Cisco Finesse Failover Mechanisms

CTI Failover

CTI failover is when the Finesse server disconnects from one CTI server and reconnects to the same or another CTI server.

The prerequisites for successful CTI failover are as follows:

- Unified Contact Center Enterprise (Unified CCE) must be configured in duplex mode.
- The B Side CTI host and port must be configured through the Finesse administration console.

In the duplex mode, if Finesse loses connection to CTI server, it attempts to connect to the server which is running. Finesse alternates between the configured servers until it makes a successful connection.

While failover is in progress, Finesse transitions to OUT_OF_SERVICE state. During this period, Finesse does not entertain client requests or send out events. Any request made during this time receives a 503 Service Unavailable error message.

After reconnecting to a CTI server and transitioning to IN_SERVICE state, Finesse responds to client requests and publishes events.

Connection to the CTI server can be lost due to the following reasons:

- Finesse misses three consecutive heartbeats from the connected CTI server (heartbeat interval is five seconds).
- Finesse socket that is opened to the CTI server fails.

After the failover is complete, the last state of call control, call data, or agent state are published as events to all clients. This allows Finesse clients to reflect an accurate view of the call control, call data, and agent state.

If an agent either makes or answers a call, and then ends that call during failover (that is, the entire call takes place during failover), the corresponding events are not published.

Note An agent or supervisor who signs in after being on an active conference with other devices (which are not associated with another agent or supervisor) may experience unpredictable behavior with the Finesse desktop due to incorrect call notifications from Unified CCE. These limitations also encompass failover scenarios where a failover occurs while the agent or supervisor is participating in a conference call. For example, an agent is in a conference call when the Finesse server fails. When the agent is redirected to the other Finesse server, that agent may see unpredictable behavior on the Finesse desktop. Examples of unpredictable behavior include, but are not limited to, the following:

- The Finesse desktop does not show all participants in a conference call.
- The Finesse desktop does not show that the signed-in agent or supervisor is in an active call.
- The Finesse receives inconsistent call notifications from Unified CCE.

Despite these limitations, the agent and supervisor can continue to perform general operations on the phone. Desktop behavior returns to usual after the agent or supervisor drops off the conference call.

Prevent Non-Voice Task RONAs during CTI Reconnect

When CTI disconnection happens, the agent state is changed to WORK, on the respective non-voice Media Routing Domain (MRD), to prevent tasks getting routed to the disconnected agents. Previous releases of Unified CCE used to change the agent states back to an available state when the CTI connection is re-established, even though the media handling gadgets and the media channels are not initialized by then.

The media handling gadgets, and the media channels are initialized only after the Finesse desktop failover completes.

Due to the significant delay in desktop failing over after the Finesse server reconnects to the CTI server, chances of occurrence of RONA (Redirection on No Answer) are high when dealing with non-voice tasks.

Unified CCE, Release12.5 (1) or later allows the agent state to remain in WORK mode after CTI reconnection. This allows the agents to change to an available state in non-voice MRD explicitly after the Finesse desktop and media channels are initialized. This avoids the task being routed to the user before the agent is ready to handle non-voice media tasks.

By default, Cisco Finesse Release 12.5(1) retains the earlier behavior, which can be modified using the **enableAutoWorkModeStateChange** property. By default, this property is set to *true*, and the administrator can set to *false* to change to the new behavior.

Note This behavior is supported from Unified CCE Release 12.5(1) onwards, and only after the relevant non-voice gadgets or custom desktop or clients support this behavior.

The agents remain in the WORK mode until they are explicitly set to active on the respective MRD using the REST API. This informs the CTI that the media channel is available (and connected) and the tasks can be routed to the respective user on that MRD.

The `Media—Change Agent from Work State to Active` API allows a user to change the agent state from WORK state to active (READY or NOT_READY), which is automatically computed by Unified CCE. Users can only use this API when an agent state is WORK.

AWDB Failover

The prerequisites for AWDB failover are as follows:

- The secondary Administrative Workstation Database (AWDB) is configured.
- The secondary AWDB host is configured through the Cisco Finesse administration console.
- Cisco Finesse can connect to the secondary AWDB host.
- The Distributor service is running on the secondary AWDB host.

