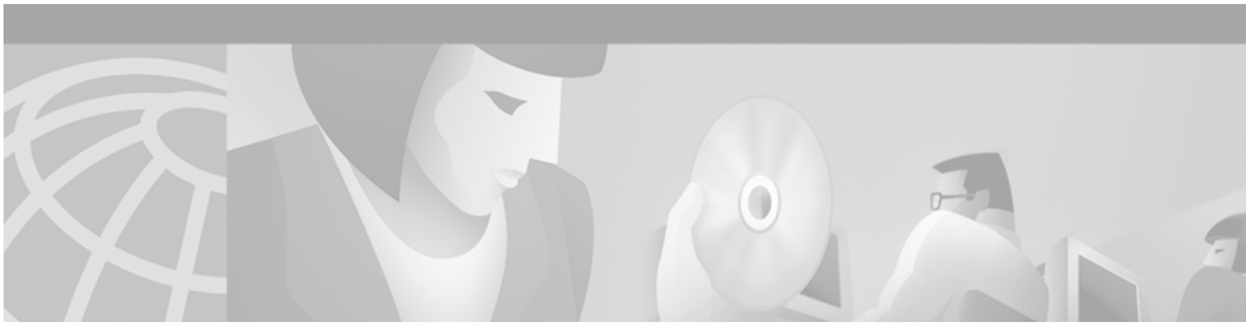


Unified Messaging

Version History

Version Number	Date	Notes
1	05/15/2001	This document was created.
2	06/27/2001	Incorporated editorial comments.

Unified Open Network Exchange (uOne) is an enhanced, IP-based software solution that gives subscribers the ability to receive voice mail, e-mail, and fax messages in a single mailbox that can be accessed via the phone or from a desktop browser or e-mail client.This document discusses various unified messaging concepts and features that apply to the Cisco uOne unified messaging (UM) solution. It also provides high-level examples showing how to deploy UM in different service provider and enterprise environments.

This document contains the following sections:

Unified Messaging Overview

Unified Communications (UC) integrates the two separate worlds of phone and Internet over a single unified network. As mentioned, Unified Open Network Exchange (uOne) is an enhanced, IP-based software solution that gives subscribers the ability to receive voice mail, e-mail, and fax messages in a single mailbox that can be accessed via the phone or from a desktop browser or e-mail client.

Unlike time division multiplex(TDM)-based proprietary messaging solutions, the Cisco UC platform is built on Open Packet Telephony (OPT), a Cisco standards-based, open-protocol voice and data architecture. The standards-based services platform is designed to carrier-class specifications, providing scalability to support millions of subscribers. It combines synchronous and asynchronous message types, including Voice over IP (VoIP), Internet fax, store and forward voice mail, and e-mail under a common message store and directory. This arrangement eliminates the need to synchronize disparate message stores and directories, such as different voice mail and e-mail systems, and dramatically reduces operational and maintenance costs. Competitive products that use old-world publich swtiched telephone networks (PSTNs) cannot offer this level of integration or scalability.

Unified Messaging Features

The features of a unified messaging solution are discussed in the following sections:

- Voice Messaging over IP, page 2
- E-Mail Messaging over IP, page 2
- Fax Messaging over IP, page 3
- Single Number Reach, page 3

Voice Messaging over IP

Voice messaging over IP allows service provider subscribers to check and access messages from any phone and to perform the following tasks:

- Create multiple personalized greetings programmed to play at different times, including times when the line is busy, when there is no answer, and when calls are received after the close of business.
- Place a new call or respond to a message without leaving the messaging system (known as the “Return Call” feature). This feature allows subscribers to respond to the message, forward it to someone else, or place a new call and return to the messaging system to continue processing other messages. All messages and calls can be handled with a single call.
- Leave messages for multiple recipients with a single call.
- Designate or prioritize messages so that subscribers can retrieve messages based on priorities.
- Locate a subscriber mailbox by name or telephone number.
- Forward voice messages as e-mail attachments to any e-mail user, enabling users of different voice-mail systems to share voice-mail messages.
- Receive message-waiting indication by pager, stutter dial tone or indicator light on telephone.

E-Mail Messaging over IP

E-mail messaging over IP allows subscribers to access e-mail messages from a phone and to perform the following tasks:

- Identify voice, e-mail, and fax messages in an e-mail inbox and save time by using one access device for all messages. Voice messages can be played as streaming audio or .wav files.
- Listen to e-mail messages from a telephone using the text to speech (TTS) feature.
- Respond to an e-mail message over the phone with an audio attachment.
- Receive paging notification on arrival of e-mail messages.

E-mail messaging over IP supports both Point of Presence (POP) and Internet Messaging Access Protocol (IMAP) clients.

Fax Messaging over IP

Fax messaging over IP allows subscribers to receive faxes anywhere by redirecting fax messages from their UM mailbox to a nearby fax machine. Fax messaging over IP also enables subscribers to perform the following tasks:

- Determine, by using their telephone, which faxes have arrived, the arrival time, and the identity of the sender.
- View faxes as .tif files from an e-mail client and save them in separate folders.
- Forward fax messages to other people as e-mail attachments.
- Receive immediate paging notification when fax messages arrive.
- Have greater privacy by printing faxes from their mailboxes when they are ready to view them.

Single Number Reach

The Single Number Reach feature improves accessibility by providing a single phone number that callers use to locate a subscriber in multiple locations. With Single Number Reach, callers can perform the following tasks:

- Use a single number to dial a subscriber's work phone, home phone, or wireless phone.
- Choose to either try to locate the subscriber or leave a message. Callers are not trapped in the system waiting for the subscriber to be located.

With Single Number Reach, subscribers can perform the following tasks:

- Decide whether to accept an incoming call or transfer it to voice mail. Callers are prompted to speak their name if they attempt to locate the subscriber. Subscribers can then choose to accept the call or transfer it to voice mail, depending on who is calling.
- Define different reach numbers for different time periods, such as business hours, nonbusiness hours, and holidays.
- Choose to be paged for incoming calls.

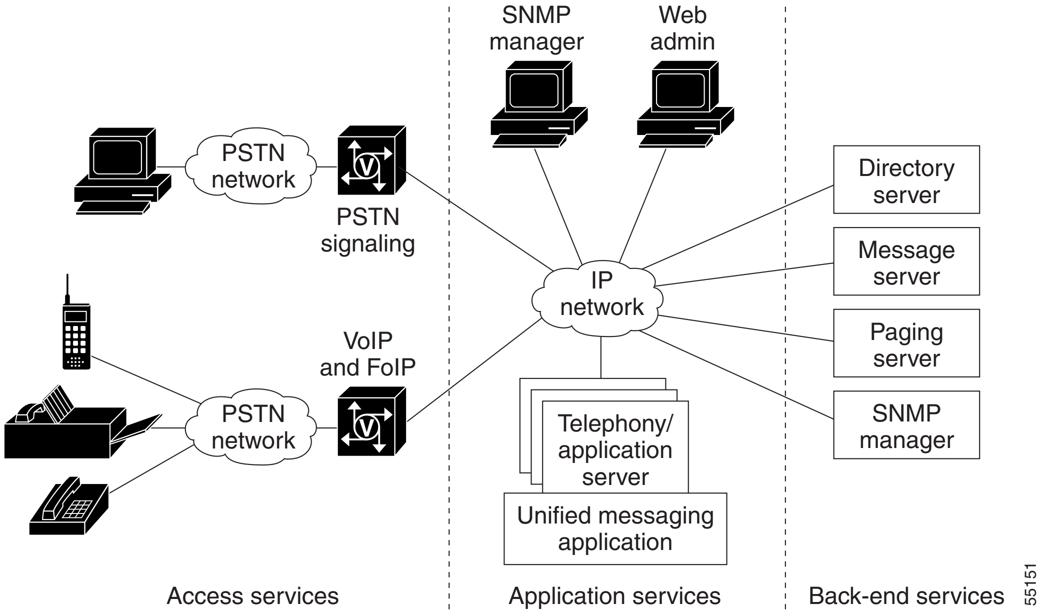
Components of a Unified Messaging System

uOne is built on a distributed agent platform (DAP), an open systems distributed computing environment that permits easy integration of new, nonproprietary voice and information processing technologies. DAP is based on a client/server model and consists of several components networked to provide all the functions of a unified messaging system.

The architecture is a distributed, object-based framework providing native support for all major industry standards such as Lightweight Directory Access Protocol (LDAP), IMAP4, Simple Mail Transport Protocol (SMTP)/Multipurpose Internet Mail Extensions (MIME), Voice Profile for INternet Messaging (VPIM), HTTP/HTML, and support for centralized Signaling Network Management Protocol (SNMP) management and web-based administration. The Cisco uOne applications reside on a gateserver that interfaces with the circuit-switched network through a Registration, Admission, and Status (RAS) gateway to any telephone, cellular phone, or fax machine. Gateserver applications then communicate over the IP network to directory services, media services, and management services. This arrangement allows uOne and other enhanced services applications to communicate with anyone, anywhere, using the IP network.

Figure 1 shows a complete unified messaging solution based on a three-tiered model of access services, application services, and back-end services.

Figure 1 The Unified Messaging Three-Tier Model



The following sections describe the components of a unified messaging system:

- Access Services, page 4
- Application Services, page 5
- LDAP Directory Services, page 6
- Messaging Server, page 8

Access Services

Access services provide access to application services and the front-end user interface of the unified messaging system. Subscribers can access messaging services with traditional telephony equipment, like phones and fax machines, or by workstations connected to an IP network. Access services provide call recognition and routing, media translation, and telco signaling.

Access services include the following components:

- PSTN and its components
- H.323 components like gateways (Cisco AS5000 series access routers) and gatekeepers

A gateway is an H.323 component that facilitates translation among various transmission formats and communication procedures (signaling). It is responsible for call setup and teardown on both the network and PSTN sides.

A gatekeeper provides call control services to the H.323 endpoints. The main functions of a gatekeeper in an H.323 network are as follows:

- Provide address translation between phone numbers and transport addresses.
- Authorize network access (admission control) using H.225 messages.
- Provide bandwidth control and zone management.

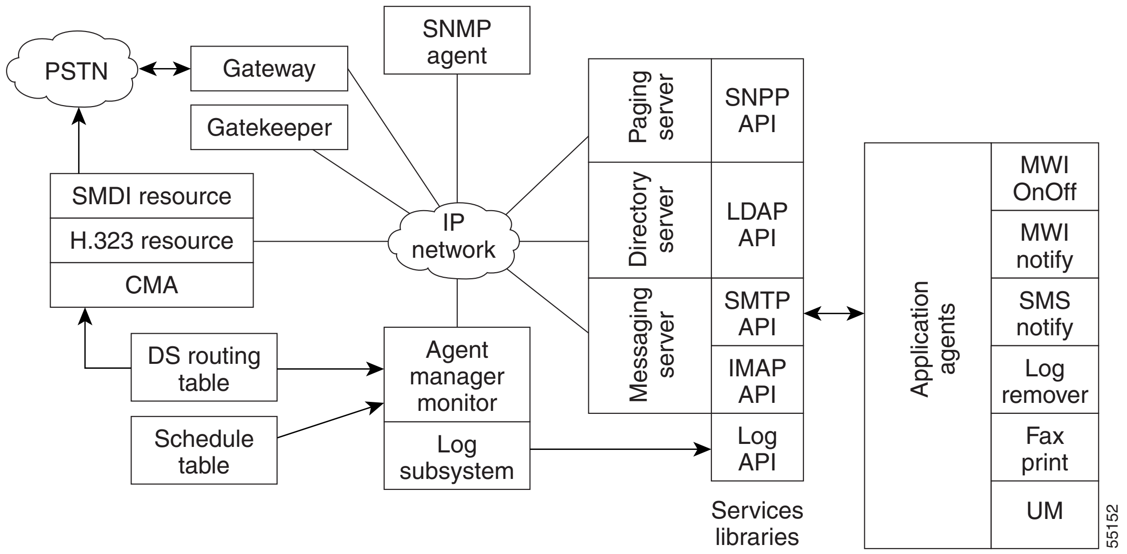
Application Services

Application services provide all the messaging logic required for the following features:

- Storing and retrieving messages (voice mail, e-mail, and fax)
- Translating among various message types
- User authentication
- Changing user profiles
- Message waiting indication
- SNMP services

Application services are the endpoints for all H.323 calls into and out of the unified messaging system. Figure 2 shows how application services are laid out.

Figure 2 Application Services



Application services have three major components: Agent Communications Broker, Call Control/Media Control Agent, and SNMP Agent.

The Agent Communications Broker (ACB) is a set of distributed software modules that provides communication services to other Directory Access Protolc (DAP) agents. The ACB includes the following features:

- The agent manager and monitor (AMM). The AMM provides scheduling, routing, launching, monitoring, and terminating services for all other DAP agent instances—it reads the schedule table, which contains information about how and when applications should be launched. Some applications are started as soon as the AMM starts; others are dynamically launched as needed, depending on a token that is passed to the AMM. Dynamic launches of agents and applications are usually triggered by external events such as an incoming call or a notification request.
- Schedule tables. The AMM provides scheduling, routing, launching, monitoring, and terminating services for all other DAP agent instances—it reads the schedule table, which contains information about how and when applications should be launched. Some applications are started as soon as the

AMM starts; others are dynamically launched as needed, depending on a token that is passed to the AMM. Dynamic launches of agents and applications are usually triggered by external events such as an incoming call or a notification request.

- Domain service routing tables. The AMM uses the information in the domain services routing table to bind agents to access specific services. The domain services routing table is used when messages need to be routed to other objects or application instances. Internal object routines use a token and the information in the domain services routing table to determine where to route the message. The AMM also monitors and manages agent instances for abnormal termination and state transition changes.
- A set of services libraries. Services libraries provide application programming interfaces (APIs) for various software services supported by unified messaging. These APIs are used to develop application agents, such as the UM and fax print agents.
- Local agent communications services (LACS). The LACS handle communications among all agents on a gateserver.
- Logging subsystems. A logging subsystem is also part of the ACB and provides centralized log management services to DAP agents.

The call control/media control agent (CMA) supports call control, media control, and the media resources. It uses H.323 call control signaling to accept, drop, and manage calls from an H.323 gateway or gatekeeper. It uses RAS to register with an H.323 gatekeeper, and it provides dual tone multi-frequency (DTMF) detection services. The CMA also provides media services such as playing, recording, and deleting messages.

The CMA and the ACB must reside on the same gateserver. The CMA uses dialed number identification service (DNIS) or redirected number (RDN) as the token to search the domain services routing table and determine which application will handle the request.

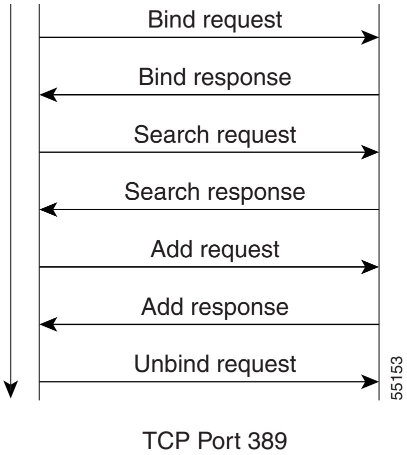
LDAP Directory Services

Lightweight directory access protocol (LDAP) is a directory service protocol that runs over TCP/IP. A directory is like a database but it usually contains more descriptive, attribute-based information. Directory information is generally read much more often than it is written. Consequently, directories do not usually use the same complicated transaction or roll-back schemes that regular databases do for high-volume, complex updates. Instead, directories are tuned to give quick response to high-volume lookup or search operations. They can replicate information widely and increase availability and reliability while reducing response time. The basic function of a directory service is to allow you to store and retrieve information about your enterprise or subscribers. You can retrieve the information by either directly searching for that information, or by searching for related, but more-easily remembered information, such as a name.

The LDAP directory service model is based on entries. An entry is a collection of attributes that has a name, called a distinguished name (DN), which is a unique reference for the entry. In LDAP, directory entries are arranged in a hierarchical tree-like structure that reflects, for example, political, geographic or organizational boundaries. LDAP defines operations for interrogating and updating entries in the directory—for adding and deleting entries from the directory, changing existing entries, and changing the names of entries. LDAP query requests permit a portion of the directory to be searched for entries that match certain criteria specified by a search filter. Information can be requested from each entry that matches the criteria.

LDAP is based on the client/server model and uses TCP as its transport protocol. One objective of LDAP is to minimize the complexity of clients, to facilitate large-scale deployment and hence scalability. Each Directory Server instance can support millions of entries and thousands of queries per second. By using replication and referrals, the Directory Server can be scaled to support even the largest of enterprises and subscriber bases, including multinational corporations and very large internet service providers (ISPs). Figure 3 shows a typical LDAP session call flow.

