Design Considerations for High Availability

- High Availability Designs, page 1
- High Availability and Virtualization, page 3
- Data Network Design Considerations, page 3
- Unified Communications Manager Design Considerations, page 5
- Unified CVP Design Considerations, page 8
- Unified IP IVR Design Considerations, page 9
- Cisco Web and E-Mail Interaction Manager Design Considerations, page 11
- Cisco Outbound Option Design Considerations, page 12
- Agent Peripheral Gateway Design Considerations, page 14
- Central Controller Design Considerations, page 19
- Common Processes That Support Failovers, page 20
- Unified CCE Failovers During Network Failures, page 21
- Unified CCE Failovers During Single-Component Failures, page 25
- Unified CCE Failovers During Multicomponent Failures, page 35
- Other Considerations for High Availability, page 37

High Availability Designs

Cisco Unified Contact Center Enterprise (Unified CCE) products incorporate high availability features in all standard deployments. Every production deployment must include redundancy for the core Unified CCE components. The redundant components are designed to fail over automatically and recover without manual intervention. Your design can include more than that basic high availability capability. A successful Unified CCE deployment requires a team with experience in data and voice internetworking, system administration, and Unified CCE application design and configuration.

Each change to promote high availability comes at a cost. That cost can include more hardware, more software components, and more network bandwidth. Balance that cost against what you gain from the change. How critical is preventing disconnects during a failover scenario? Is it acceptable for customers to spend a few
extra minutes on hold while part of the system recovers? Would the customer accept losing context for some calls during a failure? Can you invest in greater fault tolerance during the initial design to position the contact center for future scalability?

This guide focuses on design of the contact center itself. Your contact center operates in a framework of other systems. This guide cannot provide complete information about every system that supports your contact center. The guide concentrates on the Cisco Unified CCE products. When the guide must discuss another system, it does not offer a comprehensive view. For more information about the complete Cisco Unified Communications product suite, see the Cisco solutions design documents at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

The following figure shows a fault-tolerant Unified CCE single-site deployment.

*Figure 1: Unified CCE Component Redundancy*

This design shows how each Unified CCE component is duplicated for redundancy. All Unified CCE deployments use redundant Unified Communications Manager, Unified CCE, and Cisco Unified Customer Voice Portal (Unified CVP) or Unified IP IVR components. Because of the redundancy, a Unified CCE deployment can lose half of its core systems and be still operational. In that state, a Unified CCE deployment can handle calls by rerouting them through Unified CVP to either a VRU session or an agent that is connected to the still-operational components. Where possible, deploy Unified CCE so that no devices, call processing, or CTI Manager services are running on the Cisco Unified Communications Manager publisher.

To enable automatic failover and recovery, pairs of redundant components interconnect over private network paths. The components use TCP keepalive messages at 100-ms intervals for failure detection. The Unified Communications Manager uses a cluster design for failover and recovery. Each cluster contains a Unified Communications Manager publisher and multiple Unified Communications Manager subscribers. Agent
phones and computers register with a primary target but automatically reregister with a backup target if the primary fails.

High Availability and Virtualization

In a virtualized deployment, place components carefully to maintain high availability. The mechanisms that support high availability are the same. But, you must distribute components to minimize multiple failovers from a single failure. When you deploy on Direct Attached Storage (DAS) only systems, consider the following points:

- Failure of a VM brings down all the components that are installed on the VM.
- Failure of a physical server brings down all the VMs that are installed on that VMware vSphere Host.

Deployments on systems with shared storage can use some of the VMware High Availability features for greater resiliency. For specific information about supported VMware features, see the Unified Communications in a Virtualized Environment at http://docwiki.cisco.com/wiki/Unified_Communications_in_a_Virtualized_Environment.

To minimize the impact of hardware failures, follow these guidelines:

- Avoid placing a primary VM and a backup VM on the same physical server, chassis, or site.
- Avoid placing all the active components in a failover group on the same physical server, chassis, or site.
- Avoid placing all VMs with the same role on the same physical server, chassis, or site.

Also consider which components can be coresident and which components must be coresident on the same VMs. For more information about placement of components in virtual environments, see the Virtualization for Unified CCE DocWiki at http://docwiki.cisco.com/wiki/Virtualization_for_Unified_CCE.

Data Network Design Considerations

Highly available contact center designs start with the network infrastructure for data, multimedia, and voice traffic. A single point of failure in your network infrastructure devalues any other high availability features that you design into the contact center. Begin from the PSTN and ensure that incoming calls have multiple paths for reaching Unified CVP for initial treatment and queuing.

Ideally, design with at least two SIP trunks each connecting to a separate Cisco Unified Border Element (Cisco UBE). If any Cisco UBE or SIP trunk fails, the PSTN can route all traffic through the remaining SIP trunks. The PSTN route either by configuring all the SIP trunks as a large trunk group or by configuring rerouting or overflow routing to the other SIP trunks. You can also connect a redundant Cisco UBE to each SIP trunk to preserve capacity when a Cisco UBE fails and the SIP trunk is still functional.

In some areas, the PSTN does not provide multiple SIP trunks to a single site. In that case, you can connect the SIP trunk to a Cisco Unified SIP Proxy. Then, you could connect multiple Cisco UBEs to the Unified SIP Proxy to provide some redundancy.

The Cisco UBE passes calls to Unified CVP for initial treatment and queuing. Register each Cisco UBE with a separate Unified CVP for load balancing. For further fault tolerance, you can register each Cisco UBE with a different Unified CVP as a backup. If a Cisco UBE cannot connect with a Unified CVP, you can also use TCL scripts to provide some call processing. A TCL script might reroute the calls to another site or dialed number or play a locally stored .wav file to the caller and end the call.
In systems that use Cisco Unified IP IVR instead of Unified CVP, the call flows are different. But your design must still support redundant paths from the call ingress point to the queuing and treatment process.


**Figure 2: High Availability Ingress Points**

Voice gateways using the Cisco Unified Survivable Remote Site Telephony (SRST) option for Unified Communications Manager follow a similar failover process. If the gateway is cut off from its controlling subscriber, the gateway fails over into SRST mode. The failover drops all voice calls and resets the gateway into SRST mode. Phones rehome to the local SRST gateway for local call control.

While running in SRST mode, Unified CCE operates as if the agents have no CTI connection from their desktops. The Unified CCE routing application detects the agents as not ready and sends no calls to these agents. When the gateway and subscriber reestablish their connection, the subscriber takes control of the gateway and phones again, allowing the agents to reconnect.

**Public and Private Network Connections**

Unified CCE components use a public network and a private network to communicate. These networks must be separate physical networks. For high availability, include redundant connections in your public network. Ideally, each connection uses a different carrier.

If QoS and bandwidth are configured correctly, your design can merge a public or private WAN link with other corporate traffic. If you use a link that merges non-contact-center traffic, keep the public and private traffic on different networks. However, never split private network traffic onto low-priority and high-priority data paths. The same link must carry all private network traffic for a given component. Sending low-priority and high-priority traffic on different links disables the component failover behavior. Similarly, all low- and
high-priority traffic from each peripheral gateway to the low- and high-priority addresses of the call router must take the same path.

