Technical Bulletin 5844

SIP Server Fallback Enhancements on Polycom® SoundPoint® IP, SoundStation® IP, and VVX® Phones

This technical bulletin provides detailed information on how the SIP software has been enhanced to support SIP server fallback.

This information applies to SoundPoint IP, SoundStation IP, and VVX phones running SIP software version 2.1 or later.

This technical bulletin is up-to-date for UC Software 4.0.x.

Introduction

Server redundancy is often required in VoIP deployments to ensure continuity of phone service for events where the call server needs to be taken offline for maintenance, the server fails, or the connection from the phone to the server fails.

Two types of redundancy are possible. In some cases, a combination of the two may be used:

- **Fail-over** - In this mode, the full phone system functionality is preserved by having a second equivalent capability call server take over from the one that has gone down/off-line. This mode of operation should be done using DNS mechanisms or IP Address Moving from the primary to the back-up server.

Your SIP server provider should be consulted for recommended methods of configuring phones and servers for fail-over configuration.

- **Fallback -** In this mode, a second less featured call server (router or gateway device) with SIP capability takes over call control to provide basic calling capability, but without some of the richer features offered by the primary call server (for example, shared lines, presence, and Message Waiting Indicator). Polycom phones support configuration of multiple servers per SIP registration for this purpose.

The topics include:

Prior to SIP 2.1, the `reg.x.server.y` parameters (refer to <reg/> in Appendix A of the *Administrator's Guide for the Polycom UC Software, available at* http://www.polycom.com/support/voice) could be used for fail-over configuration. The older behavior is no longer supported. Customers that are using the `reg.x.server.y`. configuration parameters where y>=2 should take care to ensure that their current deployments are not adversely affected. For example the phone will only support advanced SIP features such as shared lines, missed calls, presence with the primary server (y=1).

You may find it useful to also read *Engineering Advisory 66546: Configuring Optional Re-Registration on Failover Behavior* at http://supportdocs.polycom.com/PolycomService/support/global/documents/support/technical/products/voice/Configuring_Optional.pdf .

Terminology

The following terms may assist in understanding the SIP server fallback feature:

- **SIP Registrations** - SoundPoint IP, SoundStation IP, and VVX 1500 phones support the ability to have multiple SIP Registrations per phone. This is often used to support multiple *Lines* on a single phone and normally the SIP server(s) used for each Registration are the same. However, they could be different.
- **Primary and Fallback Servers -** Each of these SIP Registrations may be configured for concurrent registration with multiple servers for fallback purposes. For example, a phone may be configured to have two SIP Registrations and each SIP Registration may be configured with two separate servers (a primary server and a fallback server). DNS mechanisms (as described in RFC 3263) may be used such that the servers are capable of resolving to multiple physical SIP servers for fail-over purposes.

The primary server is the only one that will be used for advanced SIP features such as shared lines, message waiting indicators, and presence. This is a change in behavior from software releases before SIP 2.1 All other configured servers are referred to as fallback servers.

- **Working Server** - The phone maintains a list of all possible servers gained from both DNS and configuration. The highest priority server which has an active registration is treated as the working server and will be the first server tried for call initiation purposes. At any time, there is only one working server recognized by the phone.

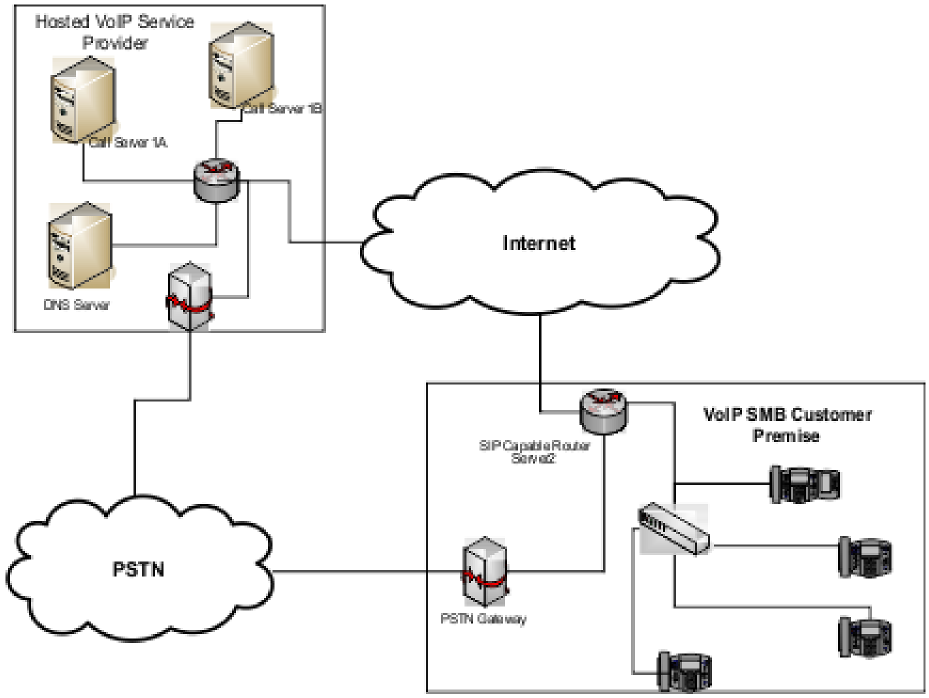
- **Registrar Server** - Servers (both primary and fallback) may be configured with registration enabled or disabled using the `reg.x.server.y.register` configuration parameter. Servers that have this parameter enabled will attempt registrations and are termed a registrar server. If a server is not a registrar server, calls will be attempted on that server if appropriate, but registration will not be attempted. Only a registrar server can become the working server.

