Unified CVP Architecture Overview

Over the past two decades, many customers have invested in TDM-based interactive voice response (IVR) applications to automate simple customer transactions such as checking account or 401K account inquiries. In addition, many TDM-based IVR platforms were based on proprietary development environments and hardware platforms, which typically meant restricting the customer's integration options with automatic speech recognition (ASR) and text-to-speech (TTS) solutions. Over the past few years there has been a dramatic shift to using VoiceXML (VXML) standards-based technology to support the next generation of IVR applications.

Because the implementation of Unified CVP is based on VXML, the discussion of Unified CVP begins with the following overview of VXML as it relates to Unified CVP.

The chapter covers the following major topics:

- VoiceXML, page 1
- Cisco Unified Customer Voice Portal, page 2
- Unified CVP Product and Solution Components, page 4
- Call Flows, page 14
- Design Process, page 16
- Quality of Service, page 22

VoiceXML

Voice eXtensible Markup Language, or VoiceXML, is a markup language similar to HTML, that is used for developing IVR services and leverages the power of web development and content delivery. VoiceXML was designed for creating audio dialogs that feature synthesized speech, digitized audio, recognition of speech or dual-tone multifrequency (DTMF) key input, and recording of spoken input. It is a common language for content providers, tool providers, and platform providers, and it promotes service portability across implementation platforms.

VoiceXML separates service logic from user interaction and presentation logic in VoiceXML voice web pages. It also shields application authors from low-level, platform-specific IVR and call control details. VoiceXML is easy to use for simple interactions, yet it provides language features to support complex IVR dialogs.
VoiceXML programs are rendered (or executed) by a VoiceXML browser, much like an HTML program is rendered using an internet browser (such as Internet Explorer). A Cisco Voice Gateway (or router) can provide the VoiceXML browser function. For small deployments, the Ingress Voice Gateway and VoiceXML Gateway are typically deployed in the same router. The Cisco IOS VoiceXML Gateway provides both gateway and VoiceXML browser functions.

In the most simple call processing scenario, a new call arrives and the voice gateway dial peer matches the call to an available VoiceXML gateway port. The VoiceXML gateway port represents a Voice over IP (VoIP) endpoint and can be logically thought of as a voice response unit (VRU) port. Upon arrival of the new call, the VoiceXML gateway (that is, the VRU) sends an HTTP request to a Cisco Unified CVP VXML Server for instruction. The URL contained in the HTTP request correlates to a specific VoiceXML doc.

In response to the HTTP request, the Unified CVP VXML Server sends the requested, dynamically generated VoiceXML doc to the VoiceXML gateway (that is, the voice browser) to be rendered. A typical VoiceXML doc is short and prompt the caller for some input, then includes the results in a new HTTP request that redirects the caller to another URL and VoiceXML doc. Because a typical call requires numerous prompts and caller inputs, there are numerous VoiceXML documents that need to be rendered and a large number of possible paths through these VoiceXML documents.

To logically link the many different VoiceXML documents that may need to be rendered and to greatly simplify the task of creating VoiceXML documents, a graphical scripting tool is often used to allow the IVR service developer to easily develop complete IVR services with conditional logic and customer relationship management (CRM) database integration. Cisco Unified Call Studio is one such scripting tool. The Cisco Unified CVP VXML Server is capable of executing scripts developed with Cisco Unified Call Studio, and both were designed to work with Cisco Unified CVP Server, Cisco Voice Gateways, Cisco VoiceXML Gateways, Cisco Unified Communications Manager, Cisco Unified Contact Center, and Cisco's VoIP-enabled LAN/WAN.

Cisco Unified Customer Voice Portal

Unified CVP is both a product and a solution. As a product, its media kit includes specific software items, as listed in the first part of Unified CVP Product and Solution Components, on page 4. As a solution, Unified CVP relies on additional Unified CVP components. The additional components are described in Additional Unified CVP Solution-Related Components, on page 8. The resulting solution provides carrier-class IVR and IP switching services on Voice over IP (VoIP) networks.