During a public network failure, you can temporarily fail over the public Unified Communications Manager traffic to the private network. Size the private network to accommodate the extra traffic. When the public traffic fails over to the private network, restore the public network as quickly as possible to return to normal operation. If the private network also fails, Unified CCE instability and data loss can occur, including the corruption of one Logger database.

A SONET fiber ring is a highly resilient network with built-in redundancy. You can send the public and private traffic over the same SONET ring under normal operations or following a network failover. A separate link for the private traffic is not required in this case. Also, two routers are required on each side of the WAN for redundancy. Under normal operations, use one router for the Unified CCE public traffic and use the other router for the Unified CCE private traffic. The other rules described in this section also apply.

Figure 3: Network Architecture with SONET Ring

Unified Communications Manager Design Considerations

After you design the data network, design the Cisco Unified Communications infrastructure. Before you can deploy any telephony applications, you need the Unified Communications Manager cluster and CTI Manager in place to dial and receive calls.

For details on the architecture of all these services, see the Cisco Collaboration System Solution Reference Network Designs at http://www.cisco.com/en/US/docs/voice_ip_comm/uc_system/design/guides/UCgoList.html.

High availability design for a cluster requires that you understand how the Unified Communications Manager, CTI Manager, and CallManager services interact. Unified Communications Manager uses the CTI Manager service to handle all its CTI resources. CTI Manager acts as an application broker that abstracts the physical binding of applications to a particular Unified Communications Manager server. The CallManager service registers and monitors all the Cisco Unified Communications devices.

The CTI Manager accepts messages from the Agent PG, a CTI application, and sends them to the appropriate resource in the cluster. The CTI Manager acts like a JTAI messaging router using the Cisco JTAI link to communicate with Agent PGs. The JTAI client library in Cisco Unified Communications Manager connects to the CTI Manager instead of connecting directly to the CallManager service.

CallManager service acts as a switch for all the Cisco Unified Communications resources and devices in the system. The CallManagers on each Unified Communications Manager server link themselves across the public network with the Signal Distribution Layer (SDL). This link keeps the cluster in sync. Each CTI Manager connects with the Unified Communications Manager and CallManager services on its server. CTI Managers do not connect directly with other CTI Managers in the cluster.

Agent PGs use a CTI-enabled user account in Unified Communications Manager, typically called the JTAI user or PG user. The Agent PGs sign in to the CTI Manager to connect to the devices that are associated to that user. If the local CallManager services the appropriate device, the CTI Manager handles the request for that device. If the device is not resident on its local subscriber, then the CallManager service forwards the request to the appropriate subscriber through the private link to the other CallManager services.
The following figure shows the connections in a cluster.

**Figure 4: Connections in Unified Communications Manager Cluster**

For high availability, distribute device registrations across all the subscribers in the cluster. If you concentrate the registrations on a single subscriber, the traffic puts a high load on that subscriber. The memory objects that the Agent PGs use to monitor registered devices also add to the device weights on the subscribers.

If the PG that is connected to that subscriber fails, the redundant PG that takes over sends all the requests to another subscriber. Then, the local CallManager service must route the CTI Manager messaging for those requests across the cluster to the original subscriber. The additional messaging in this failover condition creates greater load on the cluster.

**Unified Communications Manager Redundancy**

Some Unified Communications Manager deployments use a 2:1 redundancy scheme. Each pair of primary subscribers shares a single backup subscriber. But, because of the higher phone usage in contact centers and to simplify upgrade processes, Unified CCE uses a 1:1 redundancy scheme for subscribers. Each primary subscriber requires its own backup subscriber.


The following figure shows different size clusters. Option 1 supports up to 150 Unified CCE agents. Options 2 to 5 illustrate increasingly larger clusters. In this figure, the value of $N$ depends on the components that your contact center uses:

- For Unified CCE deployments with Unified Communications Manager and Unified CVP, $N$ is equal to 2000.
• For deployments with Unified IP IVR, N is equal to 500.

**Figure 5: Redundancy Configuration Options**

---

**Unified Communications Manager Load Balancing**

The 1:1 redundancy scheme for Unified Communications Manager subscribers lets you balance the devices over the primary and backup subscriber pairs. Normally, a backup subscriber has no devices registered unless its primary subscriber is unavailable.

You can enable load balancing through Unified Communications Manager redundancy groups and device pool settings. You can move up to half of the device load from the primary to the secondary subscriber. In this way, you can reduce by half the impact of any server becoming unavailable. To minimize the effect of any outage, distribute all devices and call volumes equally across all active subscribers.
Unified CVP Design Considerations

The Contact Center Enterprise Reference Designs use Unified CVP for call treatment and queuing. Unified CVP uses SIP for call control, rather than relying on Unified Communications Manager for JTAPI call control.

Unified CVP can use the following system components:

- Cisco Unified Border Element (CUBE) supports the transition to SIP trunking. CUBE provides interworking, demarcation, and security services between the PSTN and your contact center.

- Cisco Voice Gateway (VG) terminates TDM PSTN trunks to transform them into IP-based calls on an IP network. CVP uses specific Cisco IOS Voice Gateways that support SIP to enable more flexible call control. VGs controlled by Unified CVP can also use the Cisco IOS built-in Voice Extensible Markup Language (VoiceXML) Browser to provide caller treatment and call queuing. CVP can also leverage the gateway's Media Resource Control Protocol (MRCP) interface to add automatic speech recognition (ASR) and text-to-speech (TTS) functions.

- Unified CVP Server provides call control signaling when calls are switched between the ingress gateway and another endpoint gateway or a Unified CCE agent. The CVP Server also provides the interface to the Unified CCE VRU Peripheral Gateway (PG). The CVP Server translates specific Unified CCE VRU commands into VoiceXML code for rendering on the Unified CVP VG. The CVP Server can communicate with the gateways using SIP as part of the solution.

- Unified CVP Media Server acts as a web server that provides predefined audio files to the voice browsers as part of their VoiceXML processing. You can cluster media servers using the Cisco Content Services Switch (CSS) products. With clustering, you can pool multiple media servers behind a single URL for access by all the voice browsers.

- Unified CVP Server hosts the Unified CVP VoiceXML runtime environment. The VoiceXML service creation environment uses an Eclipse toolkit browser in the Unified CVP Call Studio Application. The runtime environment executes the dynamic VoiceXML applications and processes Java and Web Services calls for external systems and database access.
• Cisco Unified SIP Proxy servers in a Unified CVP deployment select voice browsers and associate them with specific dialed numbers. When a call comes into the network, the VG queries the Unified SIP Proxy to determine where to send the call based on the dialed number.

**Important**

Unified CCE deployments do not support Unified CVP's Enhanced Location Call Admission Control feature.

These methods can increase the high availability of Unified CVP:

• To provide automatic call balancing across the Unified CVP Servers, add redundant Unified CVP Servers under control of the Unified CCE PGs.

• To handle conditions where the gateway cannot contact the Unified CVP Server, add survivability TCL scripts to the gateway. For example, you can redirect calls to another Unified CVP Server on another Unified CVP-controlled gateway.

• To load balance audio file requests across multiple Unified CVP Media Servers and VoiceXML URL access across multiple servers, add a Cisco Content Server.