Figure 3 LDAP Session Call Flow



Unified messaging uses the directory server primarily to store and retrieve user profile information. You perform administrative tasks on the directory server by using vendor-supplied tools, like the Netscape console and admin server. UM subscribers interact with the directory server using Cisco web-based tools such as Personal Mailbox Administration (PMA), which permits subscribers to administer their personal preferences. Unified Messaging System Administration (UMSA) is the Cisco web-based tool that you can use to create new classes of service, add subscribers, manage broadcast lists, and manage user mailboxes.

You can use communities of interest (COI) as a mechanism to split a large directory into smaller, more manageable directories, each of which has its own access control and well-defined search base that restricts the view of the directory. COI usually defines a subscriber group that subscribes to a customized set of services under a single administrative authority. Service providers can use the same set of shared resources to create multiple communities of interest. The COI is based on the directory tree structure on the directory server and is defined by a specific node in the tree. Users within a COI have restricted visibility to everything below their node in the hierarchical directory tree structure.

Referrals in LDAP are a redirection mechanism that is used by a directory service to scale the service beyond the millions of users that can be supported with a single server. When an LDAP client queries a directory service and the query does not match any of the directory suffixes it supports, the server can return a referral to the client, requesting it to direct the query to a different LDAP server. Upon receipt of the referral, the client reformats the original LDAP request to fit the boundaries set by the referral, and reissues the request to the new server. Referrals are not returned if the directory names do not match, or if the client attempts to modify an entry that does not exist.

LDAP version 3 supports smart referrals, which allow you to map your directory entries to specific LDAP URLs. Smart referrals permit a directory server to refer the query to another server that services the same name space, or to refer it to a different name space within the same server. With smart referrals, if a client attempts to modify a directory entry and is referred elsewhere, the client will reformat the modification request to fit the boundaries set by the referral, and reissue the request to the new server. If the client has sufficient privileges, the operation is performed without the user ever knowing that the activity occurred on a remote server.

The Cisco unified messaging server uses LDAP version 2 APIs. The server does not process any LDAP referrals because there is no easy way to permit directory entry modification across multiple directory servers with referrals in LDAP version 2. The current version of UM will not support a model with multiple LDAP servers even though it is possible to handle directory changes by processing smart referrals. However, the unified messaging application is fully compatible with both versions of LDAP in the single directory service model.

Messaging Server

Messaging server is a messaging service that provides open, standards-based, flexible, cross-platform e-mail and messaging solutions, scalable to many thousands of simultaneous users. Messaging server provides the UM application with a common message store and allows access to its message store by standard Internet protocols such as IMAP4, POP3 and HTTP. Messaging server is an LDAP client, and uses the directory server as the centralized store for mail-user account storage, authentication, and mail routing control. The messaging server also provides a facility for specialized HTTP service for web-based e-mail. HTTP clients send mail to the specialized HTTP service, which then transfers the requests to a mail transfer agent.

The Cisco unified messaging application uses SMTP to store e-mail, voice mail, and fax mail messages on the common message store provided by the messaging server. The messages are stored in MIME format. UM subscribers use IMAP4, POP3, or HTTP to retrieve these messages from the message store.

Messaging servers use SMTP to accept and route messages. The following steps summarize how the messaging server accepts and routes a message:

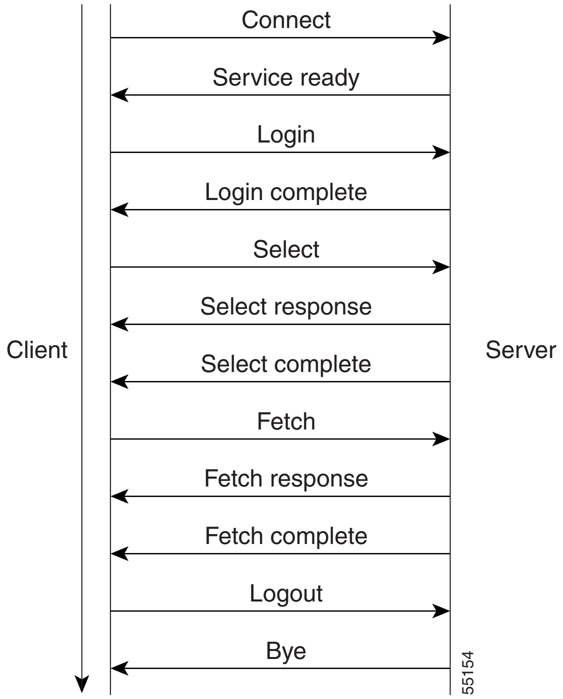
1. The messaging server queries the directory server (LDAP) to determine whether the recipient is local or remote.
2. If the recipient is local, the messaging server delivers the message, typically placing it in the message store.
3. If the recipient is remote, the messaging server performs the followign tasks:
   a. Queries the Domain Name System (DNS) to find the mail exchange (MX) servers for the remote domain.
   b. Queries DNS to find the IP address of the remote messaging server (resolves the server name from step a to an IP address).
   c. Establishes a TCP/IP connection to TCP port 25 of the remote messaging server.
   d. Optionally establishes a Secure Sockets Layer (SSL) connection to the remote messaging server.
   e. Sends the message to the remote messaging server (SMTP-Deliver).

To retrieve a message, the client must know the IP address of the messaging server, establish a connection to the server, then retrieve the message using one of the retrieval protocols: POP3, IMAP4, or HTTP. The following steps summarize how the client retrieves a message:

1. Queries DNS to find the IP address of the server.
2. Establishes a TCP/IP connection to the server.
3. Optionally establishes an SSL connection to the server.
4. Establishes a POP3, IMAP4, or HTTP connection to the server to retrieve the message.

uOne uses IMAP4 for storage and retrieval of messages from the messaging server. A typical IMAP session is summarized in Figure 4.

Figure 4 IMAP Session Call Flow



Typical uOne Call Flows

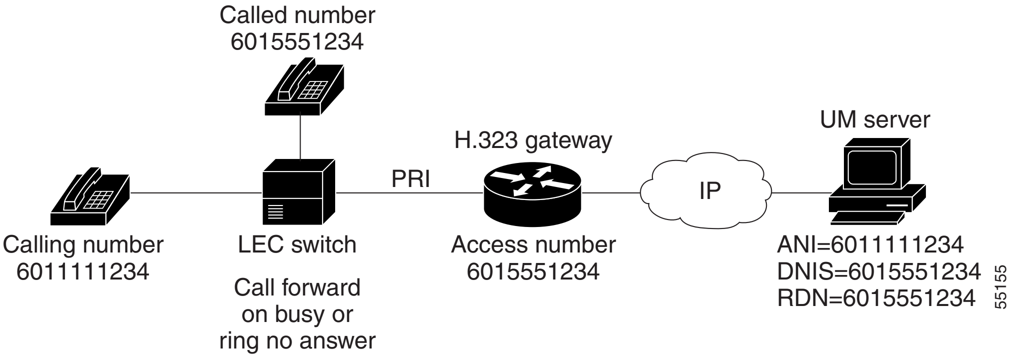
Call flows typical of uOne operation are described in the following sections:

- Subscriber Does Not Answer Call, page 9
- Caller Leaves a Message for a Subscriber, page 10
- Subscriber Is Notified to Retrieve Messages, page 12
- Subscriber Calls the UM Server to Retrieve Messages, page 13
- Inbound Fax Message to a Subscriber, page 16
- Printing a Fax Message from a Subscriber Mailbox to an Alternate Fax Number, page 17
- Overall uOne Protocol Flow Sequence, page 18

Subscriber Does Not Answer Call

When someone calls a subscriber and there is no answer, the call is forwarded to the gateserver. When the local exchange carrier (LEC) switch detects an incoming call that is destined for a busy or nonanswering party, the switch formulates a Q.931 setup message with the redirected number (RDN) field set to the original destination number, and sends it to the gateway. The called-party number of the setup message is set to one of the DNIS access numbers of the gateway. The original called number is then the RDN, and the number that was called to access the server is the DNIS. Whenever the RDN field is populated, the UM application uses it to retrieve (using LDAP) the subscriber profile from a directory server. If there is a matching subscriber, UM retrieves and plays the subscriber's personal greeting. Figure 5 shows an example of how this process works.

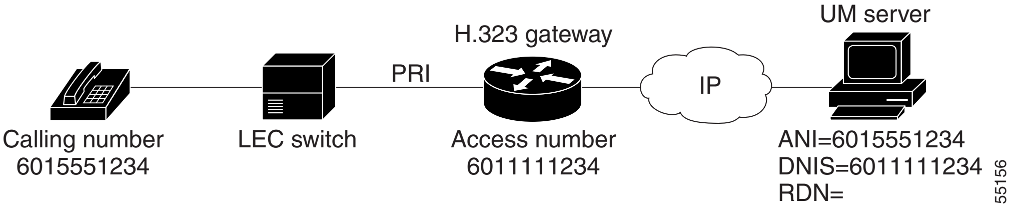
Figure 5 Retrieve Subscriber Personal Greeting



In this example, the presence of RDN indicates a call to the subscriber. The UM searches for the subscriber profile using 6015551234, retrieves it, and plays the personal greeting.

When a subscriber calls the UM server to access messages, automatic number identification (ANI) is set to the calling number, DNIS is set to the called number, and RDN is not populated. In this case, UM plays the general welcome message and requests the caller phone number and personal identification number (PIN). A subscriber calling from his or her own phone simply can press the pound (#) key, in which case UM uses the ANI to retrieve the subscriber profile, as shown in Figure 6.

Figure 6 ANI Profile Retrieval

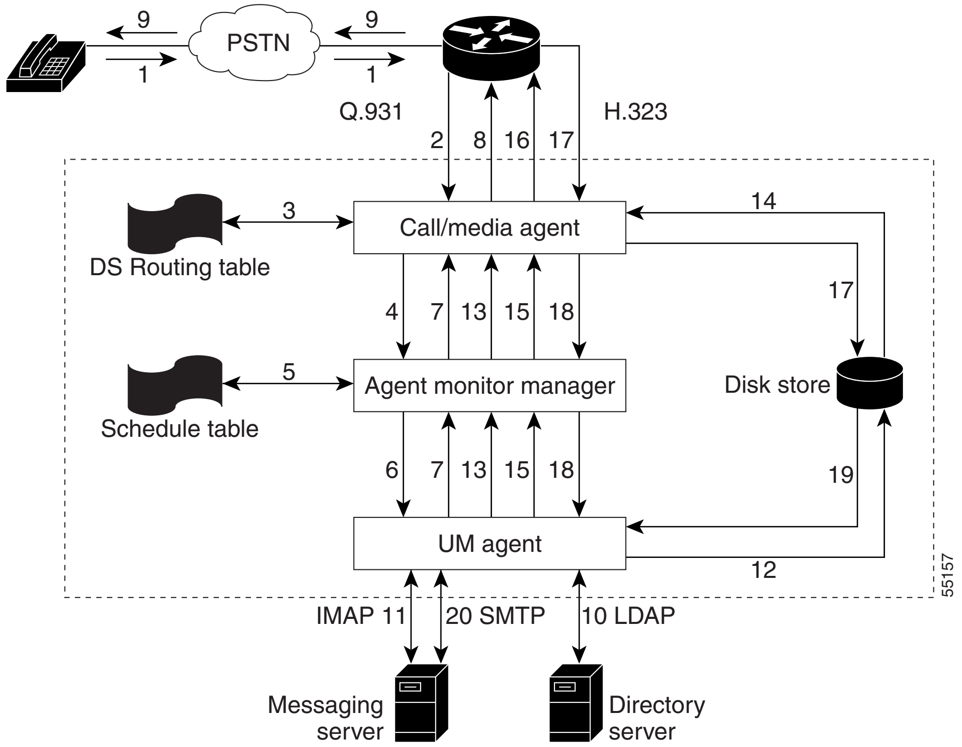


In this example, an unpopulated RDN field indicates a call from the subscriber to retrieve messages. The UM requests that the subscriber enter his or her phone number or simply press #. If the subscriber enters a phone number, it is used in a directory search (LDAP). If the subscriber enters #, 6015551234 is used to search the directory for his or her profile.

Caller Leaves a Message for a Subscriber

When someone calls a subscriber phone number and does not get an answer, the subscriber's switch forwards the call to the Cisco AS5300 gateway. Figure 7 shows how the messaging server accepts and routes a message to the subscriber.

Figure 7 User Calls and Leaves a Message



The following describes each step in the call flow diagram shown in Figure 7:

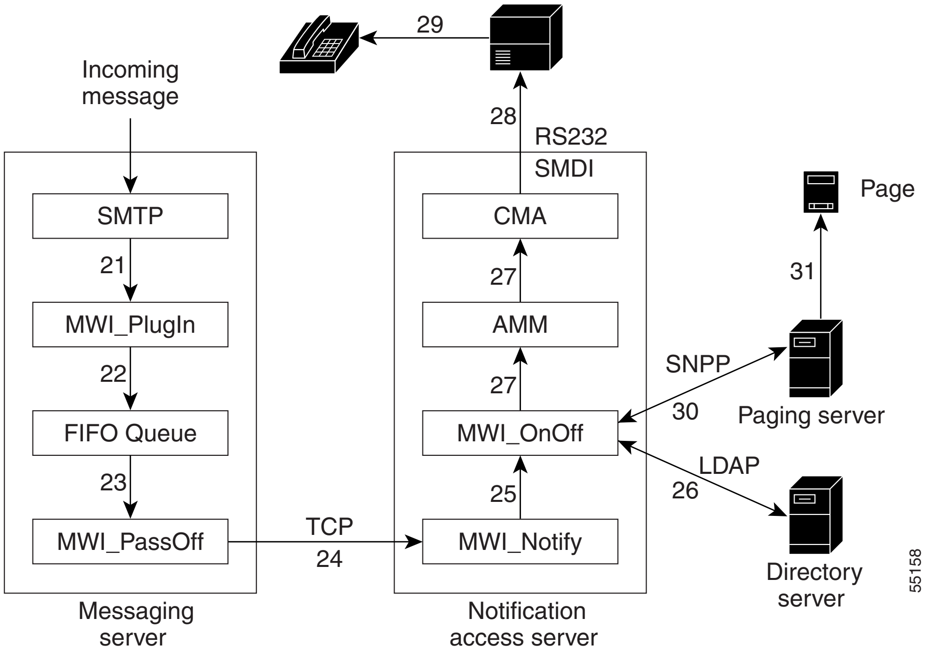
1. A caller makes a call to the subscriber phone number and does not get an answer. The call is forwarded across the PSTN to the gateway (Cisco AS5300). DNIS is the number that was called to reach the gateway, and the redirected number(RDN) is set to the original called number.
2. The gateway, based on its configuration (matching dial peers), selects the session target IP address as the call recipient. It sends an H.225 setup message to the target IP address.
3. The target IP address is that of the gateserver (CMA), which looks in its DS routing table to determine which AMM to contact.
4. The CMA then sends a “start app” command to the appropriate AMM.
5. The targeted AMM looks in the schedule table to determine which application agent to activate. In this case, the application agent is UM.
6. The AMM forks and executes a new UM process to handle this call. (An instance of the UM agent is executed for each incoming call.)
7. The new UM agent sends a message to the CMA via the AMM to accept the call.
8. The CMA sends an H.225 connect message to the gateway, requesting it to connect the call.
9. The gateway sends a Q.931 connect message to the PSTN and connects the call to the gateserver (CMA).
10. Using RDN, the UM agent gets the subscriber profile from the directory server and determines which greetings are active and their locations—on which messaging server they reside.
11. Subscriber greetings are stored as an audio file in an e-mail attachment in the e-mail account of the greeting administrator. The UM retrieves the greeting from the messaging server using IMAP.

12. The UM application detaches the greeting audio file and stores it on the file system.
13. The UM application provides a pointer to the greeting location on the file system and then issues a command to the CMA (via the AMM) to play the greeting.
14. The CMA loads the audio file from the file system and plays the greeting.
15. The UM application sends a message to the CMA to record a message from the caller.
16. The CMA plays the “beep” to start recording the caller message.
17. The caller leaves a message for the subscriber, which is stored by the CMA as an audio file on the file system.
18. The CMA uses the AMM to send a “record complete” notification to the UM application.
19. The UM application retrieves the message from the file system and, using the subscriber e-mail address, composes an e-mail message and attaches the audio file to it. While composing the e-mail message, the UM application sets the content-type attribute to voice mail, as specified in the Voice Profile for Internet Mail version 2 (VPIM v2) specification.
20. Using SMTP, the UM agent sends this e-mail to the subscriber messaging server. The messaging server deposits the message in the subscriber mailbox.

Subscriber Is Notified to Retrieve Messages

For a message waiting indicator or stutter dial tone, the gateserver must have an EIA/TIA-232 connection to a switch that has access to the telephone handset. For paging services, a Hylafax Simple Network Paging Protocol (SNPP) server must be installed and accessible to the gateserver. Figure 8 shows the subscriber notification process.

