Network Working Group
Request for Comments: 4597
Category: Informational

R. Even
Polycom
N. Ismail
Cisco Systems, Inc.

July 2006

Conferencing Scenarios

Status of This Memo

Copyright Notice

Abstract

This document describes multimedia conferencing scenarios. It describes both basic and advanced conferencing scenarios involving voice, video, text, and interactive text sessions. These scenarios will help with the definition and evaluation of the protocols being developed in the centralized conferencing XCON working group.

Table of Contents

1. Introduction

This document describes multimedia conferencing scenarios. The development of these scenarios is intended to help with the definition and evaluation of the requirements for the centralized conferencing (XCON) working group. Although this document uses some definitions and conventions described in the SIP Conferencing Framework document [1], these scenarios are not specific to SIP. The document describes basic and advanced conferencing scenarios. The advanced scenarios assume that the user agents support the set of XCON protocols, identified in the Framework and Data Model for Centralized Conferencing [3], in order to take advantage of the conference functionality. However, note that many of these features can be implemented today by using an interactive voice response (IVR) or web interface to control the conferencing application.

The entities comprising the Conferencing System are the conference that is the center point for signaling and the participants. The participant who initiated the conference is called the initiating participant.

The scenarios described here demonstrate different conferencing services. These services can be offered in a multimedia environment that benefits from having some support in the user agents that enable more robust and easier-to-use conferencing services. It is up to the conferencing system manufacturers and the conferencing service provider to decide what services can be built and which services are offered to the end users.

The scenarios describe multimedia examples, but they are applicable to audio only as well as to audio and video conferences.

Multimedia conferences may include any combination of different media types such as audio, video, text, interactive text, or presentation graphics. The conference scenarios are similar, but the media handling may be dependent on the media type.

2. Basic Conferencing Scenarios

These scenarios enable a conference-unaware participant to create, join, and participate in a conference. The participant may use out-of-band signaling to participate in a conference, but this is not mandatory. The Conferencing System has all the functionality it needs in order to supply the service offered to the participants. Typical minimum requirements are that the participant support dual-tone multi-frequency (DTMF) tones/signal or provide voice responses to an IVR system.

2.1. Ad Hoc Conferences

A participant has a service provisioned to him that enables him to start an ad hoc conference when he calls the Conferencing System. When the participant wants to start a conference, he calls the conference service. The participant may be identified by different means, including request destination, authenticated identity, or an IVR system using DTMF. The conference is created automatically with the predefined functionality. The participant who has such a service notifies the other participants how to call the conference via external means such as instant message or email.

The participant may have Conferencing System functionality and thus can create an ad hoc conference using his own user agent. An example of such a conference is an audio conference initiated by a participant who has a conference service that enables him to start a conference when he calls a specific URI. The conference may be created by the first person calling this URI, or it may be created only after the owner is authenticated using an IVR system. In the latter case, the other participants may get an announcement and are placed on hold if they call the conference before the owner.

2.2. Extension of a Point-to-Point Call to a Multipoint Call

This is a basic case. The initiating participant (PA) is in a point-to-point call with another participant (PB). PA wants to add a third participant (PC) to the call. PA cannot provide the Conferencing System functionality on his user agent nor can the other participant PB. PA and PB do not support call transfer. PA has a conferencing service that uses the methods described in 2.1. PA conveys the conference information to PB in the point-to-point call. Both participants disconnect and call the Conferencing System. The Conferencing System may support dial-out (for example, via DTMF), allowing the initiating participant PA to call the third party PC through the Conferencing System.

2.3. Reserved Conferences

The reservation for this type of conference is typically done by an out-of-band mechanism in advance of the actual conference time. The conference identification, which may be a URI or a phone number with a pin number, is allocated by the reservation system. It is sent to all participants through email, IM, etc. The participants join by using the conference identification. The conference identification must be routable, enabling the allocation of a conference with free resources at the time when the conference actually runs. The Conferencing System can also dial out to the conference participants. The participants may not be informed that they are in a conference,

since their User Agent is not conference aware. The participants may know, via announcement from the Conferencing System, that they are in a conference and who the other participants are.

