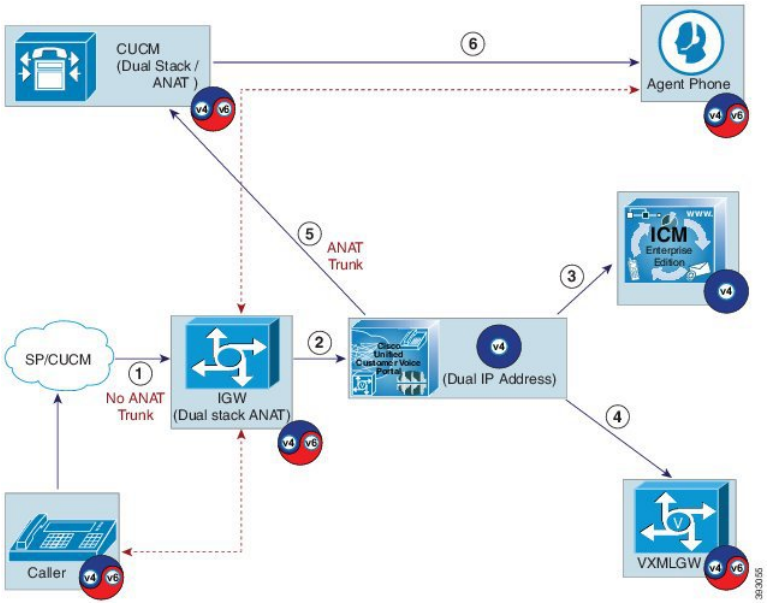
IPv6 Architecture for Unified CVP

Beginning with Unified CVP Release 11.0(1), you can use either IPv6-only or a mix of IPv4 and IPv6 endpoints. Servers that communicate with those endpoints can now accept IPv6 connections, in addition to IPv4 connections. Communication between servers continues to use IPv4 connections.

Note Ensure that you use a separate Ingress Gateway and VXML Gateway for both the Standalone and Comprehensive call flow models.

The following figure displays the Comprehensive call flow model for IPV6. The solid lines indicate the signal paths and dashed lines indicate the media path.

Figure 2: Comprehensive call flow model for IPv6



A typical Comprehensive call flow for IPV6 is as follows:

1. A customer call arrives from the service provider SIP trunk or internal CUCM trunk to Ingress Gateway (either IPv4 or IPv6 trunk).
2. The Ingress Gateway sends the call to Unified CVP.

Note The Ingress Gateway to CVP connection is on IPV4 with Alternative Network Address Types (ANAT) enabled.

3. Unified CVP sends the incoming call request to ICM in IPv4 and gets a VRU label.

4. Unified CVP sends the call to the VoiceXML Gateway. The caller hears the IVR.
5. Unified CVP sends the call to an agent with IPv4 signal and IPv4 and IPv6 media.

Note The Unified CVP to Unified Communications Manager trunk is dual stack with ANAT enabled.

6. Unified CM Communications Manager connects the caller to an available agent. The agent can be either on IPv4 or IPv6 phone configuration.

Note Cisco VVB does not support IPv6.

For information about IPv6 configuration for CVP, see the *Configuration Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-installation-and-configuration-guides-list.html.

IPv6 Design Considerations

Following are the design considerations for IPv6:

- For Call Survivability function: The incoming trunk to Gateway must be configured as IPv4 only.
- For Courtesy Callback and Refer feature: If the same incoming trunk is used for outbound dialing, then the session target in Ingress Gateway dialpeer must be same (either IPv4 or IPv6) as the incoming trunk.

Video Endpoints Design Considerations

Following are the design considerations for Video Endpoints:

- Configure the incoming trunk to Gateway in IPv4 mode only.
- Disable ANAT in the Ingress Gateway.
- Configure the Agent endpoint either in IPv4 or dual IP mode only.

Typical SIP Unified CVP Call Flow

The description follows a typical SIP Unified CVP Call Flow (comprehensive call flow model). However, it is an illustration and is not an actual solution. This can only be considered as an introduction to the overall flow of information in a Unified CVP solution.

The call flow consists of an incoming call requiring initial self-service, followed by a queue treatment, and finally delivery to a Unified ICM agent. The following diagram presents a general SIP-based solution.

The following is a typical SIP Unified CVP call flow:

1. The call arrives at an Ingress Voice Gateway and sends an invite message to the SIP Proxy Server that forwards the message to the SIP Service.
2. The Proxy Server determines the IP address of the Unified CVP Server for the dialed number and forwards the invite to the selected Unified CVP Server SIP service.

3. The SIP service consults Unified ICM through the Unified CVP Server ICM Service, which causes Unified ICM to run a routing script.
4. The routing script typically initiates a transfer of the call to a VoiceXML Gateway port through the SIP service.
5. The VoiceXML Gateway sends a message to the IVR service, which requests scripted instructions from Unified ICM.
6. Unified ICM exchanges VRU instructions with the VoiceXML gateway through the IVR service. The instructions can include requests to invoke more sophisticated applications on the Unified CVP VXML server. These requests result in multiple exchanges between the Unified CVP VXML Server and the VoiceXML Gateway to provide self-service.
7. To transfer to a live agent, the Unified ICM routing script queues the caller for an available agent. While waiting for an available agent, Unified ICM provides additional instructions to the VoiceXML Gateway to provide queueing treatment to the caller.
8. When an agent becomes available, Unified ICM sends a message to the Unified CVP Server SIP Service, which forwards a message through the SIP Proxy Server to the Ingress Gateway and to the Unified CM to transfer the call away from the VoiceXML Gateway port and deliver it to the Unified CM agent IP phone.

Design Process

When designing a Unified CVP deployment consider the following high-level steps:

1. Choose a call flow model for your functional deployment.
2. Determine where the Unified CVP components are going to be deployed (in the data center or at a branch).
3. Choose the amount of availability and resiliency that is justifiable or required.
4. Size the deployment to provide the justifiable or required grade of service for the initial deployment and near-term growth.

Note SIP is the only supported call control signaling protocol.

Call Flow Models

The first step in the design process is to determine the functionality you need. Unified CVP offers a number of call flow models to support differing needs. The deployment model you choose depends on the call flow preferences, geographic distribution requirements, and hardware configurations that best satisfy your company's needs.

- Unified CVP VXML Server (standalone) — Provides a standalone VRU with no integration to Unified ICM for queuing control or subsequent call control. Used to deploy self-service VXML applications.
- Call Director — Provides IP switching services only.

  This model is useful if you want to:

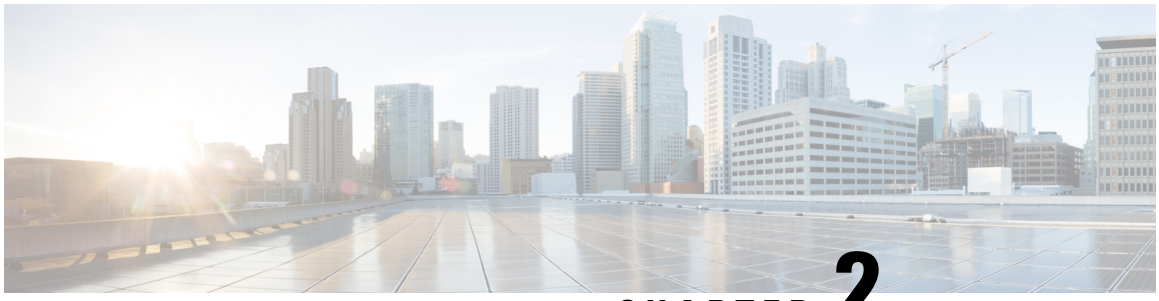
- Only use Unified CVP to provide Unified ICM with VoIP call switching.
- Prompt and collect data using third-party VRUs and ACDs.
- Avoid using an Unified CVP VXML Server.

- Comprehensive — Provides IVR services, queue treatment, and IP switching services. This model is useful if you want to,
  - Use Unified CVP to provide Unified ICM with VoIP call switching capabilities.
  - Use Unified CVP to provide Unified ICM with VRU services, including integrated self-service applications, queuing, and initial prompt and collect.
  - Use the video IVR, video queuing, and video agent capabilities.
  - Use an optional Unified CVP VXML Server.
  - Prompt or collect data using optional ASR and TTS services.

- VRU Only — Provides IVR services, queuing treatment, and switching for PSTN endpoints. This model relies on the PSTN to transfer calls between call termination endpoints.

  This model is useful if you want to:
  - Use Unified CVP to provide Unified ICM with VRU services including integrated self-service applications and initial prompt and collect.
  - Avoid using an Unified CVP for switching calls.
  - Use an optional Unified CVP VXML Server.
  - Prompt or collect data using optional ASR and TTS services.

For more details and design considerations for each of these functional deployment models, see Functional Deployment, on page 19 for details.



CHAPTER 2

Functional Deployment

Overview

This chapter describes the functional deployment models with the customer requirements, the required and optional components, and a step-by-step call flow. To deploy each model, all of the components should be located in a single site.

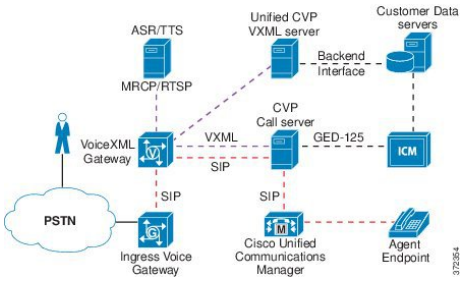
Note

- This chapter does not cover failover in high availability. For high-availability deployment options, see Unified CVP Design for High Availability, on page 93.
- For information on distributed deployment scenarios, see Distributed Deployment, on page 31.

VXML Server (Standalone)

The VXML Server (standalone) functional deployment model provides organizations with a standalone IVR solution for automated self-service. Callers can access Unified CVP from a local, long-distance, or toll-free numbers that terminate at Unified CVP Ingress Voice Gateways. Callers can also access Unified CVP from VoIP endpoints number that terminates at Unified CVP Ingress Voice Gateways. Callers can also access Unified CVP from VoIP endpoints. The following figure illustrates this model.

Figure 3: VXML Server (Standalone) Functional Deployment Model



This model includes the following optional components:

- Ingress Voice Gateway
- VoiceXML Gateway
- VXML Server
- Unified Call Studio
- Operations Console Server

Following are the optional components of this model:

- Automatic Speech Recognition/Text-to-Speech (ASR/TTS) Server
- Third-Party Media Server
- Application Control Engine (ACE)
- Egress Voice Gateway
- Reporting Server

Protocol-Level Call Flow

1. A call arrives at the Ingress VoiceXML Gateway through TDM or SIP. The gateway performs inbound Plain Old Telephone Service (POTS) or VoIP dial-peer matching.

Note Ingress Voice Gateway and VoiceXML Gateway deployment can either be separate entities or co-located (using IOS VoiceXML Gateway). The self-service application is invoked on VoiceXML Gateway.

2. The selected VoiceXML Gateway port invokes the Unified CVP self-service script.
3. The self-service script invokes the Unified CVP standalone bootstrap VoiceXML document, which sends an HTTP request to the configured IP address of the VXML Server.
4. The VXML service function, which resides in the CVP Server, runs the application that is specified in the HTTP URL and returns a dynamically generated VoiceXML document to the VoiceXML Gateway. The Unified CVP VXML Service may access back-end systems to incorporate personalized data into the VoiceXML document that is sent to the VoiceXML gateway.

5. The VoiceXML Gateway parses and renders the VoiceXML document. For spoken output, the VoiceXML Gateway either retrieves and plays back prerecorded audio files that are referenced in the VoiceXML document, or it streams media from a text-to-speech (TTS) server. DTMF is detected on VoiceXML Gateway.
6. The VoiceXML Gateway submits an HTTP request containing the results of the caller input to the VXML Server. This server runs the application that is specified in the HTTP URL again and returns a dynamically generated VoiceXML document to the VoiceXML Gateway. The dialog continues through repetition of Steps 5 and 6.
7. The IVR dialog ends when the caller hangs up, the application releases, or the application initiates a transfer.

Transfers and Subsequent Call Control

You can use the VXML Server (Standalone) functional deployment to transfer callers to another endpoint. Transfers to either VoIP (for example, Cisco Unified Communications Manager) or TDM (for example, Egress Voice Gateway to PSTN or TDM ACD). IVR application data cannot be passed to the new endpoint with this deployment model. If the endpoint is a TDM ACD, the agent screen popup window does not appear.

This model supports the following call transfers:

- VoiceXML Bridged Transfer
- VoiceXML Blind Transfer
- Release Trunk Transfer (TNT, hookflash, TBCT, and SIP Refer)

Note Cisco VVB supports only Blind Transfer feature.

For information about Call Transfer, see Call Transfer Options, on page 171 and VoiceXML Transfer, on page 176.

Call Director

The Call Director functional deployment model provides an organization with a method to route and transfer calls across a VoIP network. For example, organizations can use this model with multiple TDM ACD and TDM IVR locations that are integrated with Unified ICM through an ACD or IVR Peripheral Gateway. The organization wants to use the Cisco Unified Intelligent Contact Management (Unified ICM) to route and transfer calls intelligently across these locations without using PSTN prerouting or Release Trunk Transfer services. In this model, Unified CVP and Unified ICM can also pass call data between these ACD and IVR locations. In this model, Unified ICM can also provide beginning-to-end reporting for all calls.

The Call Director model is often the first step in the migration from a TDM-based contact center to a VoIP-based contact center. When an organization is ready to implement CVP-based IVR services and Cisco Unified Contact Center Enterprise (Unified CCE), the organization can migrate their Unified CVP deployment to the Comprehensive functional deployment model.

Callers can access Unified CVP from local, long-distance, or toll-free numbers that terminate at Unified CVP Ingress Voice Gateways. Callers can also access Unified CVP from VoIP endpoints.

Following are the required components of this model:

- Ingress Voice Gateway
- Egress Voice Gateway
- CVP Server
- Operations Console Server
- Cisco Unified ICM Enterprise
- SIP Proxy Server (for SIP deployments)

The reporting server is an optional component of this model because there is very little call information stored in the Unified CVP reporting database.

SIP-Level Call Flow

VoIP-Based Prerouting

1. A call arrives at the Ingress Voice Gateway and sends a SIP INVITE/SIP/SIP message to the SIP Proxy Server, which forwards the request to the CVP Server SIP Service.
2. The SIP Service sends a route request to Unified ICM using the CVP Server ICM Service and the VRU Peripheral Gateway. This route request invokes Unified ICM to run a routing script based on the dialed number and other criteria.
3. The Unified ICM routing script selects a target and returns a translation route label to the CVP Server SIP Service. The Server SIP Service then signals through the SIP Proxy Server to the Egress Voice Gateway (which connects to the TDM termination point) and the Ingress Voice Gateway to enable the call to be set up between the Ingress and Egress Voice Gateways. While the RTP stream flows directly between the Ingress and Egress Voice Gateways, call control signaling flows through Unified CVP to allow subsequent call control.
4. When the call arrives at the selected termination, the termination equipment sends a request to its Peripheral Gateway for routing instructions. This step resolves the translation route and allows any call data from the previous Unified ICM script to be passed to the selected termination. If the selected termination is a TDM IVR platform, self-service is provided and the caller can either release or request to be transferred to a live agent. If the selected termination is a TDM ACD platform, then the caller is queued until an available agent is selected by the TDM ACD. Call data can then be displayed on the agent screen. After receiving live assistance, the caller can either release or request to be transferred to another agent.

VoIP-Based Transfer

1. Regardless of whether the call is initially routed to a TDM IVR or ACD location, a caller can request a call to be transferred to another location. When the transfer occurs, the TDM IVR or ACD sends a postroute request with call data (with its Peripheral Gateway) to Unified ICM.
2. When Unified ICM receives this postroute request, it runs a routing script based on the transfer dialed number and other criteria. The Unified ICM routing script selects a new target for the call and then signals to the CVP Server SIP Service to release the call leg to the originally selected termination and to extend the call to a new termination.
3. When the call arrives at the new termination, the termination equipment sends a request to its PG for routing instructions. This step resolves a translation route that is allocated for this call to this new

termination location, and it allows any call data from the previous location (IVR port or agent) to be passed to the new termination. Calls can continue to be transferred between locations using this same VoIP-based transfer call flow.

Transfers and Subsequent Call Control

In addition to transfers that are manager by Unified ICM, the Comprehensive deployment model can transfer calls to non-ICM terminations or invoke a Release Trunk Transfer in the PSTN. If a call is transferred to a non-ICM termination, no call data can be passed to the termination and no further call control is possible for that call. The call reporting that Unified ICM captures is completed. In the case of a Release Trunk Transfer, the Ingress Voice Gateway port is released, no call data can be passed to the termination, and no further call control is possible for that call. If the Release Trunk Transfer is translation routed to another ICM peripheral, call data and reporting can be maintained. For more information about Call Transfer, see Call Transfer Options, on page 171.

Call Server cancels the transfer request and sends a transfer failure indication to Unified ICM. Following are the reasons for it:

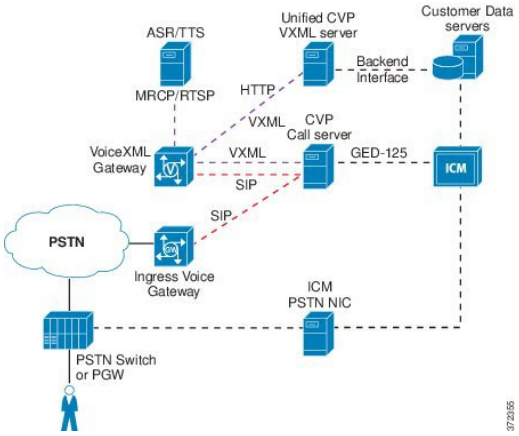
- A selected termination (for either a new or transferred call) returns a connection failure or busy status.
- The destination phone rings until it exceeds the ring-no-answer (RNA) timeout setting of Call Server.

These scenario causes a Router Requery operation. The Unified ICM routing script recovers control and selects a different target or takes other remedial action.

Comprehensive

The Comprehensive functional deployment model provides organizations with a method to route and transfer calls across a VoIP network, to offer IVR services, and to queue calls before they are routed to a selected agent. The most common usage scenario for this functional deployment model is when organizations want a pure IP-based contact center. Callers are provided IVR services initially and then upon request are provided queue treatment and are transferred to a selected Unified CCE agent. On request callers can also be transferred between Unified CCE agents. In this model, Unified CVP and Unified ICM pass call data between these endpoints and provide reporting for all calls. The following figure illustrates this model.

Figure 4: Comprehensive Functional Deployment Model



This model has the following features:

- Allows callers to access Unified CVP through local, long distance, or toll-free numbers terminating at the Unified CVP ingress voice gateways, and from VoIP endpoints.
- Provides the capabilities of the VXML Server (Standalone) and Call Director functional deployment models
- Provides the ability to route and queue calls to UCCE agents.
- Can utilize SIP.

Following are the required components of this model:

- Ingress Voice Gateway
- VoiceXML Gateway
- CVP Server
- Unified Call Studio
- SIP Proxy Server (for SIP deployments)

Following are the optional components of this model:

- Automatic Speech Recognition/Text-to-Speech (ASR/TTS) Server
- Third-Party Media Server
- Supported Load Balancers
- Egress Voice Gateway
- Reporting Server

SIP-Level Call Flow

Initial Call Treatment and Self-Service

1. A call arrives at the Ingress Voice Gateway and sends a SIP invite message to the SIP Proxy Server, which forwards the request to the Unified CVP Server SIP service.
2. The SIP service sends a route request to Unified ICM through the Unified CVP Server ICM service and the VRU Peripheral Gateway. This route request invokes Cisco Unified ICM to run a routing script based on the dialed number and other criteria.
3. The Unified ICM routing script utilizes a Send to VRU node to return a label to the SIP service and send a call to a VoiceXML Gateway. The Unified CVP Server SIP service sends an invite message to the VoiceXML Gateway using the SIP Proxy Server, which translates the label DN to the IP address of the VoiceXML Gateway.
4. The Voice XML Gateway sends an HTTP new call message to the VXML Server IVR Service with the label DN provided by Unified ICM. The IVR service then sends a route request message to Unified ICM (using the Unified ICM service), which then allows Unified ICM to re-enter the previously started routing script. You should reenter the routing script at the successful exit path of the Send to VRU node. The Unified ICM routing script uses Run Script nodes to instruct the IVR service about the desired call treatment. If call treatment requires complex IVR self-services, service control can be redirected to a VXML Server application. Upon completion of the VXML Server application or a request by the caller to transfer to a live agent, service control is returned to the VXML Server IVR service. The initial call

treatment is simple by using few prompts, then the IVR service can utilize Unified CVP microapplications to generate VoiceXML documents for the VoiceXML Gateway. VXML Server is not required.

Caller Requests to Transfer to Live Agent

1. When the caller requests to transfer to a live agent, the Unified ICM routing script queues the caller for an appropriate skill group and sends RunExternalScript messages to the IVR Service to have queue treatment provided (assuming that no agent is available).
2. When a Unified CCE agent becomes available, Unified ICM requests the Unified CVP Server SIP service to transfer the call to selected agent.
3. The SIP service transfers the caller to the dialed number of the selected agent. The SIP service sends a SIP INVITE/SIP message to the SIP Proxy Server, which finds the Unified CM SIP Trunk IP address associated with this agent DN and then forwards the SIP INVITE/SIP message to Unified Communications Manager.
4. Unified Communications Manager accepts the incoming SIP Trunk call and routes it to the selected agent.

Caller Requests to be Transferred to a Second Skill Group

1. If the caller requests to be transferred to a second agent, then the first agent initiates a transfer from the Unified CCE agent desktop application. This action generates a route request from the agent Peripheral Gateway to the Unified ICM central controller. Unified ICM executes a routing script that queues the call to another skill group. Assuming that no agent is available, the Unified ICM script uses the Send to VRU node, which signals to the SIP service to release the call leg to the Unified CM SIP Trunk and connect the call back to a VoiceXML Gateway.
2. The VoiceXML Gateway sends an HTTP new call request to the VXML server, which forwards that request to Unified ICM to allow the routing script to be reentered at the exit of the Send to VRU node. The Unified ICM sends RunExternalScript messages to the VXML server to allow queue treatment to be provided to the caller while the caller waits for a second agent.
3. When a second Unified CCE agent becomes available, Unified ICM requests the Unified CVP Server SIP service to transfer the call to the selected agent.
4. The SIP service transfers the caller to the dialed number of the selected agent. The SIP Service sends a SIP INVITE/SIP message to the SIP Proxy Server, which finds the Unified CM SIP Trunk IP address associated with the second agent DN and then forwards the SIP INVITE/SIP message to Unified Communication Manager.
5. Unified Communication Manager accepts the incoming SIP trunk call and routes it to the second agent.

Transfers and Subsequent Call Control

In addition to the transfers managed by Unified ICM, the Call Director deployment model can transfer calls to non-ICM terminations or invoke a Release Trunk Transfer in the PSTN. If a call is transferred to a non-ICM termination, then no call data can be passed to the termination; no further call control is possible for that call; and the cradle-to-grave call reporting, that Unified ICM captures, is completed. In the case of a Release Trunk Transfer, the Ingress Voice Gateway port is released; no call data can be passed to the termination; and no further call control is possible for that call. If the Release Trunk Transfer is translation-routed to another ICM peripheral, call data and cradle-to-grave reporting can be maintained. For information on call transfers, see Call Transfer Options, on page 171.

Call Server cancels the transfer request and sends a transfer failure indication to Unified ICM. Following are the reasons for it:

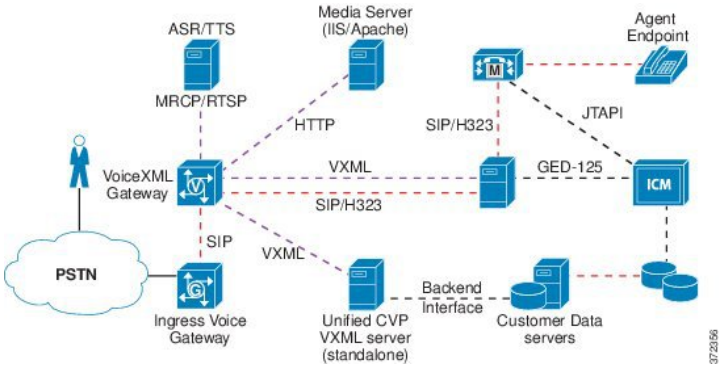
- A selected termination (for either a new or transferred call) returns a connection failure or busy status.
- The destination phone rings until it exceeds the ring-no-answer (RNA) timeout setting of Call Server.

These scenario invokes the Router Requery operation. The Unified ICM routing script then recovers control and selects a different target or takes a remedial action.

VRU-Only

The VRU-Only functional deployment model provides self-service applications and queueing treatment for organizations that use advanced PSTN switching services that are controlled using a Cisco Unified ICM PSTN Network Interface Controller (NIC). Two Unified ICM PSTN NICs allow subsequent call control of calls in the PSTN: the NIC and the Carrier Routing Service Protocol (CRSP) NIC. These NICs allow Unified ICM to route calls intelligently to Unified ICM peripherals, such as ACDs and IVRs. They also allow Unified ICM to invoke mid-call transfers in the PSTN. The following figure illustrates this model.

Figure 5: VRU-Only Functional Deployment Model



In the VRU-Only functional deployment model:

- Unified ICM routes a call before it is routed to another calls to Ingress Voice Gateway for call treatment and queuing. When an agent becomes available, Unified ICM instructs the PSTN to transfer the call to that agent. The agents can be Cisco Unified Contact Center Enterprise agents, Cisco Unified Contact Center Express agents, or ACD agents. If necessary, Unified ICM can request the PSTN (using the NIC) to transfer the call, just as Unified ICM can request Unified CVP to transfer the call.
- Ingress Voice Gateway is a Unified ICM-managed PSTN termination point that provides VRU services using a VoiceXML Gateway, the VXML server, the ICM Service, and Unified ICM.
- The SIP Service is not used for call control. All call control and switching is controlled by Unified ICM and the PSTN.
- Unified ICM can pass call data between these termination points (for a pop-up window or other intelligent treatment) and provide reporting for all calls.

Following are the required components of this model:

- Ingress Voice Gateway
- VoiceXML Gateway

- CVP Server
- Unified Call Studio
- Cisco Unified ICM Enterprise and NIC (CRSP)

Following are the optional components of this model:

- ASR/TTS Server
- Third-Party Media Server
- Application Control Engine (ACE)
- SIP Proxy Server (for SIP deployments)
- Reporting Server

Protocol Call Flows

The following are the protocol-level call flows for calls originated by Unified CM in each of the deployment models:

Model #1: Standalone Self-Service, on page 41

Model #2: Call Director, on page 41

Model #3a: Comprehensive Using ICM Micro-apps , on page 42

Model #3b: Comprehensive Using VXML Server, on page 43

Note Model #4, VRU Only with NIC Controlled Routing, is not discussed here because no NIC is involved with calls originated by Unified CM.

Video

The Video service is an extension of the Comprehensive deployment model that allows for a video caller to interact with a video agent. IVR and queuing are audio-only.

The following video endpoints are supported when using the Unified CVP Video:

- Cisco Unified Video Advantage
- Cisco TelePresence

The Video service supports the following call flows:

- A TelePresence caller dials into Unified CVP, receives audio-only IVR queuing treatment, and then is transferred to an agent on a second TelePresence unit.
- The TelePresence agent can conference in a second agent on an audio-only IP phone by dialing a direct extension from the TelePresence phone.

- The TelePresence agent can conference in a Unified CVP dialed number that results in audio queuing, followed by connecting to a second agent on an audio-only IP phone.
- A TelePresence caller dials into Unified CVP, receives audio-only IVR queuing treatment, and then is transferred to an agent on an audio-only IP phone. Enable Media Transfer Protocol (MTP) on the SIP trunk to listen to audio both ways.

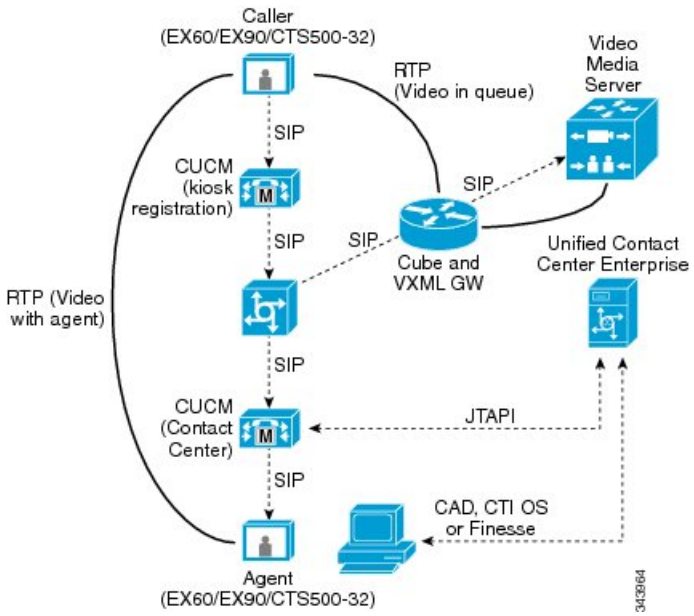
Note Because Video is an extension of the SIP-based Comprehensive deployment model, the required components and SIP protocol-level call flow details remain same. See Comprehensive, on page 23 for details.

Video in Queue

Video in Queue (VIQ) is an optional Basic Video feature in Unified CVP. It allows the caller to interact through high-definition video prompt or navigate a video menu using DTMF keys. The following figure displays the topology and call flow for Basic Video.

Note Cisco VVB does not support Video in Queue.

Figure 6: Video in Queue



The Unified CVP Studio VideoConnect element allows the specific video prompt to be played for video endpoints. It also allows the DTMF input during video-prompt playback to be collected and integrated with the Unified Call Studio or Unified CCE scripting environment.

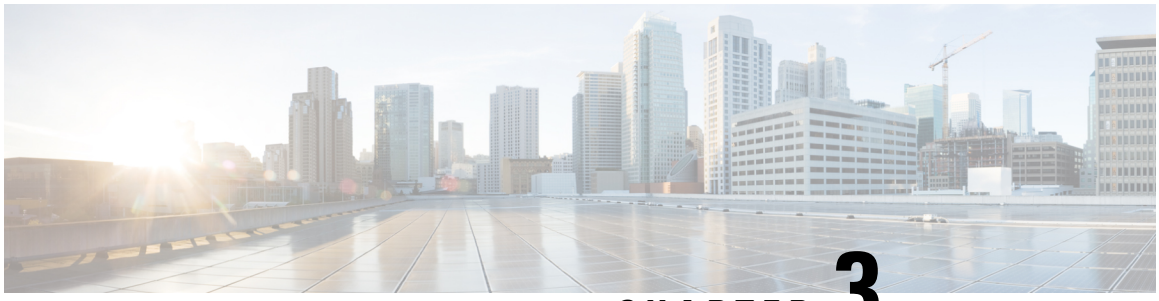
Note Video in Queue is not played during a CUCM failover.

See the *Configuration Guide for Cisco Unified Customer Voice Portal* for specific Cisco Unified Border Element or VXML Gateway configuration information for VideoConnect.

See the *Element Specifications for Cisco Unified CVP VXML Server and Cisco Unified Call Studio* for using the VideoConnect element.

See "Incoming Call Configuration and Media File Management" in the MediaSense User Guide to use media files.

Note When configuring the Video in Queue for Unified CVP, set the MediaSense **Incoming Call Configuration** > **Action** to play once.



CHAPTER 3

Distributed Deployment

Distributed Gateways

Unified CVP can use different types of gateways depending on the deployment model. This section discusses each type of voice gateway and their effects in a distributed deployment.

Ingress or Egress Voice Gateway at the Branch

In this deployment model, Ingress Voice Gateways located at a branch office are typically used to provide callers with access using local phone numbers instead of using centralized or non-geographic numbers. This capability is important in international deployments spanning multiple countries. Egress Gateways are located at branches either for localized PSTN breakout or for integration of decentralized TDM platforms into the Unified CVP switching solution. All other Unified CVP components are centrally located, and WAN links provide data connectivity from each branch location to the central data center.

Ingress or VoiceXML Gateway at the Branch

Consider other voice services that run at the branch that can affect Ingress or VoiceXML Gateways. For example, the branch is a remote Cisco Unified Communications Manager (Unified CM) site supporting both ACD (Agent Desktop provides call control capabilities ready/not ready, wrap up.) agent and non-agent phones. In this model, the PSTN gateway is used for ingress of Unified CVP calls as well as Ingress and Egress of normal user calls. In circumstances when the VoiceXML and Voice Gateway functions reside at the same branch location but on separate devices, special attention has to be given to the dial plan to ensure that the VRU leg is sent to the local VoiceXML resource. This is because the Unified CVP Call Server **settransferlabel** label applies only to coresident VoiceXML and Voice Gateway configurations.

When the Ingress Voice Gateway and the VoiceXML Gateway at a branch do not reside on the same Gateway, there are two ways to ensure that the calls are handled within the branch and not sent through the WAN to a different VoiceXML Gateway:

- Configure Unified ICM with multiple customers, one Unified ICM configuration.

  The Unified ICM configuration differentiates between calls based on the Dialed Number. The Dialed Number is associated with a customer representing the branch site. When a NetworkVRU is needed, the NetworkVRU associated with the customer in Unified ICM is selected and the caller is sent to that NetworkVRU. This method allows you to have multiple NetworkVRUs, each with a unique label. The disadvantage of this method is that each NetworkVRU requires its own VRU scripts in Unified ICM.

- Configure Unified CVP using the SigDigits feature.

  The SigDigits feature allows you to use the dial plan on the SIP Proxy to route calls to the correct site. When the call arrives at an Ingress Voice Gateway, the gateway prepends digits before sending the call to Unified CVP. Those prepended digits are unique to that site for a dial plan.

  When the call arrives at Unified CVP, Unified CVP strips the prepended digits and stores them in memory, resulting in the original DID on which the call arrived. Unified CVP then notifies Unified ICM of the call arrival using the original DID and matches a Dialed Number in Unified ICM.

  When Unified ICM returns a label to Unified CVP to transfer the call to a VoiceXML gateway for IVR treatment or to transfer the call to an agent phone, Unified CVP prepends the digits that it stored in memory before initiating the transfer. The dial plan in the SIP Proxy must be configured with the prepended digits to ensure that the calls with a certain prepended digit string are sent to specific VoiceXML Gateways or Egress Gateways.

  When the VoiceXML Gateway receives the call, the CVP bootstrap service is configured to strip the digits again, so that when the IVR leg of the call is set up, the original DN is used on the incoming VoiceXML request.

  Note The digits can be prepended to translation route DNs, and that the egress or receiving component (such as Unified CM) may need to strip digits to see the original DN.

  The term SigDigits is used to describe this feature because the command in Unified CVP to turn on the feature and specify how many significant digits should be stripped is called Prepend Digits for SIP in the operations console.

  This method is preferred because it involves the least amount of Unified ICM configuration overhead: a single NetworkVRU and single set of VRU scripts and Unified ICM routing scripts. This allows all of the Unified CVP Servers and VoiceXML Gateways to function as a single network-wide virtual IVR from the perspective of Unified ICM.

  The SigDigits feature can also be used to solve multicluster call admission control problems. (See Call Admission Control Considerations , on page 35, for more information.)

Colocated VXML Servers and VoiceXML Gateways

Either all gateways and servers are centralized or each site has its own set of colocated Unified CVP VXML Servers and VoiceXML Gateways.

Colocation has the following advantages:

- A WAN outage does not impact self-service applications.
- No WAN bandwidth is required for VoiceXML.

Colocation has the following disadvantages:

- Extra Unified CVP VXML Servers are required when using replicated branch offices.
- Additional overhead is required when deploying applications to multiple Unified CVP VXML Servers.

Gateways at Branch with Centralized VXML Server

Advantages of centralized VoiceXML:

- Administration and reporting are centralized.
- Unified CVP VXML Server capacity can be shared among branch offices.

Disadvantages of centralized VoiceXML:

- Branch survivability is limited.
- WAN bandwidth must be sized for additional VoiceXML over HTTP traffic.

Cisco Unified Communications Manager

In a Unified CVP environment, Unified CM can be an Ingress or Egress Gateway. It is more common for Unified CM to be an Egress Gateway because calls typically are from the PSTN, queued by Unified CVP, and then switched to Unified CM for handling by an agent. If the call is not from the PSTN, but from an IP phone, the Unified CM is an Ingress Voice Gateway from the perspective of Unified CVP.

Unified CM as an Egress Gateway

To deploy Unified CM with Unified CVP, you must use Unified CM call admission control for calls between the Ingress Voice Gateway and the agent IP phone. Therefore, Unified CM recognizes the call coming from the centralized Unified CVP Call Server instead of from the Remote Ingress Voice Gateway.

Unified CM as an Ingress Voice Gateway

When an IP phone initiates a call to Unified CVP, the Unified CM acts as the Ingress Voice Gateway to Unified CVP. A SIP trunk is used to send calls to Unified CVP. For more information on Unified CVP call flows, see Calls Originated by Cisco Unified Communications Manager, on page 39.

Multicast Music-on-Hold

Multicasting is used for Music-on-Hold (MOH) with supplementary services on Unified CM as an alternative to the unicast MOH. There are two ways to deploy MOH using this feature:

- With Unified CM multicasting the packets on the local LAN
- With the branch gateway multicasting on their local LAN

Use the latter method when survivable remote site telephony (SRST) is configured on the gateway. This method enables the deployment to use MOH locally and avoid MOH streaming over the WAN link.

Note Refer to the following location for information about configuring MOH on the Call Manager Enterprise (CME):

http://www.cisco.com/en/US/docs/voice_ip_comm/cucme/admin/configuration/guide/cmemoh.html#wpmkr1022205

Multicast MOH Usage Guidelines

The following guidelines apply when using Multicast MOH:

- Do not use this setting globally, or on a dial peer on the Ingress or Egress Gateway:

  ```
  modem passthrough nse codec g711ulaw
  ```

  This setting might cause Unified CM to stop the MOH after a timeout period of 10 to12 seconds.
- Do not set media inactivity on the Ingress Voice Gateway because multicast MOH does not send RTP or RTCP, and the call might get disconnected due to media-inactivity configuration. The setting media-inactivity criteria does not support multicast traffic.
- SIP-based multicast MOH is not supported on a 5400 platform because CCM-manager-based MOH subsystems are not supported on 5400 platform. This limitation also affects the ability of a TDM caller to hear multicast packets broadcasted from the Unified CM MOH server.

Call Survivability in Distributed Deployments

Distributed deployments require design guidelines for other voice services that are being run at the branch. For example, the branch is a remote Unified CM site supporting both ACD agent and nonagent phones. This deployment also implies that the PSTN Gateway is used not only for ingress of Unified CVP calls but for ingress or egress of the regular non-ACD phone calls.

Branch reliability in WANs may be an issue in a centralized Unified CVP model because they are typically less reliable than LAN links. The call survivability function must be considered for both the Unified CVP and non-CVP calls. For Unified CM endpoint phones, survivability is accomplished by using a Cisco IOS feature known as Survivable Remote Site Telephony (SRST). For further details on SRST, see the latest version of the *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*, available at:

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

For Unified CVP calls, survivability is handled by a combination of services from a TCL script (survivability.tcl) and SRST functions. The survivability TCL script monitors the SIP connection for all calls that ingress through the remote gateway. If a signaling failure occurs, the TCL script takes control of the call and redirects it to a configurable destination. The destination choices for the TCL script are configured as parameters in the Cisco IOS Gateway configuration.

Note When the called number is in "E164" format, the survivability script removes the "+" sign from the called number before forwarding it to Unified CVP. This is because Unified CVP or ICM does not support the "+" sign in the beginning of DNIS.

Alternative destinations for this transfer include another IP destination (including the SRST call agent at the remote site), *8 TNT, or hookflash. With transfers to the SRST call agent at the remote site, the most common target is an SRST alias or a basic ACD hunt group. For further information about these SRST functions, see the *Cisco Unified Communications Solution Reference Network Design (SRND) based on Cisco Unified Communications Manager*.

Voice mail and recording servers do not send Real-Time Control Protocol (RTCP) packets in reverse direction toward the caller (TDM Voice Gateway), which can falsely trigger the media inactivity timer of the survivability script. It is important to apply the survivability.tcl script carefully to the dial peers because a call might drop if it goes to the voice mail or to a recording element. One method is to use a separate dial peer for voice mail or recording calls, and do not associate the Unified CVP survivability script for those dial peers. Another method is to disable the media inactivity on the survivability script associated with the voice mail or recording dial peers.

For further information on configuration and application of these transfer methods, see the latest version of *Configuration Guide for Cisco Unified Customer Voice Portal*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

You can also refer to CUBE Deployment with SIP Trunks, on page 73.

Note To take advantage of alternate routing on signaling failures, you must use the survivability service on all gateways pointing to Unified CVP. Always use this service, unless you have a specific implementation that prevents using it.

Router requery is not supported when using SIP REFER with Unified CVP Comprehensive Call Flow when the survivability service is handling the REFER message from Unified CVP. Router requery with REFER can be supported in other call flows when Cisco IOS is handling the REFER without the survivability service or if Unified CM is handling the REFER. For third-party SIP trunks, the support of router requery with REFER is dependent on their implementation and support for SIP REFER.

Call Admission Control Considerations

Call admission control can be considered as a solution and not just a Unified CVP component. These considerations are most evident in the distributed branch office model where there are other voice services, such as Unified CM, sharing the same gateways with Unified CVP and the amount of bandwidth between the sites is limited. Be sure that, call admission control methods are in place on the network so that the same call admission control method is used for all the calls traversing the WAN from that site. If two call admission control methods can admit four calls each and the WAN link can handle only four calls, then it is possible for both call admission control entities to admit four calls onto the WAN simultaneously. This control method impairs the voice quality. If a single call admission method cannot be implemented, then each call admission control method must have bandwidth allocated to it. This situation is not desirable because it leads to inefficient bandwidth overprovisioning.

Two call admission control methods can be used in a Unified CVP environment: Unified CM Locations and Unified CM RSVP Agent. In a single-site deployment, call admission control is not necessary.

Unified CM performs call admission by assigning devices to certain locations and keeping track of the number of calls that are active between these locations. Unified CM knows the number of calls that are active and the codec use for each call, so that it can calculate the bandwidth used and limit the number of calls allowed.

A thorough conceptual understanding of call admission control features is important. These features are explained in the *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*, available at:

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

Unified CM Call Admission Control

If Unified CM sends or receives calls from Unified CVP and there are Unified CVP gateways and IP phone agents collocated at remote sites, it is important to understand the call flows in order to design and configure call admission control correctly.

SIP Call Flows

With SIP-based call flows, Cisco Unified CM Release 6.0 (and earlier releases) can look at only the source IP address of the incoming SIP INVITE from Unified CVP. This limitation causes a problem with call admission control because Unified CM cannot identify the gateway that the Unified CVP call originated.