Figure 8 The Subscriber Notification Process



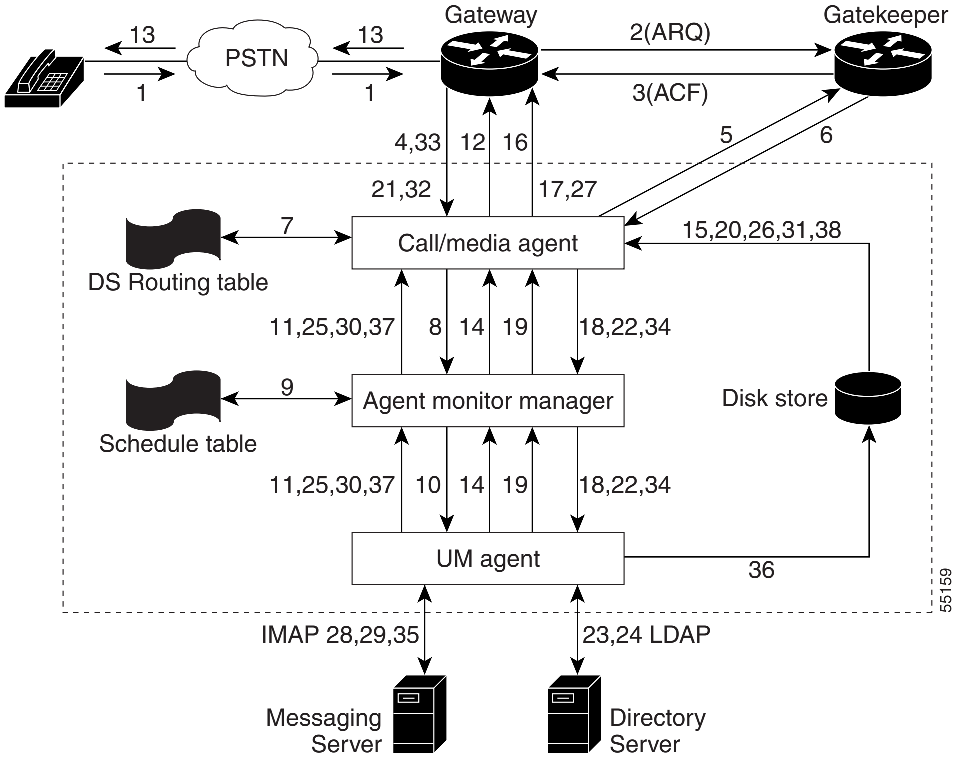
The following describes each step in the subscriber notification process flow diagram in Figure 8:

21. When the messaging server accepts a new message, it calls a configured message waiting indicator (MWI) plug-in. This plug-in must be installed as an additional messaging server component during installation.
22. The MWI plug-in inserts a notification message in a local queue (First in, first out (FIFO)).
23. MWI_PassOff monitors the queue and receives the notification request.
24. Using a TCP connection, MWI_PassOff forwards the notification request to the MWI_Notify process, which is resident on a notification access server. Typically, this is the UM server where the CMA and AMM components are running.
25. MWI_Notify receives the request and uses AMM to forward it to the MWI_OnOff process.
26. Using LDAP, MWI_OnOff retrieves the subscriber profile from the directory server and determines the type of notification to use for that subscriber.
27. If the subscriber notification requires an MWI light or dial tone stutter, the MWI_OnOff process issues a command to the CMA, using the AMM for Simplified Message Desk Interface (SMDI) signaling.
28. Using SMDI signaling, the CMA sends the appropriate notification message to the switch.
29. The switch turns on the stutter tone (by sending an SMDI message to the central office switch) or MWI light on the handset as appropriate.
30. If the subscriber has requested to be notified by a page, MWI_OnOff issues a command to the paging server using Simple Network Paging Protocol (SNPP). SNPP is an Internet standard (RFC 1861) for sending one-way or two-way wireless messages to pagers.
31. The paging server notifies the paging provider to send a page using Telocator Alphanumeric Protocol (TAP).

Subscriber Calls the UM Server to Retrieve Messages

After being notified by an MWI or a page, the subscriber can retrieve messages. Figure 9 shows how the subscriber retrieves messages.

Figure 9 Subscriber Calls to Retrieve Messages



The following describes each step in the message retrieval flow diagram shown in Figure 9:

1. The subscriber makes a call to access the UM server. DNIS is set to the called number, and ANI is set to the calling number (the phone number that the subscriber is calling from).
2. The gateway has a matching dial peer for the called number with the session target set to RAS. It sends an admission request (ARQ) to the gatekeeper.
3. The gatekeeper looks at all its registered gateways and, in an admission confirm message (ACF), returns the IP address of the gateway to which this call must be forwarded.
4. The target IP address is that of the gateserver (CMA) that is registered with the gatekeeper. The gateway sends an H.225 setup message to the gateserver.
5. The CMA sends an ARQ to the gatekeeper for permission to accept the call.
6. The CMA receives an ACF from the gatekeeper, permitting it to accept the call.
7. The CMA looks in its directory server routing table to determine which AMM to contact.
8. The CMA then sends a “start app” command to the appropriate AMM.
9. The targeted AMM looks in the schedule table to determine which application agent to activate. In this case, the application agent is UM.
10. The AMM forks and executes a new UM process to handle this call. An instance of the UM agent is executed for each incoming call.
11. The new UM agent sends a message to the CMA via the AMM to accept the call.
12. The CMA sends an H.225 connect message to the gateway, requesting it to connect the call.

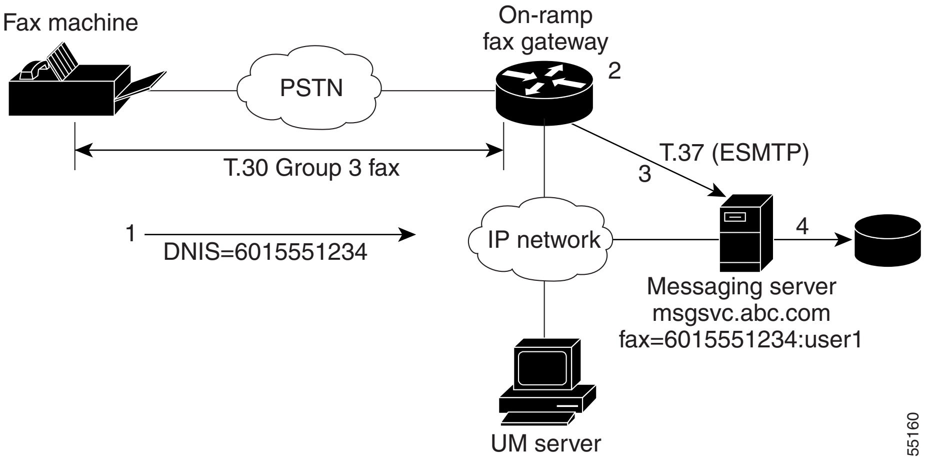
13. The gateway sends a Q.931 connect message to the PSTN, and connects the call to the UM server (CMA).
14. Because RDN is unpopulated, the UM agent sends a message to CMA to play the message that asks for the caller phone number, and collects the DTMF.
15. The CMA retrieves the message from the file system.
16. The CMA plays the message. The subscriber hears “Good morning, please enter your …”.
17. The subscriber enters a phone number followed by a #, or presses the # key if calling from his or her own phone, or does nothing (times out). DTMF is transported across the H.323 network to the CMA using the Cisco Real-Time Transport Protocol (RTP) encapsulation.
18. The CMA uses the AMM to pass DTMF information to the UM application.
19. The UM application sends a message to the CMA to play the message, prompt the user for a password, and collect the DTMF.
20. The CMA retrieves and plays the message.
21. The subscriber password is keyed in. The DTMF is transported to CMA using RTP encapsulation.
22. The CMA uses the AMM to pass the DTMF information to the UM application.
23. The UM application requests user profile information from the directory server. The subscriber profile is retrieved using the keyed-in phone number, or the ANI (calling number) if the caller simply pressed #.
24. The directory server returns the entire profile and authentication to the UM application. The UM application verifies the caller as a valid subscriber.
25. The UM application sends a “Play prompt” message to the CMA via the AMM.
26. The CMA retrieves the welcome-message.wav file from disk storage.
27. The CMA plays the prompt to the caller and the caller hears the welcome message.
28. The UM agent determines the messaging server for the subscriber (based on the messaging server host name specified in the subscriber profile) and sets up an IMAP4 connection to it using information from the subscriber profile.
29. The UM application retrieves the message headers and inventories the subscriber mailbox.
30. If the subscriber has urgent messages, the UM application passes the urgent messages as a .wav file to the CMA via AMM. If there are no urgent messages, the UM application sends the inventory prompt command to the CMA.
31. The CMA retrieves the prompt from the file system.
32. The CMA plays the prompt. The caller hears something like “You have one voice message and three e-mail messages…”.
33. The subscriber enters a “1” to retrieve the messages.
34. The digit is collected by the gateway and sent to the CMA, which uses the AMM to pass it on to the UM application.
35. Using IMAP, the UM application retrieves any urgent messages for the subscriber from the subscriber messaging server.
36. Depending on whether headers are on or off in the subscriber profile, the UM application retrieves and stores just the message body .wav file or both the message body and header .wav files.
37. The UM application sends a command to the CMA to play the audio files.
38. The CMA retrieves the .wav files from the file system and plays them.

Inbound Fax Message to a Subscriber

The gateserver does not participate in handling incoming fax messages. When a fax account is created on the UM server, it creates an alias file on the messaging server, where it maps the subscriber fax number to his or her e-mail address. This alias is used in Step 4 of the fax delivery process described later in this section.

With the Store and Forward Fax feature, the Cisco AS5300 acts as an on-ramp gateway, which receives faxes from end users and converts them into Tag Image File Format-Fax (TIFF-F) files. It attaches this TIFF-F file to a MIME e-mail message and forwards it to a designated SMTP server where the e-mail is stored. Figure 10 shows how the fax delivery process works.

Figure 10 The Fax Delivery Process



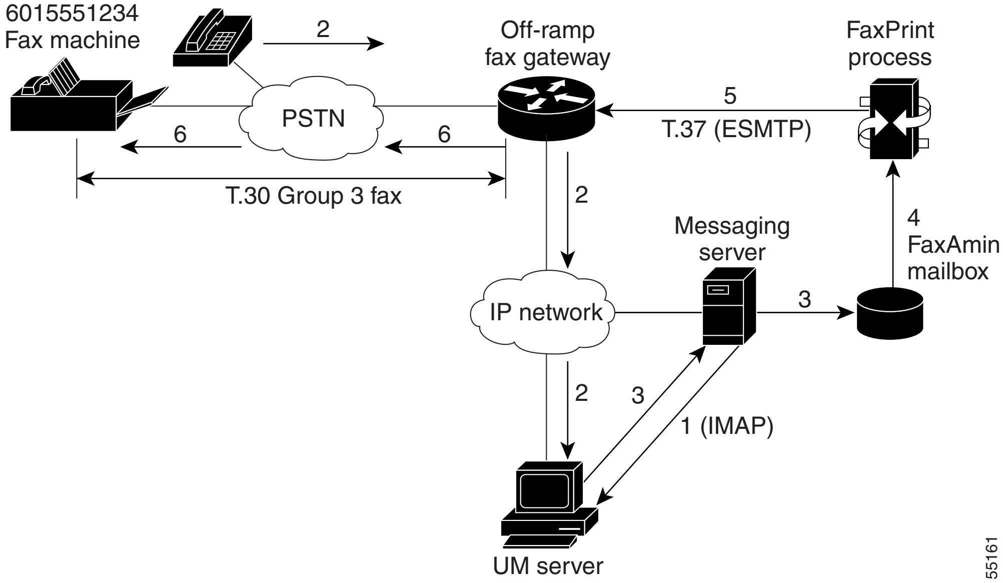
The following describes each step in the fax delivery process flow diagram shown in Figure 10:

1. A fax is sent to the subscriber telephone number (6015551234). The fax machine connects to a fax gateway (Cisco AS5300).
2. The fax gateway receives the call. The incoming call is determined to be a fax call because the DNIS matches a fax inbound dial peer (dial-peer voice 1 mmoip). The gateway converts the T.30 Group 3 fax to a .tif file. Because the dial peer that it matches identifies the call as a fax or a voice call, two separate numbers need to be used for fax and voice mail for each subscriber.
3. The gateway creates a mail message, attaches the .tif file, and delivers it to the messaging server using Extended SMTP (ESMTP). The session target statement under the fax dial peer determines the delivery e-mail address. The statement “session target mailto:$d$@mailserver.com” sets the destination e-mail address to <DNIS>@mailserver.com. In this case, the destination e-mail address is set to fax=6015551234@msgsvc.abc.com. The **mta send server msgsvc.abc.com** command specifies the messaging server to which this e-mail with the .tif attachment is sent.
4. The messaging server contains a list of aliases that map phone numbers to valid e-mail addresses on the server. For example, fax=6015551234 is mapped to faxuser@msgsvc.abc.com. The server accepts the e-mail from the gateway, looks up the alias file, and deposits the fax in the subscriber mailbox. The receipt to e-mail address is the DNIS-based e-mail alias (fax=6015551234@msgsvc.abc.com). The recept to e-mail address enables the UM server to determine that this is a fax message when retrieving messages from the message store.

Printing a Fax Message from a Subscriber Mailbox to an Alternate Fax Number

After successfully logging in using a telephone, the subscriber can choose to retrieve faxes or e-mail messages containing faxes and redirect these messages to another fax number to be printed. Figure 11 shows how the subscriber retrieves fax messages by printing them to an alternate fax number.

Figure 11 Printing to an Alternate Fax Number



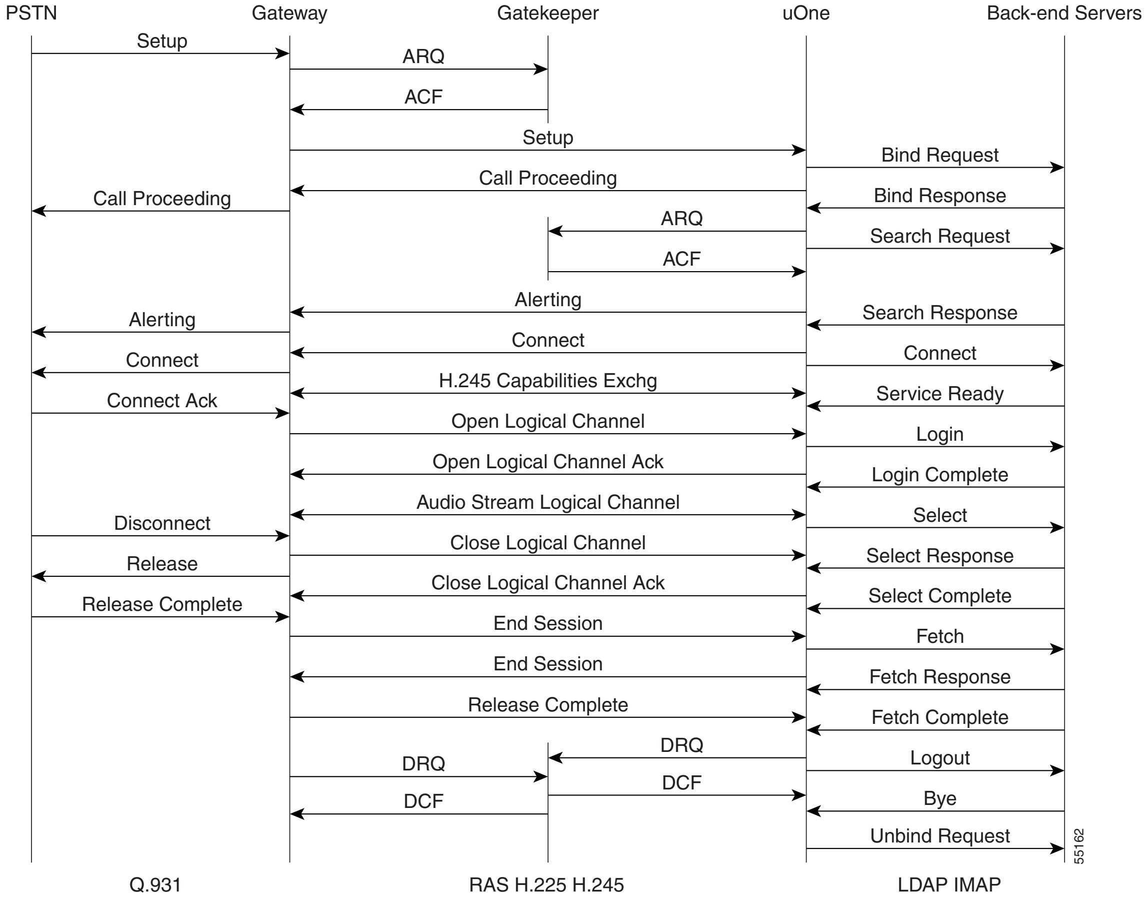
The following describes each step in the redirect fax printing process flow diagram shown in Figure 11:

1. The UM application uses the subscriber information from the directory server to log in to the subscriber mailbox, and uses IMAP to retrieve the fax or e-mail message from the messaging server.
2. The subscriber chooses the option to print the message (redirect it to a fax machine close by). The subscriber keys in the phone number of the fax machine where the message is to be sent—for example, 6015551234.
3. Every subscriber mailbox is associated with a faxadmin account. The UM application adds the destination fax information to the message and uses SMTP to forward it to the subscriber faxadmin e-mail account.
4. The FaxPrint application, which runs on the messaging server, constantly monitors the faxadmin mailbox for new messages. It uses IMAP to retrieve the message sent in the previous step.
5. The FaxPrint application addresses the message to the destination fax machine and sends the message to the off-ramp fax gateway using ESMTP (T.37). The faxprint.ini and dialmap.ini files define the gateway to use. The destination e-mail address would be fax=6015551234@gateway.abc.com.
6. The fax gateway extracts the destination phone number from this e-mail address, converts any text to .tif format, and sends the fax to the destination as a T.30 Group 3 fax.