Agents and supervisors are authenticated against the AWDB database. When an agent or supervisor makes a successful API request (such as a sign in or call control request), the credentials are cached in Cisco Finesse for 30 minutes from the time of the request. After a user is authenticated, that user continues to be authenticated until 30 minutes pass, even if both AWDBs are down. Cisco Finesse attempts to reauthenticate the user against the AWDB only after the cache expires.

If Cisco Finesse loses connection to the primary Administration & Data server, and the preceding prerequisites have been implemented, AWDB failover occurs. After Cisco Finesse loses connection to the primary Administration & Data server, it tries to reconnect to the secondary server.

Cisco Finesse repeats this process for every API request until it can connect to one of the Administration & Data servers. During failover, Cisco Finesse does not process any requests, but clients can still receive events.

If Cisco Finesse cannot connect to either of the Administration & Data servers and the cache has expired, the systems returns the following errors:

- Agents and supervisors who attempt to sign in to the Finesse desktop receive an “Invalid user ID or password” error message.
- Administrators cannot update or retrieve settings in the Cisco Finesse administration console.
- Users who are already signed in to Cisco Finesse receive an “Operation timed out” error message.
- Users who make API requests receive an 401 “Unauthorized” HTTP error message.

If Cisco Finesse loses connection to one AWDB and then receives requests, these requests may time out before Cisco Finesse can detect that the connection is down and connect to the alternate AWDB. In this scenario, the user (administrator, agent, or supervisor) may need to retry the operation for it to succeed.

Finesse Desktop Failover

Desktop failover can occur for the following reasons:

- When the Finesse desktop loses network connectivity to the Finesse Notification Service.
- When the Finesse Tomcat Service becomes *Unavailable*
- When the Finesse REST API Service becomes *Unavailable*

- When the Finesse Notification Service becomes *Unavailable*
- When the Finesse loses connection to CTI servers

Note

- After the failover, the pending state of an agent will not be displayed once the agent fails over to the subscriber. The pending state change is lost during the failover, as the agent will be logged out, and logged in again.
- Finesse is IN_SERVICE, coordinates the distribution of desktop reloads, such that failover and consequent desktop reloads are evenly distributed to prevent overwhelming of the Finesse server. Configuration data such as reason codes, workflows and so on are not reloaded during failover to improve the performance.

If the server that an agent is connected transitions to OUT_OF_SERVICE, the agent receives a notification that the connection with the server is lost. The Finesse desktop:

- Checks whether the subscriber is available and IN_SERVICE.
- Continues to check whether the publisher recovers its state.

If the subscriber is available, then the desktop automatically signs the agent into the subscriber. If the publisher recovers its state, the desktop notifies the agent that it has reconnected.

The failover logic has three triggers to detect desktop failure:

- The Finesse desktop receives a SystemInfo event that the publisher is OUT_OF_SERVICE.
- The Finesse Notification Service is disconnected.
- The XMPP presence of “Finesse” user changes to *Unavailable*.

No matter which trigger is detected, the desktop reconnection logic is as follows:

1. Poll SystemInfo for publisher every 20 seconds.
2. If SystemInfo API reports Finesse is IN_SERVICE, check the Finesse Notification Service.
3. If SystemInfo is IN_SERVICE, check whether the last CTI heartbeat status of the side being connected is a success.

Note The last CTI heartbeat status is checked to ensure that the subscriber is healthy before failover, and thus does not immediately transition to OUT_OF_SERVICE after the client has failed over. This may occur in CTI failure, since both Finesse servers connect to the same PG and CTI server, and a CTI failure can cause both Finesse servers to disconnect and connect to the alternate PG. Depending on the network topology the subscriber might be slower to sense a network disconnect.

4. If XMPP is disconnected, make the Finesse Notification Service request.
5. If the Finesse Notification Service is successful and Finesse service is IN_SERVICE, refresh the data.

The failover logic prefers to stay with the publisher. If the failover logic detects that the subscriber is available, it checks the publisher one more time. If the publisher has recovered, the desktop reconnects to the publisher. If the publisher is still down, the desktop connects the agent to the subscriber. In this case, the agent does not automatically reconnect to the failed server after it recovers, but instead remains connected to the subscriber.