You can use the SIP trunk feature to look beyond the source IP address and to inspect information in the SIP header when determining the device that originated a call. This enhancement allows the SIP trunk to be dynamically selected by the original source IP address instead of the remote port on Unified CVP. The SIP profiles and settings can be used on the source trunks that are different from the Unified CVP trunk.

The Call-Info header in the SIP INVITE specifies the originating device in the following format:

<sip: *IPAddress:port*>;purpose=x-c isco-origIP

The *IPAddress:port* value indicates the originating device and its SIP signaling port.

This source IP SIP trunk selection feature does not impact the bandwidth monitoring for call admission control. In Unified CM, bandwidth monitoring is performed with SIP using locations configuration on Unified CVP and Unified CM. The following header is used by the location server in Unified CM to manipulate bandwidth information for call admission control:

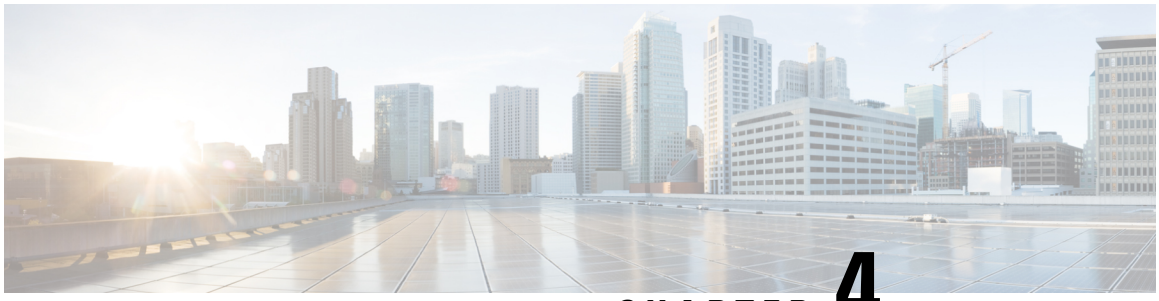
Call-Info: [urn:x-cisco-remotecc:callinfo];x-cisco-loc-id="PKID";x-cisco-loc-name="Loc-NAME"

Resource Reservation Protocol

Resource Reservation Protocol (RSVP) is used for call admission control, and it is used by the routers in the network to reserve bandwidth for calls. RSVP is not qualified for call control signaling through the Unified CVP Call Server in SIP. The recommended solution for CAC is to use the Locations configuration on Unified CVP and in Unified CM.

For more information on RSVP, see the latest version of the *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*, available at:

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



CHAPTER 4

Calls Originated by Cisco Unified Communications Manager

Overview

A call originated by the Cisco Unified Communications Manager (Unified CM) first enters the Unified Intelligent Contact Management (ICM) system when someone dials a Unified CM route point that is associated with the JTAPI interface into Unified ICM. These calls initiate a Unified ICM routing script that can be used to place the caller into queue or into a self-service application, select an available agent, or invoke Application Gateway. A call invoked through the JTAPI interface to the Unified ICM is a typical post-route request, This call provides a dialed number, ANI, variables, and returns a label. The Unified CM then delivers the call to the destination specified by the returned label. As with other ACD post-route requests, the exchange ends at the destination. Unified ICM cannot send a subsequent label to that Unified CM unless Unified CM issues another post-route request.

This limitation creates one difference between calls originated by Unified CM and calls originated through a Ingress Voice Gateway. Unified CVP can continue to route and reroute the call as many times as necessary. When calls are originated from Unified CM, routing client responsibilities should be handed off to Unified CVP as soon as possible.

Another difference is when a calls are transferred to a VRU. The ACD routing clients such as Unified CM can be transferred only by using a TranslationRouteToVRU label. When Unified CVP is the routing client, it can handle Translation Route labels as well as the Correlation ID labels that are generated by SendToVRU nodes.

The next sections provide more details on these differences.

Customer Call Flows

The following types of calls originated by Unified CM must be treated differently from calls originated by Unified CVP:

Unified ICM Outbound Calls with IVR Transfer

The Cisco Unified CCE Outbound Dialer introduces an outbound call by impersonating a Skinny Client Control Protocol (SCCP) phone and places the outbound call from the Unified CM. When a person answers the call, Unified CM transfers the call to a Unified CCE destination, taking itself out of the loop. If the customer requirement is to provide a Unified CVP message or a self-service application to the called party, then the call is transferred to Unified CVP using a Unified CM route point. This process fits the definition of a call originated by Unified CM.

Internal Help Desk Calls

Enterprises that use IP phones often want to provide their employees with the capability to call into a self-service application, for example, an application that allows employees to sign up for health benefits. The employee also might be trying to reach an agent, such as the IT help desk, and ends up waiting in queue. Both of these scenarios result in calls originating from Unified CM to Unified CVP.

The internal caller can also dial into a self-service application hosted on a Unified CVP VXML Server that is deployed using Model #1, Standalone Self-Service. No ICM is involved in this scenario, but it still qualifies as a call originated by Unified CM.

Warm Consultative Transfers and Conferences

In a typical contact center call flow, most companies want to provide their agents with the ability to transfer calls to a second agent, who might or might not currently be available. There are two ways to transfer: blind transfer or warm consultative transfer (or conference).

In a blind transfer, the first agent dials a number and hangs up; the caller then gets connected to the second agent or placed into a queue if necessary. This type of transfer does not involve a call originated by Unified CM, and it is called Network Transfer. Network Transfer is also discussed in the section on ICM Managed Transfer, on page 174.

In a warm transfer or conference, the agent dials a number and is connected to the second agent while the caller is placed on hold. The two agents can talk, then they can conference in the caller, and the first agent can then drop off. If the second agent is not available, it is the first agent (and not the caller) who is placed into a queue. All of this processing can take place without involving Unified CVP, unless the first agent needs to be queued. In that case, the first agent's call must be transferred to Unified CVP, which creates a call originated by Unified CM.

Protocol Call Flows

The following are the protocol-level call flows for calls originated by Unified CM in each of the deployment models:

Model #2: Call Director, on page 41

Model #3a: Comprehensive Using ICM Micro-apps , on page 42

Model #3b: Comprehensive Using VXML Server, on page 43

Note Model #4, VRU Only with NIC Controlled Routing, is not discussed here because no NIC is involved with calls originated by Unified CM.

Model #1: Standalone Self-Service

Model #1 does not involve Unified ICM. This model is implemented when a Unified CM user dials a directory number that connects to a VoiceXML Gateway and invokes a Unified CVP VXML Server application. The VoiceXML Gateway is configured in Unified CM as a SIP trunk. The call flow for this model is as follows:

1. A caller dials a route pattern.
2. Unified CM directs the call to the VoiceXML Gateway.
3. The VoiceXML Gateway invokes a voice browser session based on the configured Unified CVP self-service application.
4. The Unified CVP self-service application makes an HTTP request to the Unified CVP VXML Server.
5. The Unified CVP VXML Server starts a self-service application.
6. The Unified CVP VXML Server and VoiceXML Gateway exchange HTTP requests and VoiceXML responses.
7. The caller hangs up.

Model #2: Call Director

Model #2 has only switching and has no VRU leg. Calls originated by Unified CM are always delivered directly to their targets or rejected. No queuing or self-service is involved.

This model assumes that the call is truly originating from Unified CM. This model excludes calls that originally arrived through a Ingress VoiceXML Gateway and were transferred to Unified CM, and are now transferred again. These situations are rare because Unified CM can usually handle those transfers itself. There are exceptions, however, such as when the target is an ACD other than Unified CM.

This model requires that the following items be configured:

- Unified CM route point that invokes a Unified ICM script
- Unified CVP configured as a Type 10 NetworkVRU
- VRU translation routes to Unified CVP
- Translation route Dialed Number Identification Service (DNIS) numbers configured in the Unified CVP Call Server
- Unified CM configured with a SIP trunk

- Unified CM route patterns for Translation Route DNIS

The call flow for this model is as follows:

1. A caller dials a route point.
2. Unified ICM invokes a routing script.
3. The routing script encounters a TranslationRouteToVRU node to transfer the call to Unified CVP. (Unified CVP is configured as a Type 10 NetworkVRU.)
4. Unified ICM returns the translation route label to Unified CM.
5. Unified CM consults the SIP Proxy to locate the Unified CVP Call Server.
6. Unified CM connects the call to the Unified CVP Call Server.
7. The routing script encounters a Select or Label node, and it selects a target label.
8. Unified ICM returns the target label to the Unified CVP Call Server (not to the device that issued the route request).
9. The Unified CVP Call Server consults the SIP Proxy to locate the destination device.
10. The Unified CVP Call Server communicates through SIP with the target device and instructs Unified CM to establish a media stream to it.

    If the target device issues another route request to Unified ICM. This part of the call flow is not possible without the initial TranslationRouteToVRU mentioned in step 3.
11. Unified ICM invokes a new routing script.
12. The routing script encounters a Select or Label node, and it selects a target label.
13. Unified ICM returns the target label to the Unified CVP Call Server (not to the device that issued the route request).
14. The Unified CVP Call Server consults the SIP Proxy to locate the destination device.
15. The Unified CVP Call Server communicates using SIP with the target device and instructs Unified CM to establish a media stream to the device.

Model #3a: Comprehensive Using ICM Micro-apps

Model #3a involves both call switching and VRU activity. This model differs from Model #2, so the calls must be transferred to the VoiceXML Gateway after they are transferred to the Unified CVP Switch leg. Queuing is possible in this model because it is basic prompt-and-collect activity.

This model requires that the following items be configured:

- Unified CM CTI route point that invokes a Unified ICM script
- Unified CVP configured as a Type 10 NetworkVRU
- The CTI route point configured in Unified ICM as a DN with a Type 10 NetworkVRU
- The NetworkVRU must have labels for the Unified CVP Switch leg routing client
- The NetworkVRU labels must be configured in a SIP Proxy to point to VoiceXML Gateways

- Unified CM configured with a SIP trunk

The call flow for this model is as follows:

1. A caller dials a route point.
2. Unified ICM invokes a routing script.
3. The routing script encounters a SendToVRU node to transfer the call to Unified CVP. (Unified CVP is configured as a Type 10 NetworkVRU.)
4. Unified ICM returns the VRU label with Correlation ID to Unified CM.
5. Unified CM consults the SIP Proxy to locate the Unified CVP Call Server.
6. The call is connected to the Unified CVP Call Server.
7. Unified ICM sends a VRU transfer label with Correlation ID to the Unified CVP Call Server.
8. The Unified CVP Call Server consults the SIP Proxy to locate the VoiceXML Gateway.
9. The Unified CVP Call Server communicates through SIP with the VoiceXML Gateway and instructs Unified CM to establish a media stream to it.
10. The routing script executes one or more Unified CVP Microapplications through RunExternalScript nodes, plays media files, requests DTMF input, and so forth.
11. While the Unified CVP Microapplications are in progress, a target agent becomes available to take the call.
12. Unified ICM determines a label for the target agent.
13. Unified ICM returns the target label to the Unified CVP Call Server.
14. The Unified CVP Call Server consults the SIP Proxy to locate the destination device.
15. The Unified CVP Call Server communicates through SIP with the target device and instructs Unified CM to establish a media stream to it, removing the VoiceXML Gateway's media stream.

If the target device later issues another route request to Unified ICM, the call flow again is performed as described. The call must again be transferred with Correlation ID through SendToVRU to the Unified CVP Call Server and VoiceXML Gateway to create the VRU leg. Microapplications might be executed, and eventually the new target label is delivered to the Unified CVP Switch leg, which transfers the call to that target.

Model #3b: Comprehensive Using VXML Server

Model #3b does not differ significantly from Model #3a regarding call control and signaling. The only difference is that the Unified CVP Microapplications executed in Model #3b might also include subdialog requests to the Unified CVP VXML Server as well. The self-service applications are not likely to be invoked during the period when the call is queued. Any agent selection nodes or queue nodes in the Unified ICM routing script are postponed until after the self-service application has completed and control has returned to the Unified ICM routing script.

Deployment Implications

This section presents guidelines for these tasks for incorporating calls originated by Unified CM into the deployment:

- Unified ICM Configuration , on page 44
- Hosted Implementations , on page 126
- Cisco Unified Communications Manager Configuration, on page 45
- Sizing, on page 57

Unified ICM Configuration

- With Cisco Unified ICM 7.0, to perform subsequent call control through Unified CVP, always use translation route to route the call to Unified CVP as a Type 2 NetworkVRU before delivering the call to its next destination. This practice passes the control to Unified CVP as an charge of subsequent call transfers because Unified CM cannot receive any further labels.
- To perform any queuing treatment, prompt and collect, or self-service applications, always follow translation route with a SendToVRU node. SendToVRU can sometimes be invoked implicitly by a Queue node or a RunExternalScript node, but you should not rely on that method. Always include an actual SendToVRU node.
- With Cisco Unified ICM 7.1, to perform subsequent call control through Unified CVP, a translation route is not necessary if you use a Type 10 NetworkVRU. The Type 10 VRU uses the Correlation ID method to perform a transfer from Unified CM to Unified CVP using a SendToVRU node. When the SendToVRU node is used with a Type 10 VRU, an initial transfer to Unified CVP hands off call control to Unified CVP, and then an automatic second transfer to the VRU leg is performed to deliver the call to a VoiceXML Gateway for IVR treatment.

Note This call flow and all others in this document assume that you are using Cisco Unified ICM 7.0(0) or later.

- For additional configuration requirements, see Protocol Call Flows, on page 27.
- When the SendToVRU node is used with a Type 10 VRU, an initial transfer to Unified CVP hands off call control to Unified CVP, and then an automatic second transfer to the VRU leg is performed to deliver the call to a VoiceXML Gateway for IVR treatment.

Hosted Implementations

Translation routes sent by one ICM router must be received by a peripheral connected to the same ICM router. Therefore, you can use a translation route to route a call from a Unified CM at the CICM level into Unified CVP only if Unified CVP is also located at the CICM level. In Hosted environments, this means you must provision Unified CVP Call Servers (Call Servers) at the CICM level even if you already have other Call Servers at the NAM level.

For more details on this subject, see the chapter on Cisco Unified ICM Interactions, on page 119.

Cisco Unified Communications Manager Configuration

The following guidelines apply to Unified CM configuration:

- Configure a SIP trunk.
- Configure the appropriate route patterns for the Translation Route DNIS or VRU Label with Correlation ID appended. The Correlation ID method is used with a Type 10 VRU, and the route pattern in Unified CM must be configured to allow the extra digits to be appended, such as adding an exclamation point (!) to the end of the route pattern.
- When configuring agent labels, consider which device is the routing client. In situations where the label will be returned directly to Unified CM, Unified CM must be the routing client. In situations where the label will be sent to Unified CVP, the labels must be associated with each of the Unified CVP Switch leg Call Servers.

SIP Proxy Dial-Plan Configuration

If you are using a SIP Proxy, the VRU label associated with the Unified CM routing client must be different from the VRU label associated with the Unified CVP routing clients. The reason is because the VRU label for a call originated by Unified CM is intended to send the call to the Unified CVP Call Server to hand off call control first. The VRU label for a call where Unified CVP is already the routing client is intended to be sent to the VoiceXML Gateway for treatment. Once the call has been sent to Unified CVP to hand off call control, Unified CVP subsequently transfers to the VRU label associated with the Unified CVP routing client and delivers the call to the VoiceXML Gateway for queuing treatment.

The dial plan in your SIP Proxy should be structured as follows:

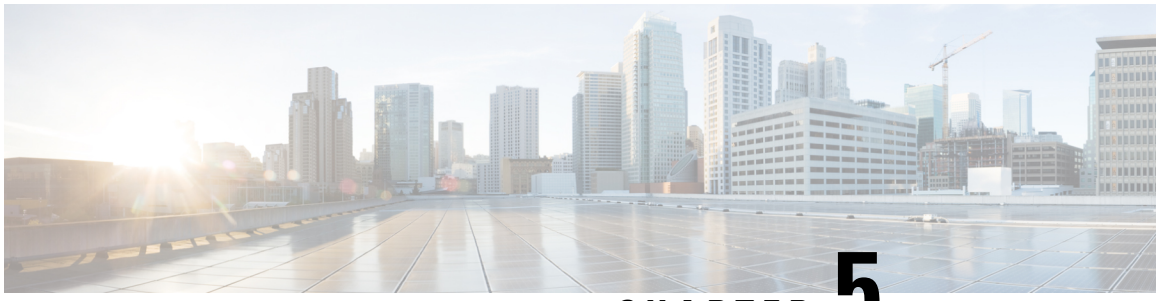
[Unified CM routing client VRU label + correlation-id]: pointing to CVP server(s)

[CVP routing client VRU label + correlation-id]: pointing to VoiceXML Gateway(s)

For a description of the Cisco SIP Proxy Server, refer to SIP Proxy Server, on page 9.

Mobile Agent in UCM

Mobile Agent Connect Tone initiated by JGW/UCM and Whisper Announcement initiated by Unified CVP can overlap because JGW/UCM has control on the LCP and RCP port. To avoid overlapping, either use the Mobile Agent Connect Tone or the CVP Whisper Announcement.



CHAPTER 5

Media File Options

Deployment and Management of Voice Prompts

You can deploy voice prompts using following approaches:

- Local File System

  The voice prompt files are stored on a local system and audio prompts are retrieved without using bandwidth. With this approach, VoiceXML Gateways do not have to retrieve audio files for playing prompts, so WAN bandwidth is not affected. However, if a prompt needs to be changed, you must change it on every VoiceXML Gateway.

  - IOS VoiceXML Gateway—prompts are deployed on flash memory.

    IOS VoiceXML Gateway can either be VoiceXML Gateway or PSTN Gateway, which has Ingress Voice Gateway and VoiceXML Gateway colocated. Store only critical prompts such as error messages or other messages that can be used when the WAN is down.

    When recorded in G.711 mu-law format, typical prompts of average duration are about 10 to 15 KB in size. When sizing gateways for such implementations, size the flash memory by factoring in the number of prompts and their sizes, and also leave space for storing the Cisco IOS image.

  - Cisco VVB—prompts are installed on local file system.

    Built-in CVP prompts are packaged with Cisco VVB product and installed during installation. You can change *Error* tone default prompt through Cisco VVB Administrator console.

- Media Server

  Each local VoiceXML Gateway, if configured properly, can cache many or all prompts, depending on the number and size of the prompts (up to 2 GB for Cisco VVB and 100 MB for IOS). The best way to test whether your Media Server is appropriately serving the media files is to use a web browser and specify the URL of a prompt on the Media Server, such as *http://10.4.33.130/en-us/sys/1.wav*. Your web browser should be able to download and play the .wav file without any authentication.

The design of Media Server deployment depends on the following factors:

- Number of media files to be played on each gateway.
- Network connectivity between the gateway and the Media Server.
- Frequency in which the media files are changed.

Media File Deployment Design Concepts

The concepts described in this section are relevant to media file deployment design.

Bandwidth Calculation for Prompt Retrieval

When prompts are stored on an HTTP media server, the refresh period for the prompts is defined on that server. The bandwidth consumed by prompts consists of the initial loading of the prompts at each VoiceXML Gateway and of the periodic updates at the expiration of the refresh interval.

To calculate the bandwidth that your prompts consume, multiply the number of prompts by average size of each prompt. As an example of determining the bandwidth consumed by prompts, assume that a deployment has 50 prompts with an average size of 50 KB (50,000 bytes) each. Also assume that the refresh period for the prompts is defined as 15 minutes (900 seconds) on the HTTP media server. The WAN bandwidth required for prompts in this deployment can be calculated as follows:

(50 prompts) * (50,000 bytes/prompt) * (8 bits/byte) = 20,000,000 bits

(20,000,000 bits) / (900 seconds) = 22.2 kbps per branch

TCP Socket Persistence

Unified CVP does not support TCP socket persistence.

WAN Acceleration

The Cisco Wide Area Application Services (WAAS) system consists of a set of devices called Wide Area Application Engines (WAEs) that work together to optimize TCP traffic over your network. Cisco WAAS uses a combination of TCP optimization techniques and application acceleration features to overcome the most common challenges associated with transporting traffic over a WAN. Cisco WAAS deployed at the periphery of the network on the VoiceXML Gateway side performs the following functions:

- Makes changes in TCP header to optimize the traffic.
- Acts as a large HTTP cache located locally.
- Reduces the traffic more using compression algorithms.
- Reduces traffic by using Data Redundancy Elimination (DRE) techniques.

Cisco WAAS is deployed in inline mode where whole data is forced to pass through the Cisco WAAS.

IOS Gateway Media File Deployment

Cisco IOS Caching and Streaming

The Cisco IOS VoiceXML Gateway uses an HTTP client, which is a part of Cisco IOS. The client fetches VoiceXML documents, audio files, and other file resources.

Caching and streaming are two key properties associated with playing audio prompts. These two properties are closely related to each other, and they can affect system performance greatly when the router is under load.

Streaming and Non-Streaming Modes

In non-streaming mode, the entire audio file must be downloaded from the HTTP server onto the router before the Media Player can start playing the prompt. This implies a delay for the caller. If the audio file is relatively small, the caller will not notice any delay because downloading a small file takes only a few milliseconds. The delay caused by loading larger files can be overcome by using either caching or streaming mode.

In streaming mode, the Media Player streams the audio in media chunks from the HTTP server to the caller. As soon as the first chunk is fetched from the server, the Media Player can start playing. The advantage of streaming mode is that there is no noticeable delay to the caller, regardless of the size of the audio prompt. The disadvantage of streaming mode is that, because of all of the back-and-forth interactions from fetching the media file in chunks, the performance deteriorates. Additionally, the ability to cache the files in memory reduces the advantage of streaming large files directly from the HTTP server.

For recommendations on when to use streaming and non-streaming mode for prompts, see section Design Considerations for Large Number of Media Files, on page 52.

Cache Types

There are two types of cache involved in storing media files: the IVR Media Player cache and the HTTP Client cache.

The HTTP Client cache is used for storing files that are downloaded from the HTTP server. In nonstreaming mode, the entire media file is stored inside the HTTP Client cache. In streaming mode, the first chunk of the media file is stored in the HTTP Client cache and in the IVR cache, and all subsequent chunks of the file are saved in the IVR cache only. The HTTP Client cache can store 100 MB of prompts, while the IVR cache is limited to 32 MB.

Use only nonstreaming mode, so that the IVR prompt cache is never used and the HTTP Client cache is the primary cache. In nonstreaming mode, the HTTP Client cache can also store 100 MB of prompts, while the IVR cache is limited to 16 MB.

To configure the HTTP Client cache, use the following Cisco IOS commands:

http client cache memory file *1-10000*

The 1–10000 value is the file size in kilobytes. The default maximum file size is 50 KB, but you can also have a file size up to 600 KB file size. Any file that is larger than the configured HTTP Client memory file size will not be cached.

http client cache memory pool *0-100000*

The 0–100000 value is the total memory size available for all prompts, expressed in kilobytes. A value of zero disables HTTP caching. The default memory pool size for the HTTP Client cache is 10 MB. The memory pool size is the total size of all prompts stored on the media server, which is up to 100 MB.

Query URL Caching

A query is a URL that has a question mark (?) followed by one or more **name=value** attribute pairs in it. The Unified CVP VXML Server uses query URLs extensively when generating the dynamic VoiceXML pages that are rendered to the caller. Because each call is unique, data retrieved from a query URL can waste cache memory and a possible security risk, because the query URL can contain information such as account numbers or PINs.

Query URL caching is disabled by default in Cisco IOS. To ensure that it is disabled, enter a **show run** command in Cisco IOS and ensure that the following Cisco IOS command does not appear:

Gateway configuration: **http client cache query**

Cisco VVB Media File Deployment

Caching and Query

Cisco VVB uses an HTTP client, which is a part of the product. The client fetches VoiceXML documents, audio files, and other file resources.

Caching property is associated with VXML resources, audio prompts, grammar and script files.

A query is a URL that has a question mark (?) followed by one or more **name=value** attribute pairs in it. By default, Query URLs are not cached.

Cache Aging

The HTTP Client manages its cache by the freshness of each cached entry. Whether a cached entry is fresh or stale depends on two numbers: Age and FreshTime. Age is the elapsed time since the file was last downloaded from the server. FreshTime is the duration that the file is expected to stay in the HTTP Client cache since the file was last downloaded.

Several variables that can affect the FreshTime of a file, such as HTTP message headers from the server and the cache refresh value configured using the command line interface (CLI).

The FreshTime of a file is determined in the following sequence:

1. When a file is downloaded from the HTTP server, if one of the HTTP message headers contains the following information, the max-age is used as the FreshTime for this file:

   Cache-Control: max-age = <*value in seconds*>

2. If Step 1 does not apply, but the following two headers are included in the HTTP message, the difference (Expires – Date) is used as the FreshTime for this file:

   Expires: <*expiration date time*>

   Date: <*Current date time*>

3. The HTTP/1.1 specification, RFC 2616 (HyperText Transport Protocol), recommends that either one of the HTTP message headers as described in Step 1 or 2 should be present. If the server fails to send both 1 and 2 in its HTTP response, then take 10 percent of the difference between Date and Last-Modified from the following message headers:

   Last-Modified: <*last-modified date time*>

   Date: <*Current date time*>

So the FreshTime for this file is calculated as:

FreshTime = 10% * ([Date] – [Last-Modified])

4. The CLI allows the user to assign a FreshTime value to the files as a provisional value:

   http client cache refresh *1-864000*

   The default refresh value is 86400 seconds (24 hours). The configured HTTP Client cache refresh has no effect on files when any of the message headers in steps 1 to 3 are present. If the resultant FreshTime from the CLI command calculation turns out to be less than the system default (which is 86400 seconds), the FreshTime will be set to the default value (86400 seconds). This command is not retroactive. That is, the newly configured refresh value applies only to new incoming files, and it has no effect on the entries already in the cache.

Note Step 4 is not applicable to Cisco VVB.

Stale files are refreshed on an as-needed basis only. A stale cached entry can stay in the cache for a long time until it is removed to make room for either a fresh copy of the same file or another file that needs its memory space in the cache.

A stale cached entry is removed on an as-needed basis when all of the following conditions are true:

- The cached entry becomes stale.
- Its refresh count is zero (0); that is, the cached entry is not being used.
- Its memory space is needed to make room for other entries.

Note When the Age exceeds the FreshTime and the file needs to be played, the HTTP Client checks with the media server to determine whether or not the file has been updated. When the HTTP Client sends a GET request to the server, it uses a conditional GET to minimize its impact on network traffic. The GET request includes an If-Modified-Since in the headers sent to the server. With this header, the server returns a 304 response code (Not Modified) or returns the entire file if the file was updated recently.

This conditional GET applies only to nonstreaming mode. In streaming mode, the HTTP Client always issues an unconditional GET. There is no If-Modified-Since header included in the GET request that results in an unconditional reload for each GET in streaming mode.

You can reload individual files into cache by entering the following command:

Gateway configuration: **test http client get http://10.0.0.130/en-us/sys/1.wav reload**

Note This command is only applicable to IOS VoiceXML Gateway.

Note HTTPS for media files is not supported.

Design Considerations for Large Number of Media Files

In situations where a large number of different media files (.wav) is played to the customers, the gateway cannot cache all the media files because of space constraints in the gateway.

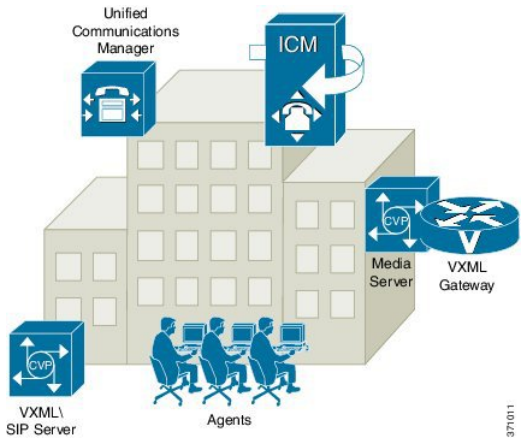
For example, consider an enterprise having a large number of agents. All of the agents have their own customized agent greeting file. It is impossible to cache all the customized agent greeting files in the gateway flash because of space constraints in the gateway.

Collocated Media Server with VoiceXML Gateway

The following section outlines the recommended solution when a Media Server and VoiceXML Gateway coexist in a LAN environment, if the bandwidth is abundant over the LAN, the prompt download should not add noticeable delay.

The following figure shows the collocated deployment for Media Server and VoiceXML Gateway.

Figure 7: Collocated Deployment for Media Server with VoiceXML Gateway

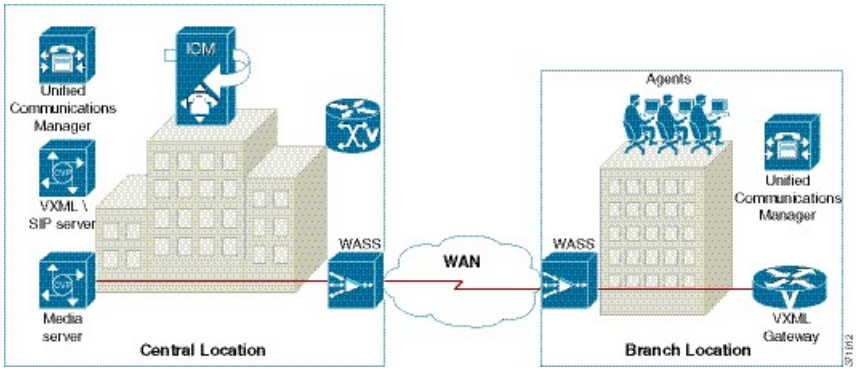


Distributed Media Server and VoiceXML Gateway Separated by a High Latency Link

This section outlines the recommended solution when a Media Server and VoiceXML Gateway are separated by a WAN.

The following figure shows the distributed deployment over WAN.

Figure 8: Distributed Deployment Over WAN



In this situation downloading the media files from Media Server across a high latency WAN to the VoiceXML Gateway can add noticeable delays to the caller. This delay will greatly impact the user experience. The delay will be proportional to the size and number of media files transported across the WAN. This delay can be optimized using Cisco Wide Area Application Services (WAAS).

Considerations for Streaming

Consider the following factors for both the LAN deployment and the WAN accelerator deployment:

- Maximum network round-trip time (RTT) delays of 200 milliseconds.

  For example, during the transfer of files from the CVP Operations Console to the Ingress or the VXML Gateway using Bulk Administration File Transfer (BAFT).

- Maximum number of streaming sessions supported per gateway with no additional overhead of video with media forking.

  For example, the maximum number of calls supported per gateway in streaming is 275. This number is valid only for the Cisco 3945E Integrated Services Router (ISR) with no additional overhead at a maximum network RTT delay of 200 milliseconds.

The following table describes the media file deployment scenarios over LAN and WAN:

Scenario	Frequency of Change	Over LAN	Over WAN
Small number of files	Rare	Cached	Cached
Small number of files	Often	Streamed or Cached	Streaming with WAAS
Large number of files	Rare	Streamed	Streaming with WAAS
Large number of files	Often	Streamed	Streaming with WAAS

Note Cisco VVB does not support Media Streaming feature.

Media Server Association with Call Server and VXML Server

Note Unified CVP Call Server, Media Server, and Unified CVP VXML Server are co-resident on the same server.

If your Unified CVP Call Server, Media Server, and Unified CVP VXML Server reside on the same hardware server and you have multiple co-resident servers, Unified CVP does not automatically use the same physical server for call control, VXML, and media file services. If the components are co-resident, no component is forced to use the other co-resident components, and Unified CVP might possibly use the components located on another server.

By default, the components are load balanced across all of the physical servers and do not attempt to use the same server for all of the services. During thousands of calls, all of the components on all of the servers are load balanced and equally utilized, but one specific call could be using several different physical servers. For example, for one particular call you can be using SIP call control on one server, VoiceXML on another server, and the media files on another server.

You can simplify management and troubleshooting by configuring Unified CVP to use the same physical server for all of these functions on a per-call basis. If there is only one server in the system, then simplification is not a concern. The instructions in the following procedures show you how to configure Unified CVP so that it uses components on the same physical server instead of load balancing and using a random server for each component.

Choose Coresident Unified CVP VXML Server in ICM Script Editor

Procedure

Step 1 Set up the **media_server** ECC variable that specifies your Unified CVP VXML Server in the ICM script by using use the Formula Editor to set the **media_server** ECC variable to **concatenate("http://",Call.RoutingClient,":7000/CVP")**.

Call.RoutingClient is the built-in call variable that ICM sets automatically for you. The routing client name in ICM is usually not the same as the Unified CVP Server's hostname.

Step 2 Apply the routing client name as a hostname in the VXML gateway. Do not use noncompliant characters such as an underscore as part of the hostname because the router cannot translate the hostname to an IP address if it contains noncomplaint characters. Use the **ip hostname strict** command in the router to prevent the use of invalid characters in the hostname. This action ensures that the hostname is acceptable to Unified CVP.

Step 3 Configure the routing client hostname for every Unified CVP Server Routing Client.

Choose Coresident Media Server in Call Studio

Procedure

Step 1 In the ICM script, set one of the **ToExtVXML[]** array variables with the call.routingclient data, such as ServerName=call.routingclient. This variable is passed to the Unified CVP VXML Server, and the variable is stored in the session data with the variable name ServerName.

Step 2 In Cisco Unified Call Studio, use a substitution to populate the Default Audio Path. Add the Application_Modifier element found in the Context folder, and specify the Default Audio Path in the Settings tab in the following format: **http://***{Data.Session.ServerName}*

Choose Coresident VXML Server Using Micro-Apps

If you are using Micro-Apps in conjunction with the Unified CVP VXML Server, pay careful attention to the **media_server** ECC variable in the ICM script because the same variable is used to specify both the Unified CVP VXML Server and the media server, but the contents of the variable use a different format depending on which server you want to specify. Use the **media_server** ECC variable as indicated in this procedure whenever you want to use a Micro-App for prompting. If you subsequently want to use the Unified CVP VXML Server, rewrite this variable by following the previous procedure.

Procedure

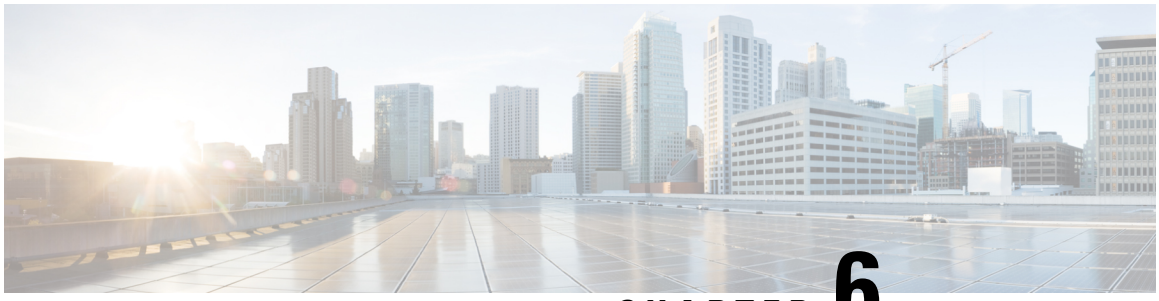
Step 1 Set up the **media_server** ECC variable that specifies your Media server in the ICM script by using the Formula Editor to set the **media_server** ECC variable to **concatenate("http://",Call.RoutingClient)**

Call.RoutingClient is the built-in call variable that ICM sets automatically for you. The routing client name in ICM usually is not the same as the Unified CVP Server hostname.

Step 2 Use the name of the routing client as a hostname in the VoiceXML Gateway.

Do not use noncompliant characters such as an underscore as part of the hostname because the router cannot translate the hostname to an IP address if it contains any noncomplaint characters. Use the **ip hostname strict** command in the router to prevent the use of invalid characters in the hostname and to ensure that the hostname is acceptable to Unified CVP.

Step 3 Configure the routing client hostname for every Unified CVP Server Routing Client.



CHAPTER 6

Sizing

Overview

An important consideration in sizing a contact center is to determine the worst-case contact center profile for the number of calls that are in each state. For instance, if you observe the contact center at its busiest instant in the busiest hour, observe how many calls you find in the following states:

- **Self-service**—Calls that are executing applications using the VXML Server.
- **Queue and collect**—Calls that are in queue for an agent or are executing prompt-and-collect type self-service applications.
- **Talking**—Calls that are connected to agents or to third-party TDM VRU applications.

In counting the number of calls that are in the talking state, count only calls that are using Unified CVP or gateway resources. To determine whether a call in talking state is using resources, consider how the call gets transferred to that VRU or agent. If the call is transferred using VoIP, it continues to use an Ingress VoiceXML Gateway port and it continues to use a Unified CVP resource, because Unified CVP continues to monitor the call and provides the ability to retrieve it and redeliver it at a later time. Unified CVP also continues to monitor calls that are sent to a TDM target, using both an incoming and an outgoing TDM port on the same gateway or on a different gateway (that is, toll bypass). Calls that are transferred to VRUs or agents in this manner are counted as talking calls.

However, if the call was transferred through *8 TNT, hookflash, Two B Channel Transfer (TBCT), or an ICM NIC, neither the gateway nor Unified CVP play any role in the call. Both components have reclaimed their resources, therefore, such calls are not counted as talking calls.

Include in the overall call counts those calls that have been transferred back into Unified CVP for queuing or self-service, using either blind or warm methods. For example, if a warm transfer is used and the agent is queued at Unified CVP during the post-route phase, the call will use two ports due to two separate call control

sessions at Unified CVP. Because these calls usually do not contribute to more than 5 percent or 10 percent of the overall call volume, you can easily overlook them.

The definitions of these call states differ from the definitions used for port licensing purposes. Similarly, the call state determination does not influence with whether the agents are Unified CCE agents or ACD agents, nor does it matter whether the customer intends to use Unified CVP to retrieve and redeliver the call to another agent or back into self-service.

Note You should size the solution for the number of ports in use for calls in a talking state to agents. Even though licenses for those ports do not have to be purchased when using Unified CCE agents, TDM agents do require a Call Director license.

In addition to the overall snapshot profile of calls in the contact center, you must also consider the busiest period call arrival rate in terms of calls per second. You need this information for the contact center as a whole because it is difficult to identify the exact maximum arrival rate. You can use statistical means to arrive at this number, except in very small implementations, which is seldom the critical factor in determining sizing.

You can begin sizing each component in the network following the information in the overview section. The overview section, deals with the number and type of physical components required to support the Unified CVP system, but it does not include any discussion of redundancy. For an understanding of how to extend these numbers to support higher reliability, see Unified CVP Design for High Availability, on page 93.

Note In Unified CVP, the Call Server, VXML Server, and Media Server are combined as one installation as CVP Server. Installing the CVP Server installs all three components. In the earlier versions, Call Server, VXML Server, and Media Server could be installed on different machines.

Call Server Sizing

Note Call Server is not used in Model #1 (Standalone Self-Service). This section does not apply to these deployments.

Unified CVP Call Servers are sized according to the number of calls they can handle, in addition to their maximum call arrival rate.

Table 1: Call Server Call Rate

Call Server	Capacity
Maximum SIP Calls	900
Sustained Calls per Second (SIP)	10

Note For UCS performance numbers, see the *Virtualization for Cisco Unified Customer Voice Portal* page at http://docwiki.cisco.com/wiki/Virtualization_for_Cisco_Unified_Customer_Voice_Portal.

Example

Consider the following Call Server calculations:

Each Call Server can handle 900 SIP calls. Each Call Server is further limited to a sustained call arrival rate of 10 call per second (cps) for SIP. However, Model #4 is exempt from this limitation because the Call Server in that model does not perform any SIP processing.

Specifically, the number of Call Servers required is the higher of these totals:

((Self Service) + (Queue and Collect) + Talking) / 900, rounded up

or

(Average call arrival rate) / 10, rounded up.

In addition, calls delivered to the Cisco Unified Communications Manager cluster should be load balanced among the subscribers in the cluster and should not exceed 2 calls per second (cps) per subscriber.

Call Server Log Directory Size Estimate

When you use the Agent Greeting feature, performance of Unified CVP Call Servers is reduced by 25 percent. This reduction occurs because the servers operate at 75 percent of calls per second (CPS) of a system that is not using the Agent Greeting feature.

Size your system using the methods detailed in the guide, then multiply the CPS by 75 percent:

- For example, 10 CPS on a UCS platform without Agent Greeting translates into 7.5 CPS on a Call Server with Agent Greeting enabled.
- Ports required are calculated based on the CPS and duration of agent greeting, and must be determined from the total supported ports of a server.

Call Server Log Directory Size Estimate

Use the following formula to calculate the estimated space per day (in gigabytes) for the Call Server Directory log file:

3.5 * R

R equals the number of calls per second.

For proper serviceability, reserve enough space to retain from five to seven days of log messages.

To set the log directory size, in the Operations Console, choose the Infrastructure tab for Call Server set up.

VXML Server Sizing

VXML Server call rate calculations are described in the table and examples in this section.

One VXML Server can handle up to 900 calls. If you are using VXML Servers, size them according to the following formula:

Calls / 900, rounded up,

Calls refers to the number of calls that are in VXML Server self-service applications at that snapshot in time.

Note For UCS performance numbers, go to this location http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CVP.

You can configure Unified CVP to use HTTPS on the VXML Server and on the Unified CVP IVR Service (IVR Service can generate basic VoiceXML documents and is part of the Call Server). Due to the large processing overhead of HTTPS, the Tomcat application server can achieve a maximum of 275 simultaneous connections, depending on the configuration.

Set Cache Command

- IOS VoiceXML Gateway—Configure the Cisco IOS VoiceXML Gateway with the HTTPS option. If this configuration setting is not available then this can severely impact the performance and sizing of the VXML Gateway and the overall solution in general with HTTPS.

  http client connection persistent

  http client cache memory pool 15000

  http client cache memory file 1000

- Cisco VVB—Configure the Cisco VVB with the HTTPS option. If this configuration setting is not available then this can severely impact the performance and sizing of the VXML browser and the overall solution in general with HTTPS. The following CLI commands are used for modifying cache properties (for more details, see *Cisco Virtualized Voice Browser Operations Guide*):

  set vvb cache browser_cache_size

  set vvb cache dom_cache_capacity

  set vvb cache enable_browser_cache

  set vvb cache enable_browser_cache_trace

  set vvb cache enable_dom_cache

The following table provides simultaneous call information for HTTPS calls using various applications and call flow models.

Table 2: HTTPS Simultaneous Calls for Unified CVP Servers

Call Server Type, Application, and Call Flow Model	Number of Simultaneous Calls
VXML ServerMax simultaneous HTTPS connections with Tomcat (Standalone Call Flow model)	275
Call Server and VXML ServerMax simultaneous HTTPS connections with Tomcat (Comprehensive Call Flow model)	275

Note In all of the above scenarios, the Reporting and Datafeed options are disabled. Also, Cisco IOS Release 12.4(15)T5 or later release is required on the gateway to support the HTTPS option. Mainline Cisco IOS is not supported.