Overall uOne Protocol Flow Sequence

Figure 12 summarizes the overall uOne protocol flow sequence.

Figure 12 Overall Protocol Flow Sequence



Deploying Unified Messaging Services in a Service Provider Environment

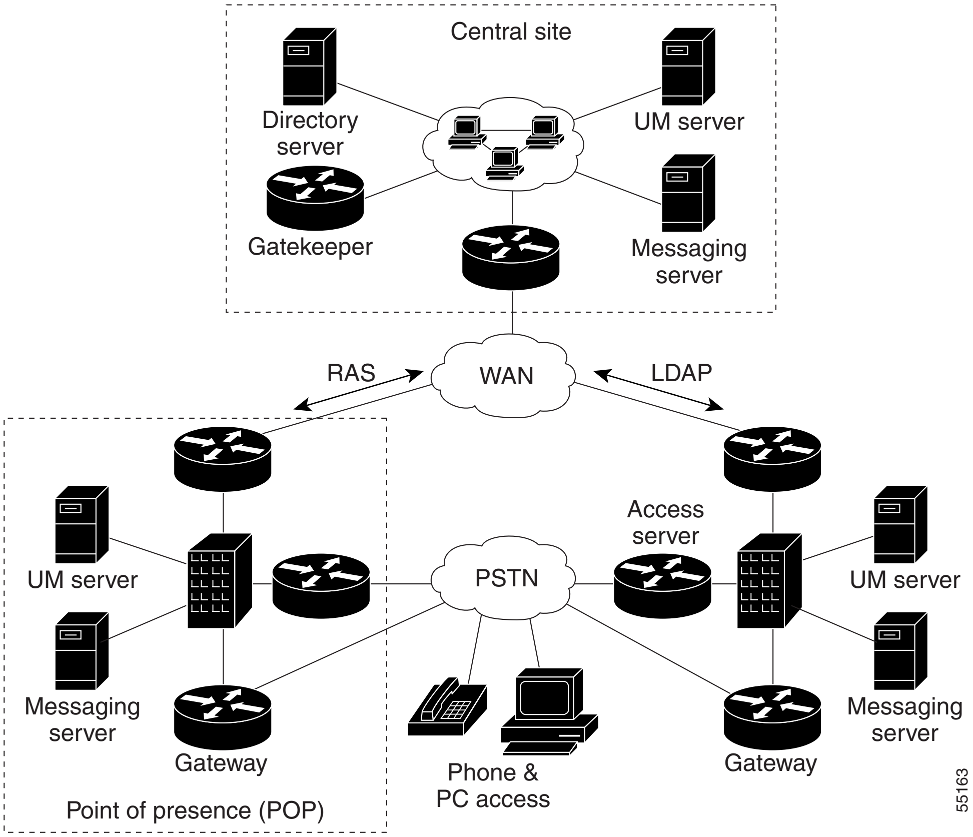
This section describes the general process of deploying unified messaging services in a service provider environment. This process typically includes the steps described in the following sections:

- Determine Optimal Design, page 20
- Create Multiple COIs, page 21
- Define Classes of Service, page 22
- Add Greeting and Fax Administrators, page 23
- Add Organizational Unit UMSA Administrators and Subscribers, page 24
- Deploy Fax Services, page 25

Service providers typically have a large set of users with varying requirements. They also provide Internet service to a number of small corporations. Unified messaging services can be deployed across the entire service provider network and sold at different levels to individual users and corporations. A typical deployment in a service provider environment would be decentralized (as shown in Figure 13) to provide for local-number access to services.

Ideally, there is one or more gateservers per point of presence (POP), with their own messaging servers connected locally. However, they all share a common centralized directory service. From a unified messaging point of view, a totally self-contained POP has a gateserver, a messaging server (for local message store), an H.323 gateway (for local access to the gateserver), and an access gateway that allows users to dial in to the service provider network.

Figure 13 ISP Deployment Scenario



In this scenario, LDAP and RAS are the only UM-related protocols that use the WAN. E-mail messages to the user are relayed (using SMTP) to messaging servers located at each POP. Depending on the subscriber base at each POP, multiple gateservers can share the same central messaging server or one that is located at one of the POP sites.

Initially, you can provide service to your subscribers at the central site, then add messaging servers and gateservers at POP installations as the subscriber base grows. If subscribers travel from one POP to another, they can still access their services with a local call. The local gateserver will be able to service all requests, but because the messaging serverof the subscribers is not local, they might notice a small degradation in service, depending on network bandwidth availability.

To subscribers who travel out of your service area, you can provide 800-number access to a gateway at the central site for an additional fee. In the scenario described, the distributed architecture allows any gateserver to service any subscriber because they all have access to a common directory server. The distributed architecture provides for complete redundancy and also helps with maintenance of the gateservers.

After the services are deployed, they can be used to support many different COIs, enabling you to sell different levels and classes of service to individual subscribers, corporations, and resellers.

To deploy unified messaging services in a service provider environment, perform the following tasks, which are described in the sections that follow:

- Determine where to place the uOne components for an optimal solution.
- Create multiple COIs.
- Define various classes of service (CoS).
- Add greeting and fax administrators.
- Add Unified Messaging System Administrators (UMSA) and subscribers.
- Deploy fax services.
- Plan for redundancy and load balancing.

Determine Optimal Design

A typical service provider services both individual subscribers (with dial Internet access) and corporations, with their own dedicated Internet access solutions. The decision where to place various components of a uOne solution depends on the subscriber base, the available bandwidth, and the quality of the unified messaging services offered. The various network components associated with a uOne solution affect service quality in different ways. The following sections describe the major components, their main functions, and how the components affect service quality.

Gateserver

In the unified communication solution, the gateserver is the termination point for an H.323 connection. Depending on its proximity to the H.323 gateway, and the available bandwidth between the gateserver and H.323 gateway, the gateserver affects call setup times and voice quality. Other factors that influence the performance of the gateserver are the number of simultaneous calls that can be handled, and available resources such as memory and CPU.

Directory Server

Directory services are used to authenticate, store, and retrieve subscriber profile information. Directory services directly influence authentication and message response times. Authentication time is the time a user must wait for the system to respond after a user ID and PIN have been entered. Directory services are also used to store subscriber mailbox and login information so that uOne can retrieve subscriber messages. Login information must be retrieved from the directory server to be able to log in to the messaging server and retrieve the message. Message response time is the time a subscriber must wait to hear the message after the message has been selected. We recommend that directory services be centralized and deployed at the core because all uOne servers in an ISP share the same directory.

Messaging Server

The messaging server is used to store and retrieve personal greetings, and voice, e-mail, and fax messages. It directly affects message response time and greeting response time. Greeting response time is the time the system takes to retrieve and play a personal greeting after it has determined which greeting to play. Other factors that can influence the performance of a messaging server are the size of the subscriber base that is served by the server, and available resources such as CPU and memory.

H.323 Gateway

The H.323 gateway serves as the protocol translator between the PSTN and the H.323 networks. Depending on the proximity of the unified messaging system to the subscriber base, subscribers might not have local number access to it. The gateway directly affects call setup times and voice quality.

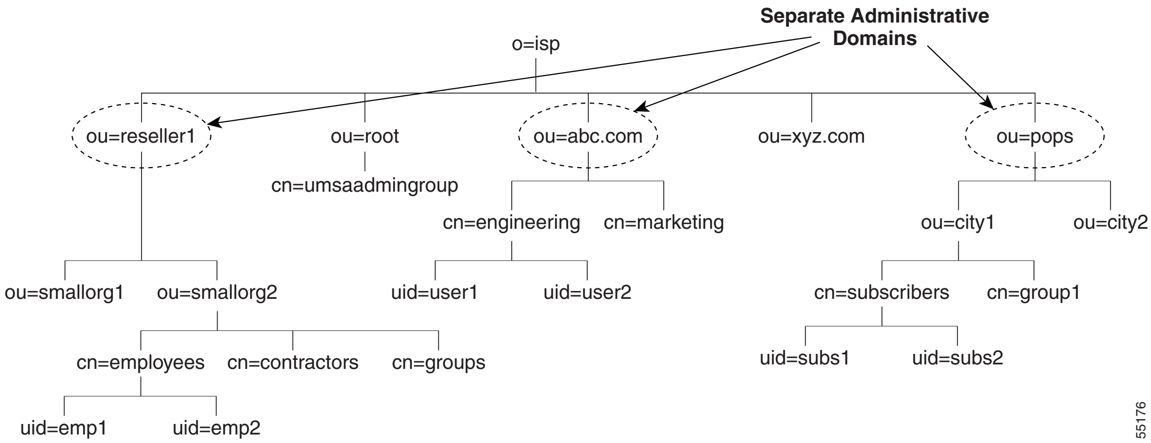
H.323 Gatekeeper

Primarily, the gatekeeper determines which gateserver will handle an incoming call. It has a direct influence on call setup times.

Create Multiple COIs

The concept of COI involves taking a large set of users and logically grouping them into smaller communities under a single administrative authority. As single administrative authority allows the same unified messaging service infrastructure to be used by multiple communities at the same time, and permits delegation of administrative tasks to the unified messaging administrators for that community. Every administrator can customize its own greetings, provide different classes of service, and perform other administrative tasks within their own COI. A COI translates to a point in the directory tree on a directory server. Figure 14 illustrates a sample directory tree for a typical service provider.

Figure 14 Sample Directory Tree



Data in a directory is hierarchical, and is represented as attribute-data pairs. The attributes used in this directory example are as follows:

- o: Organization name.
- ou: Organizational unit. This attribute is typically used to represent smaller divisions within your enterprise.
- cn: Group. “cn” stands for common name.
- uid: User ID.

The top level “o=isp” and the administrative account are created at installation time. The top-level administrator (admin) can create more organizations and organizational units, groups, and users. uOne requires an organizational unit called “root” and a group under “root” for Unified Messaging System Administration (UMSA) administrators. The administrator accesses the LDAP directory server using a web interface at http:\\directoryServer:2500.

Once logged in as the admin, you can create more organizational units and groups, as shown in Figure 14. You should refer to the directory server user guide for details on how to create additional organizational units, groups and users. Creating organizational units, groups, and a few sample users will create database entries with directory branch points.

If you export the directory, the resulting LDAP Data Interchange Format (LDIF) file will have the format of individual entries in the directory database. You can then uses this LDIF file as a template for creating a large number of entries in the directory. For example, you can use an existing subscriber database as a source to create a large number of directory entries by using a scripting language to automate the creation of the LDIF file, which can then be imported into the directory.

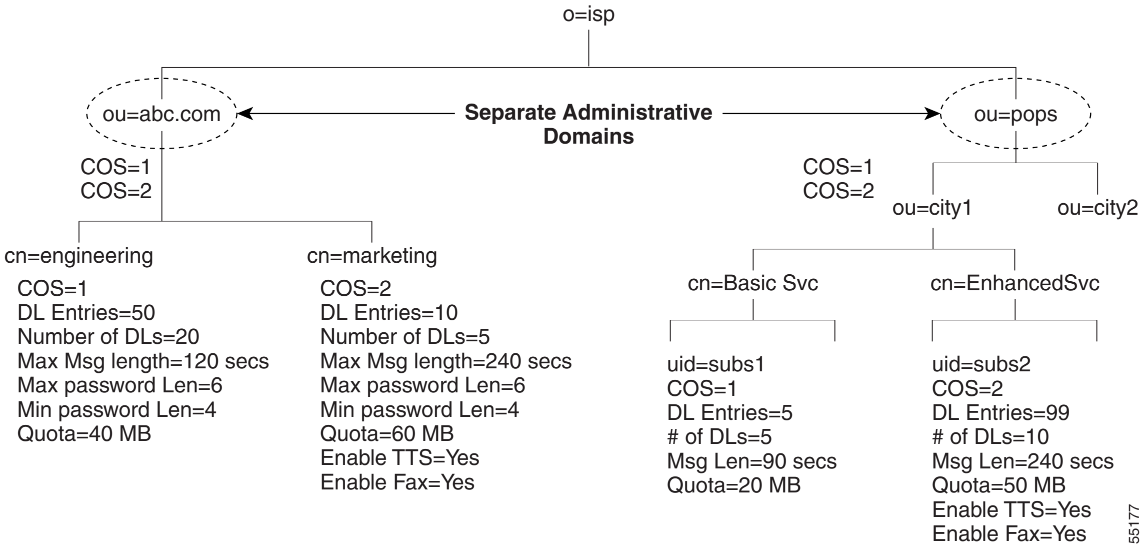
You can also use the bulk add tool that comes with the UM server to add a large number of entries to your directory. Once again, you should refer to the directory user and deployment guides that came with the directory server for details about directory design.

Define Classes of Service

A CoS defines a common set of unified communication services for a group of subscribers that is administered by a central authority. Subscriber groups use COSs to bundle various feature sets into distinct packages that facilitate administration of common features. A COS is unique within a COI and is identified by a number (for example, COS=1). COSs are defined by identifying sets of features that you can market as different levels of services to subscribers and resellers.

In the example in Figure 15, “DL” stands for directory listing, “DL Entries” specifies how many listings are permitted per list, and “Number of DLs” specifies the maximum number of lists that the user can create. Because a CoS is unique per organizational unit, it is possible to have the same COS number under different organizational units.

Figure 15 Classes of Service



Two CoSs are defined for subscribers at POP sites. Basic services include voice mail and e-mail, with each voice message restricted to a maximum length of 90 seconds. Enhanced services include basic services and permit fax and Text To Speech (TTS) services, and increase the maximum length of voice messages to 240 seconds. Also, enhanced services subscribers can create more and larger distribution lists.

Every organizational unit needs one COS defined for each feature set being offered to users. You can define COSs by using the web-based administration tool under "COS Administration." The entry in the "DN" field in "add a new COS" associates the COS with an organizational unit. In the Figure 15 example, the Distinguished Name (DN) entry for abc.com would be "ou=abc.com,o=isp." The complete entry is:

```
Entry DN: ou=abc.com, o=isp
Class of Service ID: 1
Class of Service Name: EngSvc
Search Base: ou=abc.com,o=isp
Personal Access UM Ini File Name: UM.ini
```

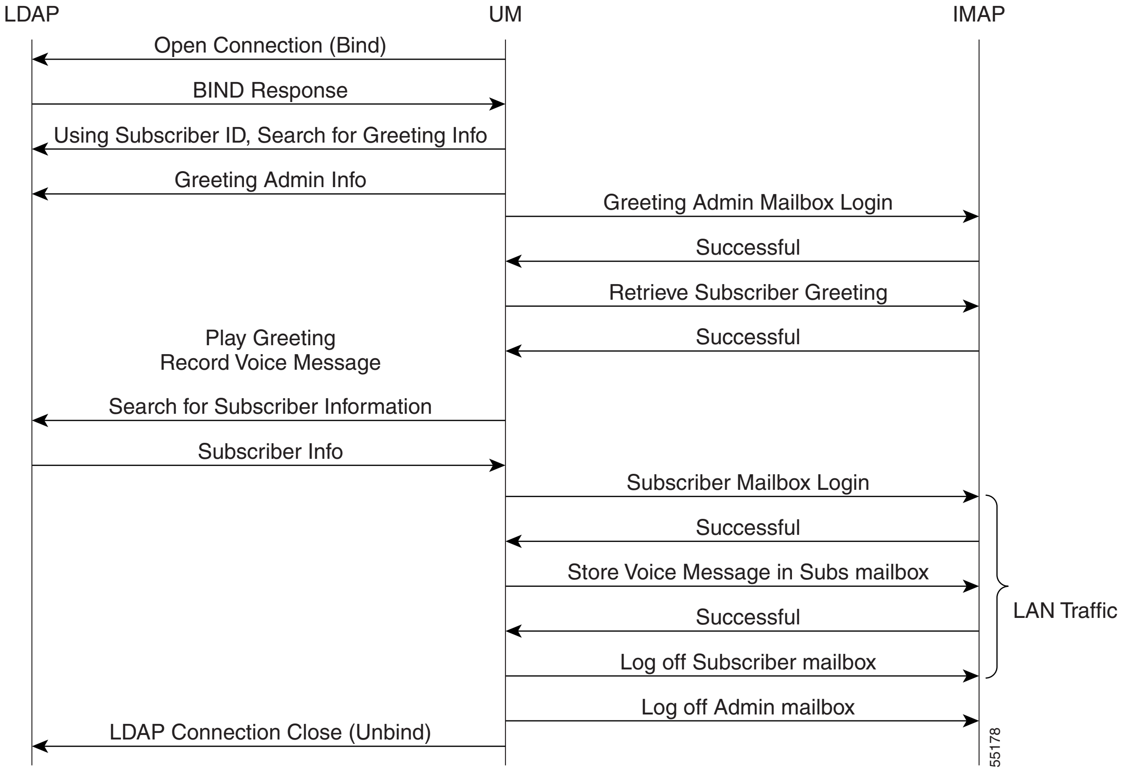
You should refer to the UM administrator guide for more details.