### Unified IP IVR Design Considerations

Cisco Unified IP IVR can establish JTAPI connections with two CTI Managers on different subscribers in the Unified Communications Manager cluster. This feature enables Unified IP IVR redundancy at the CTI Manager level. You can gain more redundancy by deploying multiple Unified IP IVR servers. Multiple Unified IP IVR servers enable call routing scripts to load balance calls between the available IP IVR resources.

The following figure shows two Unified IP IVR servers set up redundantly with a cluster. Set up the Unified IP IVR group with each server connected to the CTI Manager on a different Unified Communications Manager.
subscriber. Then, add a second CTI Manager as a backup to each Unified IP IVR server. If the primary CTI Manager fails, the Unified IP IVR server fails over to the backup CTI Manager.

**Figure 7: IP IVR High Availability Deployment**

High Availability Through Call Forwarding

You can use the following call forwarding features in Unified Communications Manager to manage Unified IP IVR port usage:

- **Forward Busy**—Forwards calls to another port or route point when Unified Communications Manager detects that the port is busy.
- **Forward No Answer**—Forwards calls to another port or route point when Unified Communications Manager detects that a port did not pick up a call within the timeout period.
- **Forward on Failure**—Forwards calls to another port or route point when Unified Communications Manager detects a port failure caused by an application error.

When using the call forwarding features, do not establish a path back to the CTI port that initiated the call forwarding. Such paths create loops when all the Unified IP IVR servers are unavailable.

High Availability Through Call Flow Routing Scripts

You can use Unified CCE call flow routing scripts to support high availability. Check the Unified IP IVR Peripheral Status with a call flow routing script before sending calls to Unified IP IVR. This check prevents calls from queuing to an inactive Unified IP IVR. For example, create a Translation Route to the Voice Response Unit (VRU) to select the Unified IP IVR with the most idle ports. This method distributes the calls evenly on a call-by-call basis. You can modify this method to load balance ports across multiple Unified IP IVRs. This method can address all the Unified IP IVRs on the cluster in the same Translation Route or Send to VRU node.
If the Unified IP IVR server fails, all calls at the Unified IP IVR are dropped. Minimize the impact of such failures by distributing calls across multiple Unified IP IVR servers. In Unified IP IVR, a default script prevents loss of calls if the Unified IP IVR loses the link to the VRU PG.

---

**Cisco Web and E-Mail Interaction Manager Design Considerations**

The Cisco Web and E-Mail Interaction Manager provides web and email interaction management through a common set of web servers and pages for agents and administrators. It integrates with the Unified CCE platform to provide universal queuing of contacts to agents from different media channels.


**Unified CCE Integration**

To integrate with Unified CCE, Unified WIM and EIM adds a Services server running the External Agent Assignment Service (EAAS) and the Listener Service. The EAAS interacts with the Media Routing Peripheral Gateway (MR PG) through the Media Routing interface. The Listener Service interacts with the Agent PG through the Agent Resource Management (ARM) interface.

The Unified WIM and EIM application server connects with the Unified CCE Administration and Data Server to import relevant configuration data. The Unified WIM and EIM application server also maps the configuration to Unified WIM and EIM objects in the Cisco Interaction Manager database.

For certain deployments of Unified CCE, the MR PG of Unified CCE can reside on the Services server.

**Load-Balancing Considerations**

You can load balance the web service component of a Unified WIM and EIM deployment to serve many agents who simultaneously access the application. You can set up the web (or web and application) servers behind the load balancer with a virtual IP address. When an agent accesses the Unified WIM and EIM with the virtual IP address, the load balancer sends a request to one of the servers behind the address. The load balancer then sends a response back to the agent. In this way, from a security perspective, the load balancer also serves as a reverse proxy server.

The load balancer must support sticky sessions with cookie-based persistence. Before allowing agents access after every scheduled maintenance task, verify that all Web and application servers are available to share the load. If you allow access without all servers being available, the sticky connection feature can cause an overload on the first Web and application server. Using other parameters, you can define a load-balancing algorithm to meet the following objectives:

- Equal load balancing
- Isolation of the primary Web and application server
• Sending fewer requests to a low-powered Web and application server

The load balancer monitors the health of all Web and application servers in the cluster. During a failure, the load balancer removes the given server from the available pool of servers.

**Failover Management**

Unified WIM and EIM supports clustered deployments. These deployments ensure high availability and performance through transparent replication, load balancing, and failover. To handle failure conditions within integrated deployments of Unified WIM and EIM and Unified CCE, use the following methods:

• Implement multiple Web and Application servers. If the primary server goes down, the load balancer can mitigate the failure by routing requests to alternate servers. The load balancer detects application server failure and redirects requests to another application server. The alternate application server creates new user sessions are created and the agents have to sign in again to the Unified Web and E-Mail Interaction Manager.

• Dynamically add or remove servers from the online cluster to accommodate external changes in demand or internal changes in infrastructure.

• Use redundant Unified CCE components, such as MR PIMs and Agent PIMs, to enable Unified WIM and EIM service failovers.

The single points of failure in Unified WIM and EIM include the following:

• A JMS server failure

• A Services server failure

• A Database server failure

**Cisco Outbound Option Design Considerations**

The Cisco Outbound Option enables your contact center to place outgoing calls for a calling campaign. The major components of the Cisco Outbound Option are the following:

• Outbound Option Campaign Manager manages the dialing lists and rules associated with the calls. This component always resides on the Side A Logger platform. You cannot install a redundant copy on the Side B Logger.

• Outbound Option Dialer performs the dialing tasks on behalf of the Campaign Manager. The Dialer emulates a set of IP phones for Unified Communications Manager to make the outbound calls. The Dialer detects the called party and manages the interaction tasks with the Cisco Finesse or CTI OS server to transfer the call to an agent. The Dialer also communicates with the Media Routing Peripheral Gateway (MR PG). Each Dialer has its own peripheral interface manager (PIM) on the MR PG.

• The Outbound Option Import component does not run as a redundant pair. Install this component with the Side A Logger.

• The MR PG accepts route requests from noninbound voice systems such as the Unified Outbound Option or the Multichannel products.

You can deploy the Dialers in either of these methods:

- Coresident on a VM with the MR PG and the Agent PG for the Unified Communications Manager.
- Coresident on a VM with just the MR PG.

The system can support multiple Dialers across the enterprise, but the central Campaign Manager controls all the Dialers. Redundant pairs of SIP Dialers operate in a warm-standby mode similar to the PG fault-tolerance model.

To improve high availability in the Cisco Outbound Option:

- Deploy the MR PGs in redundant pairs.
- Deploy a redundant pair of SIP Dialers for each redundant Agent PG pair. Use the redundant SIP Dialer in the Campaign Manager to enable automatic fault recovery during a failure.
- Deploy redundant Voice Gateways for outbound dialing. The redundant gateways ensure that the Dialers have enough trunks available to place calls if a gateway fails. In some instances where outbound calling is the primary application, you can dedicate these gateways to outbound calling only.

SCCP Dialers do not function as a redundant pair as a PG does, with a pair of Dialers under control of the Campaign Manager. But, you can set up each Dialer to ensure that calls continue on a backup dialer automatically during a failure. During an SCCP Dialer failure, any calls that are already connected to agents remain connected.