3. Advanced Conferencing Scenarios

These scenarios assume user agents that support at least call transfer service and a way to communicate information on events from the Conferencing System to the user agent. The Conferencing System may have the ability to discover the capabilities of the participants, for example, whether they support call transfer. This section specifies the dependencies in each scenario. An advanced conference can be initiated only by a user agent that has advanced features, but some user agents in the conference may have less functionality.

3.1. Extending a Point-to-Point Call to a Multipoint Call

The initiating participant PA is in a point-to-point call and wants to add a third participant. PA can start a multipoint call on a conferencing bridge known to him. The extension can be without consultation, which means that PA moves the point-to-point call to the Conferencing System and then adds the third party (this can be done in various ways). Alternatively the extension can be done with consultation, which means that PA puts his current party on hold, calls the third party, asks him to join the conference, and then transfers all the participants to the Conferencing System.

3.2. Lecture Mode Conferences

This conference scenario enables a conference with a lecturer who presents a topic and can allow questions. The lecturer needs to know who the participants are and needs to be able to give them the right to speak. The right to speak can be based on floor control or an out-of-band mechanism.

In general, the lecturer is seen/heard by the conference participants and often shares a presentation or application with the other participants.

A participant joining this type of conference can get the identity of the lecturer and often the identities of the audience participants.

This type of conference may have multiple media streams. For example, if simultaneous language translation is available, a participant has the option of selecting the appropriate language audio stream. Multiple video streams could include the speaker's face and a whiteboard/demonstration stream.

3.3. Conference with Conference-Aware and Unaware Participants

A conference can include a mix of participants that are conference-aware and unaware. Conference-unaware participants may be using a proxy function that proxies the advanced functionality between the different protocols and the Conferencing System. For example, an IVR system or a web page interface can be used to provide additional functionality.

3.4. A Reserved or Ad Hoc Conference with Conference-Aware Participants

In order to start the conference, the initiating participant calls the Conferencing System using, for example, a unique identifier. The Conferencing System may use some authenticating method to qualify the participant. The other participants may call the Conferencing System and join the conference. The Conferencing System is able to find the capabilities of the participants. In case of a reserved conference, the Conferencing System starts the conference at the scheduled time. The participants may join by calling the conference URI, or the Conferencing System may call them. The conference may have privilege levels associated with a specific conference or participant. The privileges are for the initiating participant and for a regular participant; the initiating participant may delegate privileges to the other participants. The privileges allow functionality as defined in the next section.

3.5. Advanced Conference Features

The following features can be used in all the advanced conferencing scenarios. In the examples given in this section, when referring to a participant that has a functionality, it means a participant with the right privileges. These scenarios may be available in the advanced conferencing scenarios and are common in many conferencing applications. This is not a requirement list, rather some examples of how specific functions may be used in a conference.

- Add Participants - A participant may add a new participant to the conference. This can be done, for example, by instructing the Conferencing System to call the participant or by the first participant calling the new participant and pointing him to the conference.

- Delete Participant - A participant may delete participants from the conference if he can identify them.

- Changing User Agent/Modes - During the course of a conference, a participant may switch between user agents with different capabilities while still remaining part of the conference. For example, a participant may initially join using a mobile phone and then switch to a desktop phone. Or a participant may join with a phone, discover that the conference has video streams available, and switch to a video phone.

- Changing Media - During the conference, a participant may be able to select different media streams than the one he had when he joined the conference. An example is a participant that initially joined the conference as an audio participant. The participant is unable to understand the conversation properly, and he learns that there is also an interactive text available. He will ask to receive the text stream also.

- Authenticate participants - A participant can authenticate other participants who want to join the conference. This can be done, for example, in a video conferencing session by creating a sidebar between the two participants, allowing the authenticating participant to talk with the new participant and verify his identity.

- Authorize participants - A participant can authorize other participants in order to allow them to join the conference. This can be done implicitly by assigning a password to the conference or to each participant and letting the Conferencing System decide if the new participant is allowed to join. The authorization can be done explicitly by directing the entered password to the initiating participant who will authorize each participant. The conferencing system may use an authentication mechanism to authenticate the participants.

- Controlling the presentation of media - During the conference, the participant may be able to manage whose media is being sent to each participant. For example, the participant may be able to decide that he wants to be the speaker and all the rest to be listeners; he may also specify whose media he wants to receive. The participant may be able to mute a media stream during the conference.