Unified CVP includes the following features:

- Carrier-class performance
  Create your solution using a reliable, redundant, and scalable platform, which enables works with service providers and large enterprise networks.

- Call switching and routing support
  Route and transfer calls between voice gateways and IP endpoints. Voice gateways provide natural integration of TDM ACDs and PBXs with the PSTN.
  After completing the routing or transfer of a call, Unified CVP maintains SIP call control to provide switching services similar to take-back-and-transfer (TNT) between IP endpoints via the Cisco Unified Intelligent Contact Management Enterprise (Unified ICME) interface.
  Supports call routing services for SIP (RFC 3261) protocol.

- IP-based IVR services
In addition to switching and transfer, Unified CVP provides classic prompt-and-collect functions, such as Press 1 for Sales.

Provides sophisticated audio and video self-service applications with CRM database integration as well as ASR and TTS integrated via Media Resource Control Protocol (MRCP). Examples include banking and brokerage account handling, and airline reservations.

Park calls for personalized prompts or hold music while waiting for a call center agent to become available. Calls can be prioritized based on their CRM profiles.

Take back a transferred call for further IVR treatment or transfer.

• VoiceXML services

Provides a platform for developing powerful, speech-driven interactive applications accessible from any phone. The VoiceXML platform includes:

- The Cisco Unified CVP VXML Server, a J2EE- and J2SE-compliant application server that dynamically drives the caller experience.
- The Cisco Unified Call Studio, a drag-and-drop graphical user interface (GUI) for the rapid creation of advanced voice applications.

• Unified CVP Operations Console Server

Centrally operate, administer, maintain, and provision the components in the Unified CVP solution from its web-based Operations Console. Integrate with Cisco Contact Center Support Tools. See the Unified CVP Operations Console Server, on page 7 for hosting information.

• VRU reporting

Access historical data using its included centralized reporting database. Design and run custom reports using its well-documented schema.

• Compatibility and integration

- Use with other Cisco Call Routing and VoIP Products, including, Cisco Unified Intelligent Contact Management Hosted or Unified ICME, Cisco Gateways, and Cisco Unified Contact Center Enterprise (Unified CCE).
- Use with Cisco Unified Communications Manager (Unified CM). Unified CM manages and switches VoIP calls among IP phones. When combined with Unified ICME, Unified CM becomes the Unified CCE product.
- Use with the Public Switch Telephone Network (PSTN). Calls can be moved onto an IP-based network for Unified CVP treatment and then moved back out to a PSTN for further call routing to a call center.
- Integration with Cisco Unified Contact Center

Unified CVP integrates with Cisco Unified Contact Center via a VRU Peripheral Gateway (PG). This integration enables Unified CCE to control Unified CVP VoIP switching and IVR services. It also enables Unified CCE to control the agent selection application and to initiate the Real-Time Transport Protocol (RTP) stream transfer from the VoiceXML gateway to the selected agent.

Unified CVP integration with Unified CCE requires that the traditional Unified CM PG be used for Unified CCE integration with Cisco Unified CM.

Unified CCE can be integrated with Unified CVP via the Cisco Unified Intelligent Contact Manager (ICM) System PG and the parent-child deployment model. This integration method provides callers...
with some simple menus and prompts by the parent Unified ICM and Unified CVP, and it intelligently routes the calls using skill groups to the best Cisco Unified Contact Center Express or Enterprise child. Queuing control and agent selection are handled by the child contact center solution. In this model, it is also easy for a TDM automatic call distributor (ACD) to play the role of a child. All call transfers between Unified CVP and children retains call data, and the ICM provides enterprise-wide browser-based consolidated reporting.

Unified CVP integration is not directly supported with the Unified CCE System PG (which is also used by System Unified CCE). The Unified CCE System PG supports only the Cisco Unified IP IVR. Unified CVP works only with System PG children via the parent-child deployment model. Unified CVP can also provide IVR services for Unified CCE outbound IVR campaigns and post-call customer surveys.