Note

The SCCP Dialer is deprecated as of Unified CCE Release 10.0(1).

SIP Dialer Design Considerations

The Outbound Option with SIP Dialer provides high availability through fault-tolerant design in the SIP Dialer, the Agent PG, and Unified SIP Proxy. Many components in the Outbound Option with SIP Dialer run as redundant pairs.

The Campaign Manager supports a single active Dialer per peripheral. Configure one SIP Dialer, but use the same Dialer Name when installing both parts of the redundant Dialer pair. The peripheral setup program allows you to input the Dialer name in the setup page for each SIP Dialer. Only register two SIP Dialers under each name. The Campaign Manager rejects the registration of any additional Dialers that use the same name.

The Campaign Manager activates one SIP Dialer in the Ready state from its registered SIP Dialer pool. If the activated SIP Dialer changes state from Ready to Not Ready or loses its connection, the Campaign Manager activates the standby SIP Dialer. The Campaign Manager returns all outstanding records to Pending status after a timeout period.

If the active SIP loses connection to the CTI Server, Agent PG, or SIP server, the SIP Dialer fails over. The SIP server can be a Voice Gateway (VG) or Unified SIP Proxy. The SIP Dialer uses a heartbeat mechanism to verify its connection to the VG or Unified SIP Proxy. Configure each SIP Dialer to connect to a different VG or Unified SIP Proxy.
During the failover, the SIP Dialer sends all active and pending customer records to the Campaign Manager. If the Campaign Manager is not available, the Dialer closes them internally.

The Unified SIP Proxy server provides weighted load balancing and redundancy in a multiple-gateway deployment by configuring each gateway as part of the Server group configuration. If a gateway is overloaded or loses its WAN link to the PSTN network, Unified SIP Proxy can resend an outbound call to the next available gateway.

Unified SIP Proxy supports the Hot Swappable Router Protocol (HSRP). This protocol provides network redundancy by allowing two Unified SIP Proxy servers to test each other for connectivity continuously.

Because the Campaign Manager and SIP Dialer already include warm-standby functionality, configuring HSRP for Unified SIP Proxy adds undesirable complexity for Outbound Option. Do not use the HSRP configuration for the Unified SIP Proxy servers that are dedicated for Outbound Option.

### SCCP Dialer Design Considerations

The Outbound Option with SCCP Dialer provides high availability by using multiple Dialers for each Unified Communications Manager cluster. You connect two Dialers to a Peripheral Gateway pair. The Campaign Manager distributes calls evenly among the Dialers. If a Dialer fails, the Campaign Manager reroutes calls to the other Dialers that support the remaining campaign contacts. The Campaign Manager marks for retry any in-progress calls on the failed Dialer.

Normally, you distribute IP phones among the Unified Communications Manager subscribers. But, you must all the ports for an SCCP Dialer on the same subscriber.

The Dialer load on a Unified Communications Manager subscriber increases when you start a campaign or allocate resources to a campaign (agents or VRU ports). Two Dialers must not use resources on the same subscriber. Because the Dialers do no coordinate their port usage, their demands can push a subscriber into a code-yellow condition. That condition negatively impacts performance across the contact center.

**Note**

The SCCP Dialer is deprecated as of Unified CCE Release 10.0(1).

### Agent Peripheral Gateway Design Considerations

The Agent PG communicates with the Unified Communications Manager cluster through the CTI Manager. An Agent PG can control agent phones and CTI route points anywhere in the cluster. The Agent PG registers with the CTI Manager on a Unified Communications Manager subscriber in the cluster. The CTI Manager accepts all JTAPI requests from the PG for the cluster. When the PG requests a phone or route point on another subscriber, the CTI Manager forwards the request to the other subscriber using the CallManager SDL links.

**Note**

This section uses *Agent PG* to describe any PG that includes the Unified Communications Manager PIM. The Agent PG can be a Generic PG or a Unified Communications Manager PG. For example, the following figure uses Generic PGs to connect to the subscribers. Those Generic PGs are acting as Agent PGs.

A fault-tolerant design deploys Unified CCE Agent PGs in a redundant configuration, because a PG only connects to the cluster through a single CTI Manager. If that CTI Manager fails, the PG cannot communicate
with the cluster. A redundant PG provides a second pathway through a different CTI Manager on a different subscriber in the cluster.

The minimum design for a high-availability cluster is one publisher and two subscribers. If the primary subscriber fails, the devices rehome to the secondary subscriber and not to the publisher for the cluster.

*Figure 8: High Availability Design for Unified Communications Manager Cluster*

The redundant PG servers keep in synchronization through a private network that is isolated from the public network. If the two servers run on different physical machines at the same site, you can create the private network by connecting an Ethernet Cross-Over Cable between their private-network NICs. If the two PG servers are geographically distributed, use a separate WAN connection for the private network. To avoid a single point of failure in the network, do not use the same circuits or network gear as for the public network.

Within the Agent PG, the JTAPI Gateway and Agent PG PIM manage the connectivity to the cluster. The JTAPI Gateway handles the JTAPI socket connection protocol and messaging between the PIM and the CTI Manager. The PIM manages the interface between Unified CCE, the JTAPI Gateway, and the cluster. It requests specific objects to monitor and handle route requests from the cluster. The PG starts the JTAPI Gateway and PIM automatically as node-managed processes. The PG monitors the processes and automatically restarts them if they fail.

During PG installation, download the JTAPI Gateway from the cluster to ensure compatibility. Whenever you upgrade either the PG or Unified Communications Manager, remove and reinstall the JTAPI Gateway. The JTAPI services from both redundant Agent PGs sign in to the CTI Manager after initialization. Agent PG Side A signs in to the primary CTI Manager; Agent PG Side B signs in to the secondary CTI Manager. Only one PG in each pair actively registers and monitors phones and CTI route points. The redundant PG
runs in hot-standby mode. The redundant PG signs into the secondary CTI Manager only to initialize the interface and make it available for a failover. This arrangement significantly decreases the time for the failover. During system start, the PG that first connects to the Unified CCE Call Router server and requests configuration information is the active PG. The Call Router ensures that the PG side that has the best connection becomes active. The nominal designations of "Side A" and "Side B" do not affect which PG becomes active. During a PG failover caused by a private link failure, a weighting mechanism chooses which PG is active to minimize the impact on the contact center.

The PIM startup process registers all CTI route points first, which is done at a rate of five route points per second. For systems with large numbers of CTI route points, it can take several minutes before the system allows any agents to sign in. You can reduce this time by distributing the devices over multiple PIM interfaces to the cluster.

If calls arrive at the CTI Route Points before the PIM is operational, the calls fail unless you set up the route points with a recovery number. Place the recovery number in their Call Forward on Unregistered or Call Forward on Failure setting. For example, you can set the recovery numbers to the Cisco Unity voicemail system for the Auto Attendant.

**Active PG Shutdowns**

Avoid shutting down an active peripheral gateway service in your production environment. This causes a service interruption of a minute or more while the other side connects and activates. The length depends on the size of the configuration and the type of peripheral. For example, the VRU peripheral can take less time. The other side for the VRU might take 30 seconds or less to reactivate.