- Giving privileges - During the conference, the participant may want to give a privilege to another participant. The assigning of privileges may be implicit when requested or explicit by asking the participant to grant a privilege.

- Side conferences or sidebars - The participant may want to create a side conference that include some of the main conference participants. When the side conference is finished, the participants return to the main conference. A sidebar may have the same functionality as the main conference. There can be several sidebar scenarios:

  1. A basic sidebar requires that two participants have the capability to have two calls at the same time, with a point-to-point call in parallel to the main conference. It is user agent implementation-specific whether both calls' streams are mixed automatically or the participants are allowed to manually switch between them.

  2. A conferencing-system-based sidebar uses the Conferencing System to create the sidebar and compose the relevant sidebar stream mixes. These mixes can include the main conference as an incoming stream to the mix. Mechanisms to signal the creation of the sidebar, invite participants, and control the mixes should be available.

     For example, participants in an audio sidebar may not be heard by the rest of the conference. However, the main conference audio may be mixed in the sidebar, but at a lower volume, or in a different channel. As another example, a sidebar can have a different media type from the main conference: a video call can have an audio sidebar where the other participants can see the sidebar participants talking but cannot hear them; or an audio or video conference may have a text sidebar.

- Conference information - When a participant joins the conference, he is announced to the participants. An announcement may be available when he leaves the conference. The participants may query the conferencing system for the current participants of a specific conference. This conference information may include other information, for example, the media streams available in the conference.

- Extending of a conference - Reserved conferences and ad hoc conferences may have a time limit. The Conferencing System informs the participants when the limit is approaching and may allow the extension of the conference.

- Adding and removing a media type to the conference - A participant may want to start a data presentation during a conference. He may want to distribute this new media to all the participants. The participant asks the Conferencing System to start the new media channel and to allow him to send data in the new channel.

- Audio-only participants - In a multimedia conference, some of the participants who want to join may have no way to send and receive all the media types. Typically, they can send and receive audio. Such participants join the conference as audio-only participants. The general case is that participants may send and receive only part of the media streams available in the multimedia conference.

- Passive participants - In a conference, some participants may be listeners to all or part of the media streams, but may be invisible to all other participants.

- Recorders - A recorder can be added to the conference. A recorder can record all streams or a subset of the streams. Recorders may be turned on and off during the conference. Recorders may be used for a "role call" scenario in order to record a participant's name. This name can be announced at a later stage automatically or based on a participant request. A recorder is a case of a passive participant.

- Whisper/Private Message - A participant can send a one-way message (text, audio, or even some other media) to another participant that is immediately rendered. This differs from a sidebar in that it is immediate and creates no long-lived session.

- Human operator - A participant may ask for assistance from a human operator during the conference.

4. Scenarios for Media Policy Control

During a conference, media streams may be controlled by authorized participants using either a media control protocol or a third-party application. This section describes some typical media control scenarios. The conference can be of any size. Some of the media control scenarios are typical of specific conference sizes. As a general rule, larger conferences scenarios tend to be more centrally managed or structured.

The mixing of media in a conference may start when the conference starts or when the initiating participant joins. In the later case, early participants may be put on hold and get "music on hold".

The scenarios apply to audio conferences as well as to multimedia conferences. In the sections below, there is some specific information about the mixed video layout and interactive text.

4.1. Video Mixing Scenarios

For video, the participant selects one of a set of predefined video presentations offered by the server. Each video presentation is identified by a textual description as well as an image specifying how the presentation appears on the screen. In this scenario, by choosing a video presentation, the participant chooses how many video streams (participants) are viewed at once and the layout of these video streams on the screen.

The contents of each sub-window can be defined by a conference policy and/or controlled by authorized participants. It may also be possible to have multiple mixes per conference, possibly as many as there are participants. (Note that the same flexibility may be afforded to audio mixes as well.)

The following is a list of typical video presentations. Other layouts are available today in commercial products.

- Single view: This presentation typically shows the video of the loudest speaker.