Unified CVP Product and Solution Components

As mentioned previously, Unified CVP is both a product and a solution. The following topics describe both the components that make up the Unified CVP product, and the additional components that make up the Unified CVP solution.

The Cisco Unified Customer Voice Portal (CVP) product consists of the following components:

- Unified CVP Call Server, on page 5
- VXML Server, on page 6
- Cisco Unified Call Studio, on page 6
- Unified CVP Reporting Server, on page 7
- Unified CVP Operations Console Server, on page 7
- Additional Unified CVP Solution-Related Components, on page 8

The following components of the Unified CVP solution are not part of the Unified CVP product but are still required for a complete solution:

- Cisco Ingress Voice Gateway, on page 8
- Cisco VoiceXML Gateway, on page 8
- Cisco Egress Gateway, on page 9
- Video Endpoints, on page 9
- Cisco Unified Communications Manager, on page 10
- Cisco Unified Contact Center, on page 10
- SIP Proxy Server, on page 11
- DNS Server, on page 11
- Media Server, on page 6
- Application Control Engine, on page 12
- Third-Party Automatic Speech Recognition and Text-to-Speech Servers, on page 13
The following sections discuss each of these components in more detail. Depending on the particular deployment model you choose, some of the above components might not be required.

### Unified CVP Product Components

The following topics describe the Cisco Unified Customer Voice Portal (CVP) product components.

#### Note

Call Server, VXML Server, and Media Server are combined as one component known as CVP Server. Installing CVP Server installs all three components. In the earlier versions, Call Server, VXML Server, and Media Server could be installed on different machines. The option to install Call Server, VXML Server, and Media Server separately is not available.

### Unified CVP Call Server

The Call Server component provides the following independent services, which all run on the same Windows 2008 R2 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 new calls to an available VoiceXML 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 via 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 initiate subsequent call control to do things such as requesting that a caller be transferred from queue to an agent or transferred from one agent to another agent.

- **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 VoiceXML pages that implement the Unified CVP Microapplications based on Run VRU Script instructions received from Unified ICM. The IVR Service functions as the VRU leg (in Unified ICM Enterprise parlance), and calls must be transferred to it from the SIP Service in order to execute Microapplications. The VoiceXML pages created by this module are sent to the VoiceXML gateway to be executed.

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

**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 and many customers add their own custom-built or off-the-shelf 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 Microapplication which must be executed from within the Unified ICME routing script.

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


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

**Media Server**

The Media Server component is a simple Web Server, like Microsoft IIS, which can provide prerecorded audio files, external VoiceXML documents, or external ASR grammars to the gateway. Some of these files can be stored in local flash memory 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 Server used is Microsoft IIS, which is 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.

As with ASR/TTS Servers, Media Servers may be deployed simplex, as a redundant pair, or with ACE in a farm. Note that the VoiceXML Gateway caches .wav files it retrieves from the Media Server. In most deployments, the Media Server encounters extremely low traffic from Unified CVP.


**Cisco Unified Call Studio**

The 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, and it provides advanced drag-and-drop
graphical editing as well as the ability to insert vendor-supplied and custom-developed plug-ins that enable applications to interact with other services in the network. The Call Studio is essentially an offline tool and 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, customers typically designate one or more data center servers for that purpose. The Cisco Unified Call Studio cannot run on machines also running a headless version of the Cisco Security Agent.


Note
Cisco Security Agent is not supported on Unified Call Studio.

Unified CVP Reporting Server

The Unified CVP Reporting Server is a Windows 2008 R2 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 prescribed 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 Unified CVP VXML Server (VXML Server). The Reporting Server depends on the Unified CVP Call Server (Call Server) to receive call records.

For Standalone Unified CVP VXML Server deployments, one Call Server is needed per Reporting Server. The Reporting Server must be local to the Call Server(s) and VXML Server(s) that it is servicing. Deploying the Reporting Server at a remote location across the WAN is not supported. Multiple Reporting Servers should be used and placed at each site when Call Server(s) and VXML Server(s) exist at multiple locations.