---

**Agent PG Deployment for Unified Communications Manager Cluster**

You can deploy Agent PGs in a Unified Communications Manager cluster in either of the following ways:

- Deploy one Agent PG pair for every two Unified Communications Manager subscribers. Each subscriber includes a CTI Manager. Each Agent PG connects to the CTI Manager running on a different subscriber.
The following diagram shows a deployment with two Agent PG pairs that are connected to a cluster with four subscribers.

*Figure 9: 2 Agent PG Pairs for Unified Communications Manager Cluster*

- Deploy a single Agent PG pair for the entire cluster. In this deployment, only the CTI Managers on two subscribers have a direct link to the Agent PG pair. You can spread agent phone registrations among all the subscribers, not just the subscribers that directly connect to the Agent PG pair. The other subscribers
send and receive Agent PG messages through a connected subscriber. The following diagram shows a deployment with a single Agent PG pair that is connected to a cluster with four subscribers.

*Figure 10: Single Agent PG Pair for Entire Unified Communications Manager Cluster*

This model reduces the PG server count for the contact center, but you must adhere to the PG sizing constraints. You can also create teams that span across many subscribers because all calls pass through a single Agent PG pair. These teams enable supervisors to monitor agent phones registered across the cluster. However, this deployment can impose slightly higher resource utilization on the cluster.
Central Controller Design Considerations

You can geographically distribute redundant Unified CCE servers or locate them at the same physical site. In a production deployment, the Call Router, Logger, and Database Server must connect over a private network that is isolated from the public network.

Figure 11: High Availability Design for Central Controller

For simplicity, this figure represents the Central Controller as a single server. Most designs have a set of servers sized to support the Unified CCE agent count and call volume. The Central Controller includes the following redundant servers:

- Call Router provides call routing instructions based on real-time conditions. Redundant pairs of Call Routers maintain synchronized records in memory.

- Logger and Database Server is the repository for all configuration and scripting information as well as recent historical data that the system collects.

The redundant Logger pairs connect with the redundant Call Router pairs: Side A to Side A, and Side B to Side B. Each Call Router reads and writes data only to its connected Logger. Because the Call Routers run in lock step through synchronized messages, the data produced and written to both Loggers is identical.
Some designs install the Call Router and Logger on the same VM. That combination is sometimes called a Rogger.

When the servers are located at the same site, configure them with a second virtual NIC for the private network connection and isolate the private connections. When the servers are geographically separated, use a separate WAN connection for the private network. To avoid a single point of failure in the network, do not use the same circuits or network gear as for the public network.

Common Processes That Support Failovers

Failover scenarios can arise either from software component failures or from network failures. The following sections describe common processes that support component failover behavior.

Failure Detection Methods

Unified CCE uses the Message Delivery Subsystem (MDS) to send synchronization messages. The private network uses TCP keepalive messages that are generated at 100-ms intervals. If no TCP keepalive messages arrive for 500 ms, the system decides that either a network or component failure occurred.

The public network uses the UDP heartbeat mechanism between PGs and the Central Controller. Redundant components generate UDP heartbeats at 100-ms intervals. Routers and PGs generate UDP heartbeats at 400-ms intervals. In both cases, the system decides a failure occurred after missing five UDP heartbeats.

Device Majority

Device majority determines whether a Call Router enters a disabled state. The Call Router checks for device majority when it loses its connection with its redundant Call Router. Each Call Router determines device majority for itself. None, one, or both Call Routers can have device majority simultaneously.

To have device majority, a Call Router must meet one of these conditions:

- The Call Router is the Side A router and it can communicate with at least half of its total enabled PGs.
- The Call Router is the Side B router and it can communicate with more than half of its total enabled PGs.

PG Weight

During a failover for a private link failure, a weighted value determines which PG becomes the enabled PG. The number and type of active components on each side determines the weighted value of the PG. The weight assigned to each component reflects the recovery time of that component and the disruption to the contact center when the component is down. Agent PIMs have higher weights than VRU PIMs and the CTI Server. The component weights are not configurable.
Record Keeping During Failovers

The call data that gets records during a failover depends on which component fails. Depending on the failure condition, some call data is lost. The router can lose access to in-progress calls because of the failure. The in-progress calls are still active, but the Call Router responds as if the calls have dropped. In most cases, the Agent PG creates a Termination Call Detail (TCD) record in the Unified CCE database at this point.

Calls that are already connected to an agent can continue during a failover. The Agent PG creates another TCD record for such calls when they end.

Call Survivability

Call survivability during failovers varies depending on your deployment and which components fail:

- In Unified CVP deployments, the routing dialog in the Central Controller stops and calls under Unified CVP control get treatment from the survivability TCL script in their ingress Voice Gateways. If the survivability scripts redirect the calls to another active Unified CCE component, the call appears as a "new call" to the system with no relationship to the original call for reporting or tracking purposes.

- In Unified IP IVR deployments, survival scripting on the VRU can do some call treatment as the call ends, but does not send the call back to another active Unified CCE component.

- During Agent PG failures, calls survive. An agent with a hard phone can manage the call if the deployment allows. If the agent signs back in while the call is still active, the agent gets reconnected to the in-progress call.

Unified CCE Failovers During Network Failures

Network failures simultaneously affect any components that send traffic across the affected network. Unified CCE components use both private and public network links to communicate.

The traffic on the private network performs these functions:

- State transfer during component startup
- Synchronization of redundant pairs of Call Routers
- Synchronization of redundant pairs of PGs

The public network carries the rest of the traffic between the contact center components: voice data, call context data, and reporting data. The public network includes all the public network links between the Unified CCE components.

Note

In virtualized contact centers, network failures can arise from failures in the virtual environment, like a virtual NIC, or from failures of physical resources.
Response to Private Network Failures

When a private network fails, the contact center quickly enters a state where, depending on system topology, one or both sides transition into isolated-enabled operation. The isolated operation continues until the Call Routers detect the restoration of the private network. The redundant pairs of Routers and PGs then resynchronize and resume normal operation.

Assume that Side A is the pair-enabled Call Router and Side B is the pair-disabled Call Router. When the private network fails, the Side A Call Router behaves as follows:

- If the Side A Call Router has device majority, the Call Router transitions to the isolated-enabled state and continues handling traffic.
- If the Side A Call Router does not have device majority, the Call Router transitions to the isolated-disabled state and stops processing traffic.

When the private network fails, the Side B Call Router behaves as follows:

- If the Side B Call Router does not have device majority, the Call Router transitions to the isolated-disabled state and does not process traffic.
- If the Side B Call Router does have device majority, the Call Router enters a test state. The Router instructs its enabled PGs to contact the Side A Call Router over the public network. Then, the Side B Call Router responds as follows:
  - If no PG can contact the Side A Call Router to determine its state, the Side B Call Router transitions to the isolated-enabled state and begins handling traffic. This case can result in both Side A and Side B running in isolated-enabled state.
  - If any PG contacts the Side A Call Router and finds the Call Router in the isolated-disabled state, the Side B Call Router transitions to the isolated-enabled state and begins handling traffic.
  - If any PG contacts the Side A Call Router and finds the Call Router in the isolated-enabled state, the Side B Call Router transitions to the isolated-disabled state and does not process traffic.