If the Finesse Notification Service is the source of failure, the desktop makes three attempts to reconnect before changing the state of the desktop to disconnected. These attempts occur before the failover logic begins.

Desktop Behavior

Cisco Finesse sends a code of 255 to the CTI server and you may see a different code on the CTI server side. The actual behavior of the desktop under these conditions depends on the setting for Logout on Agent Disconnect (LOAD) in Unified CCE. By default, the CTI server places the agent in Not Ready state.

Note Finesse takes up to 120 seconds to detect when an agent closes the browser. If the browser crashes, Finesse waits 60 seconds before sending a forced logout request to the CTI server. Under these conditions, Finesse can take up to 180 seconds to sign out the agent.

The following table lists the conditions under which Finesse sends this code to the CTI server.

Scenario	Desktop Behavior	Server Action	Results
The agent closes the browser, the browser crashes, or the agent clicks the Back button on the browser.	Finesse desktop makes a best-effort attempt to notify the server.	Finesse receives a presence notification of *Unavailable* from the client. Finesse waits 60 seconds, and then sends a forced logout request to the CTI server.	**Race Conditions** **1.** The agent closes the browser window. Finesse receives a presence notification of *Unavailable* for the user. Finesse tries to sign the agent out; however, that agent is already signed out. **2.** If the browser crashes, it can take the Finesse server up to 120 seconds to detect that the client is gone and send a presence notification to Finesse. A situation can occur where the client signs into the subscriber before the publisher

			receives the presence notification caused by the browser crash. In this case, the agent may be signed out or put into Not Ready state on the subscriber. **3.** If the Finesse desktop is running over a slower network connection, Finesse may not always receive an *Unavailable* presence notification from the client browser. In this situation, the behavior mimics a browser crash, as described in the preceding condition. **4.** If agent is in Not Ready State before Failover, agent moves to Not Ready — Connection Failure after CTI Disconnect or Reconnect. If agent is in Ready State before Failover, agent moves to Not Ready — Connection Failure upon his next state change to Not Ready.
The client refreshes the browser	—	Finesse receives a presence notification of *Unavailable* from	—

		the client. Finesse waits 60 seconds before sending a forced logout request to the CTI server to allow the browser to reconnect after the refresh.	
The client encounters a network glitch (Finesse is IN_SERVICE)	Connection to the Finesse server temporarily goes down, consequently the client fails over to the subscriber.	The publisher receives a presence notification of *Unavailable* from the client. Finesse is IN_SERVICE, so it sends a forced logout request to the CTI server for the agent.	**Race Conditions** A situation can occur where the forced logout does not happen before the client signs in to the subscriber. If the agent is on a call, the publisher sends the forced logout request after the call ends. The agent will be signed out or put into Not Ready state when the call ends, even though the client is already signed in to the subscriber. If agent is in Not Ready State before Failover, agent moves to Not Ready — Connection Failure after CTI Disconnect or Reconnect. If agent is in Ready State before Failover, agent moves to Not Ready — Connection Failure upon the next state change to Not Ready.
The Refresh Token has expired. For more information on tokens, see https://developer.cisco.com/docs/finesse/#single-sign-on-apis.	Finesse desktop sends a forced logout request to the CTI server.	The Finesse server forwards the forced logout request to the CTI server.	The session expiry warning appears 10 minutes and 5 minutes before the Refresh Token expires. In the last minute, a countdown timer appears till the Refresh Token expires. The agent is forcefully

			logged out when the timer reaches zero and must log in again. For Unified CCE, the state of the agent changes to Log Out or Not Ready based on the Load parameter set as below. **Load parameter = 0** • When the agent's current state is Not Ready, Ready, or Wrap-Up, the agent's state after force logout is changed to Not Ready – Connection Failure. • When the agent's current state is Talking, the Agent goes into Not Ready – Connection Failure state after the call ends. **Load parameter = 1** • When the agent's current state is Not Ready, Ready, or Wrap-Up, the agent goes to Logged Out – System Failure. • When the agent's current state is Talking, the Agent goes to Logged Out – System Failure immediately even though the call is still active.