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

Unified CVP Operations Console Server

The Unified CVP Operations Console Server is a Windows 2008 R2 server that provides an Operations Console for the browser-based administration and configuration for all Unified CVP product components, and it offers shortcuts into 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 physical machine 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 itself be configured with a map of the deployed solution network. It can then collect and maintain configuration information from each deployed component. Both the network map and the configuration information are stored locally on the server, where it can be backed up by off-the-shelf backup tools. A web-based user interface, the Operations Console, provides the ability to both display and modify the network map and the 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 are deployed to the device when a Save and Deploy option is executed.

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) comes online, the user can apply the configured settings to that server. These settings will become active when that server's services also come online. Only then will they be 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 Unified CVP Solution-Related Components

The following additional components are used in the various call flow models (solutions) described in Call Flows, on page 14.

Cisco Ingress Voice Gateway

The Cisco Ingress Voice Gateway is the point at which an incoming call enters the Unified CVP system. It terminates 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. Therefore, WAN bandwidth is conserved because no hairpinning of the media stream occurs. It also provides for sophisticated call switching capabilities at the command of other Unified CVP solution components.

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

For the most current list of supported gateways, see the Gateway Choices. For approved gateway/software combinations 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:


The Ingress Gateway can be deployed separately from the VoiceXML Gateway, but in most implementations they are the same: one gateway performs both functions. Gateways are often deployed in farms, for Centralized deployment models. In Branch deployment models, one combined gateway is usually located at each branch office.

Cisco VoiceXML Gateway

The VoiceXML Gateway hosts the Cisco IOS Voice Browser. This component interprets VoiceXML pages from either the Unified CVP Server IVR Service or the Unified CVP VXML Server. The VoiceXML Gateway encodes .wav files and accepts DTMF input. It then returns the results to the controlling application and waits for further instructions.

The Cisco VoiceXML Gateway can be deployed on the same router as the Unified CVP Ingress Voice Gateway. This model is typically desirable in deployments with small branch offices. But the VoiceXML Gateway can also run on a separate router platform, and this model is typically desirable in deployments with large or multiple voice gateways, where only a small percentage of the traffic is for Unified CVP. This model enables
an organization to share PSTN trunks between normal office users and contact center agents and to route calls based upon the dialed number.

The Cisco VoiceXML Gateway can encode .wav files stored in flash memory or on a third-party media server. Prompts retrieved from a third-party media server can be cached in the router to reduce WAN bandwidth and prevent poor voice quality. The VoiceXML doc provides a pointer to the location of the .wav file to be played or it provides the address of a TTS server to generate a .wav file. The VoiceXML Gateway interacts with ASR and TTS servers via MRCP.

For the supported VoiceXML Gateways, see the Hardware and System Software Specification for Cisco Unified Customer Voice Portal (formerly called the Bill of Materials), available at:


Unless it is combined with the Ingress Gateway (described in the previous topic), the VoiceXML Gateway does not require any TDM hardware. All its interfaces are 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 Gateways, VoiceXML Gateways are often deployed in farms for Centralized deployment models, or one each office in Branch deployments.

**Cisco Egress Gateway**

The Egress Voice Gateway is used only when calls need to be extended to TDM networks or equipment such as the PSTN or a TDM ACD. While the RTP stream goes between the ingress and egress voice gateway ports, the signaling stream logically goes through the Unified CVP Server and ICM in order to allow subsequent call control (such as transfers).

**Cisco Integrated Services Router Generation 3 Gateway**


**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**

When using the Unified CVP Basic Video Service, the following video endpoints are supported:

- Cisco Unified Video Advantage
- Cisco TelePresence
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 Unified ICME to form Cisco Unified Contact Center Enterprise (Unified CCE). Unified CVP interacts with Unified CM primarily as a means for sending PSTN-originated calls to Unified CCE agents. SIP gateway calls are routed to the available Unified CM SIP.