During Call Router failover processing, any Route Requests that are sent to the Call Router are queued until the surviving Call Router is in isolated-enabled state. A Call Router failure does not affect any in-progress calls that have already been routed to a VRU or an agent.

The Logger shuts down when its Call Router goes idle. Each Logger communicates only with its own Call Router. If the private network connection is restored within 12 hours, the isolated-enabled Call Router's Logger uses its data to resynchronize the other Logger. If the private network connection remains down for more than 12 hours, manually resynchronize the Loggers using the process described in the Administration Guide for Cisco Unified Contact Center Enterprise & Hosted at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-maintenance-guides-list.html.

In each redundant pair of PGs, there is also an enabled PG and a disabled PG. At system start, the first PG to connect becomes the enabled PG. However, after a private network failure, the PG with the greatest weight in the redundant pair becomes the enabled PG. The other PG becomes the disabled PG.
Response to Public Network Failures

Highly available networks generally include redundant channels for the public network. When one channel fails, another channel takes over seamlessly. The contact center detects a public network failure only when all channels fail between two components.

Note

In contact centers without redundant public networks, the contact center detects a failure when the single channel fails.

How the contact center responds to a public network failure depends on number and function of the sites and how the sites are linked. The following sections look at some of the more common or significant scenarios.

Failures Between Unified Communications Managers

The scenario that can cause the most problems involves the subscribers losing their public link. Because the functioning private network keeps the Call Routers and Agent PGs in synch, the Call Routers can still detect all agent devices. In this situation, a Call Router can choose an agent device that is registered on the subscriber on the other side of the public network failure. The local CVP cannot pass the connection information to the agent device on the other side of the public network failure. The call fails, but the routing client marks the call as routed to the agent device on the remote subscriber.

Failures Between Data Centers in Clustering over WAN

In a clustering over the WAN deployment, you need a highly available, highly resilient WAN with low latency and sufficient bandwidth. The public network is a critical part of Unified CCE fault tolerance. A highly available WAN is fully redundant with no single points of failure, usually across separate carriers. During a partial failure of the WAN, the redundant link needs the capability to handle the full load for the data centers within the QoS parameters. As an alternative to redundant WANs, you can employ a SONET fiber ring. For more information about designing highly available, highly resilient WANs, see the Overall WAN Architecture page in the Cisco Design Zone.

If the public network fails between the data center locations, the system responds in this manner:

1. The Unified Communications Manager subscribers detect the failure. The subscribers continue to function locally with no impact to local call processing and call control. However, any calls that were set up over the public network fail.

2. The Call Routers and Agent PGs detect the failure. The Agent PGs automatically realign their data communication stream to their local Call Router. The local Call Router then passes data to the Call Router on the other side over the private network to continue call processing. The altered data path does not cause a failover of the Agent PGs or the Call Router.

The impact of the public network failure on agents depends on where their phones and desktops registered:

- The most common case is that the agent desktop and agent phone are both registered to the Agent PG and a subscriber on the same side (Side A for this example). When the public link between the data centers fails, the agent can continue handling calls normally.

- In some cases, the agent desktop (Side A for this example) and the agent phone (Side B for this example) can end up registered on different sides. In those cases, the CTI Manager directs phone events over the
public network to the Agent PG on the opposite side. When the public network between the data centers fails, the phone does not rehome to Side A of the cluster. The phone remains operational on Side B. The Agent PG on Side A cannot detect this phone. Because the Unified CCE can no longer direct calls to the agent phone, Unified CCE automatically signs out the agent.

• Normally, the redundant desktop server pair load balances agent desktop connections. So, half of the desktops register on a desktop server that connects to the active CTI Server Peripheral Gateway (CG) across the public network. When the public network fails, the desktop server loses connection with the remote CG. The desktop server disconnects the active agent desktops to force them to rehome to the redundant desktop server at the remote site. The agent desktop automatically uses the redundant desktop server. The agent desktop remains disabled until it connects to the redundant desktop server.

Failures to Agent Sites in Clustering over WAN

The Unified CCE model for clustering over the WAN assumes that the Unified CCE agents are remotely located at multiple sites. Each agent site requires access to both data centers through the public network for redundancy. In a complete network failure, these connections also provide basic SRST functionality, so that the agent site can still make emergency (911) calls.

If the Unified CCE agent site loses the public network connection to one of the data centers, the system responds in this manner:

1. Any IP phones that are homed to the Unified Communications Manager subscribers at the disconnected data center automatically rehome to subscribers at the other data center. To use the rehoming behavior, configure a redundancy group.

2. Agent desktops that are connected to the desktop server at that disconnected data center automatically realign to the redundant server at the other data center. (Agent desktops are disabled during the realignment process.)

If the Unified CCE agent site loses the public network connection to both of the data centers, the system responds in this manner:

1. The local Voice Gateway (VG) detects the failure of the communications path to the cluster. The VG then goes into SRST mode to provide local dial-tone functionality.

2. With Unified CVP, the VGs detect the loss of connection to the Unified CVP Server. Then, the VGs execute their local survivability TCL script to reroute the inbound calls.

3. If an active call came in to the disconnected agent site on a local PSTN connection, the call remains active. But, the Agent PG loses access to the call and creates a TCD record.

4. The Finesse server (or CTIOS server) detects the loss of connectivity to the agent desktop and automatically signs the agent out of the system. While the IP phones are in SRST mode, they cannot function as Unified CCE agents.

Response to Failures of Both Networks

Individually, parts of the public and private networks can fail with limited impact to the Unified CCE agents and calls. However, if both of these networks fail at the same time, the system retains only limited functionality. This failure is considered catastrophic. You can avoid such failures by careful WAN design with built-in backup and resiliency.
A simultaneous failure of both networks within a site shuts down the site. If both the public and private networks simultaneously fail between two sites, the system responds in this manner:

1. Both Call Routers check for device majority. Each router enters isolated-enabled mode if the router has device majority or isolated-disabled mode if the router does not have device majority.

2. The PGs automatically realign their data communications, if necessary, to their local Call Router. A PG that cannot connect to an active Call Router becomes inactive.

3. The Unified Communications Manager subscribers detect the failure and continue to function locally with no impact to local call processing and call control.

4. Any in-progress calls that are sending active voice path media over the public WAN link fail with the link. When the call fails, the PG creates a TCD record for the call.

5. In a clustering over the WAN deployment, the Unified Communications Manager subscribers on each side operate with access only to local components.

6. The Unified CCE call routing scripts automatically route around the off-line devices using peripheral-on-line status checks.

7. Agents with both their phones and desktops registered with local subscribers are not affected. All other agents lose some or all functionality while their phones and desktops rehome. Those agents might also find themselves signed out, depending on the exact system configuration.

8. Unified CCE does not route new calls that come into the disabled side. But, you can redirect or handle those calls with the standard Unified Communication Manager redirect on failure for their CTI route points or with the Unified CVP survivability TCL script in the ingress Voice Gateways.

Unified CCE Failovers During Single-Component Failures

Unified CCE components run in redundant pairs. If a component fails, its counterpart takes over processing. These failovers happen automatically and often have little or no impact on active agents. The following sections discuss how Unified CCE responds to component failures.