IOS VoiceXML Gateway

VoiceXML Gateway Sizing for Agent Greeting

The additional VoiceXML Gateway ports required are calculated based on calls per second (CPS) and the duration of the agent greeting. The agent greeting is counted as one additional call to the VoiceXML Gateway.

Use the following formula to determine the additional ports required for the Agent Greeting feature:

```
Total ports = Inbound ports + [( Agent Greeting Duration / Total call duration) *
Inbound ports]
```

For example, if you estimate 120 calls, each with a 60-second call duration, you have 2 CPS and a requirement of 120 inbound ports. If you assume that the agent greeting duration is 5 seconds on every call, then the overall calls per second is 4 CPS, but the number of ports required is 130.

```
Total Ports = 120 inbound ports + [(5-second agent greeting duration / 60-second total call duration) *
120 inbound ports] = 130 total ports.
```

VoiceXML Gateway Agent Greeting Prompt Cache Sizing

For sizing agent greeting prompt cache, consider the following example:

The following calculation shows that a 1-minute long file in the g711uLaw codec uses approximately 1/2 MB:

```
64 kbits/sec = 8 kbytes/sec (bit rate for g711uLaw codec)
8 kb/sec * 60 seconds - 480 kb (~ 0.5 MB
```

- The maximum memory used for prompt cache in a Cisco IOS router is 100 MB and the maximum size of a single file should be 600 KB.
- Number of Agent Greetings cached using the above sizing numbers are provided here:
  - 5-second greeting—40 KB, that is, approximately 25 greetings per MB. This typical use case scenario provides caching for approximately 80*25 agent = 2000 agent greetings with 80 percent space reserved for Agent Greeting.
  - 60-second greeting—480 KB, that is, approximately 2 greeting per MB. The worst case scenario provides caching for approximately 50*2 agent = 100 agents with 50 percent space used for Agent Greeting.

Note For Cisco VVB, maximum cache size is 2 GB and the above calculation can be based on 2 GB instead of 100 MB.

Media Server Sizing for Agent Greeting

Media Server sizing usually is not provided due to the following reasons:

- Diverse requirements of a media server based on specific deployment requirements
- Wide range of hardware that is used for media servers

However, the following sizing profile is for a Media Server that is used with the Agent Greeting feature.

Example load:

- 700 agents
- 15-second greeting (118-kb greeting file)
- 30-minute content expiration

Media Server hardware equivalent to the following (or better) is required to handle the above profile:

- UCS platform with RAID 5 (media server only)
- UCS platform with RAID 5 (media server only)
- UCS platform with RAID 5 (collocated media/call server)

Unified CVP Coresidency

Self-service means that a call requires SIP call control and runs an application on the VXML Server. Queue and collect means that a call requires SIP call control and runs an application using Microapps only on the Call Server.

The following example applies for VoiceXML and HTTP sessions only. The same values apply to both coresident and distributed deployments of Call Servers and VXML Servers.

The number of servers required using SIP call control would be as follows:

((Self Service) + (Queue and Collect) + Talking) / 900, rounded up

((900) + (500) + 3700) / 900 = 6 servers

If you use the Cisco Unified Border Element as a Session Border Controller (SBC) for flow-through calls to handle VoiceXML requirements, then you must use the sizing information provided in the example. The Cisco Unified Border Element is limited to the maximum number of simultaneous VoiceXML sessions or calls as outlined provided in the example for the particular situation and hardware platform.

If you use the Cisco Unified Border Element as a Session Border Controller (SBC) to handle flow-through calls only (no VoiceXML), then consider Voice Activity Detection (VAD) and see the sizing information in

the *Cisco Unified Border Element Ordering Guide*, available at http://www.cisco.com/en/US/prod/collateral/voicesw/ps6790/gatecont/ps5640/order_guide_c07_462222.html.

Coresident Unified CVP Reporting Server and Unified CVP Call Server

Reporting Server can be coresident with the Call Server, but only for Standalone VoiceXML deployments. A Call Server usually is not needed in a Standalone VoiceXML deployment, but if reporting is desired, a Call Server is required to send the reporting data from the VXML Server to the Reporting Server. When the Reporting Server is coresident with a Call Server, the Call Server is not processing any SIP calls, but is relaying reporting data from the VXML Server.

The coresident Call Server does not have a significant impact on performance in this model, therefore, the sizing information in the Reporting Server section does not change.

Note If Unified Border Element is to be used as a Session Border Controller (SBC), use these procedures:

- To handle flow-through or flow-around calls only (no VXML), including VAD, use the Unified Border Element Ordering Guide for sizing.
- For flow-through or flow-around calls and with VXML requirements, use the sizing information in the *Design Guide for Cisco Unified Customer Voice Portal*. Unified Border Element is limited to the maximum number of simultaneous VXML sessions and calls as outlined for the particular situation and hardware platform.

Cisco Unified SIP Proxy

Unified CVP supports only the Cisco Unified SIP Proxy (CUSP) Server.

Information on CUSP architecture, feature, configuration, and data sheets is available at http://wwwin.cisco.com/artg/products/voice_video/cusp/.

The CUSP baseline tests are done in isolation on the proxy, and capacity numbers (450 to 500 cps) should be used as the highest benchmark. In a Unified CVP deployment, a CUSP proxy sees incoming calls from the TDM Gateway, from Unified CVP, and from the UCM SIP trunk. With a SIP back-to-back user agent in CVP, the initial call setup from the proxy, involves an inbound call immediately followed by an outbound call (whether for IVR or to ACD). Later in the call, CVP may transfer the call to an agent, which involves an outbound leg, and reinvites to the inbound leg. A ringtone service setup is also available which also involves a separate outbound call and a reinvite to the caller. Reinvites on the caller leg occur at CVP transfer or during supplementary services.

Note If the Proxy Server Record Route is set to on, it impacts the performance of the Proxy Server (as shown in the CUSP baseline matrix) and it also breaks the high availability model because the proxy becomes a single point of failure for the call. Always turn the Record Route setting of the proxy server to off to avoid a single point of failure, to allow fault tolerance routing, and to increase the performance of the Proxy Server.

A CVP call from the Proxy Server, involves four separate SIP calls on an average:

- Caller inbound leg

- VXML outbound leg
- Ringtone outbound leg
- Agent outbound leg

The standard for Unified CVP and CUSP proxy sizing is to define four SIP calls for every one CVP call, so the CPS rate is 500 / 4 = 125. The overall number of active calls is a function of Call Rate (CPS) * call handle time (CHT). Assuming an average call center call duration of 180 seconds (CHT), you get an overall active call value of 22,500 calls. Because one Call Server can handle approximately 900 simultaneous calls, it allows a single CUSP proxy to handle the load of 18 CVP Call Servers. A customer deployment should include consideration of the CPS and the CHT to size the proxy for their solution.

The standard for Unified CVP and CUSP proxy sizing is to define four SIP calls for every one CVP call, so the CPS rate is 500 / 4 = 125. The overall number of active calls is a function of Call Rate (CPS) * call handle time (CHT). Assuming an average call center call duration of 180 seconds (CHT), you get an overall active calls value of 22,500 calls. Because one Call Server can handle approximately 900 simultaneous calls, it allows a single CUSP proxy to handle the load of 18 CVP Call Servers. A customer deployment should include consideration of the CPS and the CHT to size the proxy for their solution.

Unified CVP Video Service

Cisco Unified CVP release 7.0 introduced capabilities for video-capable agents of Cisco Unified Contact Center Enterprise (Unified CCE).

The same Unified CVP Call Server can be used to service both video calls and traditional audio calls as long as the audio calls are handled using the Unified CVP comprehensive call flow. If any model other than the Comprehensive Model is used for the audio calls, then separate Call Servers must be used for the video and audio calls.

Basic Video Service Sizing

The Unified CVP Basic Video Service uses the comprehensive call flow, and it requires Call Server, VXML Server, and IOS VoiceXML Gateways. Sizing of these components for the Basic Video Service is done in the same manner as for traditional audio applications.

Cisco Unified Video conferencing hardware, Radvision IVP, and Radvision iContact are not required for the Basic Video Service.

Note Video call is not supported in Cisco VVB.

Reporting Server Sizing

There are many variables to consider for sizing the Reporting Server. Different VoiceXML applications have different characteristics, and those characteristics influence the amount of reporting data generated. Some of these factors are:

- The types of elements used in the application

- The granularity of data required
- The call flow that the users take through the application
- The length of calls
- The number of calls

To size the Reporting Server, you must first estimate how much reporting data is generated by your VoiceXML application. The example applications and the tables in subsequent sections of this chapter help you to determine the number of reporting messages generated for your application.

Once you have determined the number of reporting messages generated by your application, complete the following steps for each VoiceXML application:

1. Estimate the calls per second that the application receives.
2. Estimate the number of reporting messages for your application.

Use the following equation to determine the number of reporting messages generated per second for each VoiceXML application:

A# = %CPS * CPS * MSG

Where:

- A# is the number of estimated reporting messages per second for an application. Complete one calculation per application (A1, A2, …, An).
- CPS is the number of calls per second.
- %CPS is the percentage of calls that use this VoiceXML application.
- MSG is the number of reporting messages this application generates. To determine the number of reporting messages generated by your application, use the information provided in the sections on Reporting Message Details and Example Applications, on page 68.

Next, estimate the total number of reporting messages that your deployment generates per second by adding the values obtained from the previous calculation for each application:

A(total) = A1+ A2+…..+An

This is the total number of reporting messages generated per second by your VoiceXML applications. Each Reporting Servers can handle 420 messages per second. If the total number of reporting messages per second for your deployment is less than 420, you can use a single Reporting Server. If the number is greater, you need to use multiple Reporting Servers and partition the VoiceXML applications to use specific Reporting Servers.

Multiple Reporting Servers

If the number of messages per second (as determined in steps 1 and 2 in the previous section) exceeds the Reporting Server capacity, then the deployment must be partitioned vertically.

When vertically partitioning to load balance reporting data, a Unified CVP system designer must consider the following requirements that apply to deployments of multiple Reporting Servers:

- Each Call Server and VXML Server can be associated with only one Reporting Server.

- Reports cannot span multiple Informix databases.

For more information on these requirements, see the *Reporting Guide for Cisco Unified Customer Voice Portal* available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html

When designing Unified CVP deployments with multiple Reporting Servers, observe the following guidelines:

- Subdivide applications that generate more combined call processing and application messages than are supported by one Reporting Server.
- VoiceXML can be filtered, and filtering out noninteresting data creates more usable data repositories that support higher message volume.
- Configure the dial plan and other available means to direct the incoming calls to the appropriate Call Server and VXML Server.

If you need to combine data from multiple databases, you can use these possible options:

- Exporting reporting data to Excel, comma-separated values (CSV) files, or another format that allows data to be combined outside of the database
- Exporting reporting data to CSV files and importing it into a customer-supplied database
- Extracting data to a customer-supplied data warehouse and running reports against that data

Reporting Message Details

The following table lists the various elements or activities and the number of reporting messages generated by them.

Table 3: Number of Reporting Messages per Element or Activity

Element or Activity	Number of Reporting Messages (Unfiltered)
Start	2
End	2
Subflow Call	2
Subflow Start	2
Subflow Return	2
Throw	2
Alert	2
Subdialog_start	2
Subdialog_return	2
Hotlink	2

Element or Activity	Number of Reporting Messages (Unfiltered)
HotEvent	2
Transfer w/o Audio	2
Currency w/o Audio	2
Flag	2
Action	2
Decision	2
Application Transfer	2
VoiceXML Error	2
CallICMInfo (per call)	2
Session Variable (per change)	2
Custom Log (per item)	2
Play (Audio file or TTS)	2
LeaveQueue	2
Callback_Disconnect_Caller	3
Callback_Add	4
Callback_Get_Status	4
Callback_Set_Queue_Defaults	4
Callback_Update_Status	4
Callback_Enter_Queue	5
Callback_Reconnect	5
Get Input (DTMF)	5
Callback_Validate	6
Get Input (ASR)	9
Form	10
Digit_with_confirm	20
Currency_with_confirm	20
ReqICMLabel	30

Note These elements are required in every application and cannot be filtered.

Example Applications

This section presents some examples of applications to estimate the number of reporting messages that are generated by your particular application.

Low Complexity

Total: 16 reporting messages per call.

Table 4: Example: Applications with Low Complexity

Element Type	Approximate Number of Reporting Messages
Start	2
Subdialog_start	2
Play element	2
Play element	2
Play element	2
Play element	2
Subdialog_end	2
End	2

Medium Complexity DTMF Only

Total: 51 reporting messages per call.

Table 5: Example: Applications with Medium Complexity Using ASR

Element Type	Approximate Number of Reporting Messages
Start	2
Subdialog_start	2
Play element	2
Get input	9
Play element	2
Get input	9

Element Type	Approximate Number of Reporting Messages
Form	10
Input	9
Transfer with audio	2
Subdialog_end	2
End	2

Medium Complexity Using Automatic Speech Recognition

Total: 39 reporting messages per call.

Table 6: Example: Applications with Medium Complexity DTMF Only

Element Type	Approximate Number of Reporting Messages
Start	2
Subdialog_start	2
Play element	2
Get input	5
Play element	2
Get input	5
Form	10
Input	5
Transfer with audio	2
Subdialog_end	2
End	2

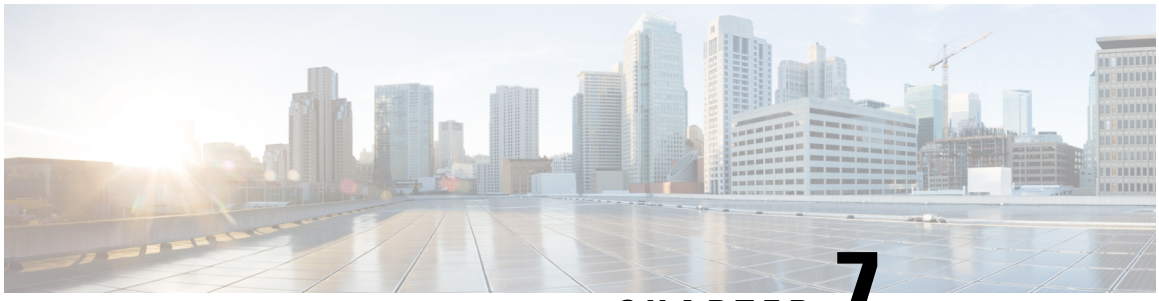
High Complexity Using Automatic Speech Recognition

Total: 107 reporting messages per call.

Table 7: Example: Applications with High Complexity Using ASR

Element Type	Approximate Number of Reporting Messages
Start	2
Subdialog_start	2
Icmrequrestlabel	30

Element Type	Approximate Number of Reporting Messages
Form	10
ASR capture	9
Digit with confirm	20
Form	10
Digit with confirm	20
Subdialog_end	2
End	2



CHAPTER 7

Design Considerations

Unified CVP Algorithm for Routing

When you set up a dial plan and call routing, you can combine Unified CVP features (such as Location Based CAC, SigDigits, SendToOriginator, LocalSRV, and Use Outbound Proxy) to achieve the required effect.

The following algorithms are used to formulate the destination SIP URI for the outbound calls from Unified CVP. This description covers CONNECT messages that include labels from the ICM (for example, VXML Gateway, and Unified Communications Manager), as well as calls to the ringtone service, recording servers, and error message playback service.

Note The following algorithm only describes calls using the SIP subsystem, which includes audio only and basic video SIP calls.

The sendtoriginator algorithm is supported only for co-located IOS VoiceXML Gateway and Ingress Voice Gateway. The sendtoriginator algorithm is not supported in Cisco VVB as the co-location concept is not applicable.

The algorithm for creating the destination SIP URI host portion for outbound calls, which include the ICM label, is as follows:

1. The ICM label is provided at the start of the algorithm. It may already have the Location siteID inserted by the ICM subsystem, or SigDigits may be prepended if used. For network VRU labels, the ICM subsystem passes in the entire prefix and correlation ID as the label.
2. If SendtoOriginator is matched for the Unified CCE label, the IP or hostname of the caller (Ingress Voice Gateway) is used by the Unified CVP algorithm, which returns the SIP URI.

The setting for SendtoOriginator only applies to callers on Cisco Ingress Voice Gateways (the SIP UserAgent header is selected), because non-Cisco IOS Gateways do not have the CVP bootstrap service used by the Cisco IOS VoiceXML Gateway.

3. If **use outbound proxy** is set, then use the host of the proxy and return SIP URI.
4. If **local static route** is found for the label and return the SIP URI.

Note If **local static route** is not found, the algorithm throws **RouteNotFoundException** exception.

The following algorithm describes planning considerations for calls using the SIP subsystem:

- To avoid complex Dialed Number strings, do not use the Sigdigits feature if Locations CAC siteIDs are used.
- An Outbound Proxy FQDN can be specified as a Server Group FQDN (local SRV FQDN). A local static route destination can also be configured as a Server Group FQDN.
- Ringtone DN (91919191), Recording Server (93939393), and Error message services (92929292) follow the same algorithm as mentioned in the procedure.
- SendToOriginator can work in conjunction with a REFER label.
- A REFER label can work with the SigDigits setting.

Distributed Network Options

After choosing a functional deployment model, you must determine where the Unified CVP components are deployed. Unified CVP deployment can use one of the following primary distributed network options:

- Combined Branch Gateways—Enables call treatment at the edge and integration of locally dialed numbers into the enterprise virtual contact center. This option can be either a combined Ingress and IOS VoiceXML Gateway, or separate gateways. Typically, both the Ingress and VoiceXML Gateways are combined when deployed in a branch. Combined Ingress and VoiceXML Gateway is available only on Cisco IOS Voice Gateway.
- Branch Ingress Voice Gateways with Centralized VoiceXML Gateways—Enables integration of locally dialed numbers and resource grouping of VoiceXML Gateways. This option can be required for organizations with many medium to large branches, with a few contact center calls in each branches. The VRU announcements in the Centralized VoiceXML Gateways traverse the WAN to the Ingress Gateway.
- Branch Egress Gateways—Enables calls to be transferred across the WAN to remote TDM terminations.
- Branch Agents—Enables a virtual contact center where agents can be located anywhere on the IP network.

You also can use a combination of these distributed options. For more details and design considerations for each of these distributed network options, see the chapter on Distributed Deployment, on page 31.

CUBE Deployment with SIP Trunks

The use of third-party SIP trunks with Unified CVP is supported by using the Cisco Unified Border Element (CUBE) product. CUBE performs the role of session border controller (SBC), for SIP normalization and interoperability.

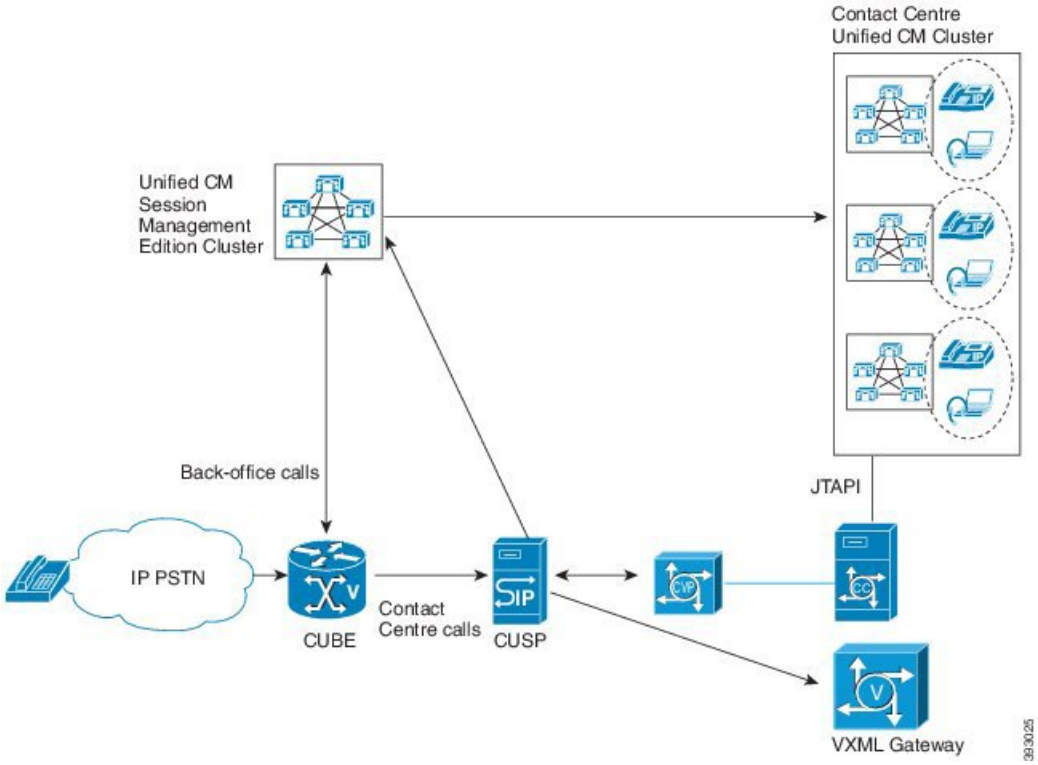
Unified CM SME Deployment

Cisco Unified Communications Manager Session Management Edition (Unified CM SME) integrates with Unified CVP as a dial peer configurator or aggregator to connect to multiple Unified Communications Manager clusters in the Cisco Unified Customer Voice Portal (Unified CVP) and Cisco Unified Contact Center Enterprise (Unified CCE) solution setup.

Unified CM SME as a back-to-back SIP user agent is configured to act as an aggregator that connects to multiple Unified Communications Manager clusters and routes the call to the appropriate cluster based on the dial plan.

The following figure illustrates the Unified CM SME deployment.

Figure 9: Unified CM SME Deployment



Unified CM SME does not support high-availability and is a single source of failure. Following are the design considerations to minimize the effect of Unified CM SME failure (either network connectivity failure or actual component failure).

- Deploy Unified CM SME in redundant clustered mode (at least 1+1 publisher subscriber) at the egress side of Unified CVP.
- Deploy Cisco Unified SIP Proxy Server (CUSP) in the egress leg between Unified CVP and Unified CM SME. This configuration is required to determine the Unified CM SME status and to hold the call session during Unified CM SME failure.
- You must configure **Session Refresh** and **Session Timer** in the Gateway/Cisco UBE. This configuration is required to clear call sessions form the gateway and to release Unified CVP Call Server ports in case of Unified CM SME failure.
- In case of Unified CM SME failure, all the call server ports are cleared after the customer drops the call.

Note Call supplementary services will not work for the already established calls once the Unified CM SME is down.

Momentary network connectivity failure to Unified CM SME results in the following limitations:

- Unified CM SME does not clear the call, when the agent hangs up the call during momentary connectivity failure to Unified CM SME. This results in a stale cached entry and ports hanging in the Unified CVP application. In such cases, the caller should drop the call to clear the stale cached entry.
- The call does not get cleared from the agent desktop and the agent will be unable to receive any incoming calls. As a result the agent remains in the talking state and is unable to clear the call from the desktop. In such cases, the call has to be manually cleared from the device or hard phone.
- Because of a delay in call clearance, the call reporting data may reflect inaccurate details for call duration and reason code.

For more information about Unified CM SME Configuration, see *Configuration Guide for Cisco Unified Customer Voice Portal* available at: http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-installation-and-configuration-guides-list.html.

CUBE or SME Deployment in Between Unified CVP and Unified CM

When CUBE or SME is deployed in between Unified CVP and Unified CM on the outbound leg, all SIP connections between Unified CVP and CUBE or SME must use TCP transport.

Scalability

After choosing the functional model and the distributed and high-availability deployment options, Unified CVP solution designers must then size their solution and select appropriate hardware. To make Unified CVP deployments larger, Unified CVP supports multiple gateways, Unified CVP Servers, and Unified CVP VXML Servers.

Note Unified CVP Servers contain the VXML Server component.

To load balance HTTP requests efficiently to multiple Unified CVP Servers, Unified CVP VXML Servers, and media stores, you can use the Application Control Engine (ACE).

For more details on choosing appropriate hardware for your deployment, see the chapter about Sizing.

Virtualization

Unified CVP may be installed and run on virtual machines (VMs) provided by VMware software. Running in a virtual environment has the potential for reducing the number of hardware equipments needed to run a Unified CVP deployment, to facilitate the deployment's administration, and to leverage your ESX infrastructure.

The following Unified CVP deployments are supported using VMware VMs:

- All SIP call flows, deployments, and features that could be installed on a physical server
- Mixed environments of physical and virtual servers

Note Deployments assume that you do not oversubscribe or overcommit the CPU and memory resources beyond what is available on the host.

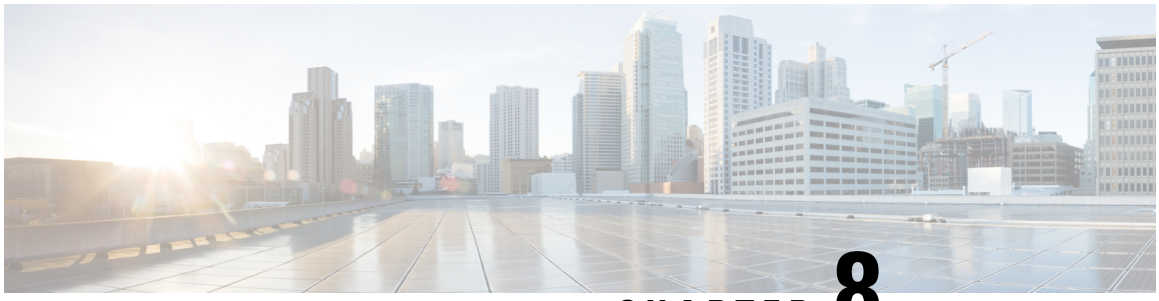
For specific information about virtualization with Unified CVP, see http://www.cisco.com/go/uc-virtualized.

Quality of Service

The quality of service (QoS) is the measure of transmission quality and service availability of a network. Unified CVP implements Layer 3 QoS defaults on all relevant network paths. Unified CVP provides a management interface through the Unified CVP Operations Console Server to modify QoS settings at each end of specifically designated data paths.

Note For instructions on configuring QoS for Unified CVP, see the *Operations Console Online Help.*

For QoS design information, see the Enterprise QoS in the *Solution Reference Network Design Guide.*



CHAPTER 8

Gateway Options

PSTN Gateway

In this type of deployment, the Ingress Voice Gateway is used as the PSTN Voice Gateway. The Ingress Voice Gateway is responsible for converting TDM speech to IP and for recognizing DTMF digits and converting them to RFC2833 events.

Note Unified CVP does not support passing SIP-Notify DTMF events.

In a centralized Unified CVP deployment, you can separate the VoiceXML functionality from the Ingress Voice Gateway to provide a separate PSTN ingress layer. The separate PSTN layer and VoiceXML enable the deployment to support many VoiceXML sessions and PSTN interfaces. For example, the Cisco AS5400XM can accept a DS3 connection, and support up to 648 DSOs. However, a gateway that is handling that many ingress calls cannot also support as many VoiceXML sessions. In such cases, the VoiceXML sessions are off-loaded to a separate farm of only VoiceXML Gateways, such as Cisco VVB.

Note You can use any TDM interface, supported by Cisco IOS gateway and Cisco IOS version, and compatible with Unified CVP.

VoiceXML Gateway with DTMF or ASR/TTS

The VoiceXML Gateway allows you to interact with the VoiceXML browser through DTMF tones or ASR/TTS. Because the gateway does not have PSTN interfaces, voice traffic is sent using Real-Time Transport Protocol (RTP) to the VoiceXML Gateway, and the RFC 2833 uses in-band signaling in RTP packets to transmit DTMF tones. An VoiceXML with DTMF or ASR and TTS allows you to increase the scale of the deployment and support hundreds of VoiceXML sessions.

In a centralized Unified CVP deployment, you could use a VoiceXML farm. For example, if you want to support 300 to 10,000 or more VoiceXML sessions, use the Cisco AS5350XM Gateway. The standalone AS5350XM can support many DTMF or ASR/TTS VoiceXML sessions per Voice Gateway. In addition, stack the AS5350XM Gateways to support large VoiceXML IVR farms. However, for better performance and higher capacity, and to avoid the need for stacking, you can use the 3945 or 3945-E Series Gateways.

In a distributed Unified CVP deployment, consider providing an extra layer of redundancy at the branch office. You can deploy a separate PSTN Gateway and a VoiceXML Gateway to provide an extra layer of redundancy. In addition, for a centralized Cisco Unified Communications Manager deployment, you must support Survivable Remote Site Telephony (SRST). The Cisco 2800 Series and 3800 Series Integrated Service routers and the newer 2900 Series and 3900 Series routers are the best choices for the Ingress Voice Gateway because they support SRST.

For a discussion of the advantages and disadvantages of each codec, see Voice Traffic, on page 140.

VoiceXML and PSTN Gateway with DTMF or ASR/TTS

The most popular Ingress Voice Gateway is the combination VoiceXML Gateway and PSTN Interface Gateway. For a centralized Cisco Unified Communications Manager deployment, Survivable Remote Site Telephony (SRST) must be supported. The Cisco 2800 Series, 3800 Series Integrated Service routers, 2900 Series, and 3900 Series routers are the best choices for the Ingress Voice Gateway, because they support SRST.

TDM Interfaces

The Cisco AS5400XM Universal Gateway offers unparalleled capacity in only 2 rack units (2 RUs) and provides best-of-class voice, fax, and remote-access services.

The Cisco AS5350XM Universal Gateway is the one rack-unit (1 RU) gateway that supports 2-, 4-,8-, or 16-port T1/12-port E1 configurations and provides universal port data, voice, and fax services on any port at any time. The Cisco AS5350XM Universal Gateway offers high performance and high reliability in a compact, modular design. This cost-effective platform is ideally suited for internet service providers (ISPs) and enterprise companies that require innovative universal services.

For the most current information about the various digital (T1/E1) and analog interfaces supported by the various voice gateways, see the latest product documentation available at the following sites:

- Cisco 2800 Series

  http://www.cisco.com/en/US/products/ps5854/tsd_products_support_series_home.html

- Cisco 3800 Series

http://www.cisco.com/en/US/products/ps5855/tsd_products_support_series_home.html

- Cisco AS5300

  http://www.cisco.com/en/US/products/hw/univgate/ps501/tsd_products_support_series_home.html

- Cisco 2900 Series
- http://www.cisco.com/en/US/products/ps10537/index.htm.
- Cisco 3900 Series
- http://www.cisco.com/en/US/products/ps10536/index.htm

Cisco Unified Border Element

The Cisco Unified Border Element (CUBE), formerly known as the Cisco Multiservice IP-to-IP Gateway is a session border controller (SBC) that provides connectivity between IP voice networks using SIP. The CUBE is supported in flow-through mode only so that all calls are routed through the CUBE.

Note Unlike flow-through mode, with flow-around mode you lose the ability to do DTMF interworking, transcoding, and other key functions, such as phone and media capabilities.

A Unified Border Element is needed when replacing a TDM voice circuit with an IP voice trunk, such as a SIP trunk, from a phone company. The CUBE serves as a feature demarcation point for connecting enterprises to service providers over IP voice trunks.

Note For outbound calls, CUBE supports Call Progress Analysis (CPA) on TDM circuits.

The CUBE has been tested with, and can be used in, any of the following scenarios:

- SIP-to-SIP connectivity between a third-party SIP device and Cisco Unified CVP over the SIP trunks certified by Cisco.
- SIP-to-SIP connectivity between Cisco Unified Communications Manager and Cisco Unified CVP.
- Coresidency of VoiceXML Gateway and CUBE for any of the above scenarios but with the limitation that the call flow does not work when the configurations listed here occur at the same time on the CUBE:
  - Survivability TCL script and incoming translation rules are configured under the same incoming dial-peer.
  - Transcoding between G.711 and G.729.
  - Header-passing between the call legs is enabled globally.

For CUBE session numbers, refer to:

http://www.cisco.com/en/US/prod/collateral/voicesw/ps6790/gatecont/ps5640/order_guide_c07_462222.html

For more information about using the CUBE with Unified CVP, including topologies and configurations, see *Cisco Unified Border Element for Contact Center Solutions* available at:

http://cisco.com/en/US/docs/voice_ip_comm/unified_communications/cubecc.html

Note Due to a limitation in Cisco IOS, the CUBE does not support midcall escalation or deescalation from audio to video, and conversely.

Using a SIP Trunk Without CUBE

When you connect to a third-party SIP device, including a SIP PSTN service provider, use a CUBE. If you do not place a CUBE between Unified CVP and the SIP device, ensure that both sides are compatible with thorough integration testing.

When connecting to a PSTN SIP Trunking service without a CUBE, carefully consider how to secure the connection between the contact center and the service provider. Also consider how to accomplish NAT and address hiding. Otherwise, the service-provider network can have full access to the contact center network. As the service-provider interconnect interface provided by Cisco, CUBE addresses both of these concerns.

Using Cisco ASR 1000 Series as Unified Border Element

Unified CVP supports Cisco IOS XE software Release 3.3.0S Enterprise with the following limitations:

- ASR 1000 Series do not support VXML. As a result, the VRU leg of the call must be routed to a separate VoiceXML Gateway. You must not use the **Send To Originator** setting on the CVP Call Server to route the IVR leg of the call back to the originating ASR CUBE Gateway, and standalone CVP calls must be routed to a separate VoiceXML Gateway.
- The global **Pass Thru SDP** setting on the ASR 1000 Series gateways is not supported with CVP deployments.
- ASR 1000 Series gateways do not support the TCP transport with SIP signaling when using the box-box hardware redundancy feature. The UDP transport is supported when failing the active ASR chassis to the standby chassis. It is important to note that the default TCP setting will not work with failover in this version of the ASR release. Therefore, UDP must be used on both the incoming and outgoing legs of the ASR CUBE for uninterrupted call control with CVP.
- Using the proxy servers to perform UDP to TCP Up-Conversion when receiving large size packet SIP messages, in a scenario where the proxy is in front of the ASR session border controller, the proxy servers should be turned off to ensure that UDP transport is used for the connection on the inbound call. Typically, a proxy server is positioned behind the session border controller in the deployment.
- A **sip-profile** configuration is needed on ASR 1000 Series for the courtesy callback feature only when deploying an IOS-XE version affected by CSCts00930. For more information on the defect, access the Bug Search Tool at https://sso.cisco.com/autho/forms/CDClogin.html. To configure the **sip-profile**, the following must be added:

  voice class sip-profiles 103

  request INVITE sip-header Call-Info add "X-Cisco-CCBProbe: <ccb param>"

ccb param is the **ccb** parameter defined in the survivability service. Add this sip-profile to the outgoing dial-peer to the CVP.

The following is a configuration example:

voice class sip-profiles 103

hoigogpoupcoioivc9iu i 8s66d8 8hxiciuvyd78zicvc8ayge

request INVITE sip-header Call-Info add "X-Cisco-CCBProbe:

id:192.168.1.50;loc:testbed04;trunks:10"

application

service survivability flash:survivability.tcl

param ccb id:192.168.1.52;loc:testbed04;trunks:10

dial-peer voice 700051 voip

description Comprehensive outbound route to CVP

destination-pattern 7000200T

session protocol sipv2

session target ipv4:192.168.1.20:5060

dtmf-relay rtp-nte

voice-class sip profiles 103

codec g711ulaw

no vad

- The following Survivability.tcl options are not applicable for use on the ASR because they are traditionally for POTS dial peers:
  - ani-dnis-split.
  - takeback-method.
  - -- *8.
  - -- hf.
  - icm-tbct.
  - digital-fxo.
- The following Survivability.tcl options are not supported: aa-name, standalone, and standalone-isntime.
  - The aa-name option is not supported because CME auto-attendant service is not supported on ASR.
  - The standalone and standalone-isntime options are not supported because there is no support for VXML on ASR.
- Due to ASR limitations, the following features are not supported:
  - Refer with Re-query
  - Legacy Transfer Connect using DTMF *8 label

- ASR 1000 does not terminate the TDM trunks. Therefore, the following TDM Gateway features do not apply to ASR 1000:
  - PSTN Gateway trunk and DS0 information for SIP calls to ICM
  - Resource Availability Indication (RAI) of DS0 trunk resources via SIP OPTIONS message to ICM

Note Because ASR 1000 represents the introduction of new equipment, to ensure success of ASR 1000 deployments, any UCCE/CVP contact center integration that utilizes the ASR 1000 requires an Assessment to Quality (A2Q) review. This review is required for new UCCE customers as well as existing UCCE customers who want to upgrade to the ASR 1000.

Note The Courtesy Call Back call flow does not work if ASR as CUBE is configured for the media flow-around instead of the media flow-through.

Using Cisco ISR as Unified Border Element

Unified CVP supports ISR with the following limitations:

- A **sip-profile** configuration is needed on ISR for the courtesy callback feature only when deploying an IOS-XE version affected by CSCts00930. For more information on the defect, access the Bug Search Tool at https://sso.cisco.com/autho/forms/CDClogin.html. To configure the **sip-profile**, the following must be added:

  voice class sip-profiles 103

  request INVITE sip-header Call-Info add "X-Cisco-CCBProbe: <ccb param>"

  ccb param is the "ccb" parameter defined in the survivability service. Add this **sip-profile** to the outgoing dial peer to the CVP.

  The following is a configuration example:

  voice class sip-profiles 103

  request INVITE sip-header Call-Info add "X-Cisco-CCBProbe:

  id:192.168.1.50;loc:testbed04;trunks:10"

  application

  service survivability flash:survivability.tcl

  param ccb id:192.168.1.52;loc:testbed04;trunks:10

  dial-peer voice 700051 voip

  description Comprehensive outbound route to CVP

  destination-pattern 7000200T

  session protocol sipv2

  session target ipv4:192.168.1.20:5060

dtmf-relay rtp-nte

voice-class sip profiles 103

codec g711ulaw

no vad

Note

- For ISR versions, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-technical-reference-list.html.
- The Courtesy Call Back call flow does not work if ISR as CUBE is configured for the media flow-around instead of the media flow-through.

Mixed G.729 and G.711 Codec Support

Transcoders (DSPs) are required if the two endpoints participating in the call cannot negotiate a common codec. Therefore, midcall codec negotiation greatly reduces the need for transcoders.

CVP supports mixed G.711 and G.729 codecs in Standalone and Comprehensive SIP deployments with Cisco Unified Border Element Enterprise Edition (CUBE) and Cisco Unified Communications Manager (Unified CM). Calls that are ingressed through a SIP trunk from the carrier to a CUBE require Cisco IOS 15.1(2)T or later for mixed codec support. You can use any combination of codecs on the legs of a call. For example, a caller can place a call using the G.729 codec, hear an IVR prompt played using the G.711 codec, be transferred to the first Agent using the G.729 codec, and then transferred to the second agent using the G.711 codec.

A typical use case where transcoders may be required is when phones in a WAN connected location only support the G729 codec, and CVP is set up for G711 support. In this case, when these phones call into CVP, Unified Communications Manager engages transcoders. For inbound calls that arrive from a gateway or CUBE can start with G711 at CVP, then later renegotiate to G729 with the agents without the need for transcoders.

Transcoders (DSPs) are controlled by CUBE and Unified Communications Manager depending on the call flow. Because most of the service providers support midcall codec negotiation, transcoders in CUBE are not necessary. You commonly need transcoders controlled by Unified Communications Manager to support call flows, in which the phone supporting G729 is calling into CVP supporting G711.

Generate G729 Prompts for Unified CVP

To generate the G.729 prompts for Unified CVP, perform the following procedure:

- Convert the audio files from G.711 to G.729 format using the Music on Hold (MOH) audio translator.
- Change the G.729 compression identifier in the file header.

Convert the Audio Files from G.711 to G.729 Format

Procedure

Step 1 Log in to the Cisco Unified CM Administration portal and select **Media Resources** > **MOH Audio File Management**.

Step 2 Click **Upload File** and select the G.711 audio files individually.

Step 3 Click **Media Resources** > **MOH Audio File Management** and check whether the audio files have been converted to G.729 format. If the conversion was successful, the recording length of audio files has a nonzero value.

Step 4 Copy the converted audio files to your Windows server using the Secure File Transfer Protocol (SFTP) Server.

Note Do not add spaces when you rename the audio files.

Step 5 Use putty to sign in to the Unified Communications Manager Server as an administrator.

Step 6 From the command prompt, run **file get activelog mohprep/*g729.wav** and provide the SFTP prompts.

Change the G.729 Compression Identifier in the File Header

The G.729 files that the Unified Communications Manager generates have a non-standard compression codec tag in the file header. The VXML Gateway cannot play these audio files, as it does not recognize the codec type. Change the compression codec type value to convert the audio files into the standard G729r8 format.

Use the following procedure to change the compression codec type number in the file header from 0x0133 to the standard 0x14db, G729r8 format.

Procedure

Step 1 Create a folder in the Unified CVP directory. Copy the G.729 audio files that have a nonstandard compression codec tag in the file header into the new folder location.

Step 2 From the command prompt, navigate to the `C:\Cisco\CVP\bin` folder.

Step 3 Perform one of these steps:

- To convert audio files individually, from the command prompt, run **<UCMHeaderFixer.exe Audio file Name>*.***.
- To perform bulk conversion of audio files, from the command prompt, run **UCMHeaderFixer.exe Folder Path**.

The script runs and the audio file is converted from name.g729.wav file into name.wav format.

Step 4 Use the Operations Console to upload the converted audio files to the IOS Gateway.

ISO Gateway Choices

Unified CVP uses IOS Gateways for two purposes: TDM ingress and VoiceXML rendering. Any Cisco gateway supported by Unified CVP can be used for either purpose or both. However, depending on your deployment model, you can use one of the following functions:

- Model #1: Standalone Self-Service

  All calls use both ingress and VoiceXML.

- Model #2: Call Director

  All calls use ingress only.

- Model #3a: Comprehensive Using Unified ICM Micro-Apps

  All calls use ingress, and some calls use VoiceXML.

- Model #3b: Comprehensive Using Unified CVP VXML Server

  All calls use ingress, and some calls use VoiceXML.

- Model #4: VRU Only with NIC Controlled Routing

  All calls use both ingress and VoiceXML.

In cases where both Ingress and VoiceXML are required, you can choose to run both functions on the same gateways or you can choose to designate some gateways for ingress and others for VoiceXML. Use the following guidelines to determine whether the functions should be combined or split:

- In classical branch office deployments, where the call is queued at the branch where it arrived, ingress and VoiceXML functions must always be combined.
- In cases where many non-CVP PSTN connections share the gateways, it is submitted to dedicated Ingress for that purpose, and use separate VoiceXML Gateways.
- VoiceXML-only gateways are less costly because they do not require DSP farms or TDM cards. Use a spreadsheet to determine which way you obtain the best price.
- With relatively low call volume, it is usually better to combine the functions for redundancy purposes. Two combined gateways are better than one of each because the loss of one gateway still allows calls to be processed, though at a lower capacity.

