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

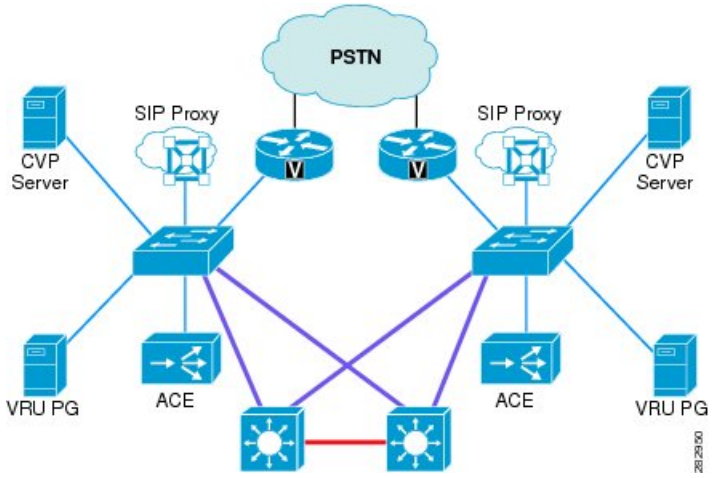
In the other chapters of this document, the Unified CVP Call Server is described as a single component because it does not need to be described in depth. When discussing Unified CVP high availability however, it is important to understand that there are actually several parts to this component:

- 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 1: 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 4
- Call Disposition, on page 13

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 11.

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 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. For a discussion of the benefits and drawbacks of each codec, refer to Voice Traffic.

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 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.