- Dual view: This presentation shows two streams. If the streams are to be multiplexed in one image (typical of centralized servers), the multiplexing can be:

  1. Side-by-side windows, with no altered aspect ratio. Thus, blanking of parts of the image might be necessary if the streams are to be combined as one image.

  2. Side-by-side windows, with altered aspect ratios. Thus, blanking parts of the image is not necessary. The mixer handles the cropping of the images.

  3. One window above the other, with no altered aspect ratio.

  4. One window above the other, with altered aspect ratios.

- Quadrate view: This presentation shows 4 streams. If the streams are multiplexed into one image (centralized server), they are arranged in a 2x2 style. Note that in this style the aspect ratios are maintained.

- 9 sub-picture view: This presentation shows 9 streams. If the streams are to be multiplexed in one image, they are arranged in a 3x3 style. In the multiplexing case, cropping is performed under the discretion of the mixer.

- 16 sub-picture view: This presentation shows 16 streams. If the streams are to be multiplexed into one image, they are arranged in a 4x4 style. In this style, the aspect ratios are maintained, and no cropping or blanking is needed.

- 5+1 sub-picture view: This presentation shows 6 streams. If the streams are to be multiplexed into one image, then the pictures are laid so that one sub-window occupies 4/9 of the screen while each of the other five occupies 1/9 of the screen.

4.2. Typical Video Conferencing Scenario

This scenario is known as voice-activated video switch. Every participant hears the N loudest participants but does not hear himself. All the participants see the loudest speaker; the loudest speaker may see the previous loudest speaker. This mode is typical for a small conference.

A participant with proper authorization can exclude one or more participants from the audio or video mix. An indication that they are not being seen/heard might be displayed to the affected participants.

A participant with proper authorization can manipulate the gain level associated with one or more audio streams in the mix.

4.3. Conference Sidebar Scenario

An authorized participant creates a sidebar. The participant selects whether the sidebar should include the media from the main conference or not and the audio gain level associated with the main conference audio.

A participant invites participants to the sidebar, and upon acceptance they start receiving the sidebar media as specified by the sidebar creator. If the new participant is not a participant of the conference, but is just a participant of the sidebar, the participant only receives the sidebar media without the media of the main conference.

A participant with the right authorization can move another participant into the sidebar with no indication, in which case the participant suddenly starts receiving the sidebar media.

Sidebar participants with the right authorization can select to hear or not to hear the main conference audio mixed with the sidebar audio.

A participant can be a participant to more than one sidebar but can only actively participate in one.

A participant can jump back and forth between the main conference and one or more sidebars.

4.4. Coaching Scenario

This is a call center or a remote training session where there is a supervisor who can monitor the conference. The supervised participants may be the call center operators or the teachers. A participant in the conference may be a supervised participant or a "customer".

The supervisor is a hidden participant and is not part of the participant roster.

The supervised participants might get an announcement/tone indicating that the supervisor has joined. The other participants do not hear the announcement.

The supervisor listens to or sees the session but can only be heard or seen by the supervised participant.

The supervisor can become a normal participant, in which case the participants see the supervisor as part of the roster and start hearing and seeing him.

4.5. Presentation and Q & A Session

An example is an earning call scenario in which a group of presenters delivers material to a group of people. After the presentation is finished, a Q & A session is opened.

The conference is created as a panel, and the panel participants are identified. Only their streams are mixed.

After the end of the presentation, the session chair changes the conference type to normal, and now streams from all participants may be mixed. Alternatively, a floor control protocol can be used. The chair can grant the right to speak by adding the participant, whose turn it is to ask a question, to the conference mix.

4.6. Presence-Enabled Ad Hoc Conference

A presence-enabled ad hoc conference, sometimes described as "walkie talkie" service, is a scenario in which a participant sends media to the other participants of the conference after receiving a confirmation of the other participants' availability. For example, a participant presses a talk button, which checks the presence of the participants to see if they are available for communication. If they are, a confirmation tone is played, and the participant can then talk; as a result, the media is sent to the other participants in the conference. These types of conferences tend to be long lived, hence the need for presence to ensure that the other participants are still available. The ad hoc nature of the conference means that the participant list can be changed at any time. Floor control can be used to allow other participants to speak, as the conference is usually half-duplex in nature.