The next decision is whether to use Cisco Integrated Service Router (ISR) Gateways (Cisco 2800 or 3800 series routers), ISR-G2 (2900 or 3900 Series routers), or the Cisco AS5x00 Series Gateways. ISR Gateways are used only in branch office sites and AS5x00 Series Gateways are used in centralized data center sites.

You might sometimes have difficulty determining what constitutes a branch office, and which gateway is used. The following guidelines can help with that determination:

- The classical definition of branch offices, for which you must use ISR Gateways, includes:
  - Multiple sites where TDM calls arrive from the PSTN.
  - Those sites are separated from the data centers where most of the Unified CVP equipment resides.
  - One gateway is used at each site.

- If you have sites where you are stacking multiple gateways for any reason, then those sites are data center sites and should use Cisco AS5x00 Series Gateways.

For more information on the Cisco AS5x00 Series Gateways, refer to the technical specifications available at http://www.cisco.com/en/US/products/hw/univgate/ps501/index.html.

For more information on the Cisco Integrated Service Routers (ISRs), refer to the documentation available at http://www.cisco.com/en/US/products/hw/routers/index.html.

IOS Gateway Sizing

Individual Cisco gateways can handle various call capacities depending on whether they are doing ingress only, VoiceXML only, or a combination of the two. IOS Voice Gateways doing VoiceXML activities also have different call capacities depending on whether they are supporting ASR or TTS activities, and on the type of VoiceXML application being executed. For instance, an intensive JavaScript application reduces call capacity. Gateways using HTTPS, have lower call capacity as compared to HTTP.

In general, gateways that perform ingress only can be sized according to the number of TDM cables that can be connected to them. For gateways that are combined or VoiceXML-only, it is important to ensure that the overall CPU usage is less than 75 percent on average. The numbers in the Maximum Number of VoiceXML Session tables are based on Unified CVP VoiceXML documents; other applications that generate more complex VoiceXML documents have a higher impact on performance. The following factors affect CPU usage:

- Calls per second (CPS)
- Maximum concurrent calls
- Maximum concurrent VoiceXML sessions

Before sizing the IOS Voice Gateways, use the Unified CCE Resource Calculator to determine the maximum number of trunks (DS0s) and VoiceXML IVR ports required to support the entire solution.

For almost all Unified CVP deployment models, sizing is based on the maximum number of concurrent VoiceXML sessions and VoIP calls. The following tables list this information for different versions of Cisco IOS.

Note The performance numbers listed in the Table 9, Table 10, and Table 11 are equivalent for MRCPv1 and MRCPv2.

Table 8: Maximum Number of VoiceXML Sessions Supported by Cisco Voice Gateways (Cisco IOS Release 15.1.4.M7 and Later)

VoiceXML Gateway CPU Capacity for Cisco IOS Release 15.1.4.M7 or Later					
Platform	**VXML Only**		**VXML + PSTN**		**Memory**
	DTMF	**ASR**	**DTMF**	**ASR**	**Recommended**
5000XM	200	135	155	104	512 MB
3825	130	85	102	68	512 MB

VoiceXML Gateway CPU Capacity for Cisco IOS Release 15.1.4.M7 or Later					
Platform	**VXML Only**		**VXML + PSTN**		**Memory**
	DTMF	**ASR**	**DTMF**	**ASR**	**Recommended**
3845	160	105	125	83	512 MB
2901	12	8	9	6	2 GB
2911	60	40	47	31	2 GB
2921	90	60	71	48	2 GB
2951	120	80	95	64	2 GB
3925	240	160	190	127	2 GB
3945	340	228	270	180	2 GB
3925E	475	450	380	375	2 GB
3945E	580	550	460	450	2 GB
Based on ISO 15.1.4.M7, G.711, basic calls, Ethernet egress, CPU NTE 75% (5000XM 80%)					

Table 9: Maximum Number of VoiceXML Sessions Supported by Cisco Voice Gateways Executing Intensive JavaScript Applications (Cisco IOS Release 15.1.4.M7 and Later)

Cisco Voice Gateway Platform	**Dedicated VoiceXML Gateway**		**Voice Gateway and VoiceXML**		
	VoiceXML and DTMF	**VoiceXML and ASR/TTS**	**VoiceXML and DTMF**	**VoiceXML and ASR/TTS**	**Memory Recommended**
AS5350XM	105	85	110	70	512 MB (default)
AS5400XM	105	85	110	70	512 MB (default)

Table 10: Maximum Number of VoiceXML Sessions Supported by Cisco Voice Gateways Using HTTPS (Cisco IOS Release 15.1.4.M7 and Later)

Cisco Voice Gateway Platform	**Dedicated VoiceXML Gateway**		**Voice Gateway and VoiceXML**		
	VoiceXML and DTMF	**VoiceXML and ASR/TTS**	**VoiceXML and DTMF**	**VoiceXML and ASR/TTS**	**Memory Recommended**
3945E	510	342	408	270	2 GB
AS5350XM	155	120	138	95	512 MB (default)
AS5400XM	155	120	138	95	512 MB (default)

Note The performance numbers listed in Table 11 are only for selected models of Cisco Voice Gateways using HTTPS. Use the HTTPS performance numbers of the 3945E router, to estimate the performance numbers for router models that are not listed in Table 11.

Note Performance numbers for the Cisco 3825 Series and 3845 Series Integrated Services Routers (ISRs) are higher when the Ingress Voice Gateway and the VoiceXML Gateway functions reside on the same router (coresident deployment). When the call is connected to the VoiceXML Gateway from the Ingress Voice Gateway, the media flows directly between the two. In a coresident deployment, the gateway does not have to spend CPU cycles to packetize and de-packetize the RTP packets. Hence, by saving these CPU cycles, the gateway can support increased VoiceXML sessions.

This note does **not** apply to Cisco IOS Release 15.0.1M and Cisco IOS 15.1.4.M7.

The numbers in Table 9, Table 10, and Table 11 assume that the only activities running on the gateway are VXML with basic routing and IP connectivity. If you intend to run extra applications such as fax, security, and normal business calls, then the capacity numbers presented here should be prorated accordingly. The numbers mentioned in the Voice Gateway and VoiceXML column mean that the indicated number of VoiceXML sessions and voice calls can be supported simultaneous on the same gateway. For example, in Table 9 the 500XM can terminate a maximum of 200 PSTN calls, and those 200 PSTN calls could have 200 corresponding VoiceXML sessions at the same time.

The numbers represent performance with scripts generated by Unified CVP Studio running on the Unified CVP VXML Server. Other VoiceXML applications might perform differently. These figures apply if the CPU utilization does not exceed more than 75 percent Voice Activity Detection (VAD) is turned off, and your system is running VoiceXML v2.0 and MRCP v2 with Cisco IOS Release 15.1.4.M7 and later.

Note These performance numbers are accurate when used with either the Cisco Call Server or Cisco Unified CVP VXML Server. Performance can, and often does, vary with different applications. Performance from external VoiceXML applications (such as Nuance OSDMs) might not be representative of the performance when interoperating with non-Cisco applications. Ensure that the CPU usage is less than 75 percent on average and that adequate memory is available on Cisco gateways at full load when running external VoiceXML applications. Contact the application provider of the desired VoiceXML application for performance and availability information. External VoiceXML applications are not provided by Cisco, and Cisco makes no claims or warranties regarding the performance, stability, or feature capabilities of the application when interoperating in a Cisco environment.

Note Cisco does not specifically test or qualify mixes of traffic because there are infinite combinations. All numbers should be seen as guidelines only and varies from one implementation to the next based on configurations and traffic patterns. The systems are required to be engineered for worst-case traffic (all ASR) if the types of calls that are offered to the VoiceXML Gateway are not known or cannot be predicted.

If you run VoiceXML on one of the Cisco 2900 and 3900 Series gateways, more licenses (FL-VXML-1 or FL-VXML-12) are required.

Consult the following links to ensure that the concurrent call load and call arrival rates do not exceed the listed capacities:

- Model comparison:

  http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-implementation-design-guides-list.html.

- Gateway sizing for Contact Center traffic:

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

Also consider how much DRAM and flash memory to order. The capacity that comes with the machine by default is usually sufficient for most purposes. However, if your application requires large numbers of distinct .wav files (as with complex self-service applications) or if your application has unusually large .wav files (as with extended voice messages or music files), you might want to increase the amount of DRAM in order to expand your flash memory. The use of DRAM for prompt caching is discussed in detail in the chapter on Media File Options, on page 47.

Note HTTP cache can only be extended to 100 MB in the current Cisco IOS releases.

Cisco VVB Sizing

Cisco VVB has call capacities is defined based on the call support for ASR or TTS activities and on the type of VoiceXML application being executed. For instance, an intensive JavaScript application reduces call capacity and VVB using HTTPS has lower call capacity as compared to HTTP.

It is important to ensure that the overall CPU usage is less than 75 percent on average. The numbers in the Maximum Number of VoiceXML Sessions tables is based on VoiceXML documents; other applications that generate more complex VoiceXML documents have a higher impact on performance. The following factors affect CPU usage:

- Calls per second (cps)
- Maximum concurrent VoiceXML sessions
- Complexity of VoiceXML applications

Before sizing the Cisco VVB, use the Unified CCE Resource Calculator to determine the maximum number of trunks (DS0s) and VoiceXML IVR ports, which is required to support the entire solution.

For almost all Unified CVP deployment models, sizing is based on the maximum number of concurrent VoiceXML sessions and VoIP calls.

Note The following performance numbers listed in ASR and TTS columns are applicable only for MRCP v1.

Table 11: Maximum Number of VoiceXML Sessions Supported by Cisco VVB

CPU Capacity for Cisco VVB					
System Specification	**DTMF**	**ASR**	**TTS**	**HTTPS**	**Standalone java script**
4 CPU, 8-GB RAM	600	400	400	480	200

Note

- Table 11 displays the numbers that represent the performance when the activities running on the gateway are only VXML with basic routing and IP connectivity. If you intend to run more activities such as security, call tracing, monitoring, then the capacity numbers presented here should be prorated accordingly.
- The numbers also represent performance with VoiceXML pages generated by Unified CVP Call Studio applications running on the Unified CVP VXML Server. Other VoiceXML applications might perform differently. These figures apply if the CPU utilization does not exceed more than 75 percent and your system is running VoiceXML v2.0 and MRCP v1.

  These performance numbers are accurate when used with Cisco Unified CVP VXML Server and moderately complex VoiceXML applications. Performance can, and often does, vary with different applications. Performance from external VoiceXML applications (such as Nuance OSDMs) is not representative of the performance when interoperating with non-Cisco applications. Ensure that the CPU usage is less than 75 percent on average and that adequate memory is available on Cisco VVB at full load when running external VoiceXML applications. Contact the application provider of the desired VoiceXML application for performance and availability information.
- External VoiceXML applications are not provided by Cisco, and Cisco makes no claims or warranties regarding the performance, stability, or feature capabilities of the application when interoperating in a Cisco environment.
- HTTP cache can be extended to 2 GB in the current Cisco VVB releases.

Using MGCP Gateways

Cisco Unified CVP requires the deployment of a SIP Gateway. However, you require the use of MGCP 0.1 voice gateways with Unified CM deployments for purposes of overlap sending, NSF, and Q.SIG support. The following design considerations apply to deploying Cisco Unified CVP in this environment:

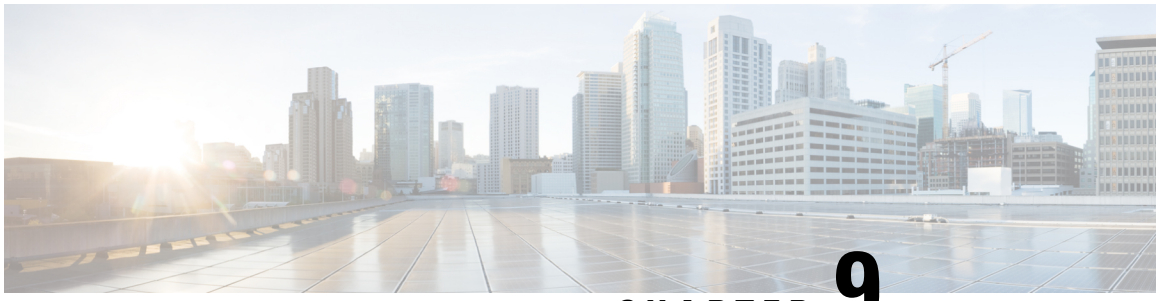
- Design and plan a phased migration of each MGCP voice gateway to SIP.
- Implement both MGCP 0.1 and SIP.

  Because of the way MGCP works, a PSTN interface using MGCP can be used for MGCP only. If you want to use MGCP for regular Unified CM calls and SIP for Unified CVP calls, you need two PSTN circuits.
- Deploy a second SIP voice gateway at each Unified CVP location.
- Send calls through the Unified CM to Unified CVP.

When sending calls through Unified CM to Unified CVP, the following guidelines apply:

- The Unified CVP survivability.tcl script cannot be used in this solution. If the remote site is disconnected from the central site, the call is dropped.
- There is an additional hit on the performance of Unified CM. This is because, in a normal Unified CVP deployment, Unified CM resources are not used until the call is sent to the agent. In this model, Unified CM resources are used for all calls to Unified CVP, even if they terminate in self-service. This is in addition to the calls that are extended to agents. If all calls are eventually extended to an agent, the performance impact on Unified CM is approximately double that of a normal Unified CVP deployment. This factor alone typically limits this scenario to small call centers only.
- In order to queue calls at the edge, use the **sigdigits** feature in Unified CVP to ensure that the calls are queued at the appropriate site or VoiceXML Gateway. For more information on how the **sigdigits** feature works, see the chapters on Distributed Deployment, on page 31 and Unified CVP Design for High Availability, on page 93.

Note The Cisco Unified CVP provides the flexibility to add, modify, remove, or deploy Unified CVP in many scenarios to facilitate interoperability with third-party devices. Not all SIP service providers support advanced features such as REFER, 302 Redirect Messages, DTMF-based take-back-and-transfer, or data transport (UUI, GTD, NSS, and so forth). Refer to the interoperability note available at the following location for information on the interoperability support for SBC when deployed in place of Cisco CUBE, http://www.cisco.com/en/US/solutions/ns340/ns414/ns728/voice_portal.html



CHAPTER 9

Unified CVP Design for High Availability

Overview

A high-availability design provides the highest level of failure protection. Your solution may vary depending upon business needs such as:

- Tolerance for call failure
- Budget
- Topological considerations

Unified CVP can be deployed in many configurations that use numerous hardware and software components. Each solution must be designed so that a failure impacts the fewest resources in the contact center. The type and number of resources impacted depends on how stringent the business requirements are and the design characteristics you choose for the various Unified CVP components. A good Unified CVP design is tolerant of most failures, but sometimes not all failures can be made transparent to the caller.

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

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

In summary, plan ahead and follow all the design guidelines and recommendations presented in this guide and in the latest version of the *Cisco Unified Communications Solution Reference Network Design (SRND) Based on Cisco Unified Communications Manager*, available at:

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

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

Unified CVP Call Server High-Availability Component Consideration

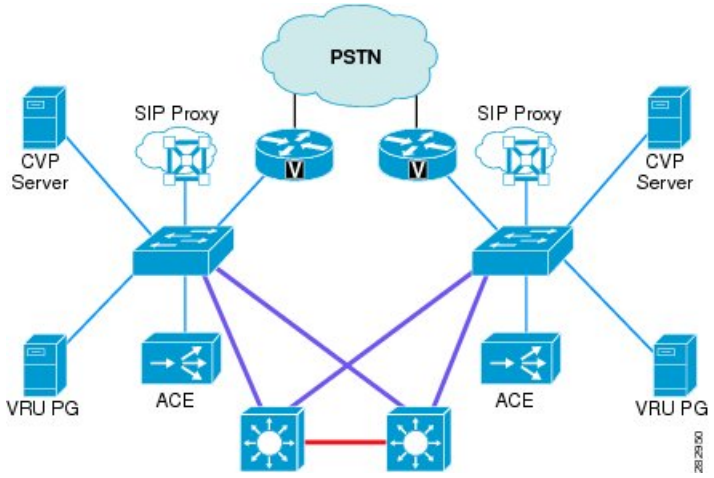
In the other chapters of this document, the Unified CVP Call Server is described as a single component because it does not need to be described in depth. When discussing Unified CVP high availability however, it is important to understand that there are actually several parts to this component:

- SIP Service—Responsible for processing incoming and outgoing calls with SIP.
- ICM Service—Responsible for the interface to ICM. The ICM Service communicates with the VRU PG using GED-125 to provide ICM with IVR control. The ICM Service was part of the Application Server in previous releases of Unified CVP, but now it is a separate component.
- IVR Service—Responsible for the conversion of Unified CVP Microapplications to VoiceXML pages, and vice versa. The IVR Service was known as the Application Server in previous Unified CVP versions.

Layer 2 Switch

Figure 10: Redundant Unified CVP System

The following illustration shows a high-level layout for a fault-tolerant Unified CVP system. Each component in the Unified CVP site is duplicated for redundancy. The quantity of each of these components varies based on the expected busy hour call attempts (BHCA) for a particular deployment.



Two switches shown in the illustration provide the first level of network redundancy for the Unified CVP Servers:

- If one switch fails, only a subset of the components becomes inaccessible. The components connected to the remaining switch can still be accessed for call processing.
- If ACE is used, its redundant partner must reside on the same VLAN in order to send keepalive messages to each other by using Virtual Router Redundancy Protocol (VRRP), a protocol similar to Hot Standby Router Protocol (HSRP). If one of the switches fails, the other ACE is still functional.

For more information on data center network design, see the *Data Center documentation* available at http://www.cisco.com/go/designzone

Note NIC teaming is not currently supported in the Unified CVP solution.

The NIC card and Ethernet switch is required to be set to 100 MB full duplex for 10/100 links, or set to auto-negotiate for gigabit links.

High Availability Options

After choosing a functional deployment model and distributed deployment options, Unified CVP solution designers must choose the amount of availability required. Unified CVP solution designers can increase solution availability in the following areas:

- Multiple gateways, Unified CVP Servers, Unified CVP VXML Servers and VRU PGs—Enables inbound and outbound call processing and IVR services to continue upon component failure.
- Multiple call processing locations—Enables call processing to continue in the event of a loss of another call processing location.
- Redundant WAN links—Enables Unified CVP call processing to occur upon failure of individual WAN links.
- ACE—Used for server load balancing and failover.

It is also possible to use a combination of these high availability options to be utilized.

Note Unified CVP VXML Server is coresident with Unified CVP Call Server.

Originating Gateway

The function of the originating gateway in a Unified CVP solution is to accept calls from the PSTN and direct them to Unified CVP for call routing and IVR treatment.

This section covers the following topics:

- Configuration, on page 96
- Call Disposition, on page 105

Configuration

For information to provide redundancy and reliability for originating gateways and T1/E1 lines, see the latest version of the *Design Guide for Cisco Unified Customer Voice Portal* available at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-implementation-design-guides-list.html.

In addition, consider the following issues when designing gateways for high availability in a Unified CVP solution:

- When used in ICM-integrated models, the originating gateway communicates with Unified CVP using SIP. Unlike MGCP, SIP does not have redundancy features built into the protocol. Instead, SIP relies on the gateways and call processing components for redundancy. The following configurations allow call signaling to operate independent of the physical interfaces. In this way, if one interface fails, the other interface can handle the traffic.
  - With dial-peer level bind, you can set up a different bind based on each dial peer. The dial peer bind eliminates the need to have a single interface reachable from all subnets. Dial peer helps in segregating the traffic from different networks (for example, SIP trunk from SP side and SIP trunk towards Unified Communications Manager or CVP). The dial peer level binding is illustrated in the following configuration example:

    ```
    Using voice-class sip bind
     dial-peer voice 1 voip
     voice-class sip bind control source-interface GigabitEthernet0/0
    ```

  - For other gateways, global binding should be used. Each gateway interface should be connected to a different physical switch to provide redundancy in the event that one switch or interface fails. Each interface on the gateway is configured with an IP address on a different subnet. The IP routers for the network are configured with redundant routes to the loopback address through the use of static routes or a routing protocol. If a routing protocol is used to review the number of routes being exchanged with the gateway, then consider using filters to limit the routing updates so that the gateway is only advertising the loopback address and not receiving routes. It is best to bind the SIP signaling to the virtual loopback interface, as illustrated in the following configuration example:

    SIP

    ```
    voice service voip
     sip
      bind control source-interface Loopback0
      bind media source-interface Loopback0
    ```

Call Disposition

If the originating gateway fails, the following conditions apply to call disposition:

- Calls in progress are dropped. These calls cannot be preserved because the PSTN switch loses the D-channel to all T1/E1 trunks on this gateway.
- New calls are directed by the PSTN carrier to a T1/E1 at an alternate gateway, provided that the PSTN switch has its trunks and dial plan configured.

SIP Proxy Server

The SIP proxy server provides dial plan resolution on behalf of SIP endpoints, allowing dial plan information to be configured in a central place instead of statically on each SIP device. A SIP proxy server is not required in a Unified CVP solution, but it is used in most solutions because of the centralized configuration and maintenance. Multiple SIP proxy servers can be deployed in the network to provide load balancing, redundancy, and regional SIP call routing services. In a Unified CVP solution, the choices for SIP call routing are:

- SIP proxy server
  - Advantages:

    Weighted load balancing and redundancy.

    Centralized dial-plan configuration.

    SIP proxy may already exist or used for other applications for dial-plan resolution or intercluster call routing.
  - Disadvantages:

    Additional server or hardware required for SIP proxy if not already deployed.
- Static routes using Server Groups (DNS SRV records) on a DNS Server
  - Advantages:

    Weighted load balancing and redundancy.
  - Disadvantages:

    Unable to use of an existing server depends on the location of the DNS server.

    The ability to share or delegate DNS server administration rights may be limited in some organizations.

    Dial-plan configuration needs to be configured on each device individually (Unified CM, Unified CVP, and gateways).

    DNS SRV lookup is performed for each and every call by Unified CVP. If the DNS server is slow to respond, is unavailable, is across the WAN, so the performance is affected.
- Static routes using local DNS SRV records
  - Advantages:

    Weighted load balancing and redundancy.

    Does not depend on an external DNS Server, which eliminates a point of failure, latency, and DNS Server performance concerns.
  - Disadvantages:

    Dial plan must be configured on each device individually (Unified CM, Unified CVP, and gateways).

Note The options for static routes using SRV with a DNS Server, or using Server Groups, can introduce some unexpected, long delays during failover and load balancing with UDP transport on the Unified CVP Call Server when the primary destination is shut down or is off the network. With UDP, when a hostname has multiple elements with different priorities in the Server Group (srv.xml), the Unified CVP attempts twice for each element, with 500 msec between each attempt. If the first element does not answer, it tries the next element one second later. The delay is on every call during failure, depending on load balancing, and is in accordance with section 17.1.1.1 of RFC 3261 regarding the T1 timer. If server group heartbeats are turned on, then the delay may only be incurred once, or not at all, depending on the status of the element. The call delay applies to TCP as well.

- Static routes using IP addresses
  - Advantages:

    Does not depend on any other device (DNS or Proxy) to deliver calls to the destination.
  - Disadvantages:

    No redundancy possible for SIP calls from Unified CVP.

    Dial plan must be configured on each device individually.

    This option is used in environments that do not have redundancy (single server) or for lab deployments.

Each device in the Unified CVP solution can use the above methods to determine where to send a call. The Unified CVP Call Server interface to the SIP network is through the Unified CVP SIP Service, which is discussed in the section on Unified CVP SIP Service , on page 103.

Cisco Unified SIP Proxy Support

Unified CVP has been validated with Cisco Unified SIP Proxy Server (CUSPServer), which implies that Unified CVP supports only CUSP proxy servers.

- CUSP is a dedicated SIP proxy server.
- CUSP server runs on the gateway.

  For additional information, see the Solution sizing tool at http://tools.cisco.com/cucst/faces/login.jsp.

Note For information on CUSP version numbers, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Cisco Unified SIP Proxy 9.0(x) Support

Cisco Unified SIP Proxy 9.0(x) virtual application supports all the features and deployments supported by Cisco Unified SIP Proxy 8.5(x). Cisco Unified SIP Proxy 9.0(x) has been tested on the following virtual machine deployment options:

- Cisco UCS B-Series Blade Servers
- ISRG2 and ISRG3 Servers

CUSP Deployment Methods

There are two deployment options supported with CUSP proxy in the CVP solution:

- Deployment Option A—Redundant SIP Proxy Servers
- Deployment Option B—Redundant SIP Proxy Servers (Double Capacity)

Deployment Option A - Redundant SIP Proxy Servers

This deployment option performs these actions:

- Two gateways are provided for redundancy, geographically separated, one proxy module each, using SRV priority for redundancy of proxies, and no HSRP.
- With Unified CVP 8.5(1) and later versions, CUSP can coreside with VXML or TDM Gateways. In earlier versions of Unified CVP due to platform validation restriction coresidency was not supported, and a dedicated ISR was required for proxy functionalities.
- TDM Gateways are configured with SRV or with Dial Peer Preferences to use the primary and secondary CUSP proxies.
- CUSP is set with Server Groups to find primary and back up Unified CVP, Unified CM, and VoiceXML Gateways.
- Unified CVP is set up with Server Group to use the primary and secondary CUSP proxies.
- Cisco Unified CM is set up with a Route Group with multiple SIP Trunks, to use the primary and secondary CUSP proxies.

In this example, ISR1 is on the east coast and ISR2 is on the west coast. The TDM Gateways will use the closest ISR, and only cross the WAN when needing to failover to the secondary priority blades.

The SRV records look like this:

```
east-coast.proxy.atmycompany.com
blade 10.10.10.10 priority 1 weight 10 (this blade is in ISR1 on east coast)
blade 10.10.10.20 priority 2 weight 10 (this blade is in ISR2 on west coast)

west-coast.proxy.atmycompany.com
blade 10.10.10.20 priority 1 weight 10 (this blade is in ISR2 on west coast)
blade 10.10.10.10 priority 2 weight 10 (this blade is in ISR1 on east coast)
```

Deployment Option B - Redundant SIP Proxy Servers (Double Capacity)

This deployment option performs the following actions:

- Two gateways are provided for redundancy, two proxy modules in each chassis. All four proxy servers are in active mode with calls being balanced between them.
- Uses SRV is used to load balance across proxies with priority.
- The ISR is dedicated to the proxy blade function and is not collocated as a VoiceXML Gateway, nor as a TDM Gateway, due to platform validation restrictions on CUSP.
- TDM Gateways are set with SRV or with Dial Peer Preferences to use the primary and secondary CUSP proxies.
- CUSP is set with Server Groups to find primary and back up Unified CVP, Unified CM, and VoiceXML Gateways.
- Unified CVP is set up with Server Group to use the primary and secondary CUSP proxies.
- Cisco Unified CM is set up with Route Group with multiple SIP Trunks to use the primary and secondary CUSP proxies.

In this example, ISR1 is on the east coast and ISR2 is on the west coast. The TDM Gateways will use the closest ISR, and only travel across the WAN when they need to failover to the secondary priority blades. The SRV records look like this:

```
east-coast.proxy.atmycompany.com
blade 10.10.10.10 priority 1 weight 10 (this blade is in ISR1 on east coast)
blade 10.10.10.20 priority 1 weight 10 (this blade is in ISR1 on east coast)
blade 10.10.10.30 priority 2 weight 10 (this blade is in ISR2 on west coast)
blade 10.10.10.40 priority 2 weight 10 (this blade is in ISR2 on west coast)

west-coast.proxy.atmycompany.com
blade 10.10.10.30 priority 1 weight 10 (this blade is in ISR2 on west coast)
blade 10.10.10.40 priority 1 weight 10 (this blade is in ISR2 on west coast)
blade 10.10.10.10 priority 2 weight 10 (this blade is in ISR1 on east coast)
blade 10.10.10.20 priority 2 weight 10 (this blade is in ISR1 on east coast)
```

Performance Matrix for CUSP Deployment

CUSP baseline tests were done in isolation on the proxy, and capacity numbers (450 TCP, 500 UDP transactions per second) should be used as the highest benchmark, and most stressed condition allowable.

A CVP call from the proxy server requires on average, four separate SIP calls:

- Caller inbound leg
- VXML outbound leg
- Ringtone outbound leg
- Agent outbound leg

When a consultation with CVP queuing occurs, an additional four SIP transactions will be incurred for the session, effectively doubling the number of calls.

CUSP Design Considerations

If the Proxy Server Record Route is set to on, it impacts the performance of the proxy server (as shown in the CUSP baseline matrix) and it also breaks the high-availability model because the proxy becomes a single point of failure for the call. Always turn the Record Route setting of the Proxy Server to off to avoid a single point of failure, to allow fault tolerance routing, and to increase the performance of the Proxy Server.

Record Route is turned off by default on CUSP.

Note **Upstream Element Routing with SIP Heartbeats**

With CUSP proxy, any response to an INVITE or OPTIONS is a good response, so CUSP will not mark an element down when it receives a response. If the response is configured in the failover response code list for the server group, then CUSP will failover to the next element in the group; otherwise, it will send the response downstream as the final response.

The standard for Unified CVP and CUSP proxy sizing is to define four SIP calls for every one CVP call, so considering there are 500 UDP transactions per second, the CPS rate is 500 / 4 = 125. The overall number of active calls is a function of Call Rate (CPS) * call handle time (CHT). Assuming an average call center call duration of 180 seconds (CHT), you get an overall active call value of 22,500 calls. Because one Call Server can handle approximately 900 simultaneous calls, it allows a single CUSP proxy to handle the load of 18 CVP Call Servers. A customer deployment should include consideration of the CPS and the CHT to size the proxy for their solution.

SIP Proxy Server Configuration

The SIP Proxy Server should be configured with static routes that point at the appropriate devices (Unified CVP Call Servers, VoiceXML Gateways, Cisco Unified Communications Manager cluster, and so forth). The SIP Proxy Server configuration allows you to specify the priority of the routes. In the case where there are multiple routes to the same destination, you can configure the SIP Proxy to load balance across the destinations with equal priority or to send the calls based on the priorities.

To reduce the impact of a Proxy Server failure, you can disable the RecordRoute header from being populated by the SIP Proxy Server. In this way, the inbound calls route through a SIP Proxy, but once they reach the Unified CVP Call Server (Call Server), the signaling is exchanged directly between the originating device and the Call Server, and a SIP Proxy failure will not affect the calls in progress.

Call Disposition

The following sections discuss configuration of Cisco IOS Gateways using SIP. It is not meant to be an exhaustive list of configuration options but only highlights certain configuration concepts.

IOS Gateway Configuration

With Cisco IOS Gateways, dial peers are used to match phone numbers, and the destination can be a SIP Proxy Server, DNS SRV, or IP address. The following example shows a Cisco IOS Gateway configuration to send calls to a SIP Proxy Server using the SIP Proxy's IP address.

```
sip-ua
 sip-server ipv4:10.4.1.100:5060
```

```
dial-peer voice 1000 voip
 session target sip-server
 ...
```

The **sip-server** command on the dial peer tells the Cisco IOS Gateway to use the globally defined SIP Server that is configured under the **sip-ua** settings. In order to configure multiple SIP Proxies for redundancy, you can change the IP address to a DNS SRV record, as shown in the following example. The DNS SRV record allows a single DNS name to be mapped to multiple Reporting Servers.

```
sip-ua
 sip-server dns:cvp.cisco.com

dial-peer voice 1000 voip
 session target sip-server
 ...
```

Alternatively, you can configure multiple dial peers to point directly at multiple SIP Proxy Servers, as shown in the following example. This configuration allows you to specify IP addresses instead of relying on DNS.

```
dial-peer voice 1000 voip
 session target ipv4:10.4.1.100
 preference 1
 ...
dial-peer voice 1000 voip
 session target ipv4:10.4.1.101
 preference 1
 ...
```

In the preceding examples, the calls are sent to the SIP Proxy Server for dial plan resolution and call routing. If there are multiple Unified CVP Call Servers, the SIP Proxy Server would be configured with multiple routes for load balancing and redundancy. It is possible for Cisco IOS Gateways to provide load balancing and redundancy without a SIP Proxy Server. The following example shows a Cisco IOS Gateway configuration with multiple dial peers so that the calls are load balanced across three Unified CVP Call Servers.

```
dial-peer voice 1001 voip
 session target ipv4:10.4.33.131
 preference 1
 ...
dial-peer voice 1002 voip
 session target ipv4:10.4.33.132
 preference 1
 ...
dial-peer voice 1003 voip
 session target ipv4:10.4.33.133
 preference 1
 ...
```

DNS SRV records allow an administrator to configure redundancy and load balancing with finer granularity than with DNS round-robin redundancy and load balancing. A DNS SRV record allows you to define which hosts should be used for a particular service (the service in this case is SIP), and it allows you to define the load balancing characteristics among those hosts. In the following example, the redundancy provided by the three dial peers configured above is replaced with a single dial peer using a DNS SRV record. Note that a DNS server is required in order to do the DNS lookups.

```
ip name-server 10.4.33.200
 dial-peer voice 1000 voip
 session target dns:cvp.cisco.com
```

With Cisco IOS Gateways, it is possible to define DNS SRV records statically, similar to static host records. This capability allows you to simplify the dial peer configuration while also providing DNS SRV load balancing and redundancy. The disadvantage of this method is that if the SRV record needs to be changed, it must be changed on each gateway instead of on a centralized DNS Server. The following example shows the

configuration of static SRV records for SIP services handled by cvp.cisco.com, and the SIP SRV records for cvp.cisco.com are configured to load balance across three servers:

```
ip host cvp4cc2.cisco.com 10.4.33.132
ip host cvp4cc3.cisco.com 10.4.33.133
ip host cvp4cc1.cisco.com 10.4.33.131
```

(SRV records for SIP/TCP)

```
ip host _sip._tcp.cvp.cisco.com srv 1 50 5060 cvp4cc3.cisco.com
ip host _sip._tcp.cvp.cisco.com srv 1 50 5060 cvp4cc2.cisco.com
ip host _sip._tcp.cvp.cisco.com srv 1 50 5060 cvp4cc1.cisco.com
```

(SRV records for SIP/UDP)

```
ip host _sip._udp.cvp.cisco.com srv 1 50 5060 cvp4cc3.cisco.com
ip host _sip._udp.cvp.cisco.com srv 1 50 5060 cvp4cc2.cisco.com
ip host _sip._udp.cvp.cisco.com srv 1 50 5060 cvp4cc1.cisco.com
```

Unified CVP SIP Service

The Unified CVP SIP service is the service on the Unified CVP Call Server that handles all incoming and outgoing SIP messaging and SIP routing. The Call Server can be configured to use a SIP Proxy Server for outbound dial plan resolution, or it can be configured to use static routes based on IP address or DNS SRV. Call Servers do not share configuration information about static routes; therefore, if a change needs to be made to a static route, then it must be made on each Call Server's SIP Service. Use a SIP Proxy Server to minimize configuration overhead.

Configuration

If only a single SIP Proxy Server is needed for outbound call routing from the Call Server, choose the SIP Proxy configuration when configuring the SIP Service. In the Unified CVP Operations Console Server, configure the following:

- Add a SIP Proxy Server and specify the IP address of the server.

Under the Call Server SIP Service settings, configure the following:

- Enable Outbound Proxy = True
- Use DNS SRV type query = False
- Outbound Proxy Host = SIP Proxy Server configured above

When using multiple SIP Proxy Servers for outbound redundancy from the Call Server, configure the SIP Proxy with a DNS name and configure DNS SRV records in order to reach the SIP Proxy Servers. The DNS SRV records can exist on an external DNS Server, or they can be configured in a local DNS SRV record on each CVP server. In the OAMP Console, configure the following:

- Add a SIP Proxy Server and specify DNS name of the server.

Under SIP Service configuration, configure the following:

- Enable Outbound Proxy = True
- Use DNS SRV type query = True

The DNS SRV record should then be configured with the list of SIP Proxy Servers.

To configure the Local DNS SRV record on each server, under the SIP service configuration, check **Resolve SRV records locally**.

To use a server group for redundant Proxy Servers:

1. Select **resolve SRV records locally** and enter the name of the server group for the outbound proxy domain name.
2. Under **System > Server Groups**, create a new server group with two proxy servers that have priority 1 and 2.
3. Deploy the server group configuration to the Call Server.

High Availability for Calls in Progress

When a Call Server fails with calls in progress, it is possible to restore all calls if certain gateway configuration steps are done. A Call Server can fail if one of the following occurs:

- The server crashes.
- The process crashes.
- The process stops.
- The network is out.

The configuration described in this section protects against all of these situations. However, if one of the following two scenarios occurs, recovery is not possible:

- Someone stops the process with calls in progress. This happens when a system administrator forgets to do a Call Server graceful shutdown. In this case, the CVP Call Server will terminate all active calls to release the licenses.
- The Call Server exceeds the recommended call rate. Although there is a limit for the absolute number of calls allowed in the Call Server, there is no limit for the call rate. In general, exceeding the recommended calls per second (cps) for an extended period of time can cause erratic and unpredictable call behavior on certain components. You must ensure that the components of the Unified CVP solution is sized correctly and balance the call load according to the weight and sizing of each call processing component. See the Call Server Sizing, on page 58 for call server call rate details.

For call survivability, configure the originating gateways as described in the latest version of the *Configuration Guide for Cisco Unified Customer Voice Portal*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html

The survivability.tcl script also contains some directions and useful information.

In the event of most downstream failures (including a Call Server failure), the call is default-routed by the originating gateway. Note that survivability is not applicable in the Unified CVP Standalone and NIC-routing models because there is no Unified CVP SIP service in those models.

There is also a method for detection of calls that have been cleared without Unified CVP's knowledge:

- Unified CVP checks every 2 minutes for inbound calls that have a duration older than a configured time (the default is 120 minutes).

- For those calls, Unified CVP sends an UPDATE message. If the message receives a rejection or is undeliverable, then the call is cleared and the license released.

The CVP SIP service can also add the Session expires header on calls so that endpoints such as the originating gateway may perform session refreshing on their own. RFC 4028 (Session Timers in the Session Initiation Protocol) contains more details on the usage of Session expires with SIP calls.

Call Disposition

Calls are handled as indicated for the following scenarios:

- Calls in progress

  If the Unified CVP SIP Service fails after the caller has been transferred (transfers include transfer to an IP phone, VoiceXML Gateway, or other Egress Gateway), then the call continues normally until a subsequent transfer activity (if applicable) is required from the Unified CVP SIP Service. If the caller is awaiting for further activity, there is a period of 9 to 18 seconds of silence before the caller is default-routed by survivability to an alternate location.

  If the call has not yet been transferred, the caller hears 9 to 18 seconds of silence before being default-routed by survivability to an alternate location. (Survivability does not apply in NIC-routing models.)

- New calls

  New calls are directed by the Unified SIP Proxy to an alternate Unified CVP Call Server. If no Call Servers are available, the call is default-routed to an alternate location by survivability. (Survivability does not apply in NIC-routing models.)

Server Group

A Server Group is a dynamic routing feature that enables the originating endpoint to know status of the destination address before attempting to send the SIP INVITE. Whether the destination is unreachable over the network, or is out of service at the application layer, the originating SIP user agent has knowledge of the status through a heartbeat method.

The Server Group features adds a heartbeat method with endpoints for SIP. This feature allows faster failover on call control by eliminating delays due to failed endpoints.

Note

- **Server Groups are not automatically created.** Server Groups are not created by the 8.0(1) upgrade. You must explicitly configure Server Groups for their deployment, and turn the feature on after upgrading, to take advantage of the feature.
- **Upgrade for customers with Local SRV.** Customers with Release 7.0(2), who already have an srv.xml file configured with local SRV, must run the import command to place their configuration into the Unified CVP Operations Console Server database before saving and deploying any new server groups to avoid overwriting your previous configuration.

The Unified CVP SIP subsystem builds on the local SRV configuration XML available with Release 7.0(1).

A Server Group consists of one or more destination addresses (endpoints), and is identified by a Server Group domain name. This domain name is also known as the SRV cluster domain name, or FQDN. The SRV method is used, but the DNS server resolution of the record is not performed. Server Groups remain the same as the Release 7.0(1) local SRV implementation (srv.xml), but the Server Group feature adds the extra heartbeat method on top of it as an option.

Note

- Server Groups in Unified CVP and SIP proxy servers functions in the same way.
- Only endpoints defined in a Server Group may have heartbeats sent to them.
- With record routes on proxy set to off, any mid-dialog SIP message, such as REFER or REINVITES, would bypass the elements defined in Server Group. These messages will be delivered directly to the other endpoint in the dialog.

You used the srv.xml configuration file to configure SRV records locally to avoid the overhead of DNS SRV querying. However, the method of configuration was manual, and could not be pushed from the Unified CVP Operations Console Server (Operations Console). Also, there was no validation on the minimum and maximum values for fields.

Unified CVP adds this configuration into the Operations Console SIP subsystem using the Server Groups concept. The Server Group term only refers to the local SRV configuration. When you turn on Server Groups with Heartbeat, you get the dynamic routing capability for Unified CVP to monitor the status of endpoints. This feature only covers outbound calls from Unified CVP. To cover the inbound calls to Unified CVP, the SIP proxy server can send similar heartbeats to Unified CVP, which can respond with status responses.

Server Group Heartbeat Settings

The Server Group heartbeat default setting tracks the ping interval between any two pings; it is not the interval between pings to the same endpoint. The Server Group does not ping at a specific interval and ping all elements because this approach would introduce a fluctuation on CPU usage. Also, it takes more resources when the system has to ping many endpoints. For example, to ping 3 elements across all groups at 30-second intervals, you have to set the ping interval at 10 seconds.

It is less deterministic for reactive mode because elements that are currently down can fluctuate, so the ping interval fluctuates with it.

Note

- **Heartbeat Behavior Settings for Server Groups.** To turn off pinging when the element is up, set the **Up Endpoint Heartbeat Interval** to zero (reactive pinging). To turn off pinging when the element is down, set the **Down Endpoint Heartbeat Interval** to zero (proactive pinging). To ping when the element is either up or down, set the heartbeat intervals to greater than zero (adaptive pinging).
- **Heartbeat Response Handling.** Any endpoint that CVP may route calls to should respond to OPTIONS with some response, either a 200 OK or some other response. Any response to a heartbeat indicates the other side is alive and reachable. A 200 OK is usually returned, but CUSP Server may return a 483 Too Many Hops response, because the max-forwards header is set to zero in an OPTIONS message. Sometimes the endpoints may not allow OPTIONS or PING, and may return 405 Method Not Allowed.

By default, Server Group heartbeats are monitored using a UDP socket connection. The transport type can be changed to TCP from the Operations Console Server Groups window.