SIP 2.1 Server Fallback Implementation

In the SIP 2.1 release, the redundancy behavior of SoundPoint IP, SoundStation IP, and VVX 1500 phones was changed and improved by adding the ability for a single SIP Registration (Line) to be registered to more than one server concurrently. In previous releases, the phone would only maintain one active server registration per SIP Registration (Line). The concurrent server registration capability adds an ability to do a faster and more efficient hand-over to an independent call server both for incoming as well as outgoing calls.

To assist in explaining the redundancy behavior, an illustrative example of how a system may be deployed is shown next.

In the example, a small medium business (SMB) customer uses a hosted IP-Centrex service from a Service provider. The Service provider has two redundant call servers at their network operations center (NOC) and uses a DNS server to resolve the IP addresses of these servers. The SMB customer also has an on-premise router which has the ability to handle SIP call traffic and has a connection to an on-site PSTN gateway. This gateway is intended to be used in conditions in which the Internet connection to the service provider is not working.



Phone Configuration

The phones at the customer site are configured as follows:

- Server 1 (the primary server) will be configured with the address of the service provider call server. The IP address of the server(s) to be used will be provided by the DNS server. For example:

  ```
  reg.1.server.1.address="voipserver.serviceprovider.com"
  ```

- Server 2 (the fallback server) will be configured to the address of the router/gateway that provides the fallback telephony support and is on-site. For example:

  ```
  reg.1.server.2.address=172.23.0.1
  ```

It is possible to configure the phone for more than two servers per registration, but you need to exercise caution when doing this to ensure that the phone and network load generated by registration refresh of multiple registrations do not become excessive. This would be of particularly concern if a phone had multiple registrations with multiple servers per registration and it is expected that some of these servers will be unavailable.

Phone Operation for Registration

After the phone has booted up, it will register to all the servers that are configured.

Server 1 is the primary server and supports greater SIP functionality than any of servers. For example, SUBSCRIBE/NOTIFY services (used for features such as shared lines, presence, and BLF) will only be established with Server 1.

Upon registration timer expiry of each server registration, the phone will attempt to re-register. If this is unsuccessful, normal SIP re-registration behavior (typically at intervals of 30 to 60 seconds) will proceed and continue until the registration is successful (for example, when the Internet link is once again operational). While the primary server registration is unavailable, the next highest priority server in the list will serve as the working server. As soon as the primary server registration succeeds, it will return to being the working server.

If `reg.x.server.y.register` is set to 0, then phone will not register to that server. However, the INVITE will fail over to that server if all higher priority servers are down.

DNS SIP Server Name Resolution

If a DNS name is given for a priority / registrar address, the IP address(es) associated with that name will be discovered as specified in RFC 3263. If a port is given, the only lookup will be an A record.

If no port is given, NAPTR and SRV records will be tried, before falling back on A records if NAPTR and SRV records return no results. If no port is given, and none is found through DNS, port 5060 will be used.

Refer to http://www.ietf.org/rfc/rfc3263.txt for an example.

Behavior When the Primary Server Connection Fails

The SIP server fallback implementation if the connection to the primary server fails is different for calls depending on if they are outgoing or incoming.

For Outgoing Calls (INVITE Fallback)

When the user initiates a call, the phone will go through the following steps to connect the call:

1. Try to make the call using the working server.

2. If the working server does not respond correctly to the INVITE, then try and make a call using the next server in the list (even if there is no current registration with these servers). This could be the case if the Internet connection has gone down, but the registration to the working server has not yet expired.
3. If the second server is also unavailable, the phone will try all possible servers (even those not currently registered) until it either succeeds in making a call or exhausts the list at which point the call will fail.

At the start of a call, server availability is determined by SIP signaling failure. SIP signaling failure depends on the SIP protocol being used as described below:

- If TCP is used, then the signaling fails if the connection fails or the Send fails.
- If UDP is used, then the signaling fails if ICMP is detected or if the signal times out. If the signaling has been attempted with all servers in the list and this is the last server then the signaling fails after the complete UDP timeout defined in RFC 3261. If it is not the last server in the list, the maximum number of retries using the configurable retry timeout is used. For more information, refer to the <voIpProt/> and <reg/> sections in Appendix A of the *Administrator's Guide for the Polycom UC Software*.