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

- A typical 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 Unified Outbound 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 protocol. PSTN calls that arrived on MGCP voice gateways registered with Unified CM can only be routed or transferred to Unified CVP via SIP (and not going through CUBE).

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. Pure TDM-based call centers using ACDs, for example, typically do not use Unified CM (except when migrating to Cisco Unified CCE), nor do strictly self-service applications using the Unified CVP Standalone self-service deployment model. The Unified CM generally is used as part of the Cisco Unified CCE solution, in which call center agents are part of an IP solution using Cisco IP Phones, or when migrating from TDM ACDs.

Only specific versions of Unified CM are compatible with Unified CVP solutions. Unified CVP is supported with SIP only if Cisco Unified CM 5.0 or later release is used. For full details on version compatibility, 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:


Cisco Unified Contact Center

Either Cisco Unified CCE or Unified ICM is a required component when advanced call control (IP switching, transfers to agents, and so forth) is required in Unified CVP. The Hosted versions of these products might also be used for this purpose. The 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. From the point of view of Unified CCE or Unified ICM, the Call Server is a voice response unit (VRU) connected to Unified CCE or Unified ICM, just as all other GED-125 VRUs are connected. Unified CVP is simply a VRU peripheral to Unified CCE or Unified ICM.
**SIP Proxy Server**

The SIP Proxy Server is the component that routes individual SIP messages among SIP endpoints. It plays a key role in Unified CVP high-availability architecture for call switching. It 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.

The SIP Proxy can be configured with multiple static routes in order to do load balancing and failover with outbound calls. The static routes can point to an IP address or a regular DNS.

DNS SRV is also supported, but is not qualified for use on the CUSP Server. It is qualified for the devices that need to reach the CUSP Server, such as Unified CVP, the Ingress Gateway, and Unified CM.

Unified CVP can also be deployed without a SIP Proxy Server depending on the design and complexity of the solution. In such cases, some of the same functions can be provided by the Unified CVP Server SIP service.

The benefits of using a SIP Proxy Server include:

- Priority and weight routing can be used with the routes for load balancing and failover.
- If a SIP Proxy Server is already used in your SIP network, Unified CVP can be an additional SIP endpoint—it fits incrementally into the existing SIP network.

If a SIP Proxy Server is not used, then Ingress Gateways and Unified CMs need to point directly to Unified CVP. In such a deployment:

- Load balancing is done using DNS SRV lookups from Gateway to DNS Server—SIP calls can be balanced using this mechanism.
- Load balancing of calls outbound from Unified CVP (outbound call leg) can be done using DNS SRV lookups.
- Failover of SIP rejections (code 503 only) can also be performed if SRV records are configured with ordered priorities.

**DNS Server**

This optional component may be installed anywhere in the network. Its purpose in general is to resolve hostnames to IP addresses. The Unified CVP, can make both Type A record lookups and SRV Type record lookups. If the DNS server is slow to respond, is unavailable or is across the WAN, so it affects the performance.

The DNS Server comes into play during SIP interactions in the following situations:

- When a call arrives at an Ingress 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 Gateway can use DNS directly to distribute inbound calls among multiple SIP Services.
- When the SIP Service is instructed by Unified CCE to transfer the call to the VRU leg, it 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 in order 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 target agent's IP address.

The use of the DNS Server for SIP routing is entirely optional in Unified CVP. It is not required to have a dedicated DNS Server, but the existing DNS server needs to handle the additional load of Unified CVP. For every call destined for Unified CVP that comes into the network, there will be approximately 3 to 4 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 could also be configured locally in the system "etc host" file. Unified CVP Server Groups can alternately be used to avoid DNS SRV lookups.

**Content Services Switch**

Support for CSS is discontinued.