Desktop Chat Failover

The following table lists the desktop chat failover scenarios:

Failover Type	Desktop Chat Behavior
Cisco IM&P server failover	The desktop chat status is retained, and all active chat sessions are lost.
Finesse server failover	The desktop chat status is retained, and all active chat sessions are lost.
CTI server failover	The desktop chat status and all chat sessions are retained.

Finesse IP Phone Agent Failover

Finesse IPPA failover can occur for the following reasons:

- The Finesse REST API Service transitions to OUT_OF_SERVICE.
- The Finesse Notification Service transitions to OUT_OF_SERVICE.
- If Finesse IPPA detects a server failure before Finesse fails over to the alternate CTI server, then Finesse IPPA declares the Finesse server OUT_OF_SERVICE.

The server that an agent is connected transitions to OUT_OF_SERVICE, the Finesse IP Phone Agent (IPPA) displays a notification that the server is unavailable. Finesse IPPA continues to check whether the current Finesse server recovers its state and notifies the agent if it reconnects.

Finesse IPPA attempts to reconnect to the server every 5 seconds and declares it OUT_OF_SERVICE after three failed attempts. The total time required for the transition to OUT_OF_SERVICE is approximately 15 seconds.

Unlike the Finesse desktop, Finesse IPPA does not check whether the subscriber is available. To connect to subscriber, the agent must exit the publisher, and manually sign into the subscriber.

Finesse IPPA failover logic has the following two triggers to detect failure:

- Finesse IPPA receives a SystemInfo event that the publisher is OUT_OF_SERVICE.

  Finesse IPPA polls SystemInfo every 5 seconds to check whether the Finesse server is IN_SERVICE. After three attempts, if the Finesse server is not IN_SERVICE, Finesse IPPA displays a server unavailable message to the agent.

- Finesse IPPA receives notification that the Finesse Notification Service is disconnected.

  Finesse IPPA tries every 5 seconds to reconnect with the XMPP server. After three attempts, if the Finesse Notification Service cannot be reestablished, Finesse IPPA displays a server unavailable message to the agent.

While the agent is still signed into the current service, Finesse IPPA continues attempting to reestablish the connections with the Finesse and XMPP servers. If they both resume service, Finesse IPPA displays the **Sign In** screen and the agent can sign in again and continue as usual.

Alternately, the agent must exit the current Finesse service and try to connect using an alternate Finesse service.

Guidelines for Optimal Desktop Failover

The following are the guidelines to optimize failover scenarios, to avoid server overload and unnecessary delays.

- Browser Configuration, on page 10
- Agent Configuration, on page 11
- Finesse Configuration, on page 11
- Agent PG Configuration, on page 12
- CUIC Configuration, on page 12
- Common Configuration Safeguards, on page 12

Note The guidelines for optimal failover ensure that desktop initialization time and general system performance is optimized.

Browser Configuration

Finesse browser failover performance depends on the number of requests made to the Finesse server. Fewer the requests, lesser the system load on the Finesse server. The following browser-specific configurations ensure that the browser does not fetch static resources unnecessarily from the server, and it is a key requirement for faster failover.

Note Clear the browser cookies before logging in to the Finesse desktop. This avoids unexpected expiry of the Refresh Token in the Single Sign-On mode for Unified CCE.

- Avoid loading the Finesse desktop with **`bypassServerCache=true&nocache`** as a query parameter in the desktop URL. The **`bypassServerCache`** is to bypass Webproxy cache, and **`nocache`** is to bypass Shindig cache.
- Host systems must have at least 200 MB of free disk space more than the free space required by the operating system (OS).
- Adequate network bandwidth must be available between the Finesse desktop and the Finesse server. Lower latency results in faster failover.

  For more information on bandwidth measurements, see *Finesse Bandwidth Calculator for Unified Contact Center Enterprise* and *Cisco Unified Contact Center Express Bandwidth Calculator* at https://www.cisco.com/c/en/us/support/customer-collaboration/finesse/products-technical-reference-list.html.