> If DNS is used to resolve the address for Servers, the DNS server is unavailable, and the TTL for the DNS records has expired, the phone will attempt to contact the DNS server to resolve the address of all servers in its list *before* initiating a call. These attempts will timeout, but the timeout mechanism can cause long delays (for example, two minutes) before the phone call proceeds "using the working server". To mitigate this issue, long TTLs should be used. It is strongly recommended that an on-site DNS server is deployed as part of the redundancy solution.

For Incoming Calls (Incoming Call Fallback)

The primary call server can use mechanisms for detecting that the Internet connection is down and route incoming calls through the PSTN link to the back-up gateway/router on-site. Since the phone is simultaneously registered to both servers, it will receive calls through the gateway even if the primary registration has not expired. This is a key advantage of the new behavior introduced in SIP 2.1.

Changes From Previous Phone Behavior (Releases Before SIP 2.1)

Before SIP 2.1	In SIP 2.1 and later
A Line maintains only one server registration.	A Line will maintains registrations with all servers that are configured as registrar servers.
If two servers are configured (for example, if both `reg.1.server.1.address="server1"` `reg.1.server.2.address="server2"` are configured), the phone will initially register with Server1 as the working server. Phone calls will be placed and received through Server1 only. If the registration to Server1 fails or expires, then the phone will attempt to register with Server2. If this registration succeeds, then incoming calls will be received using this server. At this point, Server2 takes over as the working server.	If two servers are configured (for example, if both `reg.1.server.1.address="server1"` `reg.1.server.2.address="server2"` are configured), the phone will register with both Server1 as the working and Server2. Phone calls will be placed through Server1, but may be received through either Server1 or Server2. If the registration to Server1 fails or expires, then the Server2 will become the working server.
The phone will continually attempt registration using SIP registration protocols with Server1. At the point that this succeeds, the registration with Server2 will expire and Server1 will resume as the working server.	The phone will continually attempt to register with Server1 and, when this is successful, will switch back to using Server1 as the working server. The Server2 registration will be maintained.
The phone attempts to maintain full SIP functionality with each server, but it is questionable how effective this is.	Only basic SIP registration for INVITE functions is maintained with servers other than the primary server.

Recommended Practices for Server Fallback Deployments

The best method for ensuring optimum server redundancy is to deploy two identical call servers and use either DNS methods or IP Address Moving together with call server recommended practices for maintaining synchronization of records between the redundant servers. This is termed fail-over (refer to Introduction on page 1). Deployment varies dependent on the SIP call server being used. Consult your SIP call server supplier for recommended practices for fail-over configuration.

In situations where server redundancy for fall-back purpose is used, the following measures should be taken to optimize the effectiveness of the solution:

1. Deploy an on-site DNS server to avoid long call initiation delays that can result if the DNS server records expire.

2. Do not use OutBoundProxy configurations on the phone if the OutBoundProxy could be unreachable when the fallback occurs. SoundPoint IP, SoundStation IP, and VVX 1500 phones can only be configured with one OutBoundProxy per registration and all traffic for that registration will be routed through this proxy for all servers attached to that registration. If Server 2 is not accessible through the configured proxy, call signaling with Server 2 will fail.
3. Avoid using too many servers as part of the redundancy configuration as each registration will generate more traffic.
4. Educate users as to the features that will not be available when in *fallback* operating mode.

Configuration File Changes

Configuration changes can performed centrally at the boot server. You may specify global primary and fallback server configuration parameters in <voIpProt/>, or specify per registration primary and fallback server configuration parameters in <reg/> that override those in <voIpProt/>.

- The <voIpProt/> attribute, located in the **sip-interop.cfg** configuration file (**sip.cfg** for SIP software version 3.2.x or earlier), includes the following:

Attribute	Permitted Values	Interpretation
voIpProt.server.x.lcs	0 (default) 1	This attribute overrides the `voIpProt.SIP.lcs`. If set to 1, the proprietary *epid* parameter is added to the From field of all requests to support Microsoft Live Communications Server.

- The `<reg/>` attribute, located in the **reg-advanced.cfg** configuration file (**phone1.cfg** for SIP software version 3.2.x or earlier), includes the following:

Attribute	Permitted Values	Interpretation
reg.x.server.y.lcs	0 (default) 1	This attribute overrides the `reg.x.lcs`. If set to 1, the Microsoft Live Communications Server is supported for registration x.

Trademarks

Disclaimer

Limitation of Liability

Customer Feedback

We are striving to improve the quality of our documentation, and we appreciate your feedback. Please send your email to VoiceDocumentationFeedback@polycom.com.

POLYCOM | Support

Visit support.polycom.com for software downloads, product documents, product licenses, troubleshooting tips, service requests, and more.