Add Greeting and Fax Administrators

The greeting administrator account is a special mailbox used to store personal greetings and distribution list names for subscribers. The greeting administrator is identified by msgadmin@<organizational unit name>. Each organizational unit requires its own greeting and fax admin accounts and can have more than one of each.

Every subscriber account has a greeting and a faxadmin associated with it. When a subscriber is added to the system, a set of folders is added to the greeting admin account to store the subscriber personal greetings and distribution list information. These folders are separate from the subscriber message mailbox, which stores their voice, fax, and e-mail messages. When a subscriber first logs in to the system and records a personal greeting, the greeting is stored in a folder under the subscriber greeting administrator. Figure 16 shows how the centralized greeting admin works.

Figure 16 Protocol Flows for Centralized Greeting Admin


LDAP
UM
IMAP
Open Connection (Bind)
BIND Response
Using Subscriber ID, Search for Greeting Info
Greeting Admin Info
Greeting Admin Mailbox Login
Successful
Retrieve Subscriber Greeting
Play Greeting
Record Voice Message
Successful
Search for Subscriber Information
Subscriber Info
Subscriber Mailbox Login
Successful
Store Voice Message in Subs mailbox
Successful
LAN Traffic
Log off Subscriber mailbox
Log off Admin mailbox
LDAP Connection Close (Unbind)
55178

When a call comes in for a subscriber, a personal greeting needs to be played to the caller. Using IMAP, the greeting is retrieved from the greeting admin account where it is stored. However, the voice mail message left by the caller will be stored in the subscriber mailbox, which can be accessed by an e-mail client. Centralized greeting admins and local subscriber message stores will result in personal greetings being retrieved across the WAN, but voice messages being stored and retrieved locally for subscribers at POPs. In Figure 16, the only traffic local to the POP is the storage of voice messages in the subscriber mailbox. If you created the greeting and faxadmin accounts on the local message store to service all local subscribers, all IMAP traffic will be local to the POP. The greeting and faxadmin accounts can be added using UMSA under “Global Administration.” We recommend that you create a greeting and faxadmin account on a messaging server to service all subscribers that have a message store on that server.

Add Organizational Unit UMSA Administrators and Subscribers

Each organizational unit requires a UMSA administrator that will manage its COI. UMSA administrators have add, change, and delete privileges over subscribers and COSs within their COI. These unique admin accounts have privileges within their own COI and are added as subscribers. Any subscriber can be made a UMSA administrator by adding the subscribers to the UMSA administrator group created under ou=root.

UMSA administrators can add subscribers within their own COI using the web-based UMSA tool. While adding subscribers, administrators can select the messaging server and greeting and fax admins that service the subscriber. All messaging servers known to the LDAP directory service, and the defined greeting and fax admins, are listed in the drop-down menu on the web interface. Selecting the appropriate message store and greeting admin is an important consideration when adding new subscribers because they define the message store for personal greetings and faxes and the message store for incoming e-mail, voice mail, and fax.

Deploy Fax Services

When you enable fax services for subscribers, the subscribersare assigned a fax number where incoming faxes will be accepted and stored in their mailbox. Subscribers also have the ability to redirect e-mail and fax messages from their mailboxes to any fax machine. Depending on the volume of subscribers wanting fax services, fax gateways can be local to the POP or centralized.

Fax services are described in detail in the integrated solutions document titled *Fax Services*.

Deploying Unified Messaging for Dial Internet Access

Four scenarios for deploying unified messaging for dial Internet access, along with associated call flows are described in the following sections:

- Completely Centralized, page 25
- Partially Centralized, page 27
- More Distributed, page 28
- Completely Distributed, page 30

Completely Centralized

The completely centralized configuration is a good starting point for deployment of uOne services where, except for an H.323 gateway to provide local-number access, all other uOne components are centrally located at your core network. A completely centralized configuration is an acceptable model when the subscriber base is small and services are just being introduced. Figure 17 shows an example of a completely centralized unified messaging deployment; Figure 18 shows the flow sequence for this deployment.

Figure 17 Completely Centralized Deployment

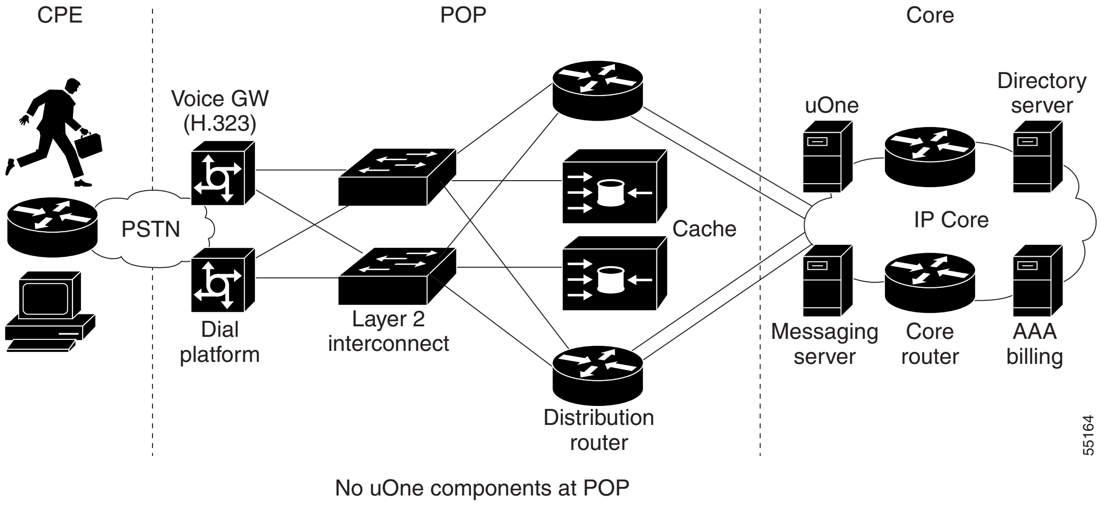
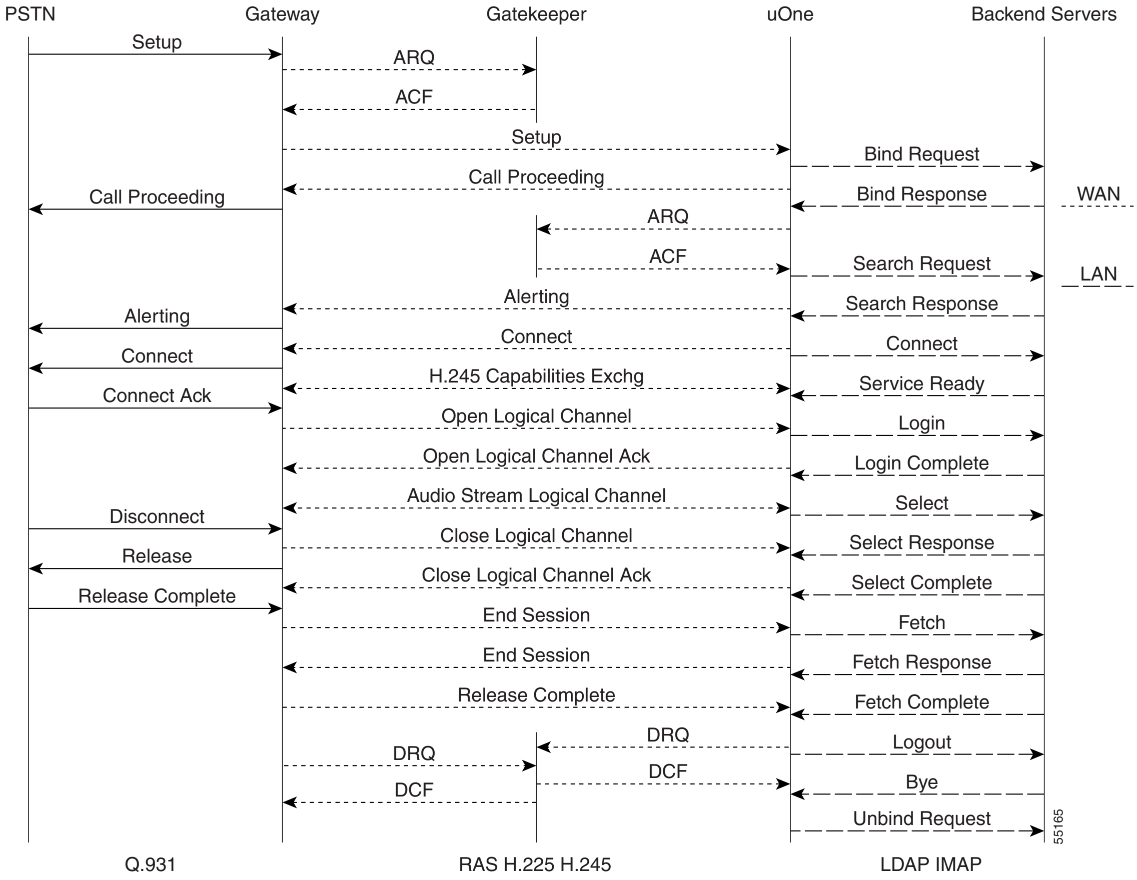


Figure 18 Completely Centralized Flow Sequence



Partially Centralized

As the subscriber base at a POP site grows, we recommend that a uOne server be dedicated to servicing the site while still maintaining back-end services at the core. Deploying the uOne server in this way improves call setup times and voice quality and is easy to deploy. The server at the POP will now service existing subscribers from the POP using the core uOne server because the gatekeeper can be configured to forward calls to the POP to a local server. No other changes to the configuration or user profile information will be necessary. Figure 19 shows an example of a partially centralized unified messaging deployment; Figure 20 shows the flow sequence for this deployment.

Figure 19 Partially Centralized Deployment

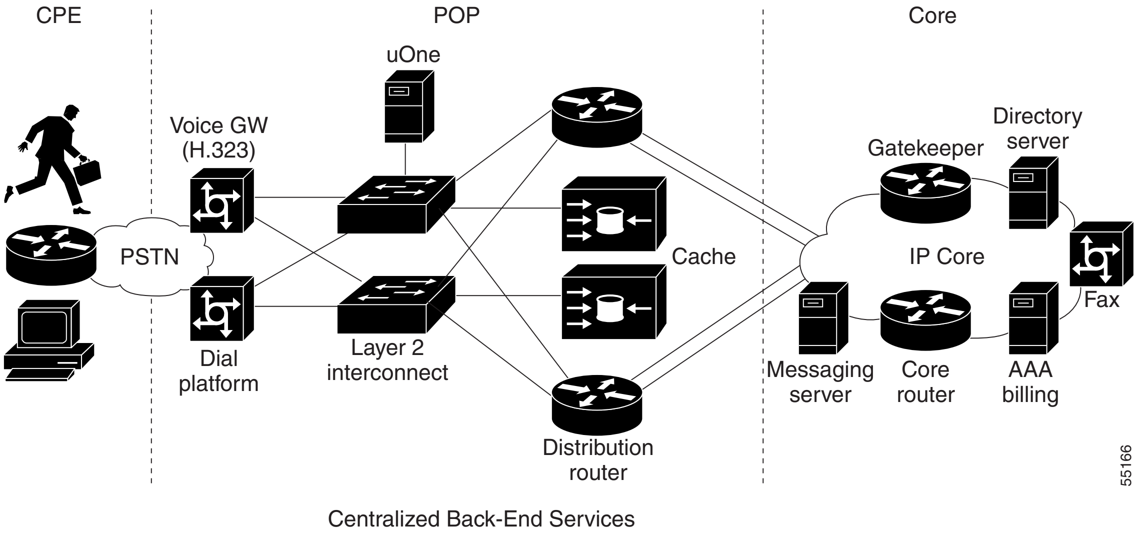
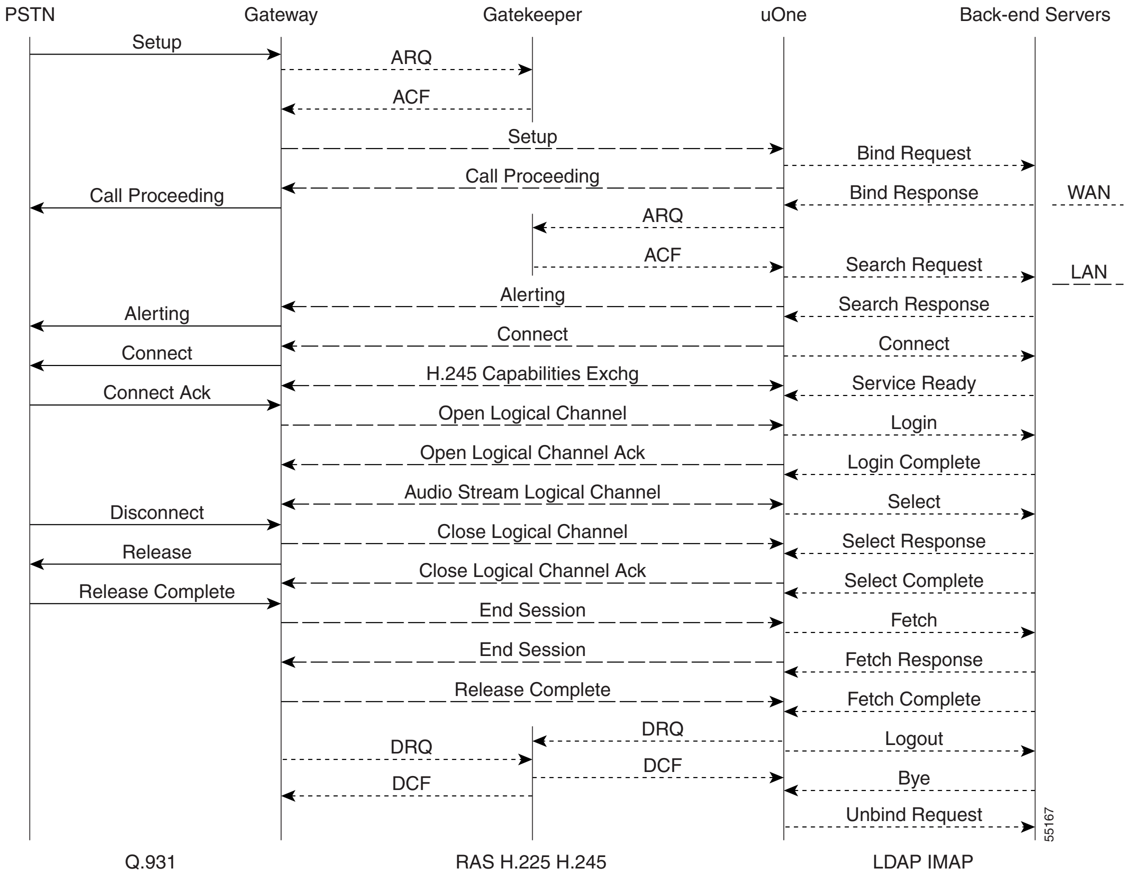


Figure 20 Partially Centralized Flow Sequence



More Distributed

As the subscriber base at a POP site continues to grow, we recommend that you dedicate a messaging server to service the site. Dedicating a messaging server greatly improves message response times and voice quality because all messages are stored and retrieved locally across the LAN. However, your subscribers could notice a slight increase in message response times if they attempt to access their messages from another POP site because the messages must be retrieved across the WAN from the messaging server at their home site. Figure 21 shows an example of this unified messaging deployment; Figure 22 shows the flow sequence for this deployment.