Note

The following sections use "Agent PG" to describe the Peripheral Gateway that has a Unified Communications Manager PIM installed. Different Unified CCE deployments can install the Unified Communications Manager PIM in various combinations. For example, you can install the PIM on a standalone PG or together with other components on a Generic PG.

Agent PG Fails

This scenario shows recovery from a PG Side A failure.

The following conditions apply to this scenario:

- Unified Communications Manager subscriber A has the primary CTI Manager.
- For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.
The following figure shows a failure on PG Side A and a failover to PG Side B. All CTI Manager and Unified Communications Manager services continue running normally.

**Figure 12: Agent PG Side A Fails**

Failure recovery occurs as follows:

1. PG Side B detects the failure of PG Side A.
2. PG Side B registers all dialed numbers and phones. Call processing continues through PG Side B.
3. Phones and gateways stay registered and operational with subscriber A; they do not fail over.
4. The in-progress calls remain active on agent phones, but the agents cannot use phone services, like transfers, until the agents sign back in.
5. During the failover to PG Side B, the states of unoccupied agents and their desktops can change depending on their configuration. Options for three-party calls can be affected. In some cases, agents have to sign back in or manually change their state after the failover completes.
6. After recovery from the failure, PG Side B remains active and uses the CTI Manager on subscriber B. The PG does not fail back to Side A, and call processing continues on the PG Side B.

**Subscriber Without CTI Manager Link to Agent PG Fails**

In a Unified Communications Manager cluster supporting 2000 agents, you have one Unified Communications Manager publisher and four Unified Communications Manager subscribers. Each subscriber can support 500 agents. Each Agent PG can support only one CTI Manager connection. While each subscriber has a CTI
Manager, only two subscribers can connect to the Agent PGs. You would have to add another pair of Agent PGs to enable all the subscribers in this cluster to connect directly to an Agent PG.

The following figure shows a failure on subscriber A, which does not have a direct connection to an Agent PG.

The following conditions apply to this scenario:

- For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.
- Subscribers C and D connect to the Agent PGs and each run a local instance of CTI Manager to provide JTAPI services for the PGs.

**Figure 13: Unified Communications Manager Without Link to Agent PG Fails**

Failure recovery occurs as follows:

1. If subscriber A fails, its registered phones and gateways rehome to the backup subscriber B.
Agent PG Side A remains active and connected to the CTI Manager on subscriber C. The PG does not fail over, because the JTAPI-to-CTI Manager connection has not failed. But, the PG detects the phone and device registrations automatically switching from subscriber A to subscriber B.

Call processing continues for any devices that are not registered to subscriber A.

While the agent phones are not registered, the Agent PG disables the agent desktops. This response prevents the agents from using the system without a subscriber connection. The Agent PG signs the agents out during this transition to avoid routing calls to them.

Call processing resumes for the phones after they reregister with their backup subscriber.

In-progress calls continue on phones that were registered to subscriber A, but the agents cannot use phone services, like transfers, until the agents sign back in.

When the in-progress call ends, that phone reregisters with the backup subscriber. The Agent PG signs the agents out during this transition to avoid routing calls to them.

When subscriber A recovers, phones and gateways rehome to it. You can set up the rehoming on subscribers to return groups of phones and devices gracefully over time. Otherwise, you can require manual intervention during a maintenance window to redistribute the phones to minimize the impact to the call center. During this rehoming process, the CTI Manager notifies the Agent PG of the registrations switching from subscriber B back to the original subscriber A.

Call processing continues normally after the phones and devices return to their original subscriber.

CTI Manager with Agent PG Link Fails

In a Unified Communications Manager cluster supporting 2000 agents, you have one Unified Communications Manager publisher and four Unified Communications Manager subscribers. Each subscriber can support 500 agents. Each Agent PG can support only one CTI Manager connection. While each subscriber has a CTI Manager, only two subscribers can connect to the Agent PGs. You would have to add another pair of Agent PGs to enable all the subscribers in this cluster to connect directly to an Agent PG.

The following figure shows the failure of a CTI Manager with a connection to the Agent PG. Only subscribers C and D are configured to connect to the Agent PGs.

The following conditions apply to this scenario:

• For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.
• The CTI Managers on subscribers C and D provide JTAPI services for the Agent PGs.

Failure recovery occurs as follows:

1 When the CTI Manager on subscriber C fails, the Agent PG Side A detects the failure and induces a failover to PG Side B.
2 Agent PG Side B registers all dialed numbers and phones with the CTI Manager on subscriber D and call processing continues.
3 In-progress calls stay active, but the agents cannot use phone services, like transfers, until the agents sign back in.
4 When the CTI Manager on subscriber C recovers, Agent PG Side B continues to be active and uses the CTI Manager on subscriber D. The Agent PG does not fail back in this model.

Unlike Unified CVP, Unified IP IVR depends on the CTI Manager for call control. In Unified IP IVR deployments, failure of a CTI Manager with an Agent PG connection causes the Unified IP IVR JTAPI
subsystem to shut down. This shutdown causes the Unified IP IVR server to drop all voice calls that the server is processing.

Then, the JTAPI subsystem restarts and connects to the CTI Manager on the backup subscriber. The Unified IP IVR reregisters all the CTI ports that are associated with the Unified IP IVR JTAPI user. After all the Unified Communications Manager devices are successfully registered, the server resumes its Voice Response Unit (VRU) functions and handles new calls.

**Voice Response Unit PG Fails**

When a Voice Response Unit (VRU) PG fails, calls in progress or queued in Unified CVP do not drop. The Survivability TCL script in the Voice Gateway redirects the calls to a secondary Unified CVP or a number in the SIP dial plan, if available.

In deployments using Unified IP IVR, all calls that are currently queued or in treatment drop unless the deployment includes one of the following:

- A default script application defined for Unified IP IVR
- Recovery numbers defined for the CTI Ports in Unified Communications Manager.

After failover, the redundant VRU PG connects to the Unified CVP or Unified IP IVR and begins processing new calls. On recovery of the failed VRU PG side, the currently running VRU PG continues to operate as the active VRU PG. Redundant VRU PGs enable Unified CVP or Unified IP IVR to function as an active queue point or to provide call treatment.

*Note*

Unless Unified IP IVR deployments have a redundant VRU PG, a VRU PG failure still blocks the use of a functional Unified IP IVR server.

*Figure 15: VRU PG Fails*

**Logger Fails**

The Unified CCE Logger and Database Server maintains the system database for the configuration (agent IDs, skill groups, call types) and scripting (call flow scripts). The server also maintains the recent historical
data from call processing. The Loggers receive data from their local Call Router. Because the Call Routers are synchronized, the Logger data is also synchronized.

A Logger failure has no immediate impact on call processing. The redundant Logger receives a complete set of call data from its local Call Router. If the failed Logger is restored within 12 hours, the Logger automatically requests all the transactions for when it was off-line from the backup Logger. The Loggers maintain a recovery key that tracks the date and time of each entry recorded in the database. The redundant Logger uses these keys to identify the missing data.

If the Logger was off line for more than 12 hours, the system does not automatically resynchronize the databases. In this case, the system administrator can manually resynchronize the Loggers using the Unified ICMDBA application. The manual process allows you to choose a convenient time to transfer the data across the private network.