Whenever an element has an unreachable or overloaded status, that element is marked as down completely, that is for both UDP and TCP transports. When the element is up again, transports are routed for both UDP and TCP.

Note TLS transport is not supported.

Duplicate Server Group Elements is not monitored because the primary element is already monitored.

Note See the *Configuration Guide for Cisco Unified Customer Voice Portal* for typical configurations for the Server Group feature, available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html

Static Routes Validation

The hostname or IP address of a static route is validated at startup and configuration deployment time with a DNS lookup resolution. If the hostname does not resolve to an A record or SRV record, then the route is disabled and a notice is printed in the Unified CVP error log. The calls cannot pass to this route in this state. If the host is in the local SRV Server Groups configuration as an SRV name, then the host is not checked, because it resolves to a local SRV name. IP addresses pass the validation.

Design Considerations

Observe the following design considerations when implementing Server Group:

- When you use the Local SRV configuration, you cannot use the DNS SRV configuration. However, elements may be declared as A record host names instead of IP addresses, and resolved through a DNS Server lookup or in the operating system host file.
- In the CUSP Proxy CLI, define the SRV cluster name (such as proxy-servers.cisco.com) in the service parameters section of the proxy configuration. Otherwise, a 404 not found rejection may result.

Diagnostics

The Unified CVP log file has traces that display endpoint status events. See the Unified CVP System CLI instructions in the *Configuration Guide for Cisco Unified Customer Voice Portal*, available at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

Unified CVP IVR Service

High availability was achieved by configuring the Unified CVP Voice Browser and VoiceXML Gateways with a list of application server IP addresses and using the ACE. With Unified CVP 4.0 and later releases, the IVR Service is combined with the SIP Service. If the IVR Service goes out of service, the SIP Service will be taken out of service so that no further calls are accepted by the Unified CVP Call Server.

Configuration

No additional configuration is required for SIP service to use IVR service. By default, the SIP service uses the IVR service that resides on the same server. It is also no longer necessary to configure the VoiceXML Gateway with the IP address of the Call Server's IVR service. When SIP is used, the SIP service inserts the URL of the Call Servers IVR service into a header in the SIP INVITE message when the call is sent to the VoiceXML Gateway. The VoiceXML Gateway extracts this information from the SIP INVITE and use this information to determine which Call Server to use. The VoiceXML Gateway examines the source IP address of the incoming call from the Call Server. This IP address is used as the address for the Call Servers IVR service.

The following example illustrates the IOS VoiceXML Gateway bootstrap service that is invoked when a call is received:

```
service bootstrap flash:bootstrap.tcl
  paramspace english index 0
  paramspace english language en
  paramspace english location flash
  paramspace english prefix en
```

Note For configuring the same feature in Cisco VVB, see section "Cisco VVB configuration for Comprehensive Call Flows".

With Unified CVP 4.0 and later releases, you have to configure the IP address of the Call Server. The bootstrap.tcl learns the IP address of the source Call Server and uses it as its Call Server. There is no need for backup Call Server configuration, because receiving a call from the Call Server means that the server is operational.

The following files in flash memory on the IOS Voice Gateway are also involved with high availability: handoff.tcl, survivability.tcl, recovery.vxml, and several .wav files. Use Trivial File Transfer Protocol (TFTP) to load the proper files into flash. Configuration information for each file can be found within the file itself. For information, see the latest version of the *Configuration Guide for Cisco Unified Customer Voice Portal*, available at:

https://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html

VoiceXML Gateway

The VoiceXML Gateway parses and renders VoiceXML documents obtained from the Unified CVP Call Server (from its IVR Service), the Unified CVP VXML Servers, or some other external VoiceXML source. Rendering a VoiceXML document consists of retrieving and playing prerecorded audio files, collecting and processing user input, or connecting to an ASR/TTS Server for voice recognition and dynamic text-to-speech conversion.

For a discussion of using mixed codecs in CVP deployments, see Mixed G.729 and G.711 Codec Support, on page 83. For a discussion of the benefits and drawbacks of each codec, refer to Voice Traffic, on page 140.

Note VoiceXML Gateway must not have a load balanced path because this route on the VoiceXML Gateway will cause a call HTTP Client Error. If the VoiceXML Gateway has a load balancing route to the CVP Call Server, it may use a different source address to send HTTP message to the CVP Call Server. CVP would return a 500 Server Error address to send HTTP message to CVP Call Server, which would cause CVP to return a 500 Server Error message. In VoiceXML Gateway, it is not possible to bind any specific interface for the HTTP Client side. If VoiceXML Gateway sends NEW_CALL using one interface and CALL_RESULT using another interface, CVP will return a 500 Server Error.

Configuration

The high-availability configuration for VoiceXML Gateways is controlled by the SIP proxy for SIP, or the Unified CVP Call Server (Call Server). Whether the VoiceXML Gateways are distributed or centralized also influences how high availability is achieved.

If a Call Server is unable to connect to a VoiceXML Gateway, an error is returned to the ICM script. In the ICM script, the Send to VRU node is separate from the first Run External script node in order to catch the VoiceXML Gateway connection error. If an END script node is used off the X-path of the Send to VRU node, the call is default-routed by survivability on the originating gateway. (Survivability does not apply in VRU-only models.) A Queue to Skill group node is effective only if there is an agent available. Otherwise, ICM tries to queue the caller, and that attempt fails because the Call Server is once again unable to connect to a VoiceXML Gateway. An END node could then also be used off the X-path of the Queue to Skill Group node to default-route the call.

Note VXML Server uses two features that assist with load balancing:

- Limiting load balancer involvement
- Enhanced HTTP probes for load balancers

See the configuration options ip_redirect and license_depletion_probe_error in the *User Guide for Cisco Unified CVP VXML Server and Cisco Unified Call Studio*, available at https://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/products-user-guide-list.html.

Centralized VoiceXML Gateways

In this configuration, the VoiceXML Gateways reside in the same data center as the Unified CVP Call Server.

SIP VoiceXML Gateways

If you are using SIP static routes on the Unified CVP Call Server, under the SIP Service configuration for the Call Server, configure a static route for each Network VRU label and gateway. If the VRU label is 5551000, the static route pattern would be 5551000>. The > is a wildcard representing one or more digits, and it is needed so that the correlation-id appended to the DNIS number can be passed to the VoiceXML Gateway correctly.

Note Other wildcard characters can be used. See the topic **Valid Formats for Dialed Numbers** in the Ops Console online help for complete wildcard format and precedence information.

In the case of both SIP proxy or Unified CVP static routes, the next-hop address of the route can be either the IP address of the gateway or a DNS SRV record. If you are using an IP address, you must create multiple static routes, one for each VoiceXML Gateway. In the case of DNS SRV, only one route for each Network VRU label is needed, and the SRV record provides for load balancing and redundancy.

High-Availability Hardware Configuration on Voice Gateways

The individual hardware components have the following high-availability options:

- Redundant power supplies
- Separate components for higher availability
- Dedicated components, which have fewer interaction issues

Example 1: Separate PSTN Gateway and VoiceXML Gateway

A PSTN Gateway and a separate VoiceXML Gateway provide greater availability for a combined PSTN and VoiceXML Gateway.

Example 2: Duplicate components for higher availability

- Two 8-T1 PSTN Gateways provide greater availability than one 16-T1 PSTN Gateway.
- Two 96-port Unified CVP VXML Servers provide greater availability than one 192-port Unified CVP VXML Server.
- Larger designs can use N+1 spares for higher availability.

Example 3: Geographic redundancy for higher availability

Geographical redundancy and high availability can be achieved by purchasing duplicate hardware for Side A and Side B.

Distributed VoiceXML Gateways

In this configuration, the gateway that processes the incoming call from the PSTN is separated from the Unified CVP servers by a low-bandwidth connection such as a WAN. The VoiceXML Gateway is different from the Ingress Gateway and can be located at the same site. The configuration keeps the media stream at the same site and without consuming bandwidth on the WAN and optimizes VoiceXML Gateway sizing when it is appropriate to separate Ingress and VoiceXML Gateways. In this case, setTransferLabel and Send to Originator cannot be used because you do not want the IVR leg of the call to go back to the Ingress Voice Gateway. It is also impractical to use a SIP Proxy to control the call routing because you would have to configure separate Network VRUs, Network VRU labels, and customers in ICM for each remote site. Instead, use SetSigDigits functionality.

With this method, the Call Server strips the leading significant digits from the incoming DNIS number. The value that is stripped is saved and prepended when subsequent transfers for the call occur.

SIP VoiceXML Gateways

When SIP is used, the significant digits are prepended to the DNIS number, and a SIP Proxy can be configured to route calls based on those prepended digits. The static routes in the SIP Proxy for the VoiceXML Gateway should have the digits prepended. Because these prepended digits were originally populated by the Ingress Gateway, the SIP Proxy can use them to determine which VoiceXML Gateway to use based on the incoming gateway. In this way, calls arriving at a particular site can always be sent back to VoiceXML treatment, with the result that no WAN bandwidth is used to carry the voice RTP stream. The Unified CVP indiscriminately prepends the **sigdigits** value to all transfers, including those to Unified CM. Therefore, when using Unified CM in this scenario, it is necessary to strip the prepended digits when the call arrives, so that the real DNIS number of the phone can be used by Unified CM to route the call, as illustrated in the following example.

Note The configurations mentioned below are only applicable to IOS Voice Gateway.

Configuration of Ingress Voice Gateway:

Apply a translation rule to the incoming DNIS to prepend the value 3333:

```
translation-rule 99
 rule 1 8002324444 33338002324444

dial-peer voice 1000 voip
 translate-outgoing called 99
```

Assuming the DNIS number is 8002324444, the final DNIS string routed to Unified CVP is 33338002324444.

Configuration of Unified CVP SIP service:

To configure the SIP service, in the Operations Console, select **Call Server > SIP**. Many of the settings are in the Advanced Configuration window.

Configuration of IOS VoiceXML Gateway:

Configure the Voice XML Gateway to match the DNIS string, including the prepended digits:

```
dial-peer voice 3000 voip
 incoming-called number 33335551000T
 service bootstrap
 ...
```

Configure the Unified CVP bootstrap.tcl application with the **sigdigits** parameter, indicating how many digits to strip off of the incoming DNIS string:

```
application
 service bootstrap flash:bootstrap.tcl
 param sigdigits 4
 ...
```

Cisco Unified CM configuration (if used):

Configure Unified CM to strip the prepended digits, either by using the Significant Digits configuration on the SIP Trunk configuration page or by using translation patterns.

SIP Proxy configuration:

Define static routes on the SIP Proxy, with the prepended digit present, to be sent to the appropriate VoiceXML Gateway. Because transfers to agents on a Unified CM cluster have prepended digits, the static routes for agent phones must also contain the prepended digits.

Summary of call routing:

1. A call arrives at Unified CVP with a DNIS number of 33338002324444.
2. Unified CVP removes four digits (3333) from the beginning of the DNIS string, leaving 8002324444.
3. The number 8002324444 is passed to ICM for call routing.
4. When it is time to transfer, ICM returns the label 5551000102. Unified CVP prepends 3333, resulting 33335551000102.
5. The SIP Service then resolves the address using the SIP Proxy or local static routes, and it sends the call to the VoiceXML Gateway.
6. The VoiceXML Gateway bootstrap.tcl removes 3333, leaving 5551000102 for the destination address.

Media Server

Audio files are stored locally in flash memory on the VoiceXML Gateway or on an HTTP/TFTP file server. Audio files stored locally are highly available. However, HTTP/TFTP file servers provide the advantage of centralized administration of audio files.

Note You cannot install the media server separately. The media server must be collocated with the Call Server and Unified CVP VXML Server.

Unified CVP Microapplication Configuration

The VoiceXML Gateway sends HTTP requests to an HTTP media server to obtain audio files. It uses the following VoiceXML Gateway configuration parameters to locate a server when not using a ACE:

```
ip host mediaserver <ip-address-of-primary-media-server>
ip host mediaserver-backup <ip-address-of-secondary-media-server>
```

The backup server is invoked only if the primary server is not accessible, and this is not a load-balancing method. Each new call attempts to connect to the primary server. If failover occurs, the backup server is used for the duration of the call; the next new call will attempt to connect to the primary server.

Note that the Media Server is not a fixed name, and it needs to match whatever name was assigned to the media_server ECC variable in the ICM script.

The VoiceXML Gateway also uses the following VoiceXML Gateway configuration parameters to locate a server when using a ACE:

```
ip host mediaserver <ip-address-of-ACE-VIP-for-media-server>
ip host mediaserver-backup <ip-address-of-ACE-VIP-for-media-server>
```

Because the ACE almost always locates a Media Server on the first request, a backup server is rarely invoked. However, you can configure the backup server when using a ACE for deployments where there are multiple data centers with ACE.

Note This feature is not required for Cisco VVB as DNS is used to resolve the hostname.

Call Dispositions

If the Media Server fails, the following conditions apply to the call disposition:

- Calls in progress should recover automatically. The high-availability configuration techniques described in the previous section (Unified CVP Microapplication Configuration) makes the failure transparent to the caller. If the media request fails, use scripting techniques to work around the error (for example, retry the request, transfer to an agent or label, or use TTS).
- New calls are directed transparently to the backup media server, and service is not affected.
- If the Media Server is located across the WAN from the VoiceXML Gateway and the WAN connection fails, the gateway continues to use prompts from the gateway cache until the requested prompt expires, at which time the gateway attempts to reacquires, the media, and the call fails if survivability is not enabled. If survivability is enabled, the calls are default-routed.

CVP Whisper Announcement and Agent Greeting Configuration

For the CVP Whisper Announcement service failover to function, duplicate the Whisper media on multiple Media Servers that are mapped by using the ACE VIP address.

For the Agent Greeting failover feature to function, configure the Agent Greeting service to duplicate the greetings recording on multiple Media Servers by configuring the default Media Server to act as a proxy. Afterwards, map the ACE VIP address to the farm of Media Servers.

For more information, see *Agent Greeting and Whisper Announcement Feature Guide for Cisco Unified Contact Center Enterprise*.

Call Studio Scripting Configuration

When scripting in Cisco Unified Call Studio, unlike with ICM scripting, there is no reverse ability for the media files. The script writer can point to **Properties** > **AudioSettings>** > **Default Audio Path URI** in the application and a single Media Server or the ACE VIP address for a farm of Media Servers.

Unified CVP VXML Server

The VoiceXML Gateway makes HTTP requests to the Unified CVP VXML Server to obtain VoiceXML documents.

Configuration

The Unified CVP VXML Server high-availability configuration and behavior is different for standalone deployments and deployments that are integrated with ICM, described in the following sections.

Standalone Self-Service Deployments

CVPPrimaryVXMLServer and CVPBackupVXMLServer gateway parameters specifically control the high-availability characteristics of the Unified CVP VXML Server. If Unified CVP VXML Server load balancing and more robust failover capabilities are desired, ACE device can be used. For configuration details, see the latest version of the *Configuration Guide for Cisco Unified Customer Voice Portal*. Load balancing can also be achieved without an ACE device by varying the primary and backup Unified CVP VXML Server configurations across multiple gateways.

Deployments Using ICM

When a Unified CVP VXML Server is used in conjunction with ICM, the ICM script passes a URL to the VoiceXML Gateway to invoke the VoiceXML applications. You can configure the ICM script to attempt first to connect to Unified CVP VXML Server A, and if the application fails out the X-path of the Unified CVP VXML Server ICM script node, try Unified CVP VXML Server B. The IP address in the URL can also represent Unified CVP VXML Server VIPs on the ACE.

Call Disposition

If the Unified CVP VXML Server fails, the following conditions apply to the call disposition:

- Calls in progress in a standalone deployment are disconnected. Calls in progress in an ICM-integrated deployment can be recovered using scripting techniques to work around the error as shown in the script (for example, retry the request, transfer to an agent or label, or force an error with an END script node to invoke survivability on the originating gateway).
- New calls are directed transparently to an alternate Unified CVP VXML Server.

Note Without an ACE device, callers might experience a delay at the beginning of the call and have to wait for the system while it tries to connect to the primary Unified CVP VXML Server.

Automatic Speech Recognition and Text-to-Speech Server

ASR and TTS in WAN Configurations

Note Cisco does not test or qualify speech applications in a WAN environment. For guidelines on design, support over WAN, and associated caveats, see the vendor-specific documentation.

The Cisco Technical Assistance Center provides limited support (as in the case of any third-party interoperability-certified products) on issues related to speech applications.

Limiting the Maximum Number of ASR or TTS-Enabled Calls

You can limit the number of calls enabled for ASR or TTS so that as soon as the limit is reached, regular DTMF prompt-and-collect can be used instead of rejecting the call altogether. In the following example, assume 5559000 is the ASR or TTS DNIS and 5559001 is the DTMF DNIS. You can configure the Ingress

Gateway to do the ASR load limiting for you by changing the DNIS when you exceed maximum connections allowed on the ASR or TTS VoIP dial peer.

Note Cisco VVB does not support this feature.

```
voice translation-rule 3 rule 3 /5559000/ /5559001/
!
voice translation-profile change
 translate called 3
!
!Primary dial-peer is ASR or TTS enabled DNIS in ICM script
dial-peer voice 9000 voip
 max-conn 6
 preference 1
 destination-pattern 55590..
 ...
!
!As soon as 'max-conn' is exceeded, next preferred dial-peer will change
the DNIS to a DTMF prompt & collect ICM script
dial-peer voice 9001 voip
 translation-profile outgoing change
 preference 2
 destination-pattern 55590..
 ...
!
```

Note 80 kbps is the rate for G.711 full-duplex with no Voice activity detection, including IP/RTP headers and no compression. The rate for G.729 full-duplex with no VAD is 24 kbps, including IP/RTP headers and no compression. For information on VoIP bandwidth usage, see *Voice Codec Bandwidth Calculator* at http://tools.cisco.com/Support/VBC/do/CodecCalc1.do.

Configuration ASR-TTS

The ASR/TTS high-availability configuration and behavior are different for standalone and ICM-integrated deployments, as described in the following sections.

Standalone Self-Service Deployments ASR-TTS

An ACE device is required in standalone deployments to provide failover capabilities for ASR/TTS. For instructions on configuring the ACE device for ASR/TTS and on configuring the ASR/TTS Server in a standalone deployment, see the latest version of the *Configuration Guide for Cisco Unified Customer Voice Portal*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html

Note If the ASR/TTS MRCP server fails, the following conditions apply to the call disposition:

- Calls in progress in standalone deployments are disconnected. Calls in progress in ICM-integrated deployments can be recovered using scripting techniques to work around the error. For example, retry the request, transfer to an agent or label, switch to the prerecorded prompts and DTMF-only input for the rest of the call, an error will occur with an END script node, to invoke survivability on the originating gateway.

  Note Cisco VVB has a built-in load-balancing mechanism that uses a round-robin technique. If the present ASR/TTS MRCP server fails, then the next request fo MRCP resource will get to the next server in the server group.

  In a call, if the selected ASR/TTS MRCP server responds with a failure to the setup request, then the VVB retries only once to set up with another server. If th VXML application has defined a preferred server for ASR dialog or TTS, then retry is not attempted.

  For configuration steps, see the section "Configure Speech Server" in *CVP Configuration Guide*.

- New calls in ICM-integrated deployments are directed transparently to an alternate ASR/TTS Server if a backup ASR/TTS Server is configured on the gateway.

Cisco Unified Communications Manager

Unified CVP transfers callers to Unified CCE agent phones or desktops using SIP. The Unified CVP Call Server receives an agent label from the ICM and routes the call using SIP proxy. The call is then sent to the appropriate Cisco Unified Communications Manager (Unified CM) in the cluster, which connects the caller to the agent. The Call Server proxies the call signaling, so it remains in the call signaling path after the transfer is completed. However, the RTP stream flows directly from the originating gateway to the phone. This fact becomes very significant in discussions of high availability.

Unified CVP also supports the Analysis Manager. See the Analysis Manager, on page 185 for more information.

Configuration

For information on providing Unified CM for high availability, see the latest version of the *Cisco Unified Contact Center Enterprise Solution Reference Network Design (SRND)* available at http://www.cisco.com/en/US/products/sw/custcosw/ps1844/products_implementation_design_guides_list.html.

Call Disposition

If the Unified CM process fails on the server that is either hosting the call or hosting the phone, the following conditions apply to the call disposition:

- Calls in progress are preserved. Skinny Client Control Protocol (SCCP) phones have the ability to preserve calls even when they detect the loss of their Unified CM. The caller-and-agent conversation continues until either the caller or agent goes on-hook. The Unified CVP Call Server recognizes that Unified CM has failed, assumes the call should be preserved, and maintains the signaling channel to the originating gateway. In this way, the originating gateway has no knowledge that Unified CM has failed. Note that additional activities in the call (such as hold, transfer, or conference) are not possible. Once the parties go on-hook, the phone is assigned to another Unified CM Server. When the agent goes on-hook, Real-Time Control Protocol (RTCP) packets cease transmitting to the originating gateway, which causes the gateway to disconnect the caller 9 to 18 seconds after the agent goes on-hook. If survivability has been configured on the gateway and the caller is waiting for some additional activity (the agent might think the caller is being blind-transferred to another destination), the caller is default-routed to an alternate location.
- New calls are directed to an alternate Unified CM Server in the cluster.

Intelligent Contact Management

Cisco Intelligent Contact Management (ICM) software provides enterprise-wide distribution of multichannel contacts (inbound/outbound telephone calls, Web collaboration requests, email messages, and chat requests) across geographically separated contact centers. ICM software is an open standards-based solution which includes routing, queuing, monitoring, and fault tolerance capabilities.

Configuration

For the most current information on configuring ICM for high availability, refer to the latest version of the *Cisco Unified Contact Center Enterprise Solution Reference Network Design (SRND)*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1844/products_implementation_design_guides_list.html

Call Disposition

There are many components in Cisco ICM, and call disposition varies depending on the component that fails. Although there are a few exceptions, the following conditions apply to the call disposition:

- If the primary router fails, calls in progress are unaffected. However, if the time for the VRU PG to realign to the other router is higher that the IVR service timeout (5 seconds default), calls in progress are default-routed by survivability on the originating gateway. If both the Side A and Side B routers fail, calls in progress are default-routed by survivability on the originating gateway.
- If the Logger fails, calls in progress are unaffected.
- If the primary router fails, calls in progress are unaffected. If both the Side A and Side B routers fail, calls in progress are default-routed by survivability on the originating gateway.
- New calls are directed to the backup ICM component.

Call Server and VXML Gateway in Different Subnets

Unified CVP shows one to two seconds delay in the Call Server when VXML gateway bootstraps the call. The delay is caused if the Call Server and VXML gateway are in different subnets.

To avoid the delay:

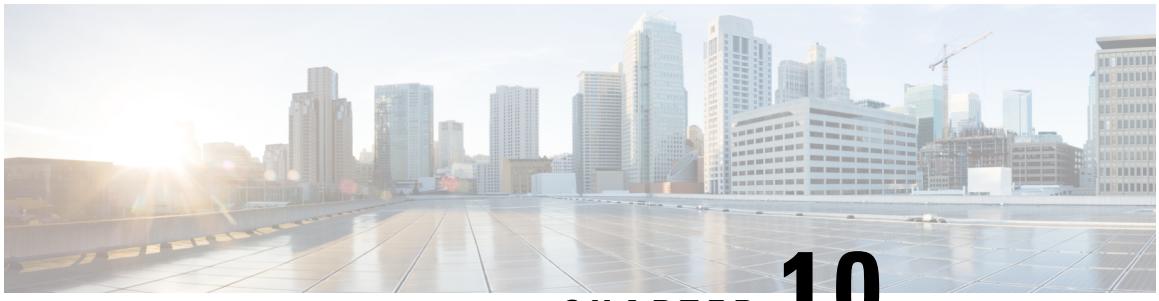
Procedure

Step 1 Open the registry of the machine.

Step 2 Navigate to the following path:
`HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\Interfaces\<Interface GUID.`

Step 3 Set **TcpAckFrequency** parameter to 1.

Step 4 Restart the windows machine.



CHAPTER 10

Cisco Unified ICM Interactions

Network VRU Types

This section discusses the Network VRU types for Unified ICM, and how Unified ICM relates to Unified CVP deployments.

This section describes the following topics:

- Unified ICM Network VRUs. See Unified ICM Network VRU, on page 119
- Unified CVP Type 10 VRU. See Unified CVP Type 10 VRU, on page 120
- Unified CVP Type 7 VRU (Correlation ID Mechanism). See Unified CVP Type 7 VRU (Correlation ID Function), on page 121
- Unified CVP Type 8 VRU (Translation Route ID Mechanism). See Unified CVP Type 8 VRU (Translation Route ID Function), on page 122

Note The terms voice response unit (VRU) and interactive voice response (IVR) are used interchangeably throughout this document.

Unified ICM Network VRU

Unified ICM perceives calls that need IVR treatment as having two portions: the Switch leg and the VRU leg. The Switch is the entity that first receives the call from the network or caller. The VRU is the entity that plays audio and preforms prompt-and-collect functions. If you use Unified ICM, Unified CVP can participate in the Switch role or the VRU role, or both. In a network deployment, multiple Unified CVP devices provide the Switch and VRU portions independently.

The call delivery to VRU can be based on either a Correlation ID or a translation route ID function, depending on the network capability to pass the call reference identification to the VRU. Call reference identification is needed because Unified ICM has to correlate the two legs of the same call in order to provide instructions for completing the call. In the Unified ICM application, the VRU supplies this call reference ID to Unified ICM when the VRU asks for instructions on how to process the incoming call that it receives from the switch. This method enables Unified ICM to retrieve the appropriate call context for this same call, which at this stage is to proceed to the IVR portion of the call.

- **Correlation ID**—This method is used if the network can pass the call reference ID to the VRU when the VRU is located in the network with the switch and the call signaling can carry this information (for example, the Correlation ID information is appended to the dialed digits when Unified ICM is used). This function usually applies to calls being transferred within the VoIP network.
- **Translation Route ID**—This method is used when the VRU is reachable across the PSTN (for example, the VRU is at the customer premise) and the network cannot carry the call reference ID information in delivering the call to the VRU. You must configure a temporary directory number (known as a translation route label) in Unified ICM to reach the VRU, and the network routes the call normally to the VRU as with other directory number routing in the PSTN. When the VRU requests instructions from Unified ICM, the VRU supplies this label (which can be a subset of the received digits), and Unified ICM can correlate the two portions of the same call. Generally, the PSTN carrier contains a set of translation route labels to be used for this purpose.

Note The deployed VRU can be located in the network (Network VRU) or at the customer premises. At the customer premises, a Network Applications Manager (NAM) is deployed in the network and a Customer ICM (CICM) is deployed at the customer premises. The corresponding Correlation ID or Translation Route ID is used, depending on the location of the VRU.

Unified CVP Type 10 VRU

Unified CVP Type 10 VRU simplifies the configuration requirements in Unified CVP Comprehensive Model deployments. Use the Type 10 VRU for new installations except for the VRU-only deployments. In deployments that need to use ICM Customers, you cannot initiate a two-step transfer from the Unified CVP VRU switch leg to a completely separate Unified CVP (for example, a two-step CVP-to-CVP transfer using SendToVRU). You are require to use a translation route for these two-step transfers to work.

Type 10 Network VRU operates as follows:

- Transferred routing client responsibilities are handed off to the Unified CVP switch leg.
- An automatic transfer to the Unified CVP VRU leg occurs resulting in a second transfer when calls are originated by the VRU, ACD, or Cisco Unified Communications Manager (Unified CM).
- For calls originated by Unified CM, the Correlation ID transfer mechanism is used. The Correlation ID is automatically added to the end of the transfer label defined in the Type 10 Network VRU configuration.
- The final transfer to the Unified CVP VRU leg is similar to a Type 7 transfer, which includes a RELEASE message to be sent to the VRU prior to any transfer.

You need to define a single Type 10 Network VRU in Unified CVP implementations without the ICM Customers feature (that is, in Unified CVP implementations with a single Network VRU), and associate all Unified ICM VRU scripts. One label for the Unified CVP Switch leg routing client, transfers the call to the

Unified CVP VRU leg. If calls are transferred to Unified CVP from Unified CM, another label for the Unified CM routing client, and this label should be different from the label used for the CVP Routing Client. This label transfers the call to the Unified CVP Switch leg. The Unified ICM Router sends this label to Unified CM with a Correlation ID concatenated to it. You must configure Unified CM to handle these arbitrary extra digits.

Configure the Unified CVP Switch leg peripheral to point to the same Type 10 Network VRU. Also, associate all incoming dialed numbers for calls that are to be transferred to Unified CVP with a Customer Instance that points to the same Type 10 Network VRU.

For calls that originate at a Call Routing Interface VRU or at a TDM ACD, a TranslationRouteToVRU node is required to transfer the call to a Unified CVP's Switch leg peripheral. For all other calls, use either a SendToVRU node, a node that contains automatic SendToVRU function (such as the queuing nodes), or a RunExternalScript.

Note Type 5 and Type 2 VRU types are not supported. Instead of these VRU types, use Type 10 VRU.

Unified CVP Type 7 VRU (Correlation ID Function)

When the VRU functions as an IVR with the Correlation ID function, Unified ICM uses Type 3 and Type 7 to designate suboperations of the VRU with the Peripheral Gateway in the Correlation ID scheme. Both Type 3 and Type 7 VRUs can be reached with the Correlation ID function, and a Peripheral Gateway is needed to control the VRU. However, the difference between these two types is in how they release the VRU leg and how they connect the call to the final destination.

In Type 3, the switch that delivers the call to the VRU can take the call from the VRU and connect it to a destination (or agent).

In Type 7, the switch cannot take the call away from the VRU. When the IVR treatment is complete, Unified ICM must disconnect or release the VRU leg before the final connect message can be sent to the Switch leg to instruct the switch to connect the call to the destination.

When used as an Intelligent Peripheral IVR, Unified CVP supports only Type 7 because it gets a positive indication from Unified ICM when its VRU leg is no longer needed (as opposed to waiting for the VoiceXML Gateway to inform it that the call has been pulled away). Type 3 has been deprecated.

A call has two legs: the Switch leg and the VRU leg. Different Unified CVP hardware can be used for each leg. A service node along with a Unified CVP for VRU leg with Peripheral Gateway acting as VRU Type 7 can be used to complete the IVR application (for example, self service and queuing).

Note Use Type 10 VRU for all new implementations of Unified CVP using Unified ICM 7.1 or greater, except as VRU Only (Model #4a).

For configuration examples of the Unified CVP application with VRU Type 7, see the *Configuration Guide for Cisco Unified Customer Voice Portal*at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

Unified CVP Type 8 VRU (Translation Route ID Function)

When the VRU functions as an IVR with the Translation Route ID function, Unified ICM uses Type 8 or Type 10 to designate suboperations of the VRU through the Peripheral Gateway in the translation route scheme. Both Type 8 and Type 10 VRUs can be reached through the Translation Route ID mechanism, and Peripheral Gateway is needed to control the VRU. However, they differ in how they connect the call to the final destination.

In Type 8, the switch that delivers the call to the VRU can take the call from the VRU and connect it to a destination or agent.

When the switch cannot take the call away from the VRU to deliver it to an agent, use Type 10. In that case, when the IVR treatment is complete, Unified ICM sends the final connect message to the VRU (rather than to the original switch) to connect the call to the destination. The VRU assumes control of the switching responsibilities when it receives the call. This process is known as handoff.

Similar to the Correlation ID, there are two legs of the call: the Switch leg and the VRU leg. Use Unified CVP for either the Switch leg or the VRU leg. For example, when Network Interface Controller (NIC), NAM, or CICM is taken, configure Unified CVP as Type 8 or Type 10 in the VRU leg.

Note Use Type 10 VRU for new implementations of Unified CVP using Unified ICM 7.1 or greater, except as VRU Only (Model #4a).

For configuration examples of the Unified CVP application with VRU Type 8 or Type 10, see *Configuration Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

Network VRU Types and Unified CVP Deployment Models

This section describes how Network VRU types relate to the Unified CVP deployment models and describes the following topics:

- Model #1: Standalone Self-Service. See Standalone Self-Service Deployments, on page 114
- Model #2: Call Director. See Model #2: Call Director, on page 124
- Model #3a: Comprehensive using ICM Micro-Apps. See Model #3a: Comprehensive Using ICM Micro-Apps, on page 124
- Model #3b: Comprehensive using VXML Server. See Model #3b: Comprehensive Using VXML Server, on page 124
- Model #4: VRU-Only. See Model #4: VRU-Only, on page 124
  - Model #4a: VRU-only with NIC Controlled Routing. See Model #4a: VRU-Only with NIC Controlled Routing, on page 124
  - Model #4b: VRU-only with NIC Controlled Pre-routing. See Model #4b: VRU-Only with NIC Controlled Prerouting, on page 125

In Unified ICM, a Network VRU is a configuration database entity that you can access by using Network VRU Explorer. A Network VRU entry contains the following information:

- Type—A number from 7, 8, and 10, which corresponds to one of the types.
- Labels—A list of labels that you use in Unified ICM to transfer a call to the particular Network VRU. These labels are relevant only for Network VRUs of Type 7 or 10 (that is, those VRU types that use the Correlation ID Mechanism to transfer calls). Each label consists of two parts:
  - A digit string, which becomes a Dialed Number Identification Service (DNIS). A SIP Proxy Server or a static route table (when SIP is used), or gateway dial peers understand DNIS.
  - A routing client, or switch leg peripheral. Each peripheral device that acts as a Switch leg must have its label, although the digit strings are the same in all cases.

Network VRU configuration entries have no value until they are associated with active calls. Unified ICM association is made at the following locations:

- In the **Advanced** tab for a given peripheral in the PG Explorer tool
- In the Customer Instance configuration in the Unified ICM Instance Explorer tool
- In every VRU Script configuration in the VRU Script List tool

Depending on the protocol-level call flow, the currently used Unified ICM Enterprise looks at either the peripheral or the Customer Instance to determine how to transfer a call to a VRU. The Unified ICM Enterprise examines the Network VRU associated with the Switch leg peripheral when the call first arrives on a Switch leg, and examines the Network VRU that is associated with the VRU leg peripheral when the call is being transferred to the VRU using the Translation Route Mechanism. The Unified ICM Enterprise examines the Network VRU that is associated with the Customer Instance when the call is being transferred to the VRU using the Correlation ID Mechanism.

Unified ICM Enterprise also checks the Network VRU that is associated with the VRU Script every time it encounters a RunExternalScript node in its routing script. If Unified ICM determines that the call is currently not connected to the designated Network VRU, the VRU Script is not executed.

Unified ICM Enterprise Release 7.1 introduced Network VRU Type 10, which simplifies the configuration of Network VRUs for Unified CVP. For most call flow models, a single Type 10 Network VRU replaces the place of the Type 2, 3, 7, or 8 Network VRUs that are associated with the Customer Instance and the switch, and VRU leg peripherals. VRU Only (Model #4a) still requires Type 7 or 8.

Note For existing deployments, the previously suggested VRU types work in the similar way, new installations are required to use Type 10. Existing deployments should switch to Type 10 on upgrade.

Model #1: Standalone Self-Service

The Standalone Self-Service model usually does not communicate with Unified ICM VRU scripts, so a Network VRU setting is not required. The Standalone Self-Service model with Unified ICM Label Lookup does not use the VRU scripts in Unified ICM. It issues a Route Request to the VRU Peripheral Gateway (PG) Routing Client, which does not require this Network VRU model.

Model #2: Call Director

In this model, Unified ICM (and also Unified CVP) is responsible for call switching only. This model does not provide queuing or self-service, so there is no VRU leg. A Network VRU setting is not required in this case.

Model #3a: Comprehensive Using ICM Micro-Apps

In this model, Unified CVP devices act as both the Switch and the VRU leg. However, the call needs to be transferred from the Switch leg to the VRU leg before any call treatment (for example, playing .wav files or accepting user input) can take place. Associate all Unified CVP peripherals with a Type 10 Network VRU in this case.

Note

- Type # 10 is available in Unified ICM 7.1 and later, and new implementations must use this configuration.
- Associate all incoming dialed numbers with a Customer Instance that is associated with a Type 10 Network VRU. You must associate all the VRU Scripts that this call executes with the same Type 10 Network VRU. Although it is not always necessary, we recommend that the Unified ICM routing script execute a SendToVRU node prior to the first RunExternalScript node.

Model #3b: Comprehensive Using VXML Server

If you consider call routing and the Network VRU, you will find this model identical to Model #3a.

Model #4: VRU-Only

In this model, the call arrives at Unified ICM through an ICM-NIC interface. Initially, Unified CVP is not responsible for the Switch leg; its only purpose is as a VRU. However, depending on the type of NIC being used, it may be required to take over the Switch leg after it receives the call.

This model has two submodels, which are described in the following sections.

Model #4a: VRU-Only with NIC Controlled Routing

Note This submodel has the following assumptions:

- A fully functional NIC can deliver the call temporarily to a Network VRU (that is, to Unified CVP's VRU leg) and then retrieving the call and delivering it to an agent when that agent is available.
- If the agent is capable of requesting that the call be retransferred to another agent or back into queue or self-service, the NIC can retrieve the call from the agent and redelivering it as requested.

Two variants of this submodel exist, depending on whether the Correlation ID or the Translation Route function is used to transfer calls to the VRU. Most NICs (most PSTN networks) cannot transfer a call to a particular destination directory number and carry an arbitrary Correlation ID along with it. The destination device can

pass back to Unified ICM to make the Correlation ID transfer mechanism function properly. For most NICs, you must use the Translation Route function.

However, a few exceptions to this rule exist, in which case the Correlation ID function can be used. The NICs transmits a Correlation ID include Call Routing Service Protocol (CRSP), and Telecom Italia Mobile (TIM). However, because this capability also depends on the PSTN devices that connect behind the NIC, check with your PSTN carrier to determine whether the Correlation ID can be passed through to the destination.

If the NIC is capable of sending the Correlation ID, the incoming dialed numbers must all be associated with a Customer Instance that is associated with a Type 7 Network VRU. The Type 7 Network VRU must contain labels that are associated to the NIC routing client, and all the VRU Scripts must also be associated with that same Type 7 Network VRU. The peripherals do not need to be associated with any Network VRU. We recommend that to execute the Unified ICM routing script SendToVRU node prior to the first RunExternalScript node.

If the NIC cannot send a Correlation ID, then the incoming dialed numbers must all be associated with a Customer Instance that is not associated with any Network VRU. However, the Unified CVP peripherals must be associated with a Network VRU of Type 8, and all the VRU Scripts must also be associated with that same Type 8 Network VRU. In this case, it is necessary to insert a TranslationRouteToVRU node in the routing script prior to the first RunExternalScript node. If the call is going to the VRU leg because it is being queued, generally the TranslationRouteToVRU node appears after the Queue node. In that way, you can avoid an unnecessary delivery and removal from Unified CVP when the requested agent is already available.

Model #4b: VRU-Only with NIC Controlled Prerouting

Note This submodel assumes a less capable NIC that can deliver the call only once, either to a VRU or to an agent. After the call is delivered to retrieve a call and then it is redelivered somewhere else, Unified CVP takes control of the switching responsibilities for the call. Unified ICM considers this process as a handoff.

Calls that fit this particular submodel must use the Translation Route function to transfer calls to the VRU. A handoff cannot be implemented by using the Correlation ID function.

To implement this model with Unified ICM 7.1 and later, the incoming dialed numbers must all be associated with a Customer Instance that is associated with a Type 10 Network VRU. The VRU labels are associated with the Unified CVP routing client, not the NIC. The Unified CVP peripherals and VRU Scripts must be associated with the Type 10 Network VRU. You need to insert a TranslationRouteToVRU node in the routing script, followed by a SendToVRU node, before the first RunExternalScript node. If the call is going to the VRU leg because it is being queued, these two nodes should appear after the Queue node. An unnecessary delivery and removal from Unified CVP can be avoided if the requested agent is already available.

Note Two different VRU transfer nodes are required. The first one transfers the call away from the NIC with a handoff, and it establishes Unified CVP as a Switch leg device for this call. Physically the call is delivered to an Ingress Gateway. The second transfer delivers the call to the VoiceXML Gateway and also establishes Unified CVP as the call's VRU device.

Hosted Implementations

Hosted implementations incorporate a two-level hierarchy of Unified ICM systems. The Network Application Manager (NAM) is at the top level, and one or more Customer ICMs (CICMs) is below it. Both the NAM and CICM are complete ICMs, with a communication link between them known as Intelligent Network Call Routing Protocol (INCRP). Each CICM functions in an isolated way; it is unaware of the other CICMs, and it is unaware that the NAM is another ICM. A CICM has no connection to the other CICMs, but its connection to the NAM is through the INCRP NIC.

Customers implement hosted setups because they are service providers. They want to provide ICM contact center services to multiple customers of their own. Each customer is hosted on its own CICM, and the NAM is responsible for routing calls. The calls are delivered to the service provider and to the appropriate customer's CICM. The individual customers run their own contact centers with their own Automatic Call Distributors (ACDs) connected to Peripheral Gateways at their own premises. The Peripheral Gateways, then are connected to their assigned CICMs at the service provider. The service provider owns and hosts the NAM and all the CICMs, but individual customers own and host all the ACDs. The Peripheral Gateways for those ACDs are owned by the service provider but are located at the customer's premises, next to the ACDs. The service provider does not necessarily operate any ACDs of its own. Those Peripheral Gateways can be connected to a CICM that is assigned to the service provider, or they can be connected to the NAM.

For ICM scripting, all incoming calls initially invoke an appropriate NAM routing script that has the primary responsibility of identifying the appropriate target customer. After NAM, these actions identify the appropriate target customer:

- The script delegates control to a routing script that is running on that customer's CICM.
- The CICM-based routing script selects the appropriate ACD to which to deliver the call, and it can return the necessary translation route label to the NAM.
- The NAM instructs its routing client to deliver the call to the designated target ACD. If the CICM routing script determines that no ACD can currently take the call or that it cannot yet identify which ACD should take the call, it can ask the NAM to place the call into queue at a Service Control VRU.
- The CICM routing script issues Network VRU Script requests through the NAM to that VRU until a routing decision is made.

Many hosted customers use this topology to get more calls or more Peripheral Gateways through their ICM setup. Other customers use CICMs, not for customer contact centers, but for external customers. In these situations, the NAM might handle the same number of calls as the CICM, and the CICM machines might be located far away from the NAM. Also, the NAM and CICM architecture was designed at a time when all contact centers ran on TDM-based ACDs. The addition of VoIP routing and ACDs based on Unified CM (that is, Unified CCE) with direct agent routing made matters more complicated.