Figure 21 More Distributed Deployment

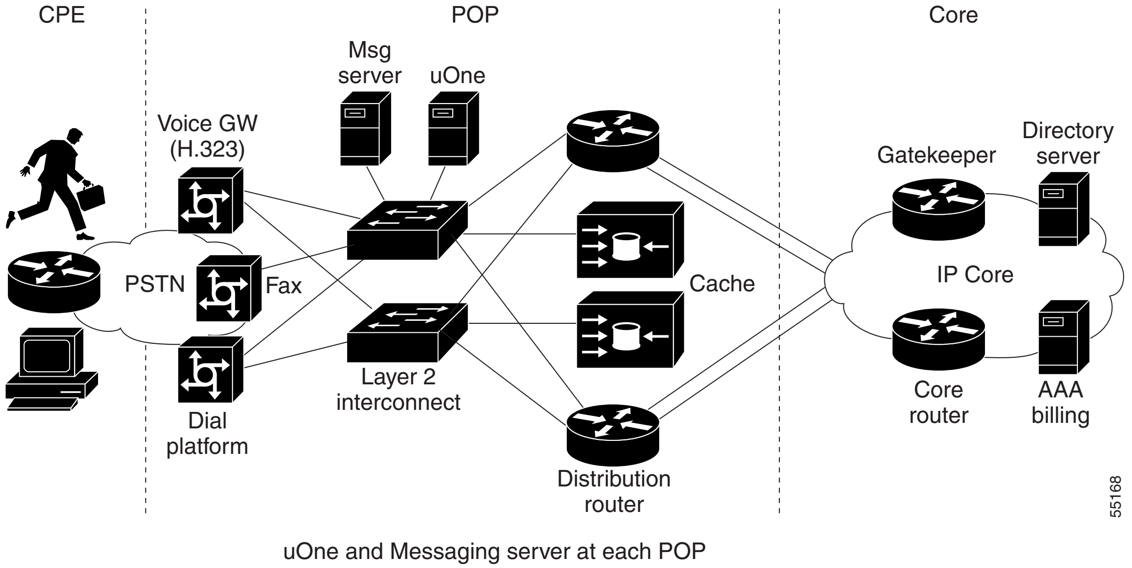
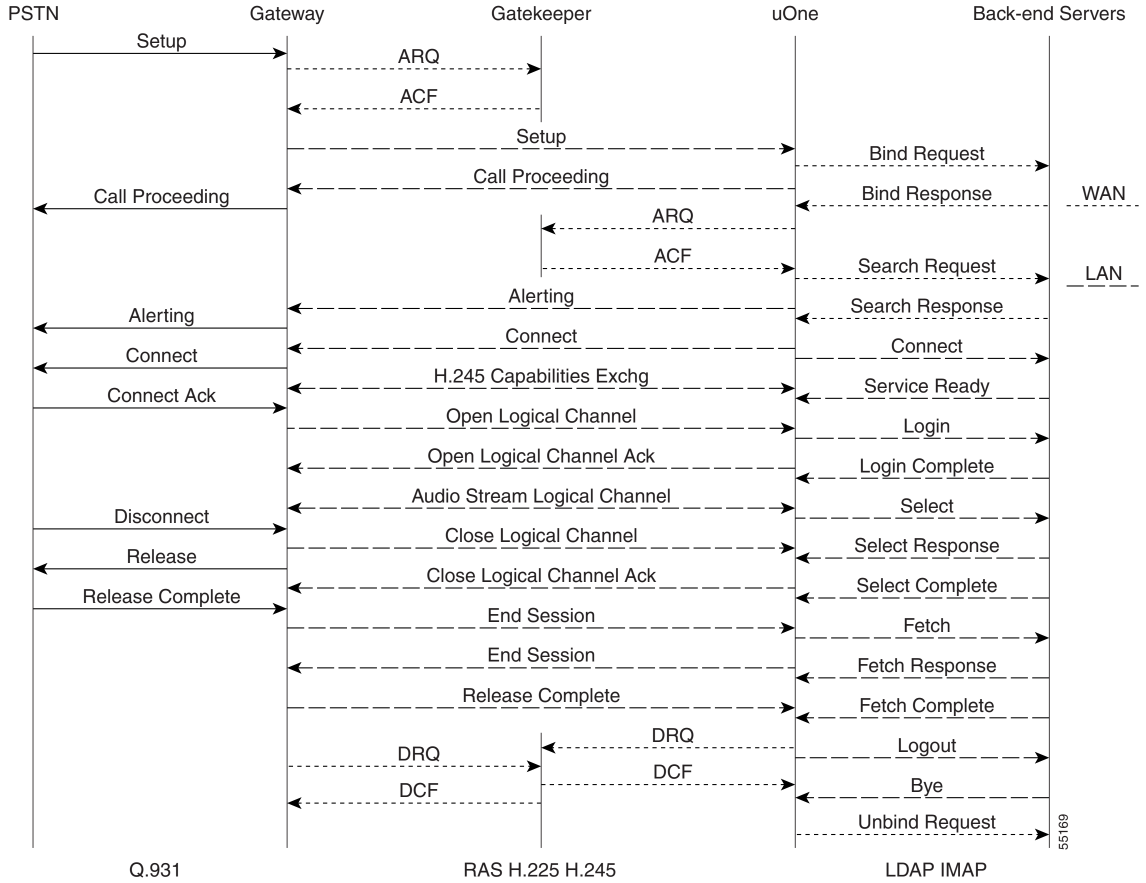


Figure 22 More Distributed Flow Sequence



Completely Distributed

In a completely distributed deployment, everything but directory services is moved to the local POP. Except for authentication and retrieving user profile information, all the other services are local to the POP. Because the gatekeeper is local as well, call setup times are very good and service quality is at its best. Each POP will have its own zone and can be designed for fault tolerance by using redundant gateservers and redundant gatekeepers running HSRP. In normal operation, both gateservers have equal priority and share the call load on a per-call basis. (Call balancing and redundancy are discussed later in this document.) Figure 23 shows an example of a completely distributed unified messaging deployment; Figure 24 shows the flow sequence for this deployment.

Figure 23 Completely Distributed Deployment

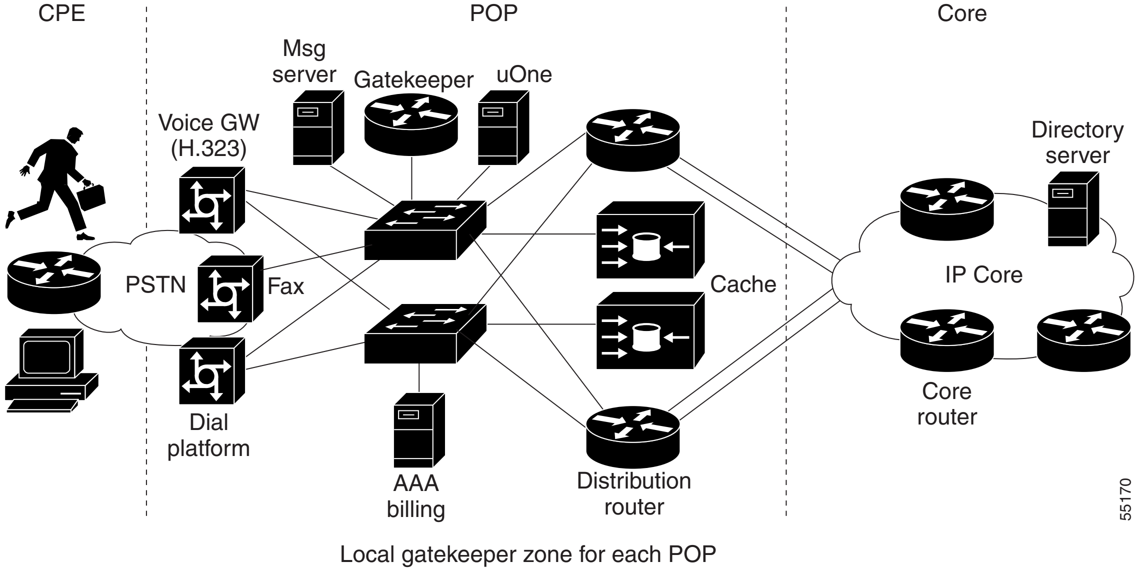


Figure 24 Completely Distributed Flow Sequence

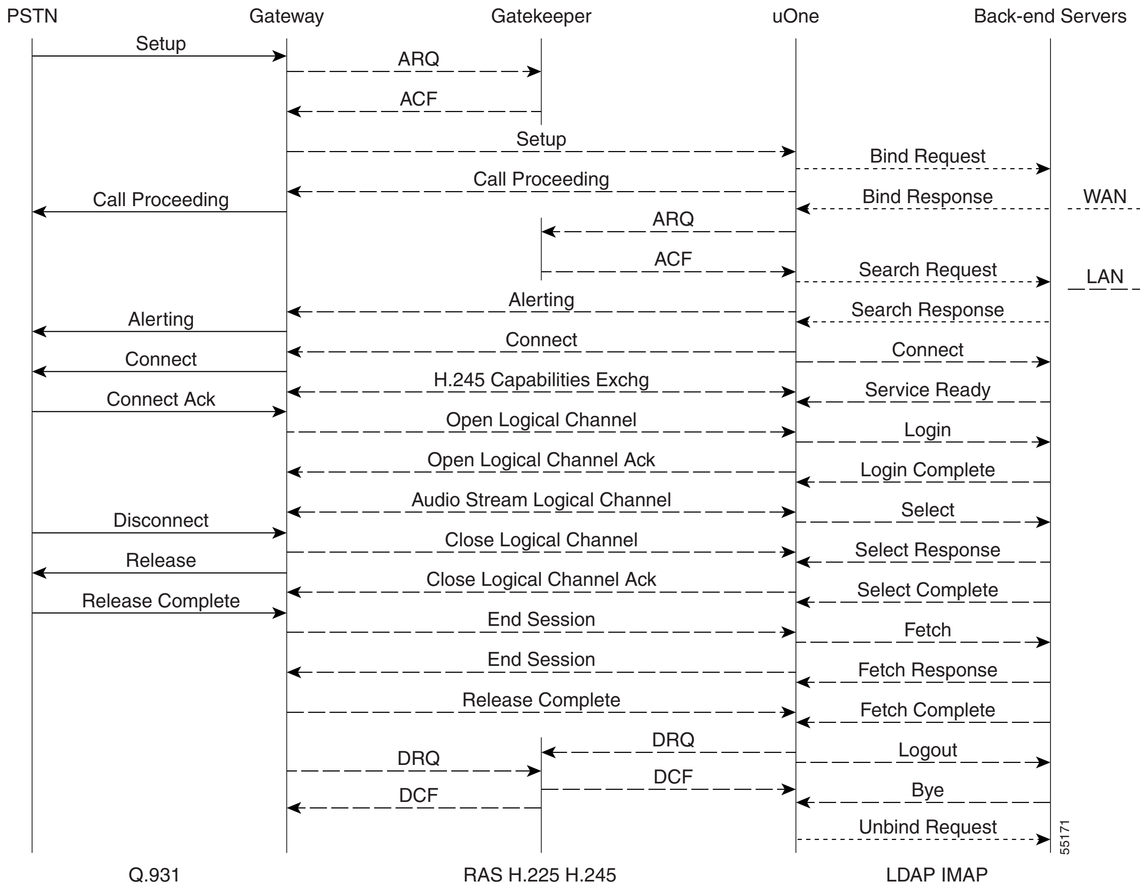


Table 1 summarizes the qualities of each of the described dial Internet access deployment scenarios.

Table 1 Deployment Summary

Quality Feature	Fully Centralized	Partially Centralized	Fully Distributed	Partially Distributed
Call Setup Time	Long	Good	Best	Good
Voice Quality	Average	Good	Good	Good
Authentication	Good	Good	Good	Good
Message Response	Acceptable	Acceptable	Good	Good

- Call Setup Time: The time taken to set up the call and hear ringing at the far end.
- Voice Quality: The quality of the messages being played back from uOne.
- Authentication: The time that the subscriber must wait for the system after entering a user ID and PIN.
- Message Response: The time that the subscriber must wait to hear a message after selecting that message.

Deploying Unified Messaging for Dedicated Internet Access

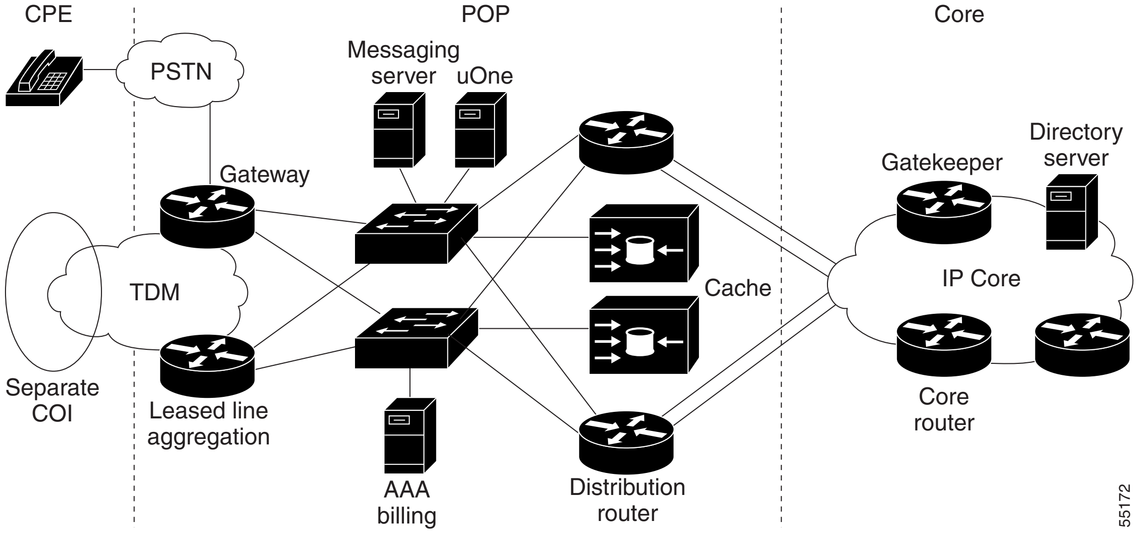
Three scenarios for deploying unified messaging for dedicated Internet access are described in the following sections:

- Sharing uOne Resources at the POP, page 32
- Local Gateway, page 32
- Dedicated uOne Resources, page 33

Sharing uOne Resources at the POP

Small and large corporations use dedicated lines to the ISP for Internet access. A separate COI is set up for each corporation, and administrative authority for this COI is delegated to a system administrator within the organization. The corporation can then set up accounts for its employees on a trial basis and share uOne resources at the POP to which they connect. Employees with unified messaging accounts can access their messages by calling the POP site. Figure 25 shows an example of sharing uOne resource at the POP to deploy unified messaging.

Figure 25 Shared POP Gateway

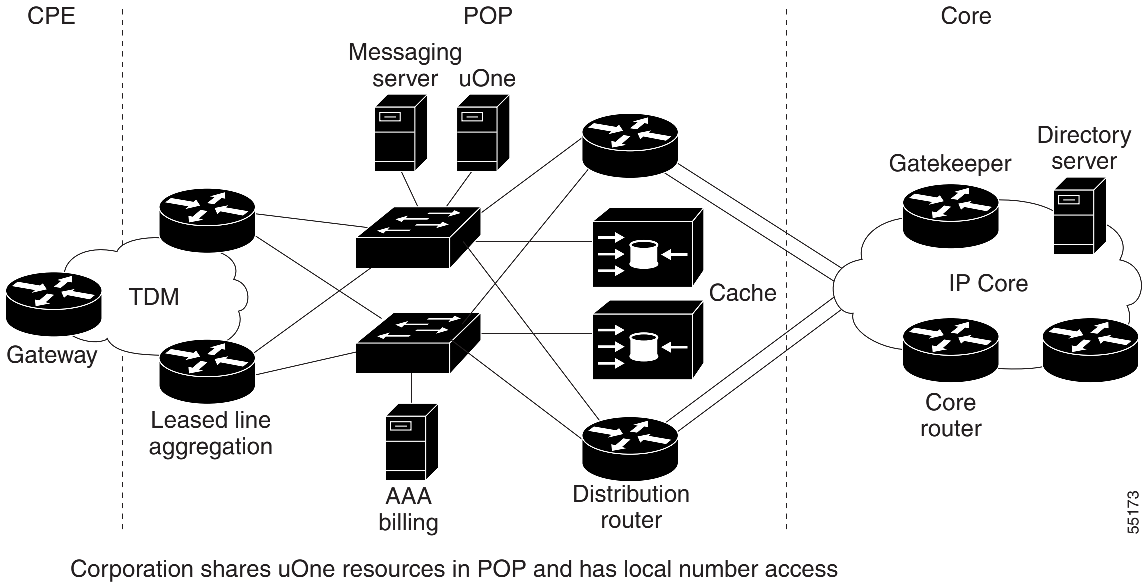


Corporation shares uOne and Gateway resources in POP

Local Gateway

As more users within the organization use unified messaging services, it is more economical for the corporation to have its own local gateway, especially if users must pay toll charges to access the services at a POP site. Figure 26 shows an example where the corporation has its own local gateway.