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

In deployments that use Cisco Outbound Option, the Campaign Manager is loaded only on the primary Logger. If that platform is out of service, any outbound calling stops while the Logger is down.

**Administration and Data Server Fails**

The Administration and Data Server provides the user interface to the system for making configuration and scripting changes. The server can also host the web-based reporting tool and the Internet Script Editor. Unlike other Unified CCE components, the Administration and Data Server does not operate in redundant pairs. If you want to provide redundancy for the functions on this server, you can include more Administration and Data Servers in your design. But, there is no automatic failover behavior.

The Administration and Data Server receives a real-time feed of data from across Unified CCE from the Call Router through a Real-Time Distributor. If you have several Administration and Data Servers at the same site, you can configure the Real-Time Distributors into a single Admin Site. The Admin Site has a primary distributor and one or more secondary distributors. The primary distributor registers with the Call Router and receives the real-time feed across the network from the router. The secondary distributors use the primary distributor as their source for the real-time feed. This arrangement reduces the number of real-time feeds that the router supports and saves bandwidth.
If the primary real-time distributor fails, the secondary real-time distributors register with the router for the real-time feed as shown in the following figure. Administration clients that cannot register with the primary or secondary Administration and Data Server cannot perform any tasks until the distributors are restored.

**Figure 16: Primary Real-Time Distributor Fails**

In some deployments, the Administration and Data Server also hosts an interface for the Cisco Unified Contact Center Management Portal (Unified CCMP). In those deployments, when the Administration and Data Server is down, any configuration changes that are made to the Unified CCE or Unified CCMP systems are not passed over the interface.

**CTI Server Fails**

The CTI Server monitors the Agent PG traffic for specific CTI messages (such as call ringing or off-hook events). The CTI Server makes those messages available to CTI clients such as the Cisco Finesse server or CTI OS server. The CTI Server also processes third-party call control messages (such as make call or answer call) from the CTI clients. The CTI Server sends those messages through the Agent PG to Unified Communications Manager for processing.
You deploy the CTI Server in redundant pairs. Each half of the redundant pair is coresident on a VM with one half of a redundant Agent PG pair. On failure of the active CTI Server, the redundant CTI Server becomes active and begins processing call events.

**Figure 17: CTI Server Fails**

Both the Finesse server and CTI OS server are clients of the CTI Server. The desktop server, rather than the CTI Server, maintains agent state during a failover. Both Finesse and CTI OS partially disable agent desktops when the CTI Server fails. In some cases, an agent must sign in again after the failover completes.

---

**Note**

If no clients are connected to the active CTI Server, a mechanism is in place to force a failover after a preset period. This failover occurs to isolate any spurious reasons that prevent the CTI clients from connecting to the active CTI Server.
Cisco Finesse Server Fails

You deploy the Cisco Finesse server in redundant pairs in dedicated virtual machines. Both Finesse servers run in active mode all the time.

Figure 18: Finesse Server Fails

When a Cisco Finesse server fails, failure recovery occurs as follows:

1. Agent desktops that are signed in to the server detect a loss of connection and fail over to the redundant server.
2. Agents are automatically signed in on the new server after the failover.
3. Third-party applications that use the Cisco Finesse REST API must perform the failover within their application logic to move to the redundant server.
4. If Cisco Tomcat failed with the Cisco Finesse server, the system attempts to restart Cisco Tomcat before restarting the Cisco Finesse server.
5. When the failed server restarts, new agent desktop sessions can sign in on that server. Agent desktops that are signed in on the redundant server remain on that server.

CTI OS Server Fails

You deploy the CTI OS server in redundant pairs. Each half of the redundant pair is coreresident on a VM with one half of a redundant Agent PG pair. Unlike the PG processes that run in hot-standby mode, both of the CTI OS server processes run in active mode all the time.

When a CTI OS server fails, failure recovery occurs as follows:

1. Agent desktops that are signed in to the server detect a loss of connection and fail over to the redundant server.
2. When the failed server restarts, new agent desktop sessions can sign in on that server. Agent desktops that are signed in on the redundant server remain on that server.
Unified CCE Failovers During Multicomponent Failures

When more than one component fails, Unified CCE might not fail over as seamlessly as during a single-component failure. The following sections discuss how Unified CCE responds to multicomponent failures.

Agent PG and CTI Manager Fail

A CTI Manager connects only with its local subscriber and a single Agent PG. There is no direct communication with the other CTI Manager in the cluster. The CTI Managers are kept in sync by data from the other components.

If the Agent PG on one side and the CTI Manager on the other side both fail, Unified CCE cannot communicate with the cluster. This scenario prevents the system from connecting to the agents on this cluster. The cluster remains disconnected until the Agent PG or the backup CTI Manager come back online.

Figure 19: Agent PG Cannot Connect to Backup CTI Manager

Unified Communications Manager Subscriber and CTI Manager Fail

The scenario shows recovery from a complete failure of the Unified Communications Manager subscriber A server.

The following conditions apply to this scenario:
• Subscriber A has the primary CTI Manager.

• For redundancy, all phones and gateways that are registered with subscriber A use subscriber B as their backup server.

**Figure 20: Unified Communications Manager and CTI Manager Fail**

Failure recovery occurs as follows:

1. When subscriber A fails, all inactive registered phones and gateways reregister to subscriber B.
2. The in-progress calls remain active, but the agents cannot use phone services, like transfers.
3. Agent PG Side A detects a failure and induces a failover to Agent PG Side B.
4. Agent PG Side B becomes active and registers all dialed numbers and phones. Call processing continues.
5. As each in-progress call ends, that agent phone and desktop reregister with the backup subscriber. The exact state of the agent desktop varies depending on the configuration and desktop.
6. When subscriber A recovers, all idle phones and gateways reregister to it. Active devices wait until they are idle before reregistering to the primary subscriber.
7. Agent PG Side B remains active using the CTI Manager on subscriber B.
8. After recovery from the failure, the Agent PG does not fail back to Side A of the redundant pair. All CTI messaging is handled using the CTI Manager on subscriber B which communicates with subscriber A to obtain phone state and call information.
Other Considerations for High Availability

A Unified CCE failover can affect other parts of the solution. Some failure scenarios can result in the loss of data that other products use.

Reporting Considerations

The Unified CCE reporting feature uses real-time, 5 minute, and reporting-interval (15 or 30 minute) data to build its reporting database. At the end of each 5 minute and reporting interval, each Peripheral Gateway gathers its local data and sends it to the Call Routers. The Call Routers process the data and send the data to their local Logger for historical data storage. The Logger replicates the historical data to the HDS/DDS database.

The PGs provide buffering (in memory and on disk) of the 5-minute data and reporting-interval data. The PGs use this buffered data to handle slow network response and automatic retransmission of data after network services are restored. If both PGs in a redundant pair fail, you can lose the 5-minute data and reporting-interval data that was not sent to the Central Controller.

When agents sign out, all their reporting statistics stop. When the agents next signin, the real-time statistics for the agents start from zero. Depending on the agent desktop and what an agent is doing during a failure, some failovers can cause the contact center to sign out agents. For more information, see the Reporting Concepts for Cisco Unified ICM/Contact Center Enterprise & Hosted at http://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-enterprise/products-user-guide-list.html.