Abstract

The current work of SIPREC will soon finish. As conferences are the key requirement for some environments, it is worth to explore several extensions and additional functionality to support multimedia conference recording. SIPREC is not sufficient to record all the conference sessions via certain interactive media channels, like multi-user chat or screen sharing.

This draft tries to show the use cases for multimedia conference recording and the requirements for how to work well under SIPREC mechanism. The requirements ask for extensions to SIP that will manage delivery of RTP media sessions, including content media defined by [RFC4796], and MSRP media sessions to a recording device. The recorded media sessions are all SIP-based.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."
1. Introduction

In general, a basic video conference has participants with video channels, audio channels and DTMF ability. An advanced multimedia conference would have extended channels like text, interactive text...
and presentation graphics [RFC4597]. These extended channels recording have the same strong needs as audio or video, especially in some conference use cases. The conference’s host and participants, even nonparticipants, would like to play back the recordings in real-time or non-real-time for different purposes, like editing summary, reviewing outlines or monitoring process. The recordings should have the ability to reconstruct the conference richly, with adequate media and metadata recorded, which are not only audio/video but also IM and shared content. Such an exhaustive reconstruction could give audiences more information and a better experience. The recorded sessions can be any RTP media sessions, including the content media (as defined by [RFC4796]) transferred as video stream, and MSRP media session which is SIP based media session.

2. Multimedia Conference Recording Overview

There is one use case (use case 11) covering the recording of a multi-channel and multimedia session in the existing use case document [RFC6341]. Aside from audio, video, DTMF (as defined by [RFC4733]) and text (as defined by [RFC4103]), it does not include other interactive channels. The limitations to the multi-channel types leads to poor support for recording multimedia conferences. A multimedia conference has various channels, including audio, video, IM, data sharing (screen/document/application), etc. SIPREC is mostly capable of recording any sort of RTP media sessions, including voice, video, DTMF (as defined by [RFC4733]) and text [RFC6341] with SDP negotiation [I-D.ietf-siprec-protocol] and certain metadata [I-D.ietf-siprec-metadata]. But it is not evident how to support the remaining media, like multi-user chat or screen sharing.

Multi-user chat is one of key cases about the IM session in multimedia conference. A multi-user chat or simple-chat session is to handle the media to relay instant messages received from one participant to the rest of the participants in the conference [I-D.ietf-simple-chat], especially for the MSRP session which is a SIP based system. The host and participants in a conference might start MSRP sessions among each other for public group chat, sidebar chat or whisper chat. These MSRP content could be replicated by SRC (might be the MSRP switch or certain MSRP replay or MSRP client) to deliver to SRS via special RS channel(s). The replicated content could be the Message/CPIM message that contains text, HTML and images, etc. The document only considers the SIP based system, recording XMPP based IM in CS is out of the scope.

The data sharing session, known as content sharing or content streams as well, in a multimedia conference’s CS has functionality like screen sharing, application sharing, document sharing, etc. These data streams could be managed as still images (snapshots with
increments) or as dynamic streams (video streams) to carry details like slide presentation, annotations, direct editing or page turning. Especially, the screen sharing would have different ways to get its data streams, like the video streams directly offered from VGA port or encoded/decoded by an application on peer’s client, or turned from multiple screenshots, or even the still images carried by MSRP channels.

One way for a conference focus/mixer to record a conference is introduced in [I-D.ietf-siprec-architecture]. This defines how the conference focus works as a SRC to deliver RTP streams and associate recording metadata with SRS. It may choose the recording RTP stream type, separated or mixed. There are more details about how to use SDP, RTP for recording by participant or by media type in [I-D.ietf-siprec-protocol]. The focus may setup different recording sessions for different media streams recorded separately, or one recording session for a mixed media stream created by the SRC, or even multiplexing different media streams in a single RTP recording session [I-D.ietf-siprec-protocol].

But more is needed to support other media streams in a multimedia conference. There is need for MSRP switch/relay/client as SRC to replicate MSRP sessions to the recorder, with a new "media stream" type in RS for delivering MSRP based streams or contents. There is another need for a new type of metadata to indicate when media streams are being used to represent screen/application/data sharing sessions, distinct from the main video streams. There is also need of a mechanism for the SRC to bound the number of media streams to be recorded, especially when the participants number in a conference is extremely large. The document does not include recording any of the extended session attributes being defined by the CLUE WG.

3. Definitions

Instant Message Stream: instant message stream refers to the streams transferred by messages between users in near real-time [RFC3248].

Data Sharing: Data sharing is to use a content channel for collaboratively working on documents, files, images, desktops, etc in real time. It is also called as content sharing, including application sharing, screen sharing, document sharing, etc.