Unified CVP in Hosted Environments

When Unified CVP is in hosted environment, it is used as a self-service or queuing platform connected to the NAM and physically located within the service provider's data center. Unified CVP enables a service provider to not only to route calls to the appropriate customer-owned ACDs but also to retain control of calls that are queued for those ACDs and to provide either basic prompt-and-collect capability or full-featured self-service applications to its customers. The latter situation typically incorporates VXML Servers into the network. Depending on the customer's needs, the service provider may host the VXML Servers or the customer may host them. The service provider also may write and own the self-service application, or the customer may

write and own them. Allowing the customer to own or host the VXML Servers is a convenient solution when the self-service application needs to reference back-end services. This solution allows the customer to retain control of that interaction within its own enterprise network while transmitting only VoiceXML over HTTP to the service provider's VoiceXML Gateway.

In many hosted environments, when the service provider is a PSTN carrier, all of the actual call routing occurs through an ICM NIC. These deployments are similar to Model #4b: VRU Only with NIC Controlled Pre-Routing (See Model #4b: VRU-Only with NIC Controlled Prerouting, on page 125). The same situation applies if a Peripheral Gateway is being used to route calls using (typically) the ICM NIC. However, quite often the service provider does not have a NIC interface at all, and all calls arrive through TDM interfaces, such as T3 or E3. In those cases, Unified CVP is used as the Switch leg as well as the VRU leg. This situation is similar to Model #3a: Comprehensive Using ICM Micro-Apps (See Model #3a: Comprehensive Using ICM Micro-apps , on page 42) or Model #3b: Comprehensive Using VXML Server (See Model #3b: Comprehensive Using VXML Server, on page 124).

Hosted Environment Unified CVP Placement and Call Routing

If Unified CVP is used in a valid Network VRU, it is connected at the NAM. However, various requirements might cause Unified CVP to be placed at the CICM level or in addition. Use the following guidelines when considering where to place Unified CVP components:

- If you place Unified CVP at the NAM, and Unified CVP handles both the Switch leg and the VRU leg, use the Correlation ID transfer function. The SendToVRU node may be executed by either the NAM or the CICM routing script. (The RunExternalScript nodes should also be in the same script that executed the SendToVRU.)
- If you place Unified CVP at the NAM and a NIC handles the Switch leg while Unified CVP handles the VRU leg, either the Correlation ID transfer function or the Translation Route transfer function may be used, depending on the capabilities of the NIC. (See Model #4a: VRU Only with NIC Controlled Routing, Model #4a: VRU-Only with NIC Controlled Routing, on page 124). In this case, the following guidelines also apply:
  - If you use Correlation ID transfer, then the SendToVRU node may be contained in either the NAM or the CICM routing script. (The RunExternalScript nodes should also be in the same script that executed the SendToVRU.)
  - If you use Translation Route transfer, then the TranslationRouteToVRU node, and all RunExternalScript nodes must be in the NAM routing script. The assumption here is that the call is queued (or treated with prompt-and-collect) before the particular CICM is selected. This configuration does not facilitate queuing. However, this configuration can be useful for service providers who want to offer initial prompt-and-collect before delegating control to the CICM.
- If you place Unified CVP at the CICM, and a NIC handles the Switch leg while Unified CVP handles the VRU leg, only the Translation Route transfer method can be used. The TranslationRouteToVRU node, together with all RunExternalScript nodes, must be in the CICM routing script.

Adding calls initiated by Unified CM or an ACD creates additional constraints. Both of these devices are considered ACDs from the ICM perspective, and they are connected at the CICM level. Assuming these are new calls (as opposed to continuations of existing calls), the route request comes from the ACD and the resulting label is returned to the ACD. Neither Unified CM nor any ACD can send a Correlation ID upon transfer. You can only use the Translation Route transfer method. This limitation also implies that the transfer destination (for example, Unified CVP) must also be connected at the CICM level, and not the NAM level.

If the calls are not new but continuations of existing calls, then they are attempts to transfer an existing inbound caller from one agent to another agent. The customers may want these transfers to be either blind network transfers or warm consultative transfers. The following guidelines apply to these transfers:

- Blind network transfers

  If the original call is introduced to the NAM through a NIC or Unified CVP Switch leg, the transfer label is passed from the CICM to the NAM to the original Switch leg device. Blind network transfers have two subcases:

  - If the Switch leg device is Unified CVP or a NIC that can handle Correlation ID, the Correlation ID transfer function can be used. The SendToVRU node and all RunExternalScript nodes must be incorporated in the CICM routing script. The Unified CVP VRU leg can be connected to the NAM. This combination of blind transfer with Correlation ID transfer is suitable for Unified CVP.
  - If the Switch leg device is a NIC that cannot handle Correlation ID, then the Translation Route transfer method must be used, which further implies that the Unified CVP VRU leg device must be connected to the CICM.

  Note In this situation, the customer may need to deploy additional dedicated Unified CVP Call Servers at the CICM level because the NAM-level Unified CVP Call Servers cannot be used.

- Warm consultative transfers

  Unified CVP provides warm consultative transfers only in the case of Unified CCE agents transferring calls to other Unified CCE agents, where Unified CVP owns the initial Switch leg for the inbound call. For TDM agents, the ACD functions are used and Unified CVP is not involved. When the incoming calls to Unified CCE agents arrive through a NIC, the Unified ICM Network Consultative Transfer facility is used and not Unified CVP.

  In the one supported case where Unified CVP owns the initial Switch leg and the transfer is among Unified CCE agents, the Translation Route transfer method must be used because Unified CM cannot handle Correlation ID transfers. The Unified CVP VRU leg device must be connected to the CICM.

Note In this situation, the customer may have to deploy additional dedicated Unified CVP Call Servers at the CICM level because the NAM-level Unified CVP Call Servers cannot be used.

Network VRU Type in Hosted Environment

In a hosted environment, there are two types of ICM systems, the NAM and the CICM. Network VRU types are configured differently in the NAM and CICM.

The NAM gets new calls either from the NIC or from Unified CVP, and is aware of the Unified CVP VRU leg device. The NAM Network VRU types must be configured exactly as an independent ICM, operating with those devices. You can ignore that the transfer labels sometimes come from CICM when configuring Network VRU types. CICM sees new calls that arrive from the Intelligent Network Call Routing Protocol (INCRP) NIC.

All of the dialed numbers that arrive from the NAM must be associated with a Customer Instance that is associated with the corresponding Network VRU on CICM. Associate that Network VRU with all VRU scripts, and provide the same label that you need in the NAM Network VRU definition, but with the INCRP NIC as its routing client. No peripherals have Network VRUs configured.

For more information on Network VRU Type support, see the *Configuration Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

Unified CM, ACD Call Deployment Models, and Sizing Implications

The information in this section applies to ACDs and Cisco Unified Communications Manager (Unified CM) integrations that use Unified CVP instead of Cisco IP IVR for queuing. If Unified CVP is considered, these devices share the following characteristics:

- Manage agents, and can be destinations for transfers.
- Can issue Route Requests, and can be Switch leg devices.
- Although they can be Switch leg devices, they cannot handle more than one transfer and they might not be able to handle the Correlation ID.

A Unified CM or ACD user issues a Route Request for one of the following reasons:

- To be connected to another agent in a particular skill group
- To reach a self-service application
- To blind-transfer a previously received call to one of the above entities

A Unified CM user might also issue a Route Request for one of the following reasons:

- To deliver a successful outbound call from the Unified ICM Outbound dialer to a self-service application based on Unified CVP
- To warm-transfer a call that the user had previously received to either a particular skill group or a self-service application

Each of the above calls invokes an Unified ICM routing script. The script searches for an available destination agent or service and if an appropriate destination is found, it sends the corresponding label either back to the ACD or, if blind-transferring an existing call, to the original caller's Switch leg device. If it needs to queue the call or if the ultimate destination is intended to be a self-service application rather than an agent or service, the script sends a VRU translation route label either back to the ACD or, if transferring an existing call through blind-transfer, to the original caller's Switch leg device.

If the above sequence results in transferring the call to Unified CVP's VRU leg device, a second transfer is done to deliver it to a VoiceXML gateway. To ensure that these events take place, the following Unified ICM configuration elements are required:

- For new calls from the ACD or warm transfers of existing calls:
  - Configure the Unified CVP peripheral to be associated with a Type 10 Network VRU.

- Associate the dialed number that the ACD dialed with a Customer Instance that is associated with a Type 10 Network VRU.
- When an ACD is not configured Unified CM, the routing script that is invoked by the ACD dialed number must contain a TranslationRouteToVRU node to get the call to Unified CVP's Switch leg, followed by a SendToVRU node to get the call to the VoiceXML gateway and Unified CVP's VRU leg.
- The routing script that is invoked by Unified CM should use a SendToVRU node to send the call to Unified CVP using the Correlation ID. The Type10 VRU performs an automatic second transfer to the VoiceXML gateway VRU leg.
- Associate all the VRU scripts that are executed by that routing script with the Type 10 Network VRU.

- For blind transfers of existing calls:
  - The Unified CVP peripheral can be associated with any Network VRU.
  - The dialed number that the ACD dialed must be associated with a Customer Instance that is associated with a Type 10 Network VRU.
  - The routing script that is invoked by the ACD dialed number must contain a SendToVRU node to send the call to the VoiceXML gateway and Unified CVP's VRU leg.
  - All the VRU scripts that are executed by that routing script must be associated with the Type 10 Network VRU.

When Unified ICM chooses an agent or ACD destination label for a call, it tries to find one that lists a routing client that can accept that label. For calls originated by an ACD or Unified CM that are not blind transfers of existing calls, the only routing client is the ACD or Unified CM, after the call is transferred to Unified CVP, because of the handoff operation, the only routing client is the Unified CVP Switch leg. However, in the case of blind transfers of existing calls, two routing clients are possible:

- The Call Server switch leg that delivered the original call.
- The ACD or Unified CM. For calls that originate through Unified CVP, you can prioritize Unified CVP labels above ACD or Unified CM labels by checking the **Network Transfer Preferred** check box in the **Unified ICM Setup** screen for the Unified CVP peripheral.

When using Unified CVP for network transfers, an agent blind-transfers the caller to a new destination with the Network Transfer Preferred option. In this scenario, the agent uses CTI Agent Desktop (and not the phone itself) to invoke the transfers. In addition to the CTI Agent Desktop, the Agent uses the Unified ICM Dialed Number Plan. If configured with the same DN as the CTI Route Point, the Unified ICM Dialed Number Plan causes Unified ICM to intercept the transfer and run the Unified ICM routing script without sending the transfer commands to Unified CM through JTAPI. When the Unified ICM script returns a label, that label is used for the Network routing client (Unified CVP), and the caller is sent directly to the new destination. This configuration avoids a timing problem that can occur if an agent uses Unified CM CTI Route Points to initiate a network transfer.

Third-Party VRU

A third-party TDM VRU can be used in any of the following ways:

- As the initial routing client (using the GED-125 Call Routing Interface)
- As a VRU (using the GED-125 Call Routing Interface)
- As a Service Control VRU (using the GED-125 Service Control Interface)

In the first and second operations, the VRU works as an ACD, as described in *Unified CM and ACD Call Deployment Models, and Sizing Implications*. Similar to ACD, the VRU can be a destination for calls that arrive from another source. Calls can even be translation-routed to such devices to carry call context information. (This operation is known as a traditional translation route, not a TranslationRouteToVRU). Also like an ACD, the VRU can issue its own Route Requests and invoke routing scripts to transfer the call to subsequent destinations or even to Unified CVP for self-service operations. These transfers almost always use the Translation Route transfer function.

In the third operation, the VRU replaces either Unified CVP's Switch leg or Unified CVP VRU leg, or it can also replace Unified CVP. Such deployments are beyond the scope of this document.

DS0 Trunk Information

Through Unified CVP, Unified ICM passes the gateway trunk and DS0 information from the arriving SIP call.

PSTN gateway trunk and DS0 information received at ICM has the following purposes:

- Reporting
- Routing in the Unified CCE Script Editor where TrunkGroupID and TrunkGroupChannelNum information is available for routing decisions.

Following message is used in the examples:

The PSTN trunk group data comes from the PSTN Gateway in the SIP INVITE as shown:

```
Via: SIP/2.0/UDP
192.168.1.79:5060;x-route-tag="tgrp:2811-b-000";x-ds0num="ISDN 0/0/0:15
0/0/0:DS1 1:DS0";branch
```

The following logic is used in Unified CVP to parse and pass the PSTN trunk group information to Unified ICM:

- For TrunkGroupID, look for **tgrp:** in the **x-route-tag** field.
  - If **tgrp:** found **TrunkGroupID=value after tgrp:** > + <**data between ISDN and :DS1 tags**>· Using the above example: **TrunkGroupID = 2811-b-000<space>0/0/0:15 0/0/0.**
  - TrunkGroupID = <**IP addr of originating device in Via header**> + <**data between ISDN and:DS1 tags**>

    Using the above example: **TrunkGroupID=192.168.1.79<space>0/0/0:15 0/0/0.**

- For TrunkGroupChannelNum, look for **DS0** in **x-ds0nun** field.
  - If found, **TrunkGroupChannelNum** = <**value before the :DS0**>· Using the above example: **TrunkGroupChannelNum = 1**
  - **TrunkGroupChannelNum** = <**max int value**> to indicate we did not find the DS0 value.

    Using the above example: **TrunkGroupChannelNum = Integer.MAX_VALUE (2^31 - 1)**

Trunk Utilization Routing and Reporting

Through the Trunk Utilization feature, a gateway is used for real-time Unified CVP routing and Unified ICM reporting and scripting. A gateway pushes the status of memory, DS0, DSP, and CPU to Unified CVP. Because this feature uses a push method to send resource data to Unified CVP, resources are monitored more closely and failover can occur faster when a device goes down or is out of resources.

This feature has the following characteristics:

- Each gateway can publish an SIP OPTIONS message with CPU, Memory, DS0, and DSP information to Unified CVP every three minutes when operation conditions are normal on the gateway.
- The push interval is configurable through the Cisco IOS CLI on the gateway.
- If a high watermark level is reached, the gateway sends the SIP OPTIONS message immediately with an **Out-Of-Service = true** indication, and does not send another OPTIONS message until the low watermark level is reached with an **Out-Of-Service = false** indication.
- Up to five Resource Availability Indication (RAI) targets can be set up on the gateway.

Trunk Utilization Routing can also be used to update trunk group status in the Unified CCE router. A PSTN call (through the ICM script) can query the router with a preroute from a NIC to use the available ingress gateway for the post route to Unified CVP.

Note DS0 is the data line that provides utilization information about the number of trunks free on a gateway.

Gateway Trunk Utilization with Server Group Pinging Combination

When you combine the Server Group element polling feature with the Cisco IOS Gateway trunk utilization feature, your solution has faster failover for high availability call signaling.

Deployment Considerations

- For Proxy Server deployment with CUSP:
  - Configure TDM originating gateways for resource allocation indication-targets (RAI-targets) to provide status in OPTIONS message to primary and secondary Unified CVP Call Servers, for reporting purposes. The data is used for reporting, and not routing so the data needs to be sent to Call Servers that have reporting enabled.
  - Configure primary and secondary CUSP proxy servers with Server Groups pinging to Unified CVP, VXML Gateways, and Unified Communications Manager elements.
  - Configure Unified CVP with Server Group that pings to both primary and secondary CUSP proxies for outbound calls.
- For a non-proxy deployment:
  - Configure TDM originating gateways for RAI-targets to provide status in OPTIONS message to primary and secondary Call Servers. Unified CVP can handle the messages for both reporting and

routing purposes. If used for routing, then the gateway must be in a server group by itself on Unified CVP.

- Configure Unified CVP with Server Groups that pings to Unified CVP, VXML Gateways, and Unified Communications Manager elements for outbound calls.
- Configure VXML gateways for RAI-targets to provide status in the OPTIONS message to primary and secondary Call Servers.

- Configure the Unified CVP Call Servers to send the same hostname in the contact header of OPTIONS requests to the gateways. This process enables a single RAI-target to be configured to all Call Servers and is important because the limit is five targets. The parameter to set is called Options Header Override.

Note See the Cisco IOS documentation for guidelines on the high and low watermark settings.

Limitations:

- RAI is not supported on Proxy Servers.

  CUSP servers do not handle the RAI header of OPTIONS messages, so they do not mark the status of elements with that information. If VXML Gateways are down, Unified CVP may send the call using the proxy, because the proxy does not handle incoming RAI headers in OPTIONS. It is possible to use a local static route scheme on Unified CVP to send all calls to the proxy except the Voice XML Gateways calls to create a server group for Voice XML Gateways and take advantage of RAI updates for routing.

Enhanced User-to-User Information

User-to-user information (UUI) is the data provided using ISDN Supplementary Services as user-to-user services. The UUI feature enables the information transfer between calling and called ISDN numbers during call setup and call disconnect with up to 128 octets of data for each message.

For calls involving Unified CVP transfers or disconnects, you can use the UUI feature to pass ISDN data provided from the PSTN, in the GTD, to the Unified ICM router, and then from Unified ICM to third-party ACDs.

The Ingress and Egress Gateways can use application specific data in the UUI field for use in CTI applications and for better third-party ACD integration.

For example, you can capture data from an external system (such as caller-entered digits from a third-party IVR) and pass that data to Unified ICM on a new call.

Unified CVP can send UUI in hex-encoded format on the outbound direction of Unified CVP, for example to the agent or even to the IVR.

While UUI is ISDN data, Unified CVP and the gateways support tunneling the ISDN data in SIP messages on the VoIP side. The data can be encapsulated in the content body of the SIP message in a Generic Type Descriptor (GTD) content type.

RTP media port and codec information is defined as a SDP body type, but the ISDN data is encapsulated in a Generic Type Descriptor body type by the Cisco IOS Gateways. When both RTP and ISDN data are sent to Unified CVP through the TDM Gateway, both body types are sent in a multipart and mixed mime type, that includes both SDP and GTD parts.

The following configuration in the gateway is required to enable the enhanced UUI feature:

```
voice service voip
        signaling forward unconditional
```

Manipulating UUS Field

You can set UUI by ICM scripts and extract it by Unified CVP to be resent in SIP messages.

UUI processing scenarios:

- When GTD (generic type descriptor) data is present in the inbound call leg of the SIP INVITE message in the mime body format for GTD, Unified CVP saves the GTD data as inbound GTD and the UUI portion (if present) is passed to Unified ICM.

  This GTD format is supported by the Cisco IOS gateways on outbound VoIP dial peers with SIP transport.

  If Unified ICM modifies the data, it sends the modified UUI back to Unified CVP. Unified CVP converts the UUI data it receives from Unified ICM into hex, modifies the UUS (if it is present), and overwrites the inbound GTD value. Only the UUS portion is modified, using the format:

  ```
  UUS,3,<converted Hex value of data from ICM>
  ```

  The rest of the GTD parameter values are preserved, saving the values as they arrived from the caller GTD.

- When GTD is not present in the inbound call leg, Unified CVP prints an informational message on the trace stating No GTD Body present in Caller Body, and the call continues as a regular call.

Note

- The modified UUI from Unified ICM is passed using the *user.microapp.uui* ECC variable, or the *Call.UsertoUserInfo* variable.
- If you use both variables, *Call.UsertoUserInfo* variable takes precedence.

Modified GTD is set in the outbound INVITE mime body from CVP SIP B2BUA, which includes IP originated callers as well as TDM callers. If a DTMF label for outpulse transfer is received on a connected call, then the BYE message is sent with the GTD only if UUI is passed by Unified ICM. The BYE message is sent immediately after the SIP INFO with DTMF.

Using UUI

Extract the UUI in your Unified ICM Script by looking at the *user.microapp.uui* Call ECC variable and the *Call.UserToUserInfo* variable, such as in the IF node. By using the SET node on either one of these variables, the variable can be set on the outbound direction of the call.

Setting *Call.UserToUserInfo* variable takes precedence over using the ECC variable.

Note Unified CVP sends a BYE message on the DTMF label only if UUI is received from Unified ICM.

If a BYE message is received, then the GTD from the received BYE is used to send it on the other leg.

Configure the Ingress Gateway with signaling forward unconditional, as in the following example, so that GTD with UUI and UUS are forwarded on the VoIP side.

Using UUI

```
voice service voip
        signaling forward unconditional
```

REFER, 302 Redirects, and UUI

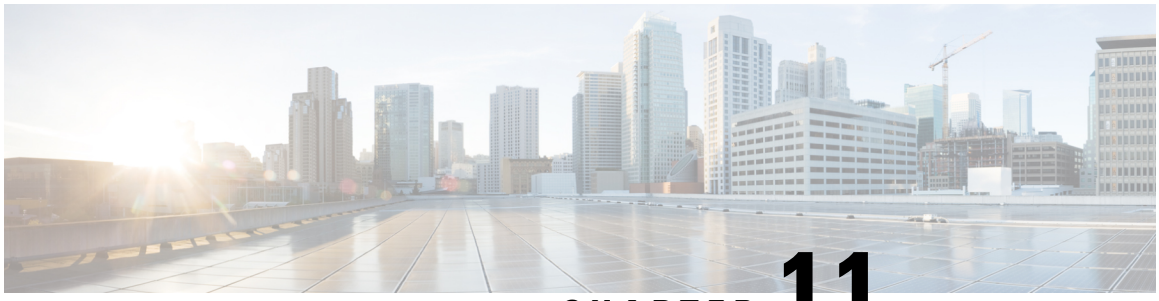
If you configure UUI in the Unified CCE script, and if you use a REFER call flow, then the UUI is placed in a mime body and hex-encoded according to an ATT IP Toll Free NSS format. This placement of UUI also applies to 302 redirect responses.

Example of NSS Mime Body Format for UUI in REFER / 302 Messages

```
VER,1.00
PRN,t1113,*,att**,1993
FAC,
UUS,0,(hex encoded UUI string here)
```

Design Considerations

You cannot use the UUI data transfer feature with Hookflash or Two B Channel Transfer (TBCT).



CHAPTER 11

VXML Server Design Implications

VoiceXML Over HTTP

Communication between the VXML Server and Voice browser is based on request-response cycles using VoiceXML over HTTP. VoiceXML documents are linked together by using the Uniform Resource Identifiers (URI), which is a standardized technology to reference resources within a network. User input is carried out by web forms similar to HTML. Forms contain input fields that the user edited and sent back to a server.

Resources for the Voice browser are located on the VXML Server. These resources are VoiceXML files, digital audio, instructions for speech recognition (Grammars), and scripts. Every communication process between the VoiceXML browser and Voice application has to be initiated by the VoiceXML browser as a request to the VXML Server. For this purpose, VoiceXML files contain grammars that specify expected words and phrases. A link contains the URL that refers to the Voice application. The browser connects to that URL as soon as it recovers a match between spoken input and one of the grammars.

Note From Unified CVP Release 9.0(1) and later release the CVP installer installs CVP Call Server, CVP VXML Server and Media Server together. On installing CVP installer, you can configure only Call Server, VXML Server, Media Server or any other combination as required.

When determining the VXML Server performance, consider the following key aspects:

- QoS and network bandwidth between the Web application server and the voice gateway

  For details, see Network Infrastructure Considerations, on page 139.

- Performance on the VXML Server

  For details, see the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

- Use of prerecorded audio versus Text-to-Speech (TTS)

Voice user-interface applications tend to use prerecorded audio files wherever possible. Recorded audio sounds better than TTS. Prerecorded audio file quality must be designed so that it does not impact download time and browser interpretation. Make recordings in 8-bit mu-law 8 kHz format.

- Audio file caching

  Ensure that the voice gateway is set to cache audio content to prevent delays from downloading files from the media source. For details about prompt management on supported gateways, see Cisco IOS Caching and Streaming, on page 49.

- Use of Grammars

  A voice application, such as any user-centric application, is prone to certain problems that might be discovered only through formal usability testing or observation of the application in use. Poor speech recognition accuracy is one type of problem common to voice applications, and a problem most often caused by poor grammar implementation. When users mispronounce words or say things that the grammar designer does not expect, the recognizer cannot match their input against the grammar. Poorly designed grammars containing many difficult-to-distinguish entries also result in many incorrectly recognized inputs, leading to decreased performance on the VXML Server. Grammar tuning is the process of improving recognition accuracy by modifying a grammar based on an analysis of its performance.

Multi-Language Support

The Cisco IOS Voice Browser or the Media Resource Control Protocol (MRCP) specification does not impose restrictions on support for multiple languages. However, there may be restrictions on the automatic speech recognition (ASR) or TTS Server. Check with your preferred ASR or TTS vendor about their support for your languages before preparing a multilingual application.

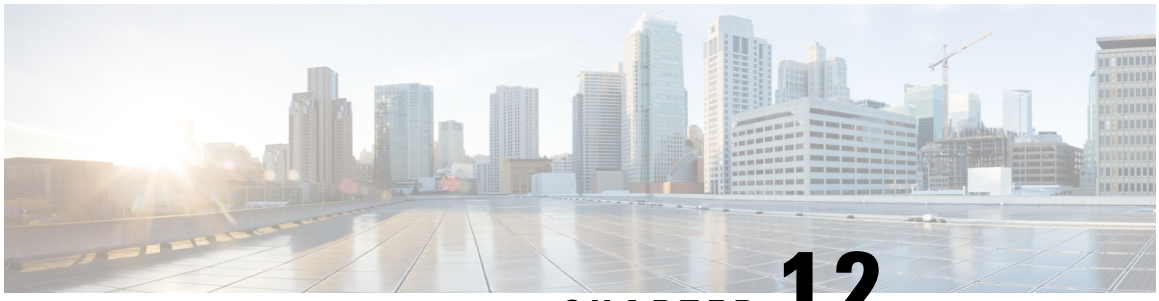
You can dynamically change the ASR server value by using the **cisco property com.cisco.asr-server** command in the VoiceVXML script. This property overrides any previous value set by the VoiceXML script.

Cisco Unified Call Studio Installation

Cisco Unified Call Studio is an Integrated Development Environment (IDE), which needs to be installed in a setup that is conducive for development, such as workstations that are used for other software development or business analysis purposes. Unified Call Studio is Eclipse-based, due to which development activities such as writing Java programs or building object models can be migrated to this tool so that developers and analysts have a common utility for their development needs.

For details on how to install Cisco Unified Call Studio, see *Installation and Upgrade Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-customer-voice-portal/tsd-products-support-series-home.html.

Note Unified Call Studio is supported only on Windows client software. Cisco does not support collocating the Cisco Unified Call Studio with the VXML Server.



CHAPTER 12

Network Infrastructure Considerations

Overview

For network infrastructure, you must consider deployment characteristics and provisioning requirements of the Unified CVP network. This chapter provides provisioning guidelines are presented for network traffic flows between remote components over the WAN, including the application of proper quality of service (QoS) to WAN traffic flows.

This chapter provides more information on network considerations, see the sections on deployment models, bandwidth, and QoS presented in *Cisco Unified Contact Center Enterprise Solution Reference Network Design (SRND)*, at http://www.cisco.com/en/US/products/sw/custcosw/ps1844/products_implementation_design_guides_list.html.

Bandwidth Provisioning and QoS Considerations

Some Unified CVP deployments have all the components centralized. Those deployments use a LAN structure, so WAN network traffic is not an issue. Consider the following scenarios when a WAN network structure is in a Unified CVP environment:

- In a distributed Unified CVP deployment when the Ingress Gateways are separated from the Unified CVP servers by a WAN.
- In Unified CVP deployments where the Ingress Gateway and the agent are separated over a WAN. The agent can be a TDM ACD agent or a Unified CCE agent.

Unified CVP considers QoS in the following way:

- CVP has no private WAN network structure. When required, WAN activity is conducted on a converged WAN network structure.

- CVP does not use separate IP addresses for high- and low-priority traffic.

Adequate bandwidth provisioning is important for Unified CVP deployments. This chapter provides bandwidth guidelines to help you plan your deployment.

Note Resource Reservation Protocol (RSVP) is used for call admission control. It is also used by the routers in the network to reserve bandwidth for calls. RSVP is not qualified for call control signaling through the Unified CVP Call Server in SIP. For Call admission Control, the solution is to employ Locations configuration on Unified CVP and Cisco Unified Communication Manager. See Enhanced Location Call Admission Control, on page 166.

Unified CVP Network Architecture

In a Unified CVP environment, group WAN and LAN traffic into voice traffic, call control traffic, and data traffic.

Voice Traffic

Voice calls consist of Real-Time Transport Protocol (RTP) packets. These packets contain voice samples that are transmitted into the following:

- Between the PSTN Ingress Gateway or originating IP phone and one of the following:
  - Another IP phone, such as an agent

    The destination phone may or may not be collocated (located on the same LAN) with the Ingress Gateway or calling IP phone, and the connection can be over a WAN or LAN.
  - An front-end Egress Gateway for a TDM ACD (for legacy ACDs or IVRs). The Egress Gateway may or may not be collocated with the Ingress Gateway, and the connection can be over a WAN or LAN.
  - A VoiceXML Gateway that performs prompt-and-collect treatment The VoiceXML Gateway can be the same or a different Ingress Gateway. In either case, both the Ingress and VoiceXML Gateways are collocated. The connection can be over a LAN or a WAN.
- Between the VoiceXML Gateway and the ASR or TTS Server. The RTP stream between the VoiceXML Gateway and ASR/TTS server must be G.711.

G.729 versus G.711 Codec Support

CVP supports mixed G.711 and G.729 codecs in Standalone and Comprehensive SIP deployments with Cisco Unified Border Element Enterprise Edition and Cisco Unified Communications Manager (Unified CM). Calls that are ingressed through a SIP trunk from the carrier to a Cisco Unified Border Element Enterprise Edition require Cisco Release IOS 15.1(2)T or later T for mixed codec support. You can use any combination of codecs on the legs of a call.

For more information on use of mixed codecs in a Unified CVP deployment, see Mixed G.729 and G.711 Codec Support, on page 83.

Table 12: Benefits and Drawbacks of G.729 and G.711 Codecs

Codec	Benefits	Drawbacks
G.711	This is the default codec shipped and works as is.	The solution requires more bandwidth over the WAN link.
G.729	Bandwidth consumption is less.	• Conversion of prompts to G.729 is required. • Audio quality of G.729 prompts is inferior to that of G.711 prompts. • ASR/TTS cannot be used. **Note** Cisco VVB does not support G.729 codec.

Call Control Traffic

A Unified CVP solution has several types of call control traffic. Call control functions are to set up, maintain, tear down, or redirect calls.

SIP

Unified CVP works in Call Control mode or Signaling mode with three types of VoIP endpoints: Cisco IOS Voice Gateways, Cisco Unified Communications Manager, and Peripheral Gateway. Call Control traffic flows between the following endpoints:

- Ingress Gateway and the Call Server

  The Ingress Gateway can be a Peripheral Gateway, Unified Communication Manager, a Cisco IOS Voice Gateway, or another SIP device. The connection can be over a WAN or a LAN.

- Call Server and Egress Gateway

  The Egress Gateway can be Unified Communication Manager or a Cisco IOS Voice Gateway. The Egress Gateway is either a VoiceXML Gateway that is used to provide prompt-and-collect treatment to the caller, or it is the target of a transfer to an agent (Unified CCE or TDM) or a legacy TDM IVR. The connection can be over a WAN or LAN.

Note Deployment designs do not support SIP for interoperability between the Peripheral Gateway and Unified CVP. If your design requires this functionality, contact the Cisco Assessment to Quality (A2Q) team.

GED-125

The Call Server and the Unified ICM VRU Peripheral Gateway communicates using the GED-125 protocol. The GED-125 protocol includes the following features:

- Notification messages, that control the caller experience when a call arrives.
- Instructions to transfer or disconnect the caller.

- Instructions that control the IVR treatment the caller experiences.

The VRU Peripheral Gateway connects to Unified CVP over a LAN connection. However, in deployments that use clustering over the WAN, Unified CVP can connect to the redundant VRU Peripheral Gateway across the WAN.

At this time, no tool exists that specifically addresses communications between the VRU Peripheral Gateway and Unified CVP. However, the bandwidth that is consumed between the Unified ICM Central Controller and VRU Peripheral Gateway is similar to the bandwidth that is consumed between the VRU Peripheral Gateway and Unified CVP.

The *VRU Peripheral Gateway to ICM Central Controller Bandwidth Calculator* tool, which is accessible through your credentials, is available through the Cisco Steps to Success Portal at http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-implementation-design-guides-list.html

You can also access the Bandwidth Calculator directly (with proper login authentication) at:

http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-implementation-design-guides-list.html

If the VRU Peripheral Gateway are split across the WAN, the total bandwidth that is required is double of what the calculator tool reports: one for Unified ICM Central Controller to VRU Peripheral Gateway and other for VRU Peripheral Gateway to Unified CVP.

Media Resource Control Protocol

The VoiceXML Gateway communicates with ASR/TTS Servers using Media Resource Control Protocol (MRCP) v1.0 and v2 only with IOS Voice Gateway. This protocol establish connections to the ASR/TTS Server, such as Nuance. The connection can be over LAN or WAN.

Note Cisco does not test or qualify speech applications in WAN environment. For guidelines on design, support over WAN and associated caveats, see the vendor-specific documentation. TAC will provide limited support (as in the case of any third-party interoperability certified products) on issues related to speech applications.

ICM Central Controller to Unified CVP VRU Peripheral Gateway

No tool exists that specifically addresses communications between the Unified ICM Central Controller and the Unified CVP VRU Peripheral Gateway. However, testing shows that the tool for calculating bandwidth that is needed between the Unified ICM Central Controller and the IP IVR Peripheral Gateway also produces accurate measurements for Unified CVP, if you perform the following substitution in one field.

For the **Average number of RUN VRU SCRIPT nodes** field, substitute the number of Unified ICM script nodes that interact with Unified CVP. Nodes that can interact with Unified CVP are Run External Script, Label, Divert Label, Queue to Skill Group, Queue to Agent, Agent, Release, Send to VRU, and Translation Route to VRU.

This bandwidth calculator tool is available (with proper login authentication) at:

http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-implementation-design-guides-list.html

The connection can be over a WAN or a LAN.

Data Traffic

Data traffic includes VoiceXML documents and prerecorded media files that are returned as a result of HTTP requests. VoiceXML Gateway executes the following requests the following:

- Media files in an HTTP request to a Media File Server—The Media File Server response returns the media file in the body of the HTTP message. The VoiceXML Gateway then converts the media files to Real-Time Transport Protocol (RTP) packets and plays them to a caller. The connection can be over a WAN or a LAN.
- VoiceXML documents from the CVP Server—In this case, the connection can be over a WAN or a LAN.

Note CVP Server includes VXML Server and CVP IVR service.

The data flows and bandwidth are used between a remote Ingress Gateway and the components for which bandwidth is required. These components are listed as follows:

- VXML Server
- Call Server IVR service
- Call Server SIP service
- IP phones
- Media Servers
- Egress Gateways
- ASR or TTS Servers
- Cisco VVB

Bandwidth Sizing

Generally, the distributed Unified CVP topology requires most of the bandwidth in a Unified CVP solution. This requirement occurs because the Ingress Gateway and VoiceXML Gateway is separated from the servers that provide it with media files, VoiceXML documents, and call control signaling.

Assume that all calls to a branch begin with 1 minute of IVR treatment. This call is followed by a single transfer to an agent that also lasts 1 minute. Each branch has 20 agents, and each agent handles 30 calls per hour for a total of 600 calls per hour per branch. The call average rate is 0.166 calls per second (cps) per branch.

Note A change in these variables can have a large impact on sizing. Remember that 0.166 calls per second is an average for the entire hour. There is no uniform pattern for calls coming across an hour, and there are usually peaks and valleys within the busy hour. Try to find the busiest traffic period and calculate the call arrival rate based on the worst-case scenario.

VoiceXML Document Types

A VoiceXML document is generated for every prompt that is played to the caller. This document is generated based on voice application scripts that you write using either Unified ICM scripts or Cisco Unified Call Studio, or both. A VoiceXML document varies in size, depending on the type of prompt being used. For example, menu prompts with multiple selections are larger in size than prompts that play announcements only.

Note The approximate size of a VoiceXML document for a Call Server or a VXML Server and the gateway is 7 kilobytes.

You can calculate bandwidth in the following ways:

- **Estimated bandwidth**—Calculate the estimated bandwidth for the number of prompts that are used per call per minute. For example, consider a VoiceXML document of 7 kilobytes and perform the following calculation:

  7,000 bytes * 8 bits = 56,000 bits per prompt

  (0.166 calls/second) * (56,000 bits/prompt) * (Number of prompts / call) = bps per branch

- **Exact Bandwidth**—Use the VoiceXML document sizes listed in the following table to calculate the required bandwidth. The document sizes in the following table are measured from the VXML Server to the VoiceXML Gateway.

Table 13: Approximate Size of VoiceXML Document Types

VoiceXML Document Type	VoiceXML Document Size (approximate)
Root document (one required at beginning of call)	19,000 bytes
Subdialog_start (at least one per call at beginning of call)	700 bytes
Query gateway for Call-ID and GUID (one required per call)	1300 bytes
Menu (increases in size with number of menu choices)	1000 bytes + 2000 bytes per menu choice
Play announcement (simple .wav file)	1100 bytes
Cleanup (one required at end of call)	4000 bytes

Note You can calculate exact bandwidth of VoiceXML document types if you use a complex application that uses multiple menu prompts.

Media File Retrieval

You can store media files also referred as prompts, locally in flash memory for IOS Voice Gateway and file system for Cisco VVB on each router. Storing them locally eliminates bandwidth considerations. However, it is difficult to maintain these prompts because a prompt that requires changes must be replaced on every router. If these prompts are stored locally on an HTTP media server (or an HTTP cache engine), the gateway can locally cache voice prompts after it retrieves the prompts. An HTTP media server can cache multiple prompts, depending on the number and size of the prompts. The refresh period for the prompts is defined on the HTTP media server. The utilized bandwidth is limited to the initial load of the prompts at each gateway, including the periodic updates after the expiration of the refresh interval.

Not caching prompts at the VoiceXML Gateway causes significant Cisco IOS performance degradation, 35 percent to 40 percent approximately, in addition to the extra bandwidth usage. For information on configuring gateway prompt caching, see *Configuration Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

For 50 prompts with an average size of 50 KB and a refresh interval of 15 minutes, the bandwidth usage is:

(50 prompts) * (50,000 bytes/prompt) * (8 bits/byte) = 20,000,000 bits

(20,000,000 bits) / (900 second) = 22.2 average kbps per branch

SIP Signaling

SIP is a text-based and signaling communications protocol that is used to control multimedia communication sessions, such as voice and video calls over Internet Protocol (IP) networks. It is also used to create, modify, and terminate sessions consisting of one or several media streams. These sessions include Internet telephone calls, multimedia distribution, and multimedia conferences. SIP can be used for two-party (unicast) or multiparty (multicast) sessions.

A typical SIP call flow uses about 17,000 bytes per call. Using the previous bandwidth formulas based on calls per second, the average bandwidth usage is:

(17,000 bytes/call) * (8 bits/byte) = 136,000 bits per call

(0.166 calls/second) * (136 kilobits/call) = 22.5 average kbps per branch

Automatic Speech Recognition and Text-to-Speech Server

ASR and TTS in WAN Configurations

Note Cisco does not test or qualify speech applications in a WAN environment. For guidelines on design, support over WAN, and associated caveats, see the vendor-specific documentation.

The Cisco Technical Assistance Center provides limited support (as in the case of any third-party interoperability-certified products) on issues related to speech applications.

Limiting the Maximum Number of ASR or TTS-Enabled Calls

You can limit the number of calls enabled for ASR or TTS so that as soon as the limit is reached, regular DTMF prompt-and-collect can be used instead of rejecting the call altogether. In the following example, assume 5559000 is the ASR or TTS DNIS and 5559001 is the DTMF DNIS. You can configure the Ingress

Gateway to do the ASR load limiting for you by changing the DNIS when you exceed maximum connections allowed on the ASR or TTS VoIP dial peer.

Note Cisco VVB does not support this feature.

```
voice translation-rule 3 rule 3 /5559000/ /5559001/
!
voice translation-profile change
 translate called 3
!
!Primary dial-peer is ASR or TTS enabled DNIS in ICM script
dial-peer voice 9000 voip
 max-conn 6
 preference 1
 destination-pattern 55590..
 ...
!
!As soon as 'max-conn' is exceeded, next preferred dial-peer will change
the DNIS to a DTMF prompt & collect ICM script
dial-peer voice 9001 voip
 translation-profile outgoing change
 preference 2
 destination-pattern 55590..
 ...
!
```

Note 80 kbps is the rate for G.711 full-duplex with no Voice activity detection, including IP/RTP headers and no compression. The rate for G.729 full-duplex with no VAD is 24 kbps, including IP/RTP headers and no compression. For information on VoIP bandwidth usage, see *Voice Codec Bandwidth Calculator* at http://tools.cisco.com/Support/VBC/do/CodecCalc1.do.

G.711 and G.729 Voice Traffic

Unified CVP supports both G.711 and G.729 codecs. However, both call legs and all IVRs on a given call must use the same voice codec. To use ASR/TTS for speech recognition, use G.711 codec because ASR/TTS server supports G.711 only. For information on voice RTP streams, see *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*, at http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-implementation-design-guides-list.html.

Port Usage and QoS Settings

The Call Server marks only the QoS DSCP for SIP messages. If QoS is needed for Unified CVP signaling and data traffic across a WAN, configure network routers for QoS using the IP address and ports to classify and mark the traffic, as described in the following table.

Neither the CVP-Data queue nor the Signaling queue is a priority queue as described in Cisco IOS router terminology. The priority queue is used for voice or other real-time traffic, while call signaling and Unified CVP traffic are reserved for a certain amount of bandwidth based on the call volume.

Table 14: Port Usage and QoS Settings

Component	Port	Queue	PHB	DSCP	Maximum Latency (Round Trip)
Media Server	TCP 80	CVP-Data	AF11	10	1 s
Unified CVP Call Server, SIP	TCP or UDP 5060	Call Signaling	CS3	24	200 ms
Unified CVP IVR Service	TCP 8000	CVP-Data	AF11	10	1 s
Unified CVP VXML Server	TCP 7000	CVP-Data	AF11	10	1 s
Ingress Voice Gateway, SIP	TCP or UDP 5060	Call Signaling	CS3	24	200 ms
VoiceXML Gateway, SIP	TCP or UDP 5060	Call Signaling	CS3	24	200 ms
SIP Proxy Server	TCP or UDP 5060	Call Signaling	CS3	24	200 ms
MRCP	TCP 554	Call Signaling	CS3	24	200 ms

Network Latency

After the proper application bandwidth and QoS policies are in place, consider the network latency in a distributed CVP deployment. With sufficient network bandwidth, the primary contributor to latency is the distance between the VoiceXML Gateway and the Call Server or VXML Server. In distributed CVP deployments, minimize the latency and understand its effect on solution performance.

Network latency affects the distributed CVP deployment in the following ways:

- Affects the end-user calling experience when the network latency is between CVP components. Call signaling latency with SIP between the Call Servers and voice gateways affects the call setup time, and may add a period of silence during this setup. It includes the initial call setup and subsequent transfers or conferences that are part of the final call flow.
- Affects the VoiceXML application document download time significantly, and has a pronounced effect on the ultimate caller experience.