- Host systems must have adequate memory and CPU without being overloaded at any point in time. A slow host slows the browser enough to cause it to fail and reload resources randomly during failover.
- External gadget hosting servers must prefer CA-signed certificates for easy integration with the browser. If they are self-signed, then import those certificates into the agent browser.

  For more information, see *Accept Security Certificates* section in *Cisco Finesse Agent and Supervisor Desktop User Guide* at https://www.cisco.com/c/en/us/support/customer-collaboration/finesse/products-user-guide-list.html.

Firefox Configurations

Disable the Race Cache With Network (RCWN) in all the agent desktops to avoid any unwanted requests to the Finesse server. If RCWN is enabled, the Firefox browser by-passes the cached data and fetches the static requests again from the server. Set the **network.http.rcwn.enabled** configuration as **`false`**.

For more information, see https://support.mozilla.org/en-US/questions/1267945.

Google Chrome, Internet Explorer, and Edge Chromium (Microsoft Edge) Configurations

Import the Finesse self-signed certificates on Google Chrome, Internet Explorer, and Microsoft Edge browsers trust store.

For more information, see the *Accept Security Certificates* section in *Cisco Finesse Agent and Supervisor Desktop User Guide for Cisco Unified Contact Center Express* at https://www.cisco.com/c/en/us/support/customer-collaboration/unified-contact-center-express/products-user-guide-list.html.

Agent Configuration

Agents configured must be evenly distributed between the publisher and subscriber. This prevents all agents from failing over when there is an outage that affects only one of the deployed Finesse server.

The number of agents failing over impacts system load and has a linear relationship with the maximum time taken for the operation to complete.

Finesse Configuration

The number of signed-in users, the gadget types, and the average number of gadgets configured per user, significantly impacts failover load.

The following are the best practices for ensuring a trouble-free failover.

- Number of Gadgets per Agent—Gadget-initiated requests constitute the bulk of the requests made during Finesse desktop failover or startup. Configuring fewer gadgets in the desktop layout results in faster desktop failover and startup. The administrator must configure the team or desktop layouts such that only the required gadgets for each team are available in the desktop layout.
- Type of Gadget—XML-based gadgets load much faster than gadgets served using a dynamic JSP-based URL. The gadget content is also cached at the WebProxy Service, which allows the Finesse server to scale further. The JSP-based gadgets take thrice the time to load than the XML-based gadgets.

  The Unified CCE deployments must ensure that Cisco Unified Intelligence Center (CUIC) 12.5(1) servers are deployed, and CUIC JSP-based URLs are replaced using the CLI **utils finesse layout updateCuicGadgetUrl** to reduce latency and improve performance.
- Finesse Server Capacity—Deployments with 2000 active users and configuring more than eight XML-gadgets per user on average, or more than six JSP-based gadgets per user on average, are recommended to deploy Cisco Finesse OVA with 8 vCPUs.

When OVA with 8 vCPUs is configured for Finesse, the time taken for CTI server/Agent PG failover and desktop failover improves by 20 percent. This configuration is supported on all deployment types including the 24000 Agent deployment type. For more information on OVA with 8 vCPUs, see *Virtualization for Cisco Finesse* at https://www.cisco.com/c/dam/en/us/td/docs/voice_ip_comm/uc_system/virtualization/virtualization-cisco-finesse.html.

- Gadget Configuration—Gadgets developers must follow certain best practices to ensure that gadgets load faster.

  For more information, see *Best Practices for Gadget Development* section of *Cisco Finesse Web Services Developer Guide* at https://developer.cisco.com/docs/finesse/#best-practices-for-gadget-development.

- Secondary Resources—The preloading of server resources reduces latency and improves performance during desktop failover. By default, the **preLoadSecondaryResources** property is enabled. Disabling this property impacts failover time. For more information, see *Desktop Properties*.

Agent PG Configuration

Upgrading to the latest Agent PG version allows faster detection of failure, which results in faster failover. When Cisco Finesse connects to Agent PG 11.6(x) and 12.0(x) versions, it incurs a delay of up to 40 seconds to detect CTI failover compared to Agent PG 12.5(1) or later.

CUIC Configuration

CUIC 12.5 (1) or later supports only XML-based gadgets corresponding to the JSP-based gadget URLs that supported the real-time reports.