4.7. Group Chat Text Conferencing

Group chat is a common scenario for text messaging in which a participant joins (or enters) a chat room in which text messages from participants are rendered in a single window and attributed to the participant that sent the message. Changes in conference membership are often announced in the text window itself (e.g., "Alice has just entered the room. Bob has just departed."). Note that a real-time transcription/closed captioning service can provide a similar window in which audio media is converted into interactive text. "Nicknames" or aliases are often chosen by participants or assigned by the Conferencing System and used as handles within the room.

4.8. Interactive Text

Interactive text uses RTP to carry text one character at a time, providing real-time interactivity, as described in RFC 4103 [2]. The interactive text session may be the main conference itself, or it may be used in conjunction with other media types. Interactive text may be used to represent the audio in the conference using some translation services. There can be more than one such stream where each text stream is in a different language. These text streams may be used as subtitles to the audio stream. The translation from to text to speech and back is done by transcoders. These transcoders have similar functionality to transcoders between different audio or video algorithms.

The conference participants should be able to select to receive text streams with the conference audio or those without it.

4.9. Moderated Group Chat

A moderated group chat scenario for text messaging is similar to group chat, but all text messages sent to the group are filtered/approved by a moderator. Note that the moderator can be a human or an application. The moderator also often has the ability to remove participants and provide feedback on their submissions (e.g., provide warnings before removal).

4.10. Text Sidebars

Interactive text or instant messaging sidebars are perhaps the most common sidebars in conferences today. Often the text sessions are separate from the conference. However, there are some advantages to having text sessions be a sidebar and as a result a part of the main conference. For example, a conference that is providing anonymity/aliases to participants can also provide anonymous/alias sidebars. A text sidebar can also benefit from other security/logging/recording services provided by the Conferencing System.

Another use of a text sidebar is a text-only conversation/discussion between two or more conference participants who are following the main conference at the same time.

4.11. Conference Announcements

The conference moderator may be able to play announcements to all the conference participants. An announcement may be prerecorded or composed by the moderator before it is sent. The announcements may be text, audio, or audio-visual. An example is a conference with several audio break-out sessions going on. At some point, the moderator wants to record an audio message like "In 5 minutes, everyone please come back to the main meeting" and then play that message to all the breakout sessions.

5. Security Considerations

Conferences generally have authorization rules about who may or may not join a conference, what type of media may or may not be used, etc. This information, sometimes called the conference policy or common conference information, is used by the Conferencing System to admit or deny participation in a conference. For the conference policy to be implemented, the Conferencing System needs to be able to authenticate potential participants. The methods used depend on the signaling protocols used by the conference. This can include a challenge/response mechanism, certificates, shared secret, asserted identity, etc.

Conferences often require that their content be confidential. In addition, secure authorization of participants is incomplete if access to the media can be gained by unauthorized participants. Functions for securing the media and for key management and distribution to authorized participants need to be provided by the Conferencing System. In some cases, the functions used for participant authorization can be leveraged for this purpose.

Privacy is an important aspect of conferencing. Users may wish to join a conference without anyone knowing that they have joined, in order to silently listen in. In other applications, a participant may wish just to hide their identity from other participants, but otherwise let them know of their presence. These functions need to be provided by the Conferencing System.

These conference-specific security requirements are discussed further in the XCON framework document.

6. Acknowledgements

Thanks to Brian Rosen for contributing conferencing scenarios.

Thanks to Alan Johnston for going over the document and adding some more scenarios; to Keith Lantz, Mary Barnes, and Dave Morgan for carefully reading the document.

7. Informative References

[1] Rosenberg, J., "A Framework for Conferencing with the Session Initiation Protocol (SIP)", RFC 4353, February 2006.

[2] Hellstrom, G. and P. Jones, "RTP Payload for Text Conversation", RFC 4103, June 2005.

[3] Barnes, M., "A Framework and Data Model for Centralized Conferencing", Work in Progress, June 2006.

Authors' Addresses

Roni Even
Polycom
94 Derech Em Hamoshavot
Petach Tikva 49130
Israel

EMail: roni.even@polycom.co.il

Nermeen Ismail
Cisco Systems, Inc.
170 West Tasman Drive
San Jose 95134
CA USA

EMail: nismail@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).