The following system configuration changes can help to minimize the effect of geographic separation of the VoiceXML Gateway from the VXML Server which results in WAN delays:

1. Provide audio to the caller during periods of silence.

   The following settings provide ringback and audio during times of dead air so that the caller does not disconnect:

   - To add a ringback tone during longer than normal call setup times with IVR, on the survivability service, keep the **wan-delay-ringback** setting at 1.

- Add the IVR subsystem settings for **IVR.FetchAudioDelay** and **IVR.FetchAudioMinimum**. These WAN Delay settings are required when root document fetch is delayed over the WAN link.
- Specify the value for **IVR.FetchAudio** as follows: **`IVR.Fetchaudio= flash:holdmusic.wav`**. Leave the default empty so that nothing is played in a normal scenario.

Note Cisco VVB does not support this feature.

Note

- Retain the default setting of 2 to avoid a blip sound in a normal network scenario.
- Set WAN Delay to zero to play holdmusic.wav immediately for a minimum of 5 seconds.
- Use ECC variables, such as **`user.microapp.fetchdelay`**, **`user.microapp.fetchminimum`**, and **`user.microapp.fetchaudio`**, to override ECC variable values in between invocations of getSpeechExternal microapps.

2. Enable Path MTU Discovery on the IOS Voice Gateways.

   On the IOS Voice Gateways, add the **ip tcp path-mtu-discovery** command.

   Path MTU Discovery method is used to maximize the use of available bandwidth in the network between the endpoints of a TCP connection.

3. Minimize round trips between the VXML Server and the ICM script.

   When control is passed from a running VXML Server application back to the ICM script, you incur a significant WAN delay.

   Aftter the VXML Server application starts to run, minimize the number of trips back to the ICM script. Each round trip between the VXML Server and the ICM script incurs a delay because it establishes two new TCP connections and HTTP retrieval of several VoiceXML documents, including the VXML Server root document.

4. Decrease the size of the VXML Server root document.

   On the VXML Server, in your gateway adapter plugin.xml file change:

   <setting name="vxml_error_handling">default</setting>

   To:

   <setting name="vxml_error_handling">minimal</setting>

   For example, the location of the plugin.xml file for the CISCO DTMF 1 GW adapter is:

   `Cisco\CVP\VXMLServer\gateways\cisco_dtmf_01\6.0.1\plugin.xml`

Note HTTP transfers VoiceXML documents and other media files that are played to the caller. For the best end-user calling experience, treat the HTTP traffic with a priority higher than that of normal HTTP traffic in an enterprise network. If possible, treat this HTTP traffic the same as CVP call signaling traffic. As a workaround for latency issues , you can move the VXML Server to the same local area as the VoiceXML Gateway, or use Wide Area Application Service (WAAS).

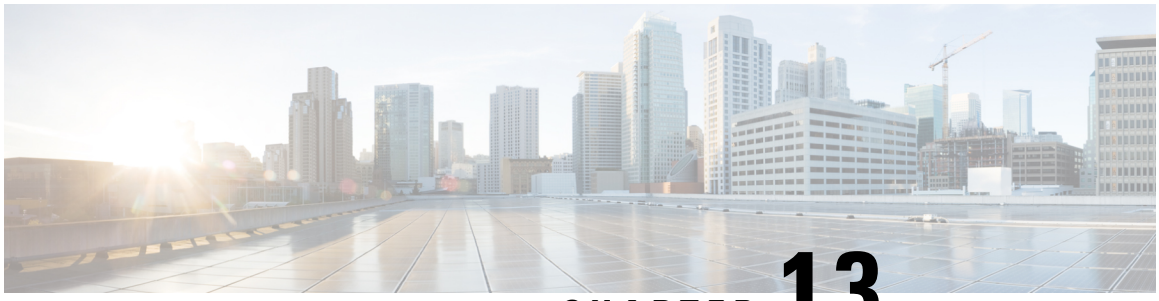
TCP/UDP Ports Used by Unified CVP, Voice, and VoiceXML Gateways

When configuring network security using firewalls or access control lists (ACLs), see the following table for information about TCP/UDP ports that are used by Unified CVP, voice gateways, and VoiceXML gateways. For a complete list of ports that Unified CVP uses, see the *Port Utilization Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/partner/products/sw/custcosw/ps1006/prod_technical_reference_list.html.

Note The Operations Console Server uses dynamic ports for communication with other components, so it cannot be deployed outside of a firewall while all other Unified CVP components reside inside the firewall.

Table 15: TCP/UDP Ports Used by Unified CVP, Voice Gateways, and VoiceXML Gateways

Source and Destination Component	Destination Port
VoiceXML Gateway to Media Server	TCP 80
VoiceXML Gateway to Unified CVP Call Server SIP	TCP or UDP 5060
VoiceXML Gateway to Unified CVP Call Server	TCP 8000 (non-SSL); TCP 8443 (SSL)
VoiceXML Gateway to Unified CVP VXML Server	TCP 7000 (non-SSL); TCP 7443 (SSL)
VoiceXML Gateway to MRCP V1 (RTSP) Server	TCP 554
IOS Voice Gateway to MRCP V2 (SIP) Server	TCP 5060
Unified CVP Call Server to Egress Voice Gateway SIP	TCP or UDP 5060
Unified CVP Call Server to VoiceXML Gateway SIP	TCP or UDP 5060
Unified CVP Call Server to SIP Proxy Server	TCP or UDP 5060



CHAPTER 13

Features and Functions

Multicast Music-on-Hold

Multicasting is used for Music-on-Hold (MOH) with supplementary services on Unified CM as an alternative to the unicast MOH. There are two ways to deploy MOH using this feature:

- With Unified CM multicasting the packets on the local LAN
- With the branch gateway multicasting on their local LAN

Use the latter method when survivable remote site telephony (SRST) is configured on the gateway. This method enables the deployment to use MOH locally and avoid MOH streaming over the WAN link.

Note Refer to the following location for information about configuring MOH on the Call Manager Enterprise (CME):

http://www.cisco.com/en/US/docs/voice_ip_comm/cucme/admin/configuration/guide/cmemoh.html#wpmkr1022205

Call Survivability in Distributed Deployments

Distributed deployments require design guidelines for other voice services that are being run at the branch. For example, the branch is a remote Unified CM site supporting both ACD agent and nonagent phones. This deployment also implies that the PSTN Gateway is used not only for ingress of Unified CVP calls but for ingress or egress of the regular non-ACD phone calls.

Branch reliability in WANs may be an issue in a centralized Unified CVP model because they are typically less reliable than LAN links. The call survivability function must be considered for both the Unified CVP and non-CVP calls. For Unified CM endpoint phones, survivability is accomplished by using a Cisco IOS feature known as Survivable Remote Site Telephony (SRST). For further details on SRST, see the latest version of the *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*, available at:

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

For Unified CVP calls, survivability is handled by a combination of services from a TCL script (survivability.tcl) and SRST functions. The survivability TCL script monitors the SIP connection for all calls that ingress through the remote gateway. If a signaling failure occurs, the TCL script takes control of the call and redirects it to a configurable destination. The destination choices for the TCL script are configured as parameters in the Cisco IOS Gateway configuration.

Note When the called number is in "E164" format, the survivability script removes the "+" sign from the called number before forwarding it to Unified CVP. This is because Unified CVP or ICM does not support the "+" sign in the beginning of DNIS.

Alternative destinations for this transfer include another IP destination (including the SRST call agent at the remote site), *8 TNT, or hookflash. With transfers to the SRST call agent at the remote site, the most common target is an SRST alias or a basic ACD hunt group. For further information about these SRST functions, see the *Cisco Unified Communications Solution Reference Network Design (SRND) based on Cisco Unified Communications Manager*.

Voice mail and recording servers do not send Real-Time Control Protocol (RTCP) packets in reverse direction toward the caller (TDM Voice Gateway), which can falsely trigger the media inactivity timer of the survivability script. It is important to apply the survivability.tcl script carefully to the dial peers because a call might drop if it goes to the voice mail or to a recording element. One method is to use a separate dial peer for voice mail or recording calls, and do not associate the Unified CVP survivability script for those dial peers. Another method is to disable the media inactivity on the survivability script associated with the voice mail or recording dial peers.

For further information on configuration and application of these transfer methods, see the latest version of *Configuration Guide for Cisco Unified Customer Voice Portal*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

You can also refer to CUBE Deployment with SIP Trunks, on page 73.

Note To take advantage of alternate routing on signaling failures, you must use the survivability service on all gateways pointing to Unified CVP. Always use this service, unless you have a specific implementation that prevents using it.

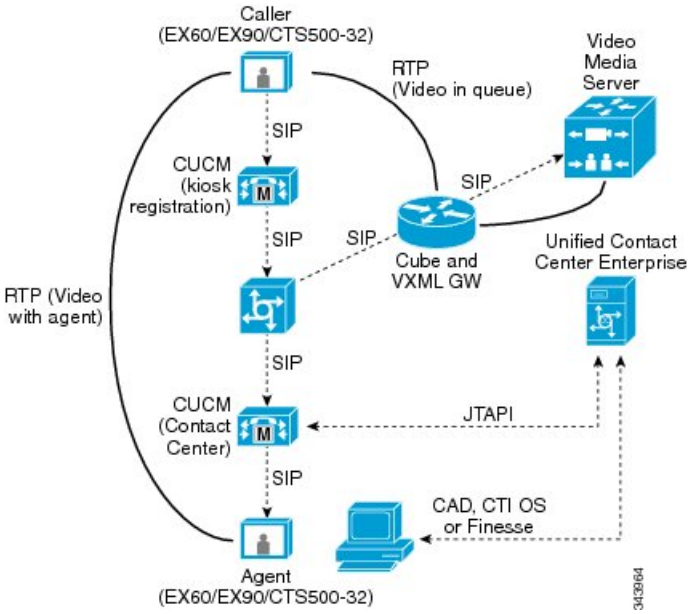
Router requery is not supported when using SIP REFER with Unified CVP Comprehensive Call Flow when the survivability service is handling the REFER message from Unified CVP. Router requery with REFER can be supported in other call flows when Cisco IOS is handling the REFER without the survivability service or if Unified CM is handling the REFER. For third-party SIP trunks, the support of router requery with REFER is dependent on their implementation and support for SIP REFER.

Video in Queue

Video in Queue (VIQ) is an optional Basic Video feature in Unified CVP. It allows the caller to interact through high-definition video prompt or navigate a video menu using DTMF keys. The following figure displays the topology and call flow for Basic Video.

Note Cisco VVB does not support Video in Queue.

Figure 11: Video in Queue



The Unified CVP Studio VideoConnect element allows the specific video prompt to be played for video endpoints. It also allows the DTMF input during video-prompt playback to be collected and integrated with the Unified Call Studio or Unified CCE scripting environment.

Note Video in Queue is not played during a CUCM failover.

See the *Configuration Guide for Cisco Unified Customer Voice Portal* for specific Cisco Unified Border Element or VXML Gateway configuration information for VideoConnect.

See the *Element Specifications for Cisco Unified CVP VXML Server and Cisco Unified Call Studio* for using the VideoConnect element.

See “Incoming Call Configuration and Media File Management” in the MediaSense User Guide to use media files.

Note When configuring the Video in Queue for Unified CVP, set the MediaSense **Incoming Call Configuration** > **Action** to play once.

Custom SIP Headers

The Custom SIP Header feature enables Unified CVP to pass selected SIP header information to and from Unified ICM for modification within ICM scripts. This feature allows much greater flexibility in providing SIP interoperability with third-party SIP trunks and gateways.

Passing Information in SIP Headers to Unified ICM

Unified CVP enables the passing of one or more SIP headers to Unified ICM for manipulation within the ICM script. Unified CVP administrator can use the Unified CVP Operations Console Server user interface (Operations Console) to select a specific header, or a header and specific parameters within that header. These SIP headers can be passed to Unified ICM in the SIPHeader field of the New Call and Request Instruction messages sent from the CVP ICM subsystem to Unified ICM.

To access the variable in the ICM script, access the Call.SIPHeader field. Setting this field causes Unified CVP to use that data in outbound SIP calls to IVR or Agents, or REFER or 302 redirect messages.

The amount of space available to send header data to Unified ICM is limited and is truncated to 255 bytes. The SIP protocol RFC provides a function to represent common header field names in an abbreviated form. The compact header format as defined in RFC 3261 (and other RFCs for newly defined headers) is used for the header titles before passing the header to Unified ICM.

Note Not all headers have a compact format. For example, P-Headers or private headers (for example P-Asserted-Identity) may not have a compact form, and the full header name is passed to ICM.

Note See the table in the RFC3261 that defines the compact header abbreviations.

String Formats and Parsing

The following example shows the formatting of a string sent to Unified ICM based on Operations Console SIP configuration screen settings:

```
"User-to-User: 123456789"
"f:Name <sip:from@127.0.0.1:6666>;param1;param2|v:SIP/2.0/UDP viaHost"
```

The delimiter is the bar character.

The data may be parsed with string manipulation syntax in the script such as this example.

Caution **No syntax checking.**

There is no syntax checking while adding or modifying headers in the Operations Console. You must be careful that the headers are in correct SIP syntax. The only characters not allowed in Operations Console input are the semicolon and the comma because these are used internally to store the configuration. Typically, if there is a problem with the header syntax, the CVP log shows that the INVITE is not sent due to a SIP stack-parsing exception, and the call is aborted. In other cases, if a mandatory SIP header is modified incorrectly, the call itself may get sent to an unexpected destination or the receiver may not be able to handle the call if the message is not conforming to RFC.

Passing of Headers from the ICM Script

The objective of this feature is to provide a scriptable option to modify SIP headers on the outgoing Unified CVP transfer. You can specify SIP header values in outgoing SIP INVITES only. The specifying can include the addition, modification, or removal of header values.

Note The SIP header modification feature is a powerful tool which can tweak SIP headers as needed. Exercise caution when applying SIP Profiles and ensure that the profile does not create interop issues, rather than solving them. Unified CVP provides the flexibility to add, modify, or remove outgoing SIP header in the INVITE message only. You can deploy Unified CVP in many scenarios to facilitate interoperability with third-party devices.

Outgoing SIP header feature do not allow you to remove or add Mandatory SIP headers. Only the modify option is available for basic mandatory headers, such as To, From, Via, CSeq, Call-Id and Max-Forwards. There is no checking for the modifications in the ICM script editor, it is actually enforced by the java SIP stack layer by throwing a DsSipParserException.

Typically, with Unified ICM, if the field is greater than 255 charcters then it is truncated. In the SIP subsystem, if there is a problem updating or adding a header with the string provided the Unified ICM script, then you either see an WARN type message in the Unified CVP log, if there is an DsSipParserException, or else sends the INVITE sends unexpected results on the receiver end.

This feature is applicable only for outgoing SIP INVITES (only the initial INVITE, not reinvites). Changes to the INVITE are applied just before it is sent out. There is no restriction on the changes that can be applied.

The header length (including header name) after modification should not exceed 255.

Examples of Unified ICM Scripting for Custom SIP Headers

In the script editor, the Set node is used to set the call variable string for SIPHeaderInfo.

In the Unified ICM script delimit the header, operation, and value with a tilde character, and use the bar character to concatenate operations.

Scripting Examples for Outbound Header Manipulations

Example	Notes
"Call-Info~add~<sip:x@y>;parm1=value1"	Adds a Call-Info header with the proper call info syntax as per RFC3261.

Example	Notes
"Via~add~SIP/2.0/UDP viaHost"	Adds a Via header to the message.
"v~add~SIP/2.0/UDP viaHost|f~mod~<sip:123@host>;parm1=value1"	Short Form notation, plus concatenated operations. Adds a Via header and modifies the From header.
"Call-Info~add~parm1=value1"	**Incorrect:** This will fail due to incorrect syntax of Call-Info header per RFC 3261. You will see a WARN message in the CVP log. This is enforced in the stack.
"From~add~<sip:x@y>;parm1=value1"	From header add and modify will do the same thing, since only one From header is allowed in a message per RFC 3261. This is enforced in the stack.
"Call-ID~add~12345@xyz"	Same as From header, only one allowed.
"Call-ID~mod~12345@abc"	Same as From header, only one allowed.
"User-To-User~mod~this is a test|P-Localization-Info~mod~1234567890"	Can be used to concatenate operations in one ICM variable Set Node.
"Call-ID~rem"	Removes the first header called Call-Id in the message.

Troubleshooting information for Unified CVP can be found on the Unified CVP Doc-Wiki Troubleshooting page: http://docwiki-dev.cisco.com/wiki/Troubleshooting_Tips_for_Unified_Customer_Voice_Portal.

Courtesy Callback

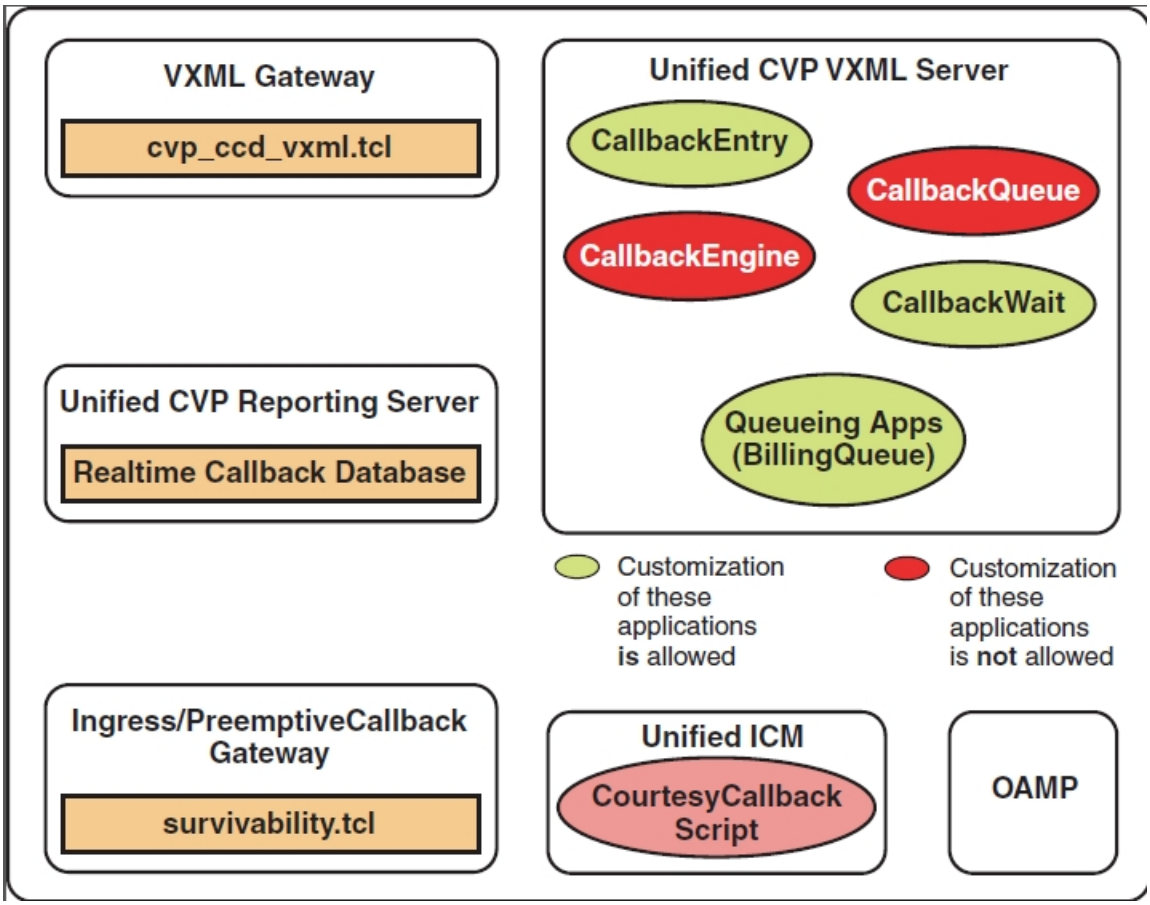
Courtesy Callback reduces the time callers have to wait on hold or in a queue. The feature enables your system to offer callers, who meet your criteria, the option to be called back by the system instead of waiting on the phone for an agent. The caller who has been queued by Unified CVP can hang up and subsequently be called back when an agent is close to becoming available (preemptive callback). This feature is provided as a courtesy to the caller so that the caller does not have to wait on the phone for an agent.

Preemptive callback does not change the time a customer must wait to be connected to an agent, but rather enables the caller to hang up and not be required to remain in queue listening to music. Callers who have remained in queue or have undergone the callback treatment will appear the same to agents answering the call.

Note Scheduling a callback to occur at a specified time is not part of this feature.

Figure 12: Courtesy Callback Components

The following illustration shows the components needed for the Courtesy Callback feature.



Note The Courtesy Callback applications on the VXML Server must not be invoked more than once for the same call.

The Courtesy Callback uses TCL service on IOS Voice Gateway and TCL built-in feature on Cisco VVB.

Typical Use Scenario

If the caller decides to be called back by the system, they leave their name and phone number. Their request remains in the system and the EWT executes when the system places a callback to the caller. The caller answers the call and confirms that they are the original caller, and the system connects the caller to the agent after a short wait.

Note Courtesy Callback is supported for IP originated calls as well.

A typical use of the Courtesy Callback feature follows this pattern:

1. The caller arrives at Unified CVP and the call is treated in the normal IVR environment.

2. The Call Studio and Unified ICM Courtesy Callback scripts determine if the caller is eligible for a callback based on the rules of your organization (such as in the prior list of conditions).
3. If a courtesy callback can be offered, the system tells the caller the approximate wait time and offers to call the customer back when an agent is available.
4. If the caller chooses not to use the callback feature, queuing continues as normal.

   Otherwise, the call continues as indicated in the remaining steps.
5. If the caller chooses to receive a callback, the system prompts the caller to record their name and to key in their phone number.
6. The system writes a database record to log the callback information.

Note If the database is not accessible, then the caller is not offered a callback and they are placed in queue.

7. The caller is disconnected from the TDM side of the call. However, the IP side of the call in Unified CVP and Unified ICM is still active. This keeps the call in the same queue position. No queue music is played, so Voice XML Gateway resources used during this time are less than if the caller had actually been in queue.
8. When an agent in the service or skill category the caller is waiting for is close to being available (as determined by your callback scripts), then the system calls the person back. The recorded name is announced when the callback is made to ensure that correct person accepts the call.
9. The system asks the caller, through an IVR session, to confirm that they are the person who was waiting for the call and that they are ready for the callback.

   If the system cannot reach the callback number provided by the caller (for example, the line is busy, RNA, network problems, etc.) or if the caller does not confirm they are the caller, then the call is not sent to an agent. The agent is always guaranteed that someone is there waiting when they take the call. The system assumes that the caller is already on the line by the time the agent gets the call.

   This feature is called preemptive callback as the system assumes that the caller is already on the line by the time the agent gets the call and that the caller has to wait minimal time in queue before speaking to an agent.
10. The system presents the call context on the agent screen-pop, as normal.

If the caller cannot be reached after a configurable maximum number and frequency of retries, the callback is aborted and the database status is updated appropriately. You can run reports to determine if any manual callbacks are necessary based on your business rules.

See the *Configuration Guide for Cisco Unified Customer Voice Portal* at https://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html that explains a call flow description of the function of the scripts providing the Courtesy Callback feature.

Determine Callback Time

The following information provides an overview of how callback time is determined, the determination process and the calculation methods employed.

These are some definitions of key terms used:

- Wait Time—The interval of time between when the call enters the queue and when the call leaves the queue.
- Reconnect Time—The interval between the point at which the callback is started and the point at which the caller has accepted the callback and is waiting for an agent.
- Callback in Queue Time—The interval between when the caller is reconnected, waiting for an agent and when the call leaves the queue
- Service Level Agreement (SLA)—Average of Callback in Queue Time. Average means that roughly 50 percent of calls are within the service level and 50 percent are outside the service level.
- Average Dequeue Time—The average number of seconds it takes for a call to leave the queue.
- Remaining Time—The number of seconds left to count down to call back the caller.

Callback in Queue Time

The average Callback in Queue Time after a callback is designed to be within an agreed service level. However, Courtesy Callback is also designed so that callers are not called back too early or too late, as both scenarios are undesirable. On the one hand, if callers are called back too early, then they are more likely to have to wait in the queue for a longer period of time, while, if the callback is made too late, there is a greater chance that call center agents could be idle and waiting for calls.

When the dynamics of a call center change, such as when more or fewer agents are available, or when the average handle time changes, it in turn causes the remaining time to change. Therefore, with Courtesy Callback, the Average Dequeue Time is calculated based on various factors such as calls in queue, average handle time, and agents in ready and talking states.

The Average Dequeue Time is updated when a call enters the queue and when it leaves the queue. This information is used for calculations for reducing the Callback in Queue Time and minimizing instances of call center agents waiting for calls.

Process Details and Calculation Methods

The following information details the process used to determine the callback time for calls in the queue. It also shows the method, or formula, used to calculate the Average Dequeue Time as well as the method used to update the remaining time for all Courtesy Callback calls in the queue.

The process for determining callback time is as follows:

1. The Average Dequeue Time (D) is calculated using the formula, D = (EWT + F)/N
    - EWT is the estimated wait time for a new Courtesy Callback call.
    - F is the number of seconds that the first call is already in position in the queue.
    - N is the number of calls in queue.

Note The Dequeue Time plays a significant role in the optimal behavior of the Courtesy Callback feature. The average Dequeue Time is calculated based on factors such as call volume, agent availability, and the average handle time for a particular skill group.

The Estimated Wait Time (EWT) is an approximation, and its accuracy is driven by the uniform average handling time and agent availability for a particular skill group. If these factors are not uniform, it may lead to a difference in the estimated wait time announced to the customer and the actual callback time. In addition, the use of microapps may insert calls into the queue that were not included in the EWT calculation. For scripting of calls that includes Courtesy Callback, all calls must be queued on the IVR using VxmlScripting, instead of microapps.

2. The remaining time for all Courtesy Callback calls in the queue is updated using the formula: R(p) = p*D - F - C:
   - p = 1, ..., N
   - R(p) is the remaining time for the pth queue position Courtesy Callback call.
   - C, the post-callback time, is the sum of the time it takes to retrieve the Courtesy Callback caller back on the phone and the SLA time.

Note Courtesy Callback can support a default wait time of 30 minutes with an exception of maximum 90 minutes.

Example Scripts and Audio Files

The courtesy callback features is implemented using Unified ICM scripts. Modifiable example scripts are provided on the Unified CVP install media in the \CVP\Downloads and Samples\ folder. These scripts determine whether or not to offer the caller a callback, depending on the criteria previously described. The files provided are:

- CourtesyCallback.ICMS, the ICM script
- CourtesyCallbackStudioScripts.zip, a collection of Call Studio scripts

Sample audio files that accompany the sample studio scripts are installed to the <CVP_HOME>\OPSConsoleServer\CCBDownloads\CCBAudioFiles.zip and also as part of the Media Files installation option.

If CCBAudioFiles.zip is used, the contents must be unzipped onto the media server. CCBAudioFiles.zip has Courtesy Callback-specific application media files under en-us\app and media files for Say It Smart under en-us\sys. If you already have media files for Say It Smart on your media server, then only the media files under en-us\app are needed.

Note The default prompts work for most of the default Call Studio scripts. However, the application designer must review and provision the Say It Smart plugin prompts for specific cases that are not covered within the default set of prompts.

The sample scripts are configured to use the default location of http://<server>:<port>/en-us/app. The default location of the sample audio files must be changed in the sample scripts to accommodate your needs (that is, substitute the media server IP address and port in <server> and <port>).

The following example scripts are provided:

- BillingQueue

  This script is responsible for playing queue music to callers that either choose to not have a callback occur or must reenter the queue for a short period after receiving a callback.

  You may customize this script to suit your business needs.

- CallbackEngine

  This script keeps the VoIP leg of a callback alive between when a caller elects to have a callback and when a caller receives the callback.

  Do **not** customize this script.

- Callback Entry

  This script handles the initial IVR when a caller enters the system and when the caller is provided with the opportunity to receive a callback.

  You may customize this script to suit your business needs.

- CallbackQueue

  This script handles the keepalive function of a call while a caller is in queue and listening to the music played within the BillingQueue script.

  Do **not** customize this script.

- CallbackWait

  This script handles the IVR portion of a call when a customer is called back.

  You may customize this script to suit your business needs.

Callback Criteria

Examples of callback criteria you can establish include:

- Number of minutes a customer is expected to be waiting in queue exceeds some maximum number of minutes (based on your average call handling time per customer)

  Note The included sample scripts use this method for determining callback eligibility.

- Assigned status of a customer (gold customers may be offered the opportunity to be called back instead of remaining on the line)
- The service a customer has requested (sales calls for example, or system upgrades may be established as callback criteria)

Courtesy Callback Design Considerations

The following design considerations apply to the Courtesy Callback (CCB) feature:

- During Courtesy Callback, callback is made using the same Ingress Gateway through which the call arrived.

  Note In Courtesy Callback, outbound calls cannot be made using any other Egress Gateway.

- Calls that allow Callback must be queued using a Unified CVP VXML Server.
- The Unified CVP Reporting Server is a prerequisite for Courtesy Callback.
- Answering machine detection is not available for this feature. During the callback, the caller is prompted with a brief IVR session message and acknowledge with DTMF that they are ready to take the call.
- Calls that are transferred to agents using DTMF *8, TBCT, or hookflash cannot use the Courtesy Callback feature.
- Courtesy Callback feature does not support Agent call transfers to CCB Queue, over a computer telephony integration (CTI) route point.
- Callbacks are a best-effort function. After a limited number of attempts to reach a caller during a callback, the callback is terminated and marked as failed.
- Customers must configure the allowed or blocked numbers that Callback is allowed to place calls through the Unified CVP Operations Console.
- Media inactivity detection feature on the VXML Gateway can affect waiting callback calls. For more information, see the *Configuration Guide for Cisco Unified Customer Voice Portal*.
- Courtesy Callback requires an accurate EWT calculation for its optimal behavior.

  Do the following to optimize the EWT, when using Precision Queues for Courtesy Callback:

  - Queue the calls to a single Precision Queue
  - Do not include a `Consider If` expression when you configure a step.
  - Do not include a wait time between steps or use only one step in the Precision Queue.

  Note Make sure that you use simple Precision Queue definitions (for example, with one step and one-to-one agent mapping). The complexity of Precision Queues makes calculating accurate EWT difficult.

Post Call Survey

A contact center typically uses a post call survey to determine whether a customer was satisfied with their call center experience. For example, did the customer find the answer they were looking for using the self service or did they have a pleasant experience with the contact center agent.

The Post Call Survey (PCS) feature enables you to configure a call flow so that after the agent hangs up, the caller is transferred to a DNIS that prompts the caller with a post call survey.

There are two responses a caller can have to a post call survey request:

1. The caller is prompted during IVR treatment as to whether they would like to participate in a post call survey. If they choose to do so, they are automatically transferred to the survey call after the agent ends the conversation.
2. The caller is prompted to participate, but declines the post call survey. A Unified ICM script writer can use an ECC variable to turn off the ability for Post Call Survey on a per-call basis. By setting the ECC variable to *n*, the call will not be transferred to the PCS DNIS.

For reporting purposes, the post call survey call has the same Call-ID and call context as the original inbound call.

Typical Uses

The caller is typically asked whether they would like to participate in a survey during the call. In some cases, the system configuration based on dialed numbers determines if the post call survey gets invoked at the end of conversation with agents. When the customer completes the conversation with an agent, the customer is automatically redirected to a survey. The post call survey gets triggered by the hang-up event from the last agent.

A customer can use the keypad on a touch tone phone and voice with ASR/TTS to respond to questions asked during the survey. From the Unified CCE point of view, the post call survey call is just like another regular call. During the post call survey, the call context information is retrieved from the original customer call.

Design Considerations

Observe the following conditions when designing the Post Call Survey feature:

- A Post Call Survey is triggered by the hang-up event from the last agent. When the agent hangs up, the call routing script launches a survey script.
- The mapping of a dialed number pattern to a Post Call Survey number enables the Post Call Survey feature for the call.
- The value of the expanded call variable **user.microapp.isPostCallSurvey** controls whether the call is transferred to the Post Call Survey number.
    - If **user.microapp.isPostCallSurvey** is set to **y** (the implied default), the call is transferred to the mapped post call survey number.
    - If **user.microapp.isPostCallSurvey** is set to **n**, the call ends.

- To route all calls in the dialed number pattern to the survey, your script does not have to set the **user.microapp.isPostCallSurvey** variable. The variable is set to **y** by default.

- REFER call flows are not supported with Post Call Survey. The two features conflict: REFER call flows remove Unified CVP from the call and Post Call Survey needs Unified CVP because the agent has already disconnected.
- For Unified CCE reporting purposes, when a survey is initiated, the call context of the customer call that was just transferred to the agent is replicated into the call context of the Post Call Survey call.

Call Admission Control

Call admission control is the function for determining if there is enough bandwidth available on the network to carry an RTP stream. Unified CM can use its own locations function or RSVP to track bandwidth between the Ingress Gateway and destination IP phone locations.

For more information about call admission control, see the chapter on Distributed Deployment, on page 31.

In networks, Resource Reservation Protocol (RSVP) is a protocol used for call admission control, and it is used by the routers in the network to reserve bandwidth for calls. RSVP is not qualified for call control signaling via the Unified CVP Call Server in SIP. As an alternative, the recommended solution for Call Admission Control is to employ locations configuration on Unified CVP and in Unified CM.

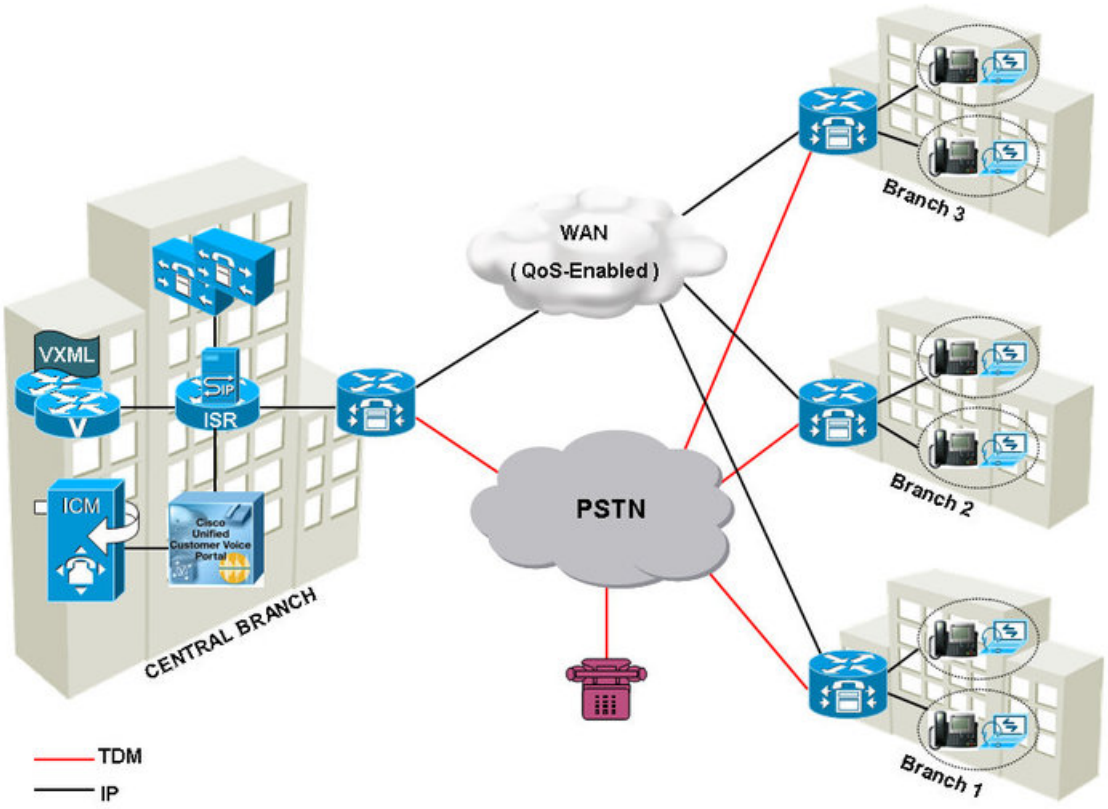
For more information on RSVP, see the latest version of the *Cisco Unified Communications SRND Based on Cisco Unified Communications Manager*, available at:

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

Queue-at-the-Edge Branch Office Deployment Model

The following figure illustrates a typical branch office deployment.

Figure 13: Typical Branch Office Deployment.



You can deploy Unified CVP in a single cluster Unified CM deployment to provide queue-at-the-edge functionality. In this deployment model, branch-located Ingress Gateways are typically used to allow callers access using local phone numbers rather than centralized or non-geographic numbers. This consideration is especially important in international deployments spanning multiple countries.

Egress Gateways are located at branches either for localized PSTN breakout or for integration of decentralized TDM platforms (ACDs) into the CVP switching solution. Apart from the gateways all other CVP components are centrally located and WAN links provide data connectivity from each branch location to the central data center. (Although the media server is centrally located, commonly used VRU media is cached at the local branch.)

In the Unified CVP branch office deployment model using queue-at-the-edge, the only equipment at the branch office is an Ingress Gateway (optionally acting as a VoiceXML Gateway as well), IP phones for Unified CCE agents, IPT (user) phones, and agent desktops.

You can configure Unified CCE Skill Groups, dial plans and routing priorities so that callers who Ingress at one branch are connected by preference to agents who are located at the same branch. In these cases, the RTP traffic flows directly from Ingress Gateway to IP phone, and does not need to traverse the WAN (although signaling and data may traverse the WAN).

The goal of this model is to first route the calls locally to an agent available in the branch office, if possible, and keep the media streams local. If the local agent is not available, only the call gets routed to the agent on another branch office over the WAN link; the originating call and the initial VRU treatment are done locally.

Another advantage of this deployment configuration is that in the event of WAN link failure, the call can still be routed locally using the CVP survivability application running on the pots dial-peer for TDM originated calls.

Enhanced Location Call Admission Control

ELCAC Concepts

The following definitions are important to the ELCAC feature:

- **Phantom Location**—A default location with unlimited bandwidth used when calculating calls that are hairpinned over a SIP trunk or when the SIP call is queued at the local branch, to enable correct bandwidth calculations. The Phantom location should be assigned to the gateway or trunk for CVP.
- **siteID**—The siteID is a string of numbers that Unified CVP appends to the label it receives from Unified ICM. Depending on the siteID the dial plan can be configured to route the call to a specific destination, such as the branch VXML Gateway or Egress Gateway, or UCM node. The siteID can be appended at the front of the label, at the end, or not at all. This configuration is separate from the Unified CM location configuration, and is specific to Unified CVP. The siteID is used to indicate the real location of the call and allow the bandwidth to be deducted from the correct location. siteID is unique across multiple Unified CM clusters. Multiple siteIDs can still route to the same branch office (if needed) by mapping the unique siteIDs to same branch gateways in proxy routes.
- **Shadow Location**—This new location is used for intercluster trunks between two Cisco Unified Communications Manager clusters. This location is not used as intercluster ELCAC is not supported in Unified CVP.

Locations are created in UCM. Unified CVP gets these locations when you synchronize the location information from the UCM on operations console. You can associate a siteID for these locations on operations console and then associate your gateways to these locations. Based on this configuration, CVP creates two hash objects. One hash would map location to a siteID and the second hash would store mapping of GW IP address to location name and siteID. These hash objects are used to route the call to appropriate GW to provide edge queuing (using siteID), and pass around the location information on the call legs for UCM to do proper CAC calculations.

For a Unified CVP branch office deployments the following considerations apply:

- Control the number of calls that goes over the WAN link to branch offices based on the available bandwidth of the WAN link.
- For the queue-at-the-edge functionality, the call originating from a specific branch office should be routed to a local VXML Gateway on priority.

When you are using the Unified CVP intracluster Enhanced Location CAC model deployment, you must control the number of calls that go over the WAN link to branch offices. The decision to admit calls is based on the CAC computations, which represent the bandwidth used by the call. These computations are valid whether the calls are IP calls between two phones within Cisco Unified Communications Manager, calls over SIP trunks, or calls originated from TDM-IP Gateway.

For queue-at-the-edge functionality, the call originating from a specific branch office must be routed to a local Unified CVP VXML Gateway based on priority. That is, always choose a local branch agent if possible.

Unified CVP supports topology modeling with Enhanced Location Call Admission Control (ELCAC) for intracluster. It does not support intercluster Enhanced Location CAC. Location Bandwidth Manager is enabled for intracluster CAC, but disabled for intercluster CAC. For more information on ELCAC topology modeling, see the Cisco Unified Communications SRND based on Cisco Unified Communications Manager, available at http://www.cisco.com/c/en/us/support/unified-communications/unified-communications-manager-callmanager/products-implementation-design-guides-list.html.

Comparison of Enhanced Location Call Admission Control Feature

The Enhanced Location Call Admission Control (ELCAC) feature addresses two important issues with the prior CAC feature:

1. Bandwidth miscalculations in CAC with IP originated callers, as well as with any post transfers from agents.
2. Inability to deterministically select a local VXML GW for VRU treatment at the branch office during warm transfers from an agent due to no correlation between the two calls at consult.

Comparison with OrigIP Trunk Feature on Unified CM

Before Unified CM implemented the phantom trunk and siteID feature for bandwidth calculation, there was the existing feature used by Unified CVP that enabled the correct trunk to be selected depending on the original ID of the caller. This feature enabled Unified CM to select to the correct trunk for the TDM gateway, instead of only using the single Unified CVP trunk, and it only applies to incoming calls on the trunk. With this feature, distinct SIP profiles and trunk settings could be used for each branch gateway without being limited to the settings of the single Unified CVP trunk. This feature has no impact on bandwidth calculations.

Router Requery with ELCAC

When a call is rejected by the UCM due to not enough bandwidth, a SIP message 488 Not Acceptable Here is returned to Unified CVP, where it triggers a router requery over the GED-125 interface to the VRU peripheral, and the UCCE Router may return another agent label if requery is configured properly.

Design Considerations

The following considerations apply when using ELCAC:

- The SIP trunk configured between Unified CVP and Unified CM should be associated with Phantom location. A new location called shadow location is added in Unified CM 9.0 for inter-cluster ELCAC, but it is not supported in Unified CVP.
- In multi-cluster CUCM deployments, consider over subscribing bandwidth on WAN links based on the anticipated peak call volume or choose a centralized branch office deployment model, as intercluster ELCAC is not supported on Unified CVP.
- In single-cluster CUCM deployments, ELCAC is supported only for Hub and Spoke topology with Unified CVP.
- A trunk configured with MTP required will not work with the ELCAC siteID feature. The reason is when MTP is inserted, the media is terminated between the end point and MTP resource, not between the two end points.