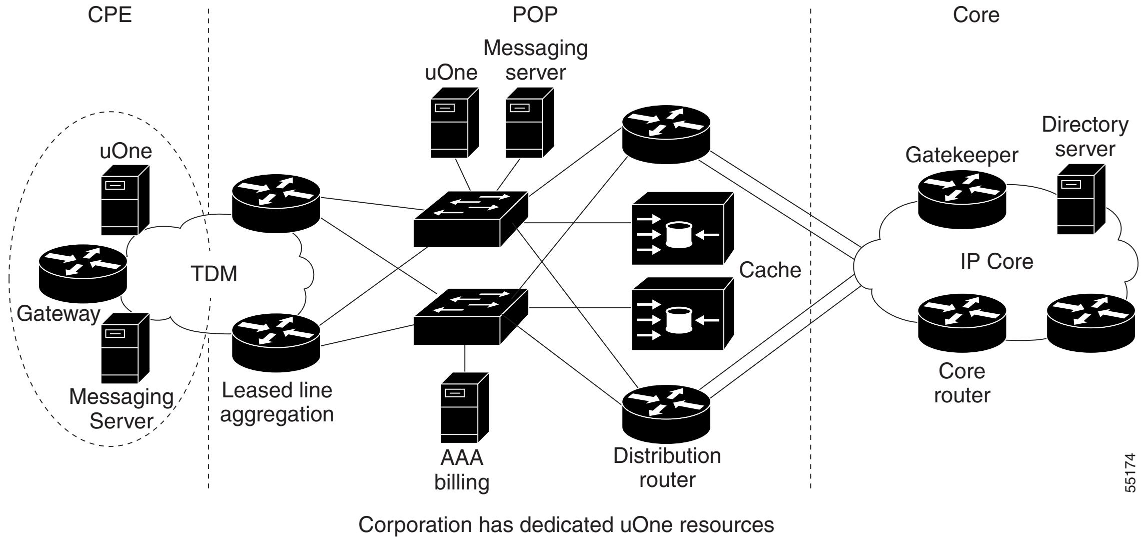
Figure 26 Local Gateway



Dedicated uOne Resources

When the subscriber base within the organization grows even more, the growth justifies dedicated uOne resources to handle all unified communication services. In this case, the uOne server is at the customer site, and messaging servers need to be integrated into existing mail servers. However, directory services is centralized at the core, and billing records can be collected at the POP site or at the core. Figure 27 shows an example of dedicated uOne resources to deploy unified messaging.

Figure 27 Dedicated uOne Resources



Redundancy and Load Balancing

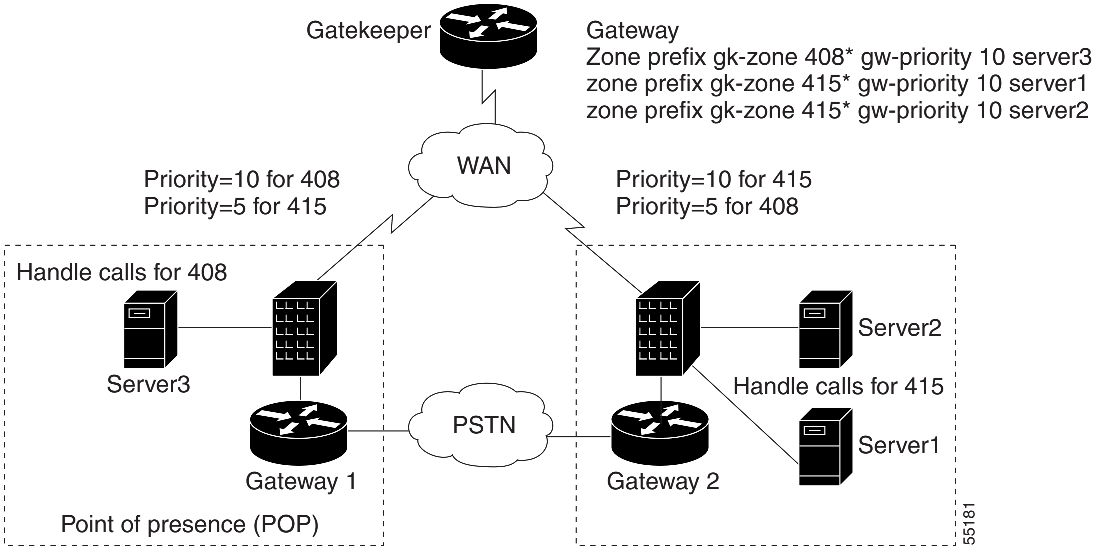
This section describes how to provide redundancy and load balancing for unified messaging services and includes the following sections:

- uOne Server Redundancy and Load Balancing, page 34
- Fax Gateway (Off-Ramp) Redundancy and Load Balancing, page 35
- H.323 Gateway Redundancy and Load Balancing, page 37
- Gatekeeper Redundancy, page 37
- A Fully Redundant Configuration, page 38

uOne Server Redundancy and Load Balancing

Gateservers register themselves as gateways with an H.323 gatekeeper. If multiple gateservers register with the same gatekeeper, and they can all handle any service call, the gatekeeper automatically rotates the calls among all the registered gateways of equal priority. Figure 28 shows an example of load balancing between two UM servers, Server1 and Server2:

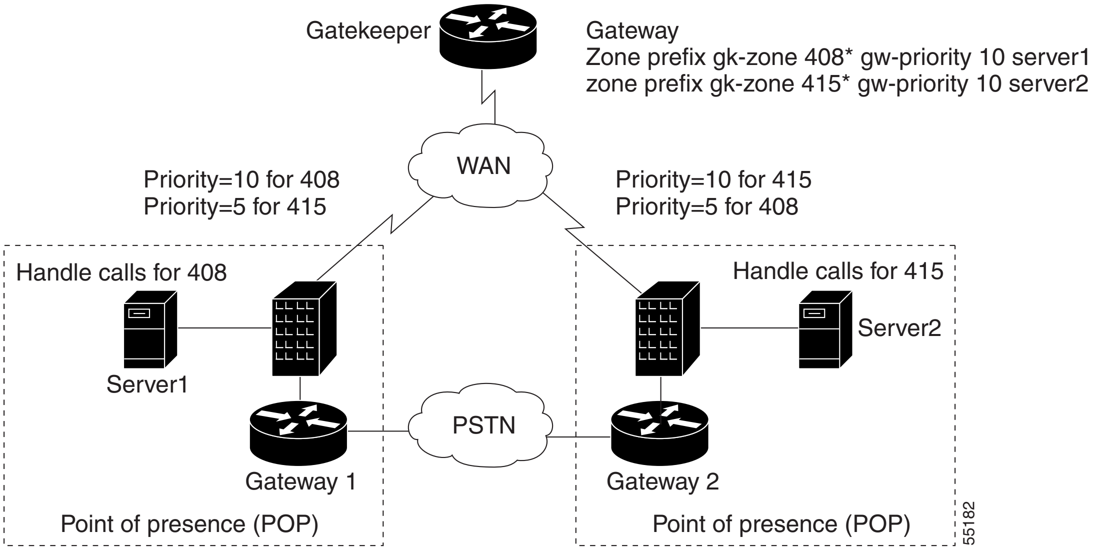
Figure 28 uOne Server Redundancy and Load Balancing



In Figure 28, calls are load-balanced between Server1 and Server2 because they have equal priority to handle calls starting with “415.” However, because all calls are not of the same duration, load balancing is only on a per-call basis. If one of the UM servers is down, it loses its registration with the gatekeeper and the other server handles all incoming calls.

Based on the geographic location of a UM server, you can configure the gatekeeper with different levels of priority for each gateserver, as illustrated in Figure 29.

Figure 29 uOne Server Redundancy Across POPs



Server1 has been assigned a priority of 10 for handling calls to area code 408. Server2 has priority 10 for area code 415. The default priority for a gateway is 5. If one of the servers fails, the server with a lower priority will take over.

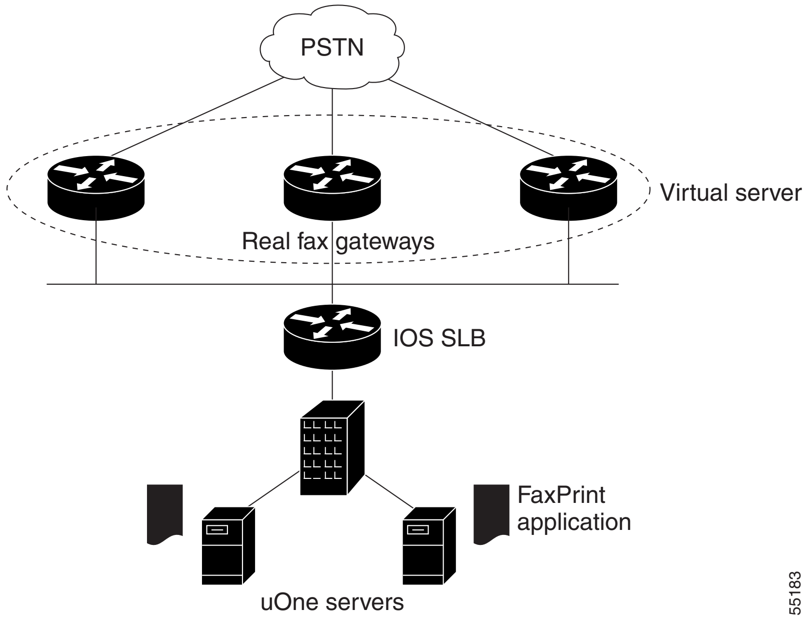
Even though a server in one geographic location can service a call in another geographic location, the server still needs to access the message store for the subscriber, which might be local to the POP. Access to this message store is across the WAN, so the subscriber might notice a slight degradation in service depending on existing traffic loads and bandwidth across the WAN.

Please note that there is no redundancy for active calls being currently serviced by a uOne server. If the server becomes unavailable, another uOne server will handle new incoming calls but existing calls will be disrupted and will need to be reestablished by the caller.

Fax Gateway (Off-Ramp) Redundancy and Load Balancing

By using the Cisco IOS Server Load Balancing (IOS SLB) feature, you can configure multiple fax gateways at a POP site to balance the outbound fax load and provide redundancy and load balancing as shown in Figure 30. IOS SLB is a Cisco IOS-based feature that provides load balancing among multiple servers.

Figure 30 Redundancy and Load Balancing with IOS SLB



A virtual server is a group of real fax gateways that can handle outbound fax calls. The virtual server is assigned an IP address, which is also configured as a secondary address on each of the constituent fax gateways. The uOne faxprint process is configured to connect to this virtual IP address in the faxprint.ini and dialmap.ini files.

When the faxprint process initiates a connection to the virtual IP address, the IOS SLB software chooses a real fax gateway to service this connection based on the configured load-balancing algorithm. IOS SLB software tracks each connection attempt to a fax gateway. If several consecutive TCP "SYN" open connections are not acknowledged, the session is assigned to a new fax gateway.

The number of connection attempts before reassigning the session is configurable. Every failed connection attempt increments a failure counter. If the failure counter exceeds a configurable threshold, the gateway is considered out of service and is removed from the list of active gateways. The failed gateway is not assigned any new connections for a specified configurable time interval called "retry timer." After the timer expires,

IOS SLB will assign the next qualified connection to the failed gateway. If it succeeds, the gateway is placed back on the list of active real gateways. If it fails again, no new connections are attempted until the retry timer expires again.

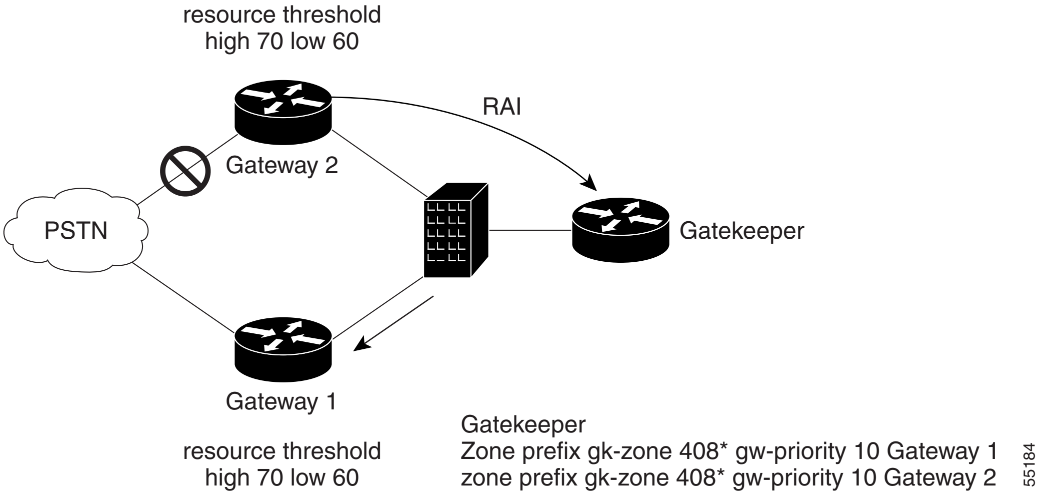
IOS SLB supports two load-balancing algorithms: weighted round robin and weighted least connections. In weighted round robin, each gateway is assigned a weight that represents its capacity to handle connections. The gateway is assigned the number of connections equal to its weight before another real gateway is chosen. In weighted least connections, the gateway chosen to service a connection request is the one with the fewest active connections. Here, also, you can assign weights to gateways. They represent the relative capacity of the gateway to service connection requests compared to the total service capacity of all the gateways that share the same virtual IP address.

H.323 Gateway Redundancy and Load Balancing

Gateways report resource availability to their gatekeepers using RAS Resource Availability Indication (RAI). Digital Signal Processor (DSP) channels can be monitored, and based on a configured threshold, gateways send an RAI to notify the gatekeeper when it is almost out of resources. When resources become available and are more than another configurable threshold, gateways send another RAI to the gatekeeper, notifying it that resources are now available.

When multiple gateways are registered with the gatekeeper, and all other factors are equal, a gatekeeper will choose a gateway with available resources over a gateway with depleted resources. Because the gateway monitors DSP resources, it will send an RAI to the gatekeeper when it loses its connection to the PSTN. When multiple resources with equal priority are registered with the gatekeeper, the gatekeeper rotates the calls with equal priority among all the registered gateways that are qualified to handle the calls, as shown in Figure 31.

Figure 31 Gateway Load Balancing and Redundancy



In this illustration, both gateways are configured to send RAIs to the gatekeeper and both have equal priority to handle calls destined to area code 408. In normal mode, calls are load balanced by turns between the two gateways. When Gateway 2 loses its connection to the PSTN, its DSP resource drops below the configured threshold and it sends an RAI to the gatekeeper, which then forwards all outbound area code 408 calls to Gateway 1.

Gatekeeper Redundancy

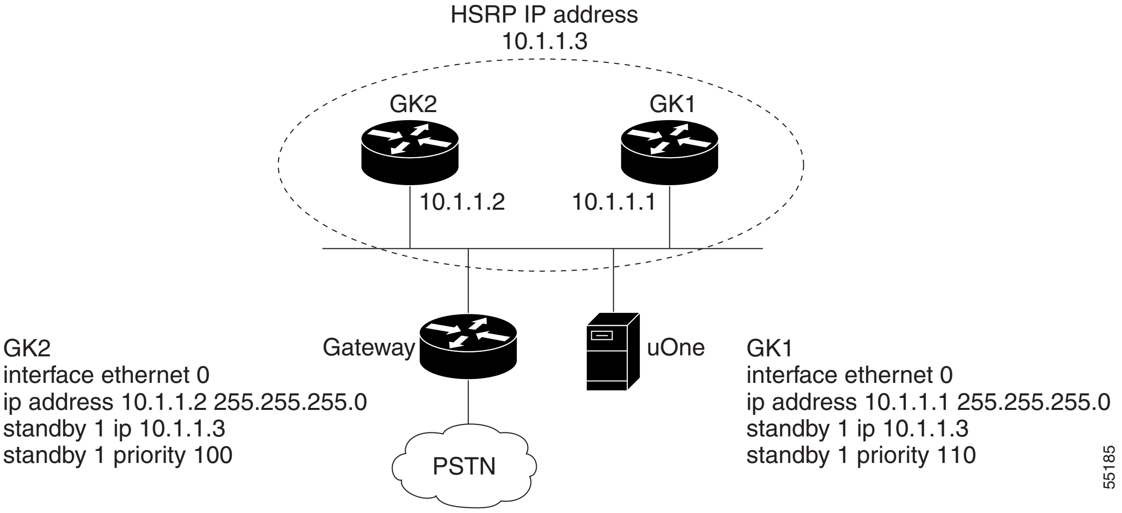
Cisco gatekeepers can be configured to use Hot Standby Routing Protocol (HSRP) so that a standby gatekeeper assumes the role if an active gatekeeper fails. A virtual HSRP IP address is configured on all gatekeepers in the HSRP group and is the common IP address to which the active gatekeeper responds. HSRP uses a priority scheme to identify one gatekeeper as active within a group. All remaining gatekeepers in the group are on standby. When the active gatekeeper fails to send a “hello” message within a configurable interval of time, the next gatekeeper in the group with the highest priority becomes the active gatekeeper and starts responding to the virtual HSRP IP address.