**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.


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

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


Third-Party Load Balancers

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 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:


Third-Party Automatic Speech Recognition and Text-to-Speech Servers

This component provides speech recognition services and text-to-speech services for the VoiceXML Gateway. Communication between the ASR/TTS servers and the VoiceXML Gateway uses Media Resource Control Protocol (MRCP). MRCP v1 can be used on the VoiceXML Gateway when the application is based on either Micro-Apps or the Unified CVP VXML Server (VXML Server). MRCP v2 can be used on the VoiceXML Gateway only with applications that are based on the Unified CVP VXML Server.

For capacity and redundancy reasons, a ACE is usually used to mediate between a VoiceXML Gateway and a farm of ASR/TTS servers. If ACE is not used, then a VoiceXML Gateway can support a maximum of two ASR/TTS Servers.
Cisco does not sell or support any ASR/TTS software or servers. Cisco does, however, 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. The simplest to implement and support is inline grammars, by which the set of acceptable customer responses is passed to the gateway. Another form is external grammars, where Unified ICM passes a pointer to an external grammar source. The Unified CVP 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 is responsible for creating the grammar file. A third type of grammar is the built-in grammar. For a complete explanation of grammar formats, consult the W3C website at:

http://www.w3.org/TR/speech-grammar/

The text for TTS is passed directly from the Unified CVP VXML Server to the 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 interfaces directly to the VoiceXML gateway via Media Resource Control Protocol (MRCP). Currently, Nuance is the only tested ASR/TTS engine. The ASR/TTS engine also supports (with limitations) voice recognition and synthesis for multiple languages.

For the latest 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. That does not, however, mean that the vendor's latest software version can be used. Unified CVP is carefully tested with specific versions of each vendor's product, and Cisco Technical Assistance Center (TAC) will not support Unified CVP customers who use different ASR/TTS versions than those which have been tested. For additional details on supported ASR and TTS products, See the Hardware and System Software Specification for Cisco Unified Customer Voice Portal (formerly called the Bill of Materials), available at:


Network Monitor

An SNMP management station can be used to monitor the solution deployment status.

Call Flows

This section describes the Unified CVP call flows for SIP.

Typical SIP Unified CVP Call Flow

This section presents a typical call flow scenario using SIP. The description roughly follows the Comprehensive call flow model. However, it is not presented as an actual solution, only 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 queue treatment, and finally delivery to a Unified ICME agent. The following diagram presents a general SIP-based solution. A detailed description of the call flow follows the diagram.
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 which 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 then forwards the invite to the selected Unified CVP Server SIP Service.

3. The SIP Service consults Unified ICME via the Unified CVP Server ICM Service, which causes Unified ICME to run a routing script.

4. The routing script typically initiates a transfer of the call to a VoiceXML Gateway port via the SIP service.

5. The VoiceXML Gateway sends a message to the IVR service, which requests scripted instructions from Unified ICME.

6. Unified ICME exchanges VRU instructions with the VoiceXML gateway via the IVR service. The instructions can include requests to invoke more sophisticated applications on the Unified CVP VXML server. Such requests result in multiple exchanges between the Unified CVP VXML Server and the VoiceXML Gateway to provide self-service.

7. If the customer wants to transfer to a live agent, the Unified ICME routing script queues the caller for an available agent. While waiting for an available agent, Unified ICME provides additional instructions to the VoiceXML Gateway to provide queueing treatment to the caller.

8. When an agent becomes available, Unified ICME sends a message to the Unified CVP Server SIP Service, which forwards a message via the SIP Proxy Server to the Ingress Gateway and to Unified CM to transfer the call away from the VoiceXML Gateway port and deliver it to the Unified CM agent IP phone.

During the VRU exchanges, the VoiceXML Gateway interacts with an ASR/TTS Server to play text as speech or recognize speech as data. It also interacts and with a Media Server (not shown in the diagram, but connected to the VoiceXML Gateway) to fetch audio files and prompts. These two devices, as well as the Unified CVP VXML Server, can be located behind a ACE, which offers sophisticated failover and redundancy capability.