Application Sharing: application sharing is the sharing of the graphical user-interface of an application amongst multiple users simultaneously in real time. The slide sharing could be one special case.
Screen Sharing: screen sharing is the sharing of a computer desktop amongst multiple users simultaneously in real time, also called desktop sharing. Comparing to application sharing, which is always a single one, screen sharing is for the whole screen.

Document Sharing: document sharing is the sharing to help multiple users work simultaneously on a single document or file to achieve a single final version. It is also called file sharing or document collaboration.

Audio/Video Conference: Audio/video conference is one sort of various conferences. In SIP, an audio/video conference is an instance of a multi-party conversation that matches the definition in [RFC4353] and the framework in [RFC5239], with the media channels as audio and video.


Multimedia Conference: multimedia conference is the multi-party conversation including any combination of different media types such as audio, video, text, interactive text, or presentation graphics [RFC4597].

4. Use cases

4.1. MSRP Instant Message Stream Recording

Use Case 1: MSRP Instant Message Stream Recording.

Instant message is the function offered to chat between/among peers. There are page mode and session mode (MSRP [RFC6914]). Here we are concerned with instant message sessions in the context of a point to point call using session mode, which treats the MSRP instant message session as a media type.

For example, in a call center or emergency (first-responder) center, a customer could use the web client to start a chat with an agent about his questions or describing the situation happened around him. Call center or emergency center would need to record those chat sessions between customers and agents.

4.2. Screen Sharing Stream Recording

Use Case 2: Screen Sharing Stream Recording.
This is also well known as desktop sharing or remote sharing between peers. This function could be also used in a point to point call directly.

In an enterprise, the softphone calls between colleagues would choose screen sharing to illustrate their views clearly if the voice discussing is not enough. The enterprise would ask to record those screen sharing sessions for security check.

Another example is the remote education training, which needs to record the screen to keep track of the whole training class.

The state of the shared screen is recorded as video (and possibly audio) media, with metadata indicating that the stream represents screen sharing.

4.3. Application Sharing Stream Recording

Use Case 3: Application Sharing Stream Recording.

Users would choose application sharing instead of screen sharing to avoid exposing the privacy content on their computer desktop, when have the point-to-point call with others or have a conference call. And the recorded streams are the content of applications shown in CS.

The state of the application window is recorded as video (and possibly audio) media, with metadata indicating that the stream represents application sharing.

4.4. Document Sharing Stream Recording

Use Case 4: Document Sharing Stream Recording.

Users would work on one document simultaneously in real time. The content of document would be recorded (which is close to use case 3).

The state of the document sharing window is recorded as video media, with metadata indicating that the stream represents data sharing.

4.5. Audio/Video Conference Recording

Use Case 5: audio/video conference Recording.

The recording for audio/video conference is basic. All the channels in a conference would be recorded as one mixed stream or as separate streams by participants. It has been supported by current SIPREC mechanism.
4.6. Chat Conference Recording

Use Case 6: chat conference Recording.

There is another Conference type, known as multi-user chat conference or chat-rooms. The chat conference would have participants in a conference to chat or text each other with nicknames and provide private chat using the Message Session Relay Protocol (MSRP) [I-D.ietf-simple-chat]. In this case, there would be a need to record the chat content and details like nicknames.

4.7. Multimedia conference Recording

Use Case 7: Multimedia conference Recording.

This is a special use case to indicate the multimedia conference recording environment. When there is a common education class or skill training conference, the audiences who are not in the conference session would prefer replaying the conference in real-time with professor or lecturer’s voice with their slides, better include the video of if available. The recording of this education conferences need to record audio, video from hosts and data sharing of theirs.

While the audiences(out-conference) need to know what the audiences (in-conference) feedback to the training, they might also want to know what those audiences(in-conference) have been discussing in IM session. Thus the recording need record the IM sessions.

5. Requirements

5.1. REQ-001

REQ-001: The mechanism MUST support recording of MSRP media sessions. This requirement derives from use cases 1,6,7.

5.2. REQ-002

REQ-002: The mechanism MUST support recording of screen sharing using audio and video media. This requirement derives from use cases 2,7.

5.3. REQ-003

REQ-004: The mechanism MUST support recording of application sharing using audio and video. This requirement derives from use cases 3,7.
5.4. REQ-004

REQ-003: The mechanism MUST support recording of document sharing using video media. This requirement derives from use cases 4,7.

5.5. REQ-005

REQ-005: The mechanism MUST support metadata or SDP to identify media streams used to record screen/application/document sharing. This requirement derives from use cases 2,3,4,7.

6. IANA Considerations

This document contains no IANA considerations.

7. Security Considerations

Not explicitly covered in this version.

8. References

8.1. Normative References


8.2. Informative References


Authors’ Addresses

Paul H. Kyzivat
Huawei

Email: pkyzivat@alum.mit.edu

Michael Yan
Huawei

Email: michael.yan@huawei.com

Simon Pietro Romano
University of Napoli

Email: spromano@unina.it