There is no load balancing among the multiple gatekeepers. Two or more gatekeepers can be grouped as an HSRP group, with the one having the highest priority being the active gatekeeper at any given time. The RAS address for all gatekeepers in the group will be the HSRP virtual address. Endpoints and gateways use this HSRP virtual address as their gatekeeper address. Using the HSRP virtual address as the gatekeeper address works even if the gateways attempt to discover the gatekeeper by using multicasting because only the active gatekeeper responds. All other gatekeepers are in standby mode and do not respond to a multicast or unicast request.

When a standby gatekeeper takes over because of the failure of an active gatekeeper, it does not have the state or the registrations of the failed gatekeeper. When a gateway or an endpoint attempts to initiate a new call by sending an ARQ, it will get an Admissions Reject (ARJ), indicating that the endpoint is not recognized. The gateways and uOne servers will need to reregister with the new gatekeeper before being able to make any calls.

Figure 32 shows an example of gatekeepers grouped together in an HSRP group to provide redundancy.

Figure 32 Redundant Gatekeepers Using HSRP



```
GK2
interface ethernet 0
ip address 10.1.1.2 255.255.255.0
standby 1 ip 10.1.1.3
standby 1 priority 100
```

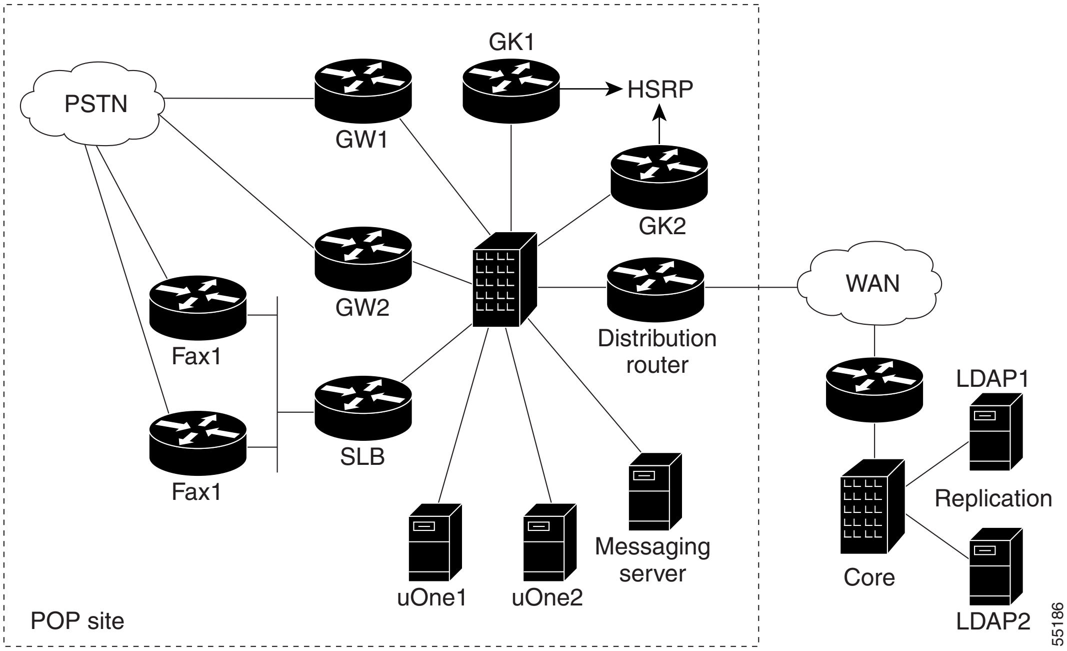
```
GK1
interface ethernet 0
ip address 10.1.1.1 255.255.255.0
standby 1 ip 10.1.1.3
standby 1 priority 110
```

In Figure 32, GK1 has higher priority than GK2, and will become active and respond to the virtual HSRP IP address 10.1.1.3. The gateway and uOne server are configured to use 10.1.1.3 as the IP address of the gatekeeper. If GK1 fails, GK2 starts responding to the virtual IP address but does not yet have the gateway or the uOne server registered as H.323 gateways. When the gateway and the uOne server make the next attempt to register with the gatekeeper by sending a Registration Request (RRQ), they will get a Registration Confirm (RCF) response from GK2.

A Fully Redundant Configuration

A fully redundant POP site has multiple gateways, gatekeepers, fax gateways, and uOne servers. By implementing replication, the LDAP directory server can be made redundant at the core. A fully redundant configuration is shown in Figure 33.

Figure 33 Fully Redundant Configuration



GW1 and GW2 are redundant gateways configured to send RAIs to the gatekeeper based on their available resources and the state of their connection to the PSTN. When a gateway loses its connection to the PSTN, it will send an RAI to the gatekeeper and force all outgoing calls to the other gateway.

GK1 and GK2 are two redundant gatekeepers configured for HSRP. They constantly monitor each other for availability, and take over its functions when the other gatekeeper is not available. The two uOne servers register with the gatekeeper with equal priority (the default is 5) and, in normal operation, handle incoming calls on a round robin basis. However, if one of the servers becomes unavailable, it loses its active registration with the gatekeeper, forcing all new calls to be handled by the remaining uOne server. The IOS SLB feature can be used for load balancing and redundancy on outbound faxes (off-ramp fax gateways).

Unified Messaging Configuration Examples

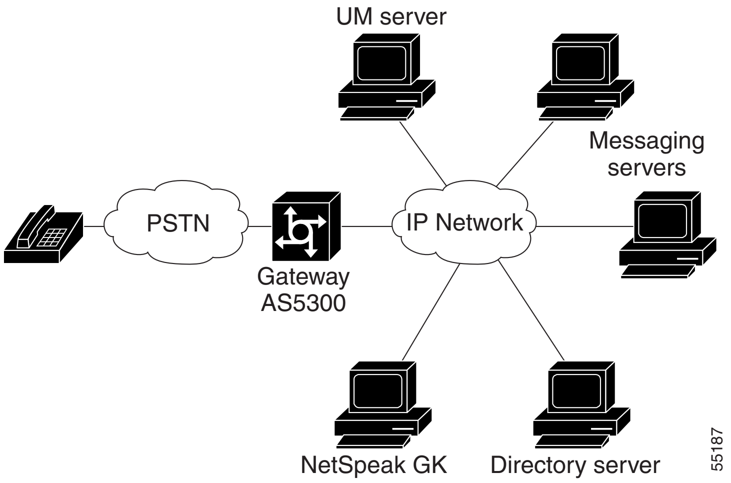
The section includes the following configuration examples:

- Interoperability with Cisco and NetSpeak Gatekeepers, page 39
- Cisco Gateway and Gatekeeper Configuration for Two-Stage Dialing, page 43

Interoperability with Cisco and NetSpeak Gatekeepers

In this configuration example, the Cisco AS5300 gateway and uOne gateserver 4.1S are configured to use a NetSpeak gatekeeper to route calls. Direct inward dial can be used at the AS5300 to accommodate single stage dialing. Figure 34 illustrates the details of the network topology.

Figure 34 Interoperability with NetSpeak Gatekeeper



NetSpeak Gatekeeper Configuration Example

The following example shows how to configure the NetSpeak gatekeeper in Figure 34:

```
From the NetSpeak control center
Select the route server (RS)
Route configuration
Gatekeeper Zones (You will see your NetSpeak Gatekeeper defined here and Online)
Associated gateways
    Add Gateway
    Primary alias (This is the H323-ID field defined in the AS5300)
    Alias type (H323)
    Vendor (Other)
    Country Code (1)
    Area Code (XXX)
    National Prefix (1)
    International Prefix (011)
    Time To Live - TTL (60)
    Number of ports (20)
Associated Hunt Groups (From the Gateways Menu after you have added a gateway)
    Group Name (Name of a hunt group you want associated with your gateway)
    Beginning port number (0)
    Ending port number (19)
    Associated Codec Compatibility
        Standard G.723.1 Audio
        Standard GSM audio
        Standard PCMCA audio
        Standard PCMU audio
    Associated Route Sets
        Route Set management
            Add a route set
    Associated routes
        Add E.164
            Country Code (1)
            Area Code (XXX)
            Beginning subscriber number (9933301)
            Ending Subscriber number (9933347)
            Number to dial (SN) - Subscriber Number
Add the newly created route set with associated routes to the newly created hunt group.
```

You need to repeat this same process for the other gateway. When you go to “Associated Gateways,” you should see both gateways on line—this is an indication that they have registered with the gatekeeper. If you do not associate codecs with a gateway, the gateway will fail registration with the NetSpeak gatekeeper.

Cisco AS5300 Gateway Configuration Example

The following example shows how to configure the Cisco AS5300 gateway in Figure 34:

```
router# show running-config
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname tokyo-5300
!
enable password xxxx
!
resource-pool disable
!
ip subnet-zero
no ip domain-lookup
isdn switch-type primary-dms100
cns event-service server
mta receive maximum-recipients 1024
!
controller T1 0
 framing esf
 clock source line primary
 linecode b8zs
 pri-group timeslots 1-24
!
controller T1 1
 framing esf
 clock source line secondary 1
 linecode b8zs
 pri-group timeslots 1-24
!
controller T1 2
 framing esf
 linecode b8zs
 pri-group timeslots 1-24
!
controller T1 3
 framing esf
 linecode b8zs
 pri-group timeslots 1-24
!
voice-port 0:D
voice-port 1:D
voice-port 2:D
voice-port 3:D
!
dial-peer voice 1 voip
 destination-pattern 9933...
 dtmf-relay cisco-rtp
 codec g711ulaw
 session target ras
!
dial-peer voice 2 pots
```

```
 incoming called-number 9933...
 direct-inward-dial
!
process-max-time 200
gateway
!
interface Ethernet0
 ip address 172.26.106.4 255.255.255.0
 no ip directed-broadcast
 h323-gateway voip interface
 h323-gateway voip id GK@hope.cisco.com ipaddr 172.26.106.10 1719
 h323-gateway voip h323-id tokyo-5300
 h323-gateway voip tech-prefix 8
!
interface Serial0:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-dms100
 isdn tei-negotiation first-call
 isdn incoming-voice modem
 fair-queue 64 256 50
 no cdp enable
!
interface Serial1:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-dms100
 isdn tei-negotiation first-call
 isdn incoming-voice modem
 fair-queue 64 256 0
 no cdp enable
!
interface Serial2:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-dms100
 isdn tei-negotiation first-call
 fair-queue 64 256 0
 no cdp enable
!
interface Serial3:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-5ess
 fair-queue 64 256 0
 no cdp enable
!
interface FastEthernet0
 no ip address
 no ip directed-broadcast
 shutdown
 duplex auto
 speed auto
!
ip classless
ip route 0.0.0.0 0.0.0.0 172.26.106.1
no ip http server
!
line con 0
 transport input none
line aux 0
line vty 0 4
 password xxxx
 login
```

```
!
end
```

Cisco Gateway and Gatekeeper Configuration for Two-Stage Dialing

In the following configuration examples, the UM server registers with the gatekeeper with a technology prefix of 4#, which also happens to be the default technology prefix defined in the gatekeeper. The gateway also must be registered with the gatekeeper, because all calls to the UM server must be routed via the gatekeeper. This particular example illustrates a two-stage dialing model, where subscribers dial a phone number to access the gateway and then use a token (265) to access the unified messaging services.

The following configuration shows how to configure the gateway described in the scenario above:

```
Gateway# show running-config
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname Gateway
!
enable password xxxx
!
resource-pool disable
!
ip subnet-zero
no ip domain-lookup
ip host hope.cisco.com 172.26.106.6
ip host faith.cisco.com 172.26.106.3
ip host charity.cisco.com 172.26.106.2
!
isdn switch-type primary-dms100
cns event-service server
mta receive maximum-recipients 1024
!
controller T1 0
 framing esf
 clock source line primary
 linecode b8zs
 pri-group timeslots 1-24
!
controller T1 1
 framing esf
 clock source line secondary 1
 linecode b8zs
 pri-group timeslots 1-24
!
controller T1 2
 framing esf
 linecode b8zs
 pri-group timeslots 1-24
!
controller T1 3
 framing esf
 linecode b8zs
 pri-group timeslots 1-24
!
voice-port 0:D
```

```
voice-port 1:D
voice-port 2:D
voice-port 3:D
!
dial-peer voice 1 voip
 destination-pattern 265
 dtmf-relay cisco-rtp
 codec g711ulaw
 session target ras
!
process-max-time 200
gateway
!
interface Ethernet0
 ip address 172.26.106.4 255.255.255.0
 no ip directed-broadcast
 h323-gateway voip interface
 h323-gateway voip id gk-splob ipaddr 172.26.106.8 1719
 h323-gateway voip h323-id tokyo-5300
 h323-gateway voip tech-prefix 8
!
interface Serial0:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-dms100
 isdn tei-negotiation first-call
 isdn incoming-voice modem
 fair-queue 64 256 50
 no cdp enable
!
interface Serial1:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-dms100
 isdn tei-negotiation first-call
 isdn incoming-voice modem
 fair-queue 64 256 0
 no cdp enable
!
interface Serial2:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-dms100
 isdn tei-negotiation first-call
 fair-queue 64 256 0
 no cdp enable
!
interface Serial3:23
 no ip address
 no ip directed-broadcast
 isdn switch-type primary-5ess
 fair-queue 64 256 0
 no cdp enable
!
interface FastEthernet0
 no ip address
 no ip directed-broadcast
 shutdown
 duplex auto
 speed auto
!
ip classless
ip route 0.0.0.0 0.0.0.0 172.26.106.1
no ip http server
```

```
!
line con 0
 transport input none
line aux 0
line vty 0 4
 password xxxx
 login
!
end
```

The following **show** command output displays the status of this gateway:

```
Gateway# show gateway

 Gateway  tokyo-5300  is registered to Gatekeeper gk-splob

Alias list (CLI configured)
 H323-ID tokyo-5300
Alias list (last RCF)
 H323-ID tokyo-5300

 H323 resource thresholding is Disabled
```

The following configuration shows how to configure the gatekeeper described in the scenario:

```
Gatekeeper# show running-config
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname Gatekeeper
!
boot system flash c2600-ix-mz.120-5.T1
boot system flash c2600-js-mz.120-5.XK1
enable password xxxx
!
ip subnet-zero
no ip domain-lookup
!
ip dvmrp route-limit 20000
!
process-max-time 200
!
interface Ethernet0/0
 ip address 172.26.106.8 255.255.255.0
 no ip directed-broadcast
!
interface Ethernet0/1
 no ip address
 no ip directed-broadcast
 shutdown
!
ip classless
ip route 0.0.0.0 0.0.0.0 172.26.106.1
no ip http server
!
gatekeeper
 zone local gk-splob cisco.com
 gw-type-prefix 4#* default-technology
 no use-proxy gk-splob default inbound-to terminal
 no shutdown
!
```

```
line con 0
 transport input none
line aux 0
line vty 0 4
 password xxxx
 login
```

The following **show gateway** command output displays the gateway technology prefix table for this gatekeeper:

```
Gatekeeper# show gateway gw-type-prefix

GATEWAY TYPE PREFIX TABLE
=========================
Prefix: 4#*    (Default gateway-technology)
    Zone gk-splob master gateway list:
    172.26.106.2:1720 Charity-UM

Prefix: 8*
    Zone gk-splob master gateway list:
    172.26.106.4:1720 tokyo-5300
```

The following **show gateway** command output displays the status of all registered endpoints for this gatekeeper:

```
Gatekeeper# show gatekeeper endpoints

Total number of active registrations = 2
       GATEKEEPER ENDPOINT REGISTRATION
       ================================
CallSignalAddr  Port  RASSignalAddr   Port  Zone Name         Type    F
--------------- ----- --------------- ----- ---------         ----    --
172.26.106.2    1720  172.26.106.2    32795 gk-splob          VOIP-GW
    H323-ID: Charity-UM
172.26.106.4    1720  172.26.106.4    1803  gk-splob          VOIP-GW
    H323-ID: tokyo-5300
```

Related Documents

- *Cisco IOS Voice, Video and Fax Configuration Guide*, Release 12.2
- *Cisco IOS Voice, Video and Fax Command Reference*, Release 12.2