Cisco Finesse 12.5 (1) provides CLI **utils finesse layout updateCuicGadgetUrl**, which automatically changes the JSP references to XML with no functional changes.

Switching to XML-based gadgets reduces latency and improves performance. After the CUIC or Coresident deployment installation, run the command to optimize the faster failover.

Common Configuration Safeguards

- Import the self-signed certificates into the browser.
- Do not disable browser caching for Finesse desktop.
- Do not clear the cache every time the browser is launched.
- Distribute the agents between the publisher and subscriber.

Failover Planning

CTI Failover

CTI failover happens when the Finesse server disconnects from the CTI server/Agent PG due to network failure or server error. In these scenarios, the Finesse server is unavailable to its clients. If the desktop is connected, it displays that the server is unavailable and tries to reconnect to the available Finesse server.

The duration required for Finesse CTI failover depends on the following factors.

- Agent PG 12.5(1) or later versions results in faster failover.

- Available bandwidth from the CTI server to the Finesse server.
- Round Trip Time (RTT) between the CTI server and the Finesse server in case of WAN deployments.
- The number of signed-in users.
- The number of gadgets configured.
- The number of active non-voice tasks.

When deployed with Agent PG 12.5(1) or later, CTI server/Agent PG failover varies approximately from 35 seconds to 75 seconds. When deployed with Agent PG 12.0(1) or 11.6(1), CTI server/Agent PG failover varies approximately from 75 seconds to 120 seconds. The numbers indicated varies depending on the customer configuration and the above-stated factors. It must be used as a guideline to determine the approximate range for failover time.

Note The time indicated does not include CTI server/Agent PG server recovery time. It only indicates the time taken for Finesse to reconnect and be IN_SERVICE, once an active Agent PG is detected.

Desktop Failover

The Finesse desktop failover happens in all failure scenarios. The Finesse desktop tries to find an active server and fails over to it, once it has located a reachable server which is IN_SERVICE.

The duration required for Finesse desktop failover depends on the following factors.

- Bandwidth available to the client to reach the Finesse server.
- RTT between the client and the Finesse server.
- The number of signed-in users.
- The number of gadgets configured in the desktop per user.
- Type of gadgets (XML) and the resources it loads. For more information on Finesse gadgets, see https://developer.cisco.com/docs/finesse/#finesse-gadgets.
- The number of vCPU configured on the Finesse server.
- Agent PG version.
- The time required for the upstream CTI and the Finesse servers to become reachable and be IN_SERVICE.

The desktop failover performance improvements in Cisco Finesse 12.5(1) are available in all releases irrespective of the Agent PG version.

When deployed with Agent PG 12.5(1) or later, desktop failover with the default desktop layout varies approximately from 50 seconds to 110 seconds. When deployed with Agent PG 12.0(1) or 11.6(1), desktop failover is approximately 40 seconds more when compared to Agent PG 12.5(1) or later.

The numbers indicated varies depending on the customer configuration and the above-stated factors. It must be used as a guideline to determine the approximate range for failover time.

The following table displays the time taken for failover for different configurations (illustration purpose only) using the out-of-the-box gadgets with Agent PG 12.5(1).

Affected Users	Number of Gadgets (XML)	LAN or WAN	Time Taken (seconds)
1000	6	LAN	50-55
2000	6	LAN	70-80
2000	6	WAN (RTT of 400 milliseconds)	80-110

Note

- The time indicated does not include the Finesse or CTI server recovery time. It only indicates the time taken for the desktop to reconnect all agents post detecting a reachable Finesse server, which is IN_SERVICE.
- Deployments with 2000 active users and configuring more than eight XML-gadgets per user on average, or more than six JSP-based gadgets per user on average, are recommended to deploy Cisco Finesse OVA with 8 vCPUs.
- During failover, agents are redirected to the subscriber and are signed in automatically, and desktop is reloaded. Expected bandwidth utilization reaches up to approximately 250 Mbps for 90 seconds (peak), to ensure all 2000 agents failover successfully from one side to another. The bandwidth requirements increase depending on the type and number of gadgets configured for teams.