During this entire process, the SIP Service, the IVR Service, and the VXML Server send a stream of reporting events to the Reporting Server (also not shown in the diagram, but connected to the Unified CVP Call Server), which processes and stores the information in a database for later reporting. All these devices use SNMP (Simple Network Management Protocol) to support a monitoring console. Cisco Unified Operations Manager can also be configured to process and forward SNMP events to higher-level monitoring stations such as HP OpenView.
All components in the solution can be managed by the Unified CVP Operations Console Server. The Operations Console is not shown in the diagram, but is connected to all the components that it manages. The Operations Console uses a variety of means to pull together the configuration, management, and monitoring of the entire solution into a single station, which can be accessed via a standard web browser.

The Unified CVP VXML Server applications are designed and built using Call Studio is used offline tool and not shown in the diagram.

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. Support for H.323 is discontinued.

Call Flow Models

The first step in the design process is to determine what 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 ICME with VoIP call switching.
* Prompt and collect data using third-party VRUs and ACDs.
* Avoiding using an Unified CVP VXML Server.

• Comprehensive — Provides IVR services, queue treatment, and IP switching services. The previously described typical call flows use this functional deployment model.

This model is useful if you want to:

* Use Unified CVP to provide Unified ICME with VoIP call switching capabilities.
* Use Unified CVP to provide Unified ICME with VRU services, including integrated self-service applications, queuing, and/or 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/TTS services.

• VRU Only — Provides IVR services, queuing treatment, and switching for SS7/IN PSTN endpoints. This model relies upon the PSTN to transfer calls between call termination endpoints.

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

For more details and design considerations for each of these functional deployment models, see the chapter on Functional Deployment Models.

Unified CVP Algorithm for Routing

When you are configuring 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 your desired effect.

The following priority and conditionals are used to formulate the destination SIP URI for the outbound calls made by Unified CVP. This description covers CONNECT messages which include labels from the ICM (for example, VXML GW, and CUCM), as well as calls to the ringtone service, recording servers, and error message playback service.

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

The algorithm for creating the destination SIP URI host portion for outbound calls, given the ICM label is as follows.

1. At the start of the algorithm, the ICM label is provided. 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 + correlation ID as the label.

2. If SendtoOriginator is matched for the Unified CCE label, the IP or hostname of the caller (ingress gateway) is used by the Unified CVP algorithm, which returns the SIP URI.

   The setting for SendtoOriginator only applies to callers who are on IOS gateways (the SIP UserAgent header is checked), because non-IOS gateways do not have the CVP "bootstrap" service used by the Cisco VXML gateway.

3. If use outbound proxy is set, then use the host of the proxy - return SIP URI.

4. If local static route is found for the label - return the SIP URI.
### Distributed Network Options

After choosing a functional deployment model, Unified CVP solution designers 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 VoiceXML gateway, or separate gateways. Typically, the gateways are combined when deployed in a branch.

- **Branch Ingress Gateways with Centralized VoiceXML Gateways** — Enables integration of locally dialed numbers and resource grouping of VoiceXML gateways. This option might be desirable for organizations with many medium to large branches, but with few contact center calls in each branch. 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.

It is also possible to 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 Deployments**.

### 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.
ASR 1000 platform is supported for CUBE with CVP Solution, according to the *Hardware and System Software Specification for Cisco Unified Customer Voice Portal*. For ASR limitations, see the *Design Guide for Cisco Unified Customer Voice Portal*. CUBE on ISR gateways is supported. Also, survivability service is supported on the CUBE gateway.

**Note**

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.

**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 1: 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.
Call supplementary services will not work for the already established calls once the Unified CM SME is down.

Note

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.


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.

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. For more details and design considerations for each of these high-availability network options, see the chapter on Unified CVP Design for High Availability.
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.

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 boxes needed to run a Unified CVP deployment, to facilitate the deployment's administration, and to leverage your ESX (tm) 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

Quality of Service

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 path. The Unified CVP provides a management interface via the Unified CVP Operations Console Server to modify QoS settings at each end of specifically designated data paths.
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.