- If a MTP/Transcoder/TRP media resource is inserted by the Unified CM media layer, the incoming location information is not used.
- If the intercluster call is not looped back to the same cluster, the former behavior of Location CAC logic will apply.
- Each site is uniquely identified by one siteID. Multiple gateways at the same site would need to align to the same siteID, but if two clusters happen to use the same location name, then two siteIDs can map to the same physical branch.
- A second Unified CM cluster may have the same location as the first cluster, but be required to use a unique siteID on Unified CVP. You can define a route in the proxy server to send those cluster calls to the common VXML Gateway at the same location, but used by both the clusters.
- Each cluster would manage the bandwidth for devices in its cluster. If two clusters happen to use the same physical location, then they would each separately manage the bandwidth for the phones that they manage.

High Availability and Failover

The following considerations apply when using LBCAC:

- During the CAC failure, Unified CVP returns a failure code to Unified CCE that triggers router requery.
- If a branch does not have a VXML Gateway, then use the VoiceXML Gateway at the Central data center.

Additional ELCAC Information

The previous version of Unified CVP provided a method of configuring CAC. This method is superseded by the ELCAC method presented here. Both configuration methods are provided in the *Configuration Guide for Cisco Unified Customer Voice Portal*, available at:

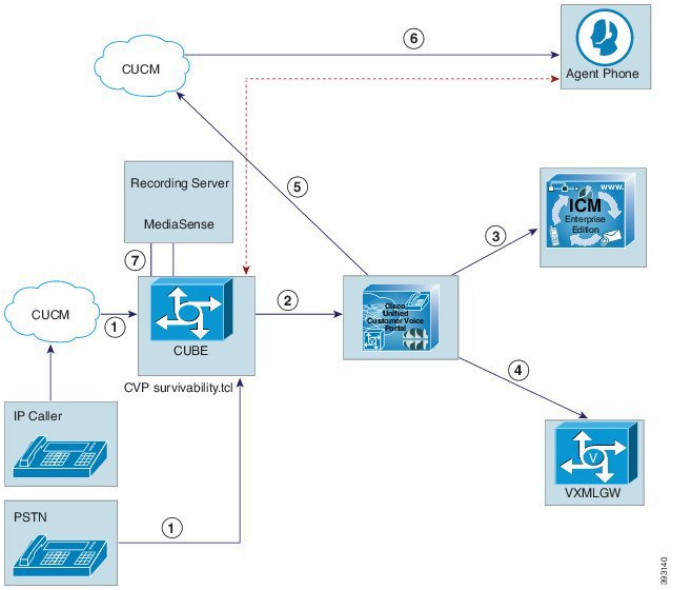
http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html

Network-Based Recording

The network-based recording (NBR) feature supports software-based forking for Real-time Transport Protocol (RTP) streams. Media forking provides the ability to create midcall multiple streams (or branches) of audio and video associated with a single call and then send the streams of data to different destinations. To enable network-based recording using CUBE, refer to the configuration guide. You can configure specific commands or use a call agent. CUBE acts as a recording client and MediaSense recorder acts a recording server.

Note Network-based recording works with the call survivability feature.

The following figure displays the topology and call flow for network-based recording.



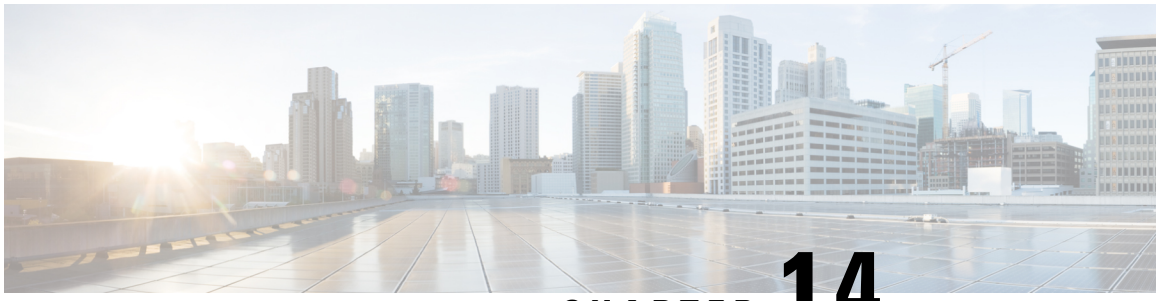
A typical call flow for network-based recording is as follows:

1. A customer calls using an IP phone or by using PSTN.
2. The Ingress Gateway sends the call to Unified CVP.
3. Unified CVP sends the incoming call request to ICM and gets a VRU label.
4. Unified CVP sends the call to the VoiceXML Gateway. The caller hears the IVR. However, the call is not recorded.
5. After the agent is available, Unified CVP connects the caller to the agent.
6. Network-based recording starts for this conversation.

Limitations

- For agent to agent call transfer, network-based recording does not work but phone-based recording does. If you want to use network-based recording, you can use an ISR gateway between Unified CVP and Cisco Unified CM.
- The NBR feature is currently supported on ISR G2 gateways, in an inbound contact center deployment only.

Note The NBR feature is supported only on selected IOS image trains. For more information about the supported IOS image trains, see the *Unified CCE Solution Compatibility Matrix*, available at http://docwiki.cisco.com/wiki/Compatibility_Matrix_for_Unified_CCE.



CHAPTER 14

Call Transfer Options

Release Trunk Transfer

Release Trunk Transfer releases the ingress trunk and removes Unified CVP and the gateway from the call control loop. These transfers have the following characteristics:

- They can be invoked by VXML Server (Standalone Call Flow Model) or using Unified ICM.
- Unified ICM Network Transfer using Unified CVP as the routing client does not work because Unified CVP can no longer control the call.
- They are blind, that is, if the transfer fails for any reason then Unified ICM does not recover control of the call. Router Requery is not supported.
- They cause the switch leg to terminate resulting in a Telephony Call Dispatcher (TCD) record being written to the database for the call even though the caller is still potentially talking to an agent. This behavior differs from other types of transfers in which the TCD record is not finalized until the caller hangs up.
- As the ingress trunk is released, you do not have to size gateways to include calls that have been transferred using Release Trunk Transfer. This behavior differs from other types of transfers in which gateway resources continue to be occupied until the caller hangs up.
- Because Unified CVP is no longer monitoring the call, you do not have to size Call Servers to include calls that have been transferred using Release Trunk Transfer. Additionally, Unified CVP Call Director port licenses are not required.

Following are the signaling methods available to trigger a release trunk transfer:

- Takeback-and-Transfer. See Takeback-and-Transfer, on page 172
- Hookflash and Wink. See Hookflash and Wink, on page 172

- Two B Channel Transfer. See Two B Channel Transfer, on page 173

Takeback-and-Transfer

Takeback-and-Transfer (TNT), also known as Transfer Connect, is a transfer method where dual tone multifrequency (DTMF) tones are outpulsed to the PSTN by Unified CVP. TNT outpulses DTMF tones to the PSTN. A typical DTMF sequence is *8xxxx, where xxxx represents a new routing label for the PSTN. Upon detection of a TNT DTMF sequence, the PSTN drops the call leg to the Ingress Gateway port, and then reroutes the caller to a new PSTN location, such as a TDM ACD location. This method is offered by a few PSTN service providers.

Customers can use TNT, if they have an existing ACD site but no IVR and want to use Unified CVP as an IVR. Over time, customers may need to transition agents from the TDM ACD to Unified CCE and use Unified CVP as an IVR, queueing point, and transfer pivot point. Using Unified CVP as more than just an IVR eliminates the need for TNT services.

In Unified CVP deployments with Unified ICM, the DTMF routing label that is outpulsed can be a Unified ICM translation routing label to enable passing of call data to another Unified ICM peripheral, such as a TDM ACD. In this scenario, Unified CVP views the call as completed, and Unified CVP call control is ended. With TNT, if the transfer to the termination point fails, Unified CVP cannot reroute the call. Using some TNT services, you can reroute the callback to Unified CVP. However, Unified CVP treats this call as a new call.

Hookflash and Wink

Hookflash and wink are signaling methods that are associated with a TDM PBX or ACD. With the Hookflash feature, Unified CVP can transfer SIP calls using a hookflash followed by the DTMF destination. This feature allows for deployments in which a PBX is in the front-end of the Unified CVP Ingress Gateway, and in which the PBX provides non-VoIP connectivity to agents. Hookflash applies to analog trunks and wink applies to digital trunks (T1 or E1 channel), although both are similar in function. Both hookflash and wink send an on-hook or off-hook signal to the PBX or ACD, which responds with dial tone (or the PBX winks back on a digital trunk). This signaling causes the Voice Gateway to send a string of routing digits to the PBX or ACD. Upon collection of the routing digits, the PBX or ACD transfers the caller to the new termination point, which can be an ACD queue or service on that same PBX or ACD.

Customers can use hookflash and wink if they have an existing ACD but no IVR, and want to use Unified CVP as an IVR that is installed on the line side of their existing PRX or ACD. Over time, the customer may need to transition agents from the TDM ACD to Cisco Unified CCE and have the Voice Gateways connected to the PSTN instead of the line side of the PBX or ACD. In Unified CVP deployments with Unified ICM, the routing label can be a Unified ICM translation routing label. This label enables passing of call data to the ACD service (and subsequently to the agent in a popup message). With hookflash and wink, if the transfer to the termination point fails, Unified CVP cannot reroute the call. Some PBX or ACD models can reroute the callback to Unified CVP. However, Unified CVP treat this call as a new call.

Note When PBXs and gateways have constrained support, hookflash transfer becomes difficult. If possible, avoid using the PBX for Unified ICM switching. Also, terminate all incoming calls on Unified CVP ingress gateways to allow Unified CVP to route calls to the PBX rather than on the ingress gateways.

Following guidelines and notes apply for hookflash transfers:

- Cisco 1700 Series Gateways are not tested with hookflash transfers.

- Cisco 2800 and 3800 Series Gateways can support Analog FXO or Digital FXO (T1/CAS). This function is considered line-side hookflash to the PBX. However, E&M is not supported at this time. You can adjust the hookflash duration with the **timing hookflash-out** command in **voice-port**. This feature is useful if you have a PBX that has a nonconfigurable hookflash duration, and it gives you the ability to adjust the hookflash duration on the gateway side.
- Cisco 5x00 Series Gateways are tested with T1/CAS and the **e&m-fgb dtmf dnis** command. E&M is considered "trunk-side hookflash" to the PBX, and not all switches support trunk-side hookflash. Additionally, the hookflash duration on the Cisco 5x00 Series Gateways is 200 ms, and you must configure the PBX for the same duration. This option varies with switch type and a proof-of-concept with the switch used.
- In Deployment Model No. 1, Standalone Self-Service, a TCL script is required to produce the hookflash. A TCL script is provided with Unified CVP.
- For Digital FXO (T1 CAS) Trunks, configure Dialed Number Identification Service (DNIS) on the gateway, based on the T1/E1 channel on which the call arrives. The PBX is programmed to route DNIS calls over T1 trunks. Configure the DNIS of the gateway because the call arrives to the gateway on that trunk.
- For Digital FXO (T1 CAS) Trunks, configure Dialed Number Identification Service (DNIS) on the gateway, based on the T1/E1 channel on which the call arrives. The PBX is programmed to route DNIS calls over T1 trunks. Configure the DNIS of the gateway because the call arrives to the gateway on that trunk.

  Note The disadvantage to this approach is that the gateway trunk allocation must be predetermined. You must know the percentage of calls that arrive to a particular DNIS so that the trunk groups on the gateway can be allocated accordingly.

  An alternate method, also known as converse on step, can be used on some PBXs where DTMF tones indicating DNIS and ANI are sent to the IVR. This method requires a single main Unified ICM routing script to input DNIS digits using a Get Data (GD) Microapplication and to invoke the correct subscript based on the collected DNIS digits. This method requires close coordination between Cisco, the PBX vendor, and the customer.
- For FGB E&M trunks in Cisco 5x100 Series Gateways, ANI and DNIS can be sent by using "*" as the delimiter, for example, *ANI*DNIS*. For configuration details, see *ANI/DNIS Delimiter for CAS Calls on CT1*, available at http://www.cisco.com/c/en/us/td/docs/ios/redirect/eol.html.

Note

- Hookflash is supported on 2X and 3X gateways only.
- Hookflash applies to TDM-originated calls only. After Unified CVP invokes hookflash, Unified CVP is no longer in control of the call.

Two B Channel Transfer

Two B Channel Transfer (TBCT) is an Integrated Services Digital Network (ISDN)-based release trunk signaling function that is offered by some public switched telephone network (PSTN) service providers. When

a TBCT is invoked, the Ingress Gateway places the initial inbound call on hold briefly while a second call leg (ISDN B Channel) is used to call the termination point. When the termination point answers the call, the gateway sends ISDN signaling to the PSTN switch to request to complete the transfer, bridge the call through the PSTN switch, and remove the call from the Ingress Gateway. As with a TNT transfer, the termination point might be a TDM PBX or ACD connected to the PSTN.

This process may be necessary for a customer with an existing ACD site but no IVR, who wants to use Unified CVP initially as just an IVR. Over time, the customer might want to transition agents from the TDM ACD to Cisco Unified CCE and use Unified CVP as an IVR, queueing point, and transfer pivot point (which eliminates the need for TBCT services and using Unified CVP to perform reroute on transfer failure).

ICM Managed Transfer

Unified CVP performs ICM Managed Transfer function, which provides gateway-based switching for Unified ICM and Unified CCE installations.

In Unified CVP deployments with Unified ICM, Unified ICM provides all call control. VoiceXML call control from the VXML Server is not supported when Unified ICM is deployed with Unified CVP.

Unified ICM Managed transfer is used to transfer the call to any of the following new termination points:

- A Cisco Unified Communications Manager phone
- An egress port on the same gateway as the ingress port
- A distant Egress Gateway that has a TDM connection to a TDM ACD or PBX (making use of toll bypass features)
- A Unified CVP VoiceXML gateway for queuing or self-service activities

To terminate a call, the Voice Gateway selects an outgoing POTS or VoIP dial peer based on the destination specified by Unified ICM. When a Unified ICM VoIP transfer occurs, the Ingress Voice Gateway port is not released. If the termination point is an Egress Voice Gateway, then a second Voice Gateway port is utilized. Unified CVP continues to monitor the call, and Unified ICM also retains control of the call and can instruct Unified CVP to transfer the call to a new destination.

This type of transfer is used when Unified CVP is used as a call treatment platform and queue point for Unified CCE agents. Unified CVP can also be used to provide call treatment to front-end calls to TDM ACD locations supported by Unified ICM. This type of transfer allows calls to be transferred between peripherals supported by Unified ICM, with full call context and without any return of the voice path.

Calls that are transferred in this way have the following characteristics:

- Unified ICM Network Transfer using Unified CVP as the routing client functions properly because Unified CVP continues to control the call.
- These transfers are supervised, meaning that if the transfer fails for any reason, the Unified ICM routing script does recover control through Router Requery method.
- The switch leg does not terminate until the caller hangs up. The TCD record that is written for the switch leg of the call encompasses the entire life of the call, from initial ingress to hang up.
- Gateways sizing is done to include calls that have been transferred using ICM Managed Transfers because the ingress trunk is not released.

- Call Servers sizing is done to include the calls that have been transferred using ICM Managed Transfers because Unified CVP continues to monitor the call. Additionally, Unified CVP Call Director port licenses are required, except for calls that are connected to Cisco Unified Communications Manager agents.

Network Transfer

Unified CVP allows Network Transfer to transfer calls to another destination after they have been answered by an agent.

When a call is transferred from Unified CVP to an agent, and that agent wants to transfer the call to another agent, the agent can make that transfer using either the agent IP phone or the agent desktop. Transfers from the IP phone are made using CTI route points that point to a Unified ICM script. Transfers from the agent desktop are made using the Dialed Number Plan.

There are two flags in Unified ICM to control the Network Transfer:

- **NetworkTransferEnabled**—This flag is part of the Unified ICM script. When enabled, it instructs the Unified ICM to save the information about the initial routing client (the routing client that sent the NewCall route request).
- **NetworkTransferPreferred**—This flag is enabled on the Unified CVP Peripheral Gateway configuration. When enabled, any route request from this routing client sends the route response to the initial routing client instead of the routing client that sent the route request.

The following points explain how you can do a network transfer:

- You can use Network Transfer to perform a blind transfer only from agent 1 to agent 2 through Unified CVP. In this case, Unified ICM instructs Unified CVP to route the call back from agent 1, and then route it either to a VoiceXML Gateway (for IVR treatment) or to another destination (for example, to agent 2).
- You cannot use Network Transfer to perform a warm transfer or conference with Unified CVP because the call leg to agent 1 must be active while agent 1 performs a consultation or conference. Unified CVP cannot route the call back from agent 1 during the warm transfer or conference.

If a caller dials the same number regardless of a blind transfer, warm transfer, or conference, then follow these best practices:

- Do not enable the **NetworkTransferEnable** flag in the Unified ICM script.
- Dial the CTI Route Point of the same Unified CCE Peripheral Gateway for any transfer or conference request to preserve the call context during the transfer. Dialing the Route Pattern or CTI Route Point of another Peripheral Gateway does not preserve the call context.
- Use **SendToVru** as the first node in the Unified ICM routing script.

Note Extra ports are used during the consultation, blind transfer, or conference calls. They are released after the originating consultation is terminated.

SIP Refer Transfer

In some scenarios, Unified CVP transfers a call to a SIP destination and does not have Unified ICM and Unified CVP retain any ability for further call control. Unified CVP can perform a SIP Refer transfer, which allows Unified CVP to remove itself from the call, and free licensed Unified CVP ports. The Ingress Voice Gateway port remains in use until the caller or the terminating equipment releases the call. SIP Refer transfers are used in both Comprehensive and Call Director deployments.

Invoke a SIP Refer transfer by any of the following methods:

- Unified ICM sends Unified CVP a routing label with a format of rfXXXX (For example, rf5551000).
- An application-controlled alternative is to set an ECC variable (user.sip.refertransfer) to the value **y** in the Unified ICM script, and then sends that variable to Unified CVP.

Note Direct Refer transfer using label works only if**Send To VRU** node is used before the Refer.

You can invoke the SIP Refer transfer after Unified CVP queue treatment has been provided to a caller. SIP Refer transfers can be made to Cisco Unified Communications Manager or other SIP endpoints, such as a SIP-enabled ACD.

Router requery on a failed SIP Refer transfer is supported using SIP with the Unified CVP, but only on calls where the survivability service is not handling the SIP Refer request.

Intelligent Network Release Trunk Transfers

Customers who use Deployment Model No. 4 (VRU Only with NIC Controlled Routing) rely on call switching methods that do not involve Unified CVP. In that scenario, all switching instructions are exchanged between a Unified ICM Network Interface Controller (NIC) and the PSTN. Examples of these NIC interfaces include Signaling System 7 and Call Routing Service Protocol (CRSP). The NIC is also used as an interface into the Peripheral Gateway in deployments that involve the device. Peripheral Gateway deployments perform Intelligent Network Release Trunk Transfers.

VoiceXML Transfer

VoiceXML call control is supported only in Standalone deployments in which call control is provided by the VXML Server. Deployment Model No. 3b, which also incorporates VXML Server, does not support VoiceXML call control. In Unified ICM integrated deployments, ICM controls all calls.

The VXML Server can invoke the following types of transfers:

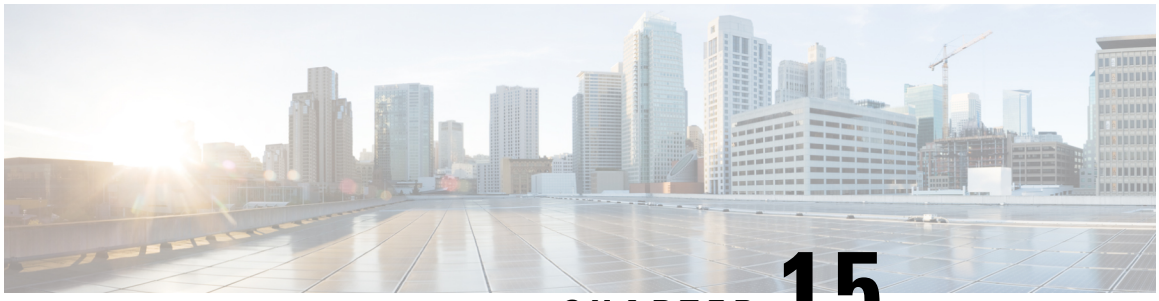
Table 16: Types of VoiceXML Transfers

Release Trunk Transfer	VoiceXML Blind Transfer	VoiceXML Bridged Transfer
Result in the incoming call being released from the Ingress Voice Gateway.	Result in the call being bridged to an Egress Voice Gateway or a VoIP endpoint. However, the VXML Server releases all subsequent call control.	Result in the call being bridged to an Egress Voice Gateway or a VoIP endpoint. However, the VXML Server retains call control so that it can return a caller to an IVR application or transfer the caller to another termination point.
Are invoked using the **subdialog_return** element. VXML Server can invoke a TNT transfer, Two B Channel transfer, HookFlash/Wink transfers, and SIP Refer Transfers. For TDM Release Trunk Transfers (TNT, TBCT and Hookflash/Wink), the VoiceXML Gateway must be combined with the Ingress Gateway for the Release Trunk Transfer to work.	Are invoked using the Transfer element in Cisco Unified Call Studio. These transfers transfer the call to any dial peer that is configured in the gateway.	Bridged transfers do not terminate the script. The VXML Server waits until either the ingress or the destination call ends. The script ends only if the ingress call leg hangs up. If the destination call leg hangs up first, the script recovers control and continues with additional self-service activity. Note that the VXML Server port license remains in use for the duration of a bridged transfer, even though the script is not actually performing any processing.

VoiceXML blind transfers differ from VoiceXML bridged transfers in the following ways:

- VoiceXML blind transfers do not support call progress supervision; bridged transfers support it. This means that if a blind transfer fails, VXML Server script does not recover control and cannot attempt a different destination or take remedial action.
- VoiceXML blind transfers cause the VXML Server script to end. Always connect the “done exit” branch from a blind transfer node to a subdialog_return and a hang up node.

Note Cisco VVB supports Blind Transfer only under VoiceXML Transfer.



CHAPTER 15

Managing, Monitoring, and Reporting Functions

Operations Console

Operations Console is a web-based interface to configure and monitor Unified CVP components and devices in the Unified CVP solution. You can manage the following Unified CVP components from the Operations Console:

- Call Server
- VXML Server
- Reporting Server

Note Operations Console is also referred to as the OAMP (Operate, Administer, Maintain, and Provision). The Operations Console manages individual components through the Unified CVP Resource Manager, which is collocated with each managed Unified CVP component. The Resource Manager is invisible to the enduser.

For details on Operations Console, see the *Configuration Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/partner/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.

DS0 Trunk Information for Reporting

Unified CVP passes the PSTN gateway trunk and DS0 information on which the SIP call arrives at Unified ICM. This information can be used for routing and for reporting.

For details, see DS0 Trunk Information, on page 131 and Trunk Utilization Routing and Reporting, on page 132.

End-to-End Individual Call Tracking

When a call arrives at a Unified CVP Ingress Gateway, Cisco IOS assigns that call a 36-digit hexadecimal Global Unique Identifier (GUID), which identifies the call. Unified CVP carries that GUID through all of the components that the call encounters, as follows:

- Ingress gateway—Shown in Cisco IOS log files.
- VoiceXML gateway—Shown in Cisco IOS log files.
- Unified CVP components—Shown in Unified CVP log files.
- Unified Intelligent Contact Management (ICM)—Shown in the Extended Call Context (ECC) variable user.media.id and stored with all Termination Call Detail (TCD) and Route Call Detail (RCD) records.
- Automatic speech recognition (ASR) and text-to-speech (TTS) servers—Shown in logs as the logging tag.
- Cisco Unified Communications Manager (Unified CM)—Appears in the detailed logs.

With proper levels of logging enabled, a call can be traced through all of the above components.

The Unified CVP logs are located in $CVP_HOME\logs. All of the Unified CVP logs roll over at 12:00 AM every night, with the date as part of the filename. The format of the date is yyyy-mm-dd. All of these logs also roll over when they reach the predefined size limit of 100 MB and have a number as part of the filename extension. The number indicates which log it was for that day. When the entire log directory reaches a predefined size, old files are purged as necessary.

For more information on Unified CVP logging, see the *Troubleshooting Guide for Cisco Unified Customer Voice Portal* and Troubleshooting wiki at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/prod_troubleshooting_guides_list.html.

Note Unified CVP components do not themselves synchronize machine times. Customers must provide a cross-component time synchronization feature, such as NTP, to assure accurate time stamps for logging and reporting.

Reporting System

Reporting Server houses the Reporting Service and hosts an IBM Informix Dynamic Server (IDS) database management system. The Reporting Service does not itself perform database administrative and maintenance activities, such as backups or purges. However, Unified CVP provides access to such maintenance tasks through the Operations Console.

The Reporting Service:

- Provides historical reporting to a distributed self-service deployment in a call center environment. This system is used to assist call center managers with call activity summary information to manage daily operations.
- Can also provide operational analysis of various IVR applications.

- Receives reporting data from the IVR Service, the SIP Service (if used), and the VXML Server.

Note To capture data from the VXML Server in the Reporting Server database, in the Operations Console, select the **CVP VXML Server** device to add the VXML Server. Selecting the **VXML Server Standalone** device option does not capture the Unified CVP Reporting data.

- Is deployed together with an Informix database management system, and it transforms and writes this reporting data into that database. The database schema conforms with Unified CVP. However, the schema is fully published so that customers can develop custom reports based on it.

A single Reporting Server may be used in a deployment. If a single Reporting Server is used, it does not necessarily represent a single point of failure, because data safety and security are provided by the database management system, and temporary outages are tolerated due to persistent buffering of information on the source components.

The following are the limitations if multiple Reporting Servers are used:

- Each Call Server can be associated with one Reporting Server only.
- Reports cannot span multiple Informix databases.

Note Unified CVP components cannot synchronize machine time themselves. Customers must provide a cross-component time synchronization feature, such as NTP, to assure accurate time stamps for logging and reporting.

Reporting Features

- Reporting Server provides increased data retention times by increasing the database space requirements. The size for Unified CVP Release 8.0(1) and above are 100 GB.

Note For Unified CVP, the 2 GB option for database size is not supported for production.

- All database backup files are compressed and stored on the Reporting Server. The backup file is called **cvp_backup_data.gz** and is stored on the %INFORMIXBACKUP% drive in the **cvp_db_backup** folder.
- Using the system CLI, you can make the request to list log files on the Reporting Server (**show log**). This request includes the Informix Database Server Engine logs. The **show tech-support** command also includes these files.
- With the debug level 3 (or 0) command from within the System CLI, you can turn on and turn off the debug. When turned on, this command generates trace files for all administrative procedures, Purge, Statistics, and Aggregator.

Note After the command is turned on, trace files place an elevated burden on the database.

- Log data for administrative procedures are written on a nightly basis to the **%CVP_HOME%\logs** folder.
- All the **StartDateTime**, **EndDateTime**, and **EventDateTime** values are stored as UTC in Reporting Server tables.
- The Reporting Server supports the Analysis Manager tool by allowing Analysis Manager to query the Reporting Server with the authenticated user, such as **cvp_dbuser**, credentials.
- Transfer Type data and Transfer Labels for SIP call events are stored in the call event table.
- The Data aggregator is introduced to aggregate Unified CVP data in 15-minute increments. Cisco Unified Intelligence Center templates are created to capture this information. Call data is summarized at 15-minute, daily, and weekly intervals. Dominant Path information is summarized at the same intervals. These summaries are stored in the **call_15**, **call_daily**, **call_weekly**, **applicationsummary_15**, **applicationsummary_daily**, and **applicationsummary_weekly** tables. Call data is summarized into the **Call_*** structure, while an aggregate of each element invoked by each application is stored in the **ApplicationSummary_*** structures.
- Summary purge results are logged in the log table.
- Three new scheduled tasks have been added to the Reporting Server scheduler:
  - **CVPSummary**, which builds summary tables.
  - **CVPCallArchive**, which archives Callback data to maintain callback database performance.
  - **CVPLogDump**, which extracts the administrative logs on a nightly basis.
- All metadata for administrative processes have been moved into the **Ciscoadmin** database. This process removes the tables from normal view of reporting users.

Note For reporting requirements related to the Courtesy Callback feature, see Courtesy Callback, on page 156.

Cisco Unified IC Templates

Customers who want to generate user friendly reports on call data stored in the database use Unified Intelligence Center templates.

See the following guides for information about the packaged Unified CVP template and the procedure to create additional templates:

- *Reporting Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.
- *Reporting Guide for Cisco Unified ICM Enterprise & Hosted* at http://www.cisco.com/en/US/products/sw/custcosw/ps1001/products_user_guide_list.html.

Backup and Restore

Unified CVP utilizes redundant array of independent disks (RAID) as protection against failure of a single drive in a mirrored pair. However, RAID 10 does not protect against the loss of a site, loss of a machine, or a loss of both mirrored drives.

Using Operations Console, you can schedule daily database backups or run database backups on-demand. You can restore database manually to the last backup time so that the worst-case scenario of losing about 24 hours worth of data is saved.

Database backups are written to the local database server. However, storing backups only on a local machine does not protect the system against server failures or the loss of a site. Copy the backup Unified CVP files to a different machine, preferably at a different location. Ensure security and backup management responsibilities.

Backups are compressed and stored on disk. During a backup, the oldest of two backups is removed and replaced with the most recent backup while a new backup is made. In the event of a hardware failure during a backup that results in a bad backup image, the older backup image can be used to replace the failed backup image. Retention of older backups is beyond the scope of the Reporting Server and should be managed by the customer.

Restore Process

In Cisco Unified CVP, there is a supported script to perform a database restore.

A backup image needs to be restored:

- In the event that older data on a backup image needs to be recovered.
- In the case of a machine that has been rebuilt after a hardware failure, where you want to recover as much data as possible.

Note Although it is possible to restore a backup image from one Reporting Server to another, such a restoration is not supported with the Unified CVP restore process.

Procedure

Step 1 Stop the CallServer process (Reporting Server).

Step 2 Execute the `%CVP_Home%\bin\cvprestore.bat` script.

Step 3 Restart the CallServer process.

Unified System CLI and Web Services Manager

Unified CVP infrastructure includes the Web Services Manager, a services layer that supports a Diagnostic Portal API.

Unified CVP Infrastructure supports the following features:

- Diagnostic Portal API service support by the Web Services Manager.

- Unified System Command Line Interface (CLI) which is a client tool that supports the diagnostic portal API and other APIs for collecting diagnostic data.
- Licensing.
  - Common Licensing for all CVP components that support FlexLM.
  - 30 ports with 30-day expiration for Call Server and VXML Server evaluation licenses.
  - 10,000 database writes for Reporting Server evaluation licenses.
  - Licenses are only valid if the license feature, **CVP_SOFTWARE**, is added. This feature is used to ensure if you are authorized to run the current version of CVP.
- Serviceability Across Products: Enhanced Log and Trace messages.

The CVP WebServices Manager (WSM) is a new component that is installed automatically on all Unified CVP Servers, including Remote Operations Manager (ROM)-only installations. WSM interacts with various subsystems and infrastructure handlers, consolidates the response, and publishes an XML response. WSM supports secure authentication and data encryption on each of the interfaces.

Analysis Manager versus Unified System CLI

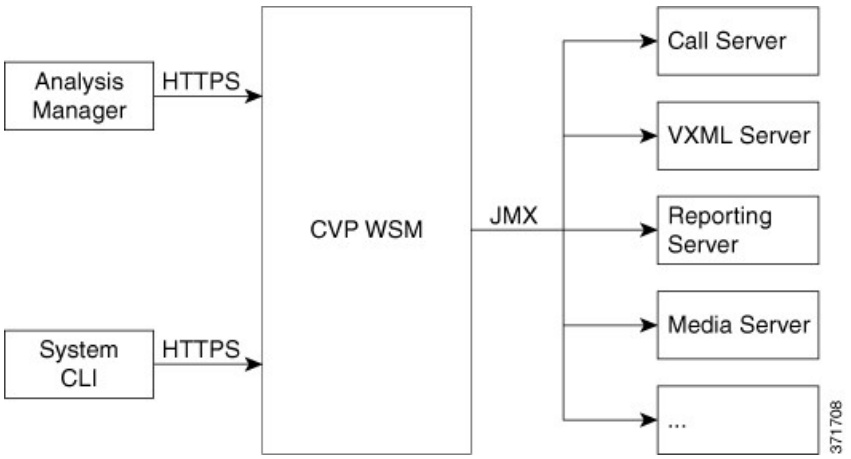
Analysis Manager and Unified System CLI access the Diagnostic Portal API. Both the Analysis Manager and the Unified System CLI have similar features, except for the differences shown in the table.

Table 17: Differences Between Analysis Manager and Unified System CLI

Analysis Manager	Unified System CLI
Is a GUI-based client that is part of the Unified CM Real-Time Monitoring Tool (RTMT). The Analysis Manager has a user-friendly interface due to its GUI-based design.	Is a command line based tool. The Unified System CLI is more flexible because it can be used in a batch file to perform more complex tasks.
Is neither bundled with CVP nor installed by Unified CVP installer.	Is bundled with Unified CVP installer, and is also bundled with the Unified CCE installer.

The following figure shows how the two interfaces interact with the Web Services Management (WSM) to provide information about Unified CVP components.

Figure 14: Typical Use of the Web Services Layer



Analysis Manager

The Web Service Manager supports all diagnostic (health and status) requests from the Analysis Manager. The Analysis Manager is part of UCM RTMT tool. It provides end users an interface for collecting health and status information for all devices in its network topology. If Unified CVP is configured as a part of the solution, you can leverage the WSM through the Analysis Manager to collect diagnostic details, such as server map, version information, licenses, configuration, components, logs, traces, performance factors, platform information for each CVP Device on a component and subcomponent level. You can set or reset debug levels using the Analysis Manager on a component and subcomponent level.

A new user with the **wsmadmin** username is created during installation with the same password as the Operations Console Server administrator user. Use **wsmadmin** to control access to the diagnostic portal services.

Note For details on Analysis Manager and the Analysis Call Path tool, see *Cisco Unified Analysis Manager* at http://www.cisco.com/en/US/docs/voice_ip_comm/cucm/service/8_0_1/rtmt/ch1_overview.html.

Unified System CLI

When an issue arises in the Unified CVP operation, use the System CLI tool to collect data to be reviewed by Cisco engineers. For example, you can use the System CLI if you suspect a call is handled incorrectly. In this case, you can use the **show tech-support** command to collect data and send the data to Cisco support.

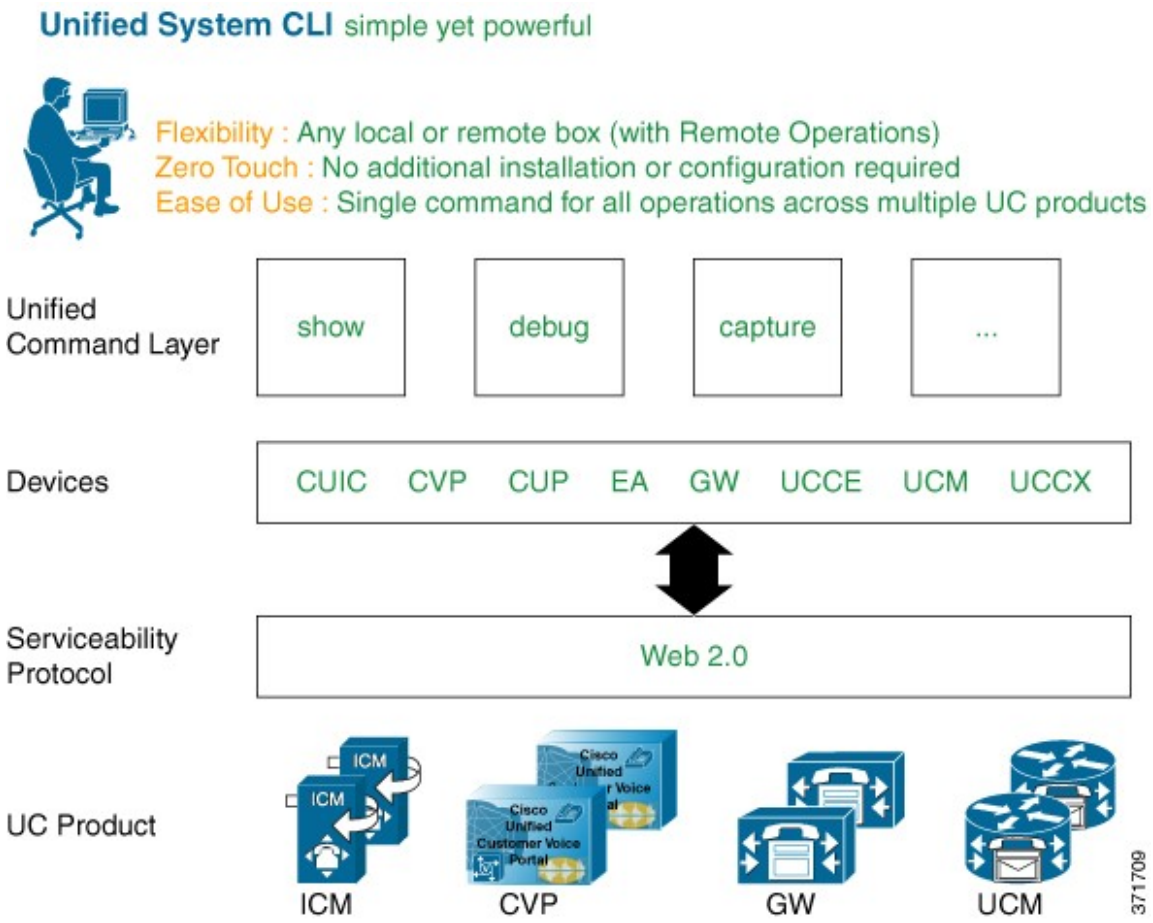
Unified System CLI has the following features:

- Is automatically installed on all Unified CVP Servers as part of the infrastructure. No additional installation is required on any Unified CVP server.
- As every Unified CVP server is aware of at least one seed device (the Operations Console Server), your entire solution topology is automatically retrieved from the Operations Console on any Unified CVP box by using System mode. No additional configuration is needed for the System mode.
- Uses a consistent command across multiple products and servers.

- Can be executed as a Windows scheduled job.

The following figure shows the high-level commands for the Unified System CLI and shows the interaction of devices and Unified Cisco products.

Figure 15: High-Level Commands for Unified System CLI



Unified System CLI Modes of Operation

The Unified System CLI operates as an interactive user interface and can also be used as a batch command. This feature allows the Unified System CLI to be used in scheduled jobs.

The Unified System CLI can operate interactively in two modes:

- **Local mode**
  - In this mode, the Unified System CLI only interacts with a single device. For example, the **show version** command shows only the version for a single device.
- **System mode**
  - In this mode, the Unified System CLI detects the Operations Console, which acts as a seed device for the CLI, and then interacts with all the devices in the device list in the Operations Console to extract the solution topology automatically.

In this mode, the **show version** command shows the version information for all devices in the device list.

- All the commands available in local mode for a single device are available in system mode.
- The command syntax remains the same in system mode.
- There are options to limit the system command option to certain device group, device type, or list of servers.

Unified System CLI FAQ

Q. Does Unified System CLI affect the performance of the devices it queries?

A. Unified System CLI runs at a low priority; it uses idle CPU time on the system. It should not affect call processing even if executed on a system running under load.

The response time from the given CLI command varies depending on the load of the system and the server response time. The response time when there is no running load should be below 5 seconds for each server for operations, such as **show version**, **show license**, **show debug**, and **show perf**. The response time when there is no running load for **show platform** operation should be below 10 seconds for each server.

However, the response time cannot be determined for commands, such as **show trace**, **show log**, **show sessions**, **show all**, and **show tech-support**. The response for these commands can vary depending on the data being transferred by the server.

Q. Can I redirect the output of a Unified System CLI command to a network drive?

A. Yes. Just specify the path to the network drive.

Q. Can I filter and include multiple components and devices?

A. Yes. Use the component and subcomponent options to filter components and subcomponents and use the server option to filter devices. You may use“ | ”symbol to select multiple components or subcomponents or devices. For example:

```
admin:show debug subcomponent cvp:SIP|cvp:ICM|cvp:IVR
Component: CallServer, subcomponent: SIP
Trace level = 0
Description:
Application data:
Component: CallServer, subcomponent: ICM
Trace level = 0
Description:
Application data:
Component: CallServer, subcomponent: IVR
Trace level = 0
Description:
Application data:
admin:
```

Q. Can turning on debug level 3 affect the performance on a production system?

A. Yes, the debug level should be set to 0 for normal production environment. Following is the definition of debug levels for reference:

Level 1 --- Low performance impact

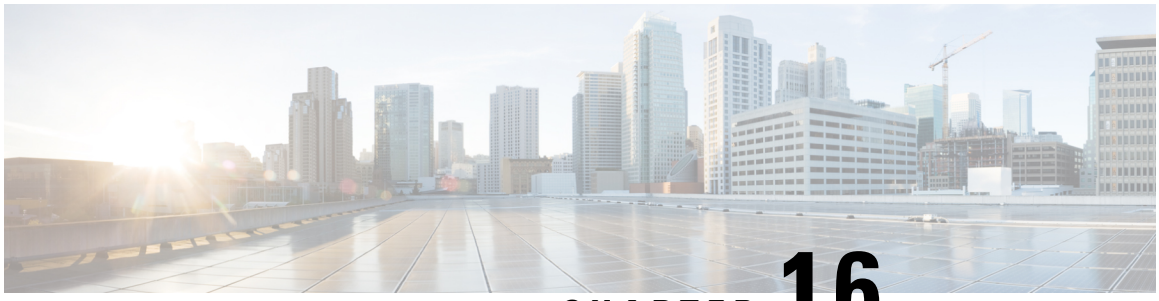
Level 2 --- Medium performance impact

Level 3 --- High performance impact

Q. How do I set the debug level to its default?

A. Set the debug level to 0.

Note For details on Unified System CLI, see *Configuration Guide for Cisco Unified Customer Voice Portal* at http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_installation_and_configuration_guides_list.html.



CHAPTER 16

Deployment Sequence

Licensing

For Unified CVP licensing information refer to the *Cisco Customer Contact Solutions Ordering Guide*. This guide is a frequently updated source for all Unified CVP licensing information. Cisco employees and partners with a valid login account can access the ordering guide at:

http://www.cisco.com/web/partners/downloads/partner/WWChannels/technology/ipc/downloads/CCBU_ordering_guide.pdf.

If you need licensing information for Unified CVP but you cannot access the Ordering Guide, contact your local Cisco Systems Engineer (SE) or Partner.

INDEX

A

C