SIPREC
Internet-Draft
Intended status: Informational
Expires: December 11, 2011

K. Rehor, Ed.
Cisco Systems
L. Portman, Ed.
NICE Systems
A. Hutton
Siemens Enterprise
Communications
R. Jain
IPC Systems
June 09, 2011

Use Cases and Requirements for SIP-based Media Recording (SIPREC) draft-ietf-siprec-req-12

Abstract

Session recording is a critical requirement in many business communications environments such as call centers and financial trading floors. In some of these environments, all calls must be recorded for regulatory and compliance reasons. In others, calls may be recorded for quality control or business analytics.

Recording is typically performed by sending a copy of the session media to the recording devices. This document specifies requirements for extensions to SIP that will manage delivery of RTP media to a recording device. This is being referred to as SIP-based Media Recording.

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."

This Internet-Draft will expire on December 11, 2011.

Copyright Notice

Table of Contents

1. Introduction

Session recording is a critical operational requirement in many businesses, especially where voice is used as a medium for commerce and customer support. A prime example where voice is used for trade is the financial industry. The call recording requirements in this industry are quite stringent. The recorded calls are used for dispute resolution and compliance. Other businesses such as customer support call centers typically employ call recording for quality control or business analytics, with different requirements.

Depending on the country and its regulatory requirements, financial trading floors typically must record all calls. In contrast, call centers typically only record a subset of the calls, and calls must not fail regardless of the availability of the recording device.

Respecting the privacy rights and wishes of users engaged in a call is of paramount importance. In many jurisdictions participants have a right to know that the session is being recorded or might be recorded, and have a right to opt out, either by terminating the call or by demanding that the call not be recorded. Therefore this document contains requirements for being able to notify users that a call is being recorded and for users to be able to request that a call not be recorded. Use cases where users participating in a call are not informed that the call is or might be recorded are outside the scope of this document. In particular, lawful intercept is outside the scope of this document.

Furthermore, one-size-fits-all model will not fit all markets where the scale and cost burdens vary widely having different needs for solution capabilities such as media injection, transcoding, and security. If a standardized solution supports all of the requirements from every recording market, but doing so would be expensive for markets with lesser needs, then proprietary solutions for those markets will continue to propagate. Care must be taken, therefore, to make a standards-based solution support optionality and flexibility.

This document specifies requirements for using SIP [RFC3261] between a Session Recording Client and a Session Recording Server to control the recording of media that has been transmitted in the context of a Communication Session. A Communication Session is the "call" between participants. The Session Recording Client is the source of the recorded media. The Session Recording Server is the sink of recorded media. It should be noted that the requirements for the protocol between a Session Recording Server and Session Recording Client have very similar requirements (such as codec and transport negotiation, encryption key interchange, firewall traversal) as compared to

regular SIP media sessions. The choice of SIP for session recording provides reuse of an existing protocol.

The recorded sessions can be any RTP media sessions including voice, DTMF (as defined by [RFC4733]), video, and text (as defined by [RFC4103]).

An archived session recording is typically comprised of the Communication Session media content and the Communication Session Metadata. The Communication Session Metadata allows recording archives to be searched and filtered at a later time and allows a session to be played back in a meaningful way, e.g., with correct synchronization between the media. The Communication Session Metadata needs to be conveyed from the Session Recording Client to the Session Recording Server.

This document only considers active recording, where the Session Recording Client purposefully streams media to a Session Recording Server. Passive recording, where a recording device detects media directly from the network, is outside the scope of this document.

2. Requirements notation

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 [RFC2119] and indicate requirement levels for compliant mechanisms.

3. Definitions

Session Recording Server (SRS): A Session Recording Server (SRS) is a SIP User Agent (UA) that is a specialized media server or collector that acts as the sink of the recorded media. An SRS is typically implemented as a multi-port device that is capable of receiving media from multiple sources simultaneously. An SRS is the sink of the recorded session metadata.

Session Recording Client (SRC): A Session Recording Client (SRC) is a SIP User Agent (UA) that acts as the source of the recorded media, sending it to the SRS. An SRC is a logical function. Its capabilities may be implemented across one or more physical devices. In practice, an SRC could be a personal device (such as a SIP phone), a SIP Media Gateway (MG), a Session Border Controller (SBC) or a SIP Media Server (MS) integrated with an Application Server (AS). This specification defines the term SRC such that all such SIP entities can be generically addressed under one definition. The SRC provides

metadata to the SRS.

Communication Session (CS): A session created between two or more SIP User Agents (UAs) that is the subject of recording.

Recording Session (RS): The SIP session created between an SRC and SRS for the purpose of recording a Communication Session.

Figure 1 pictorially represents the relationship between a Recording Session and Communication Session.

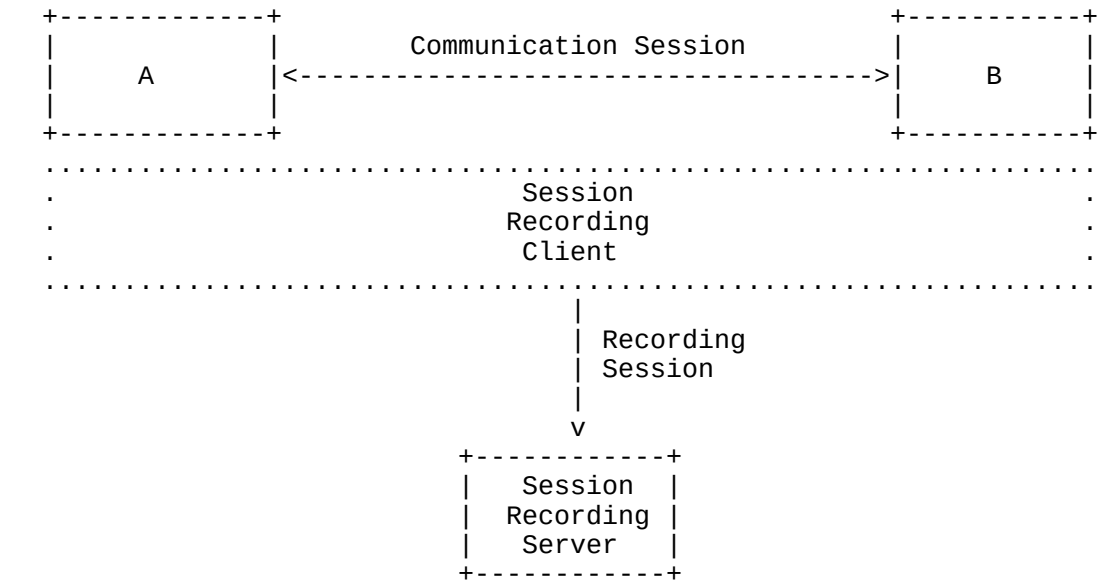


Figure 1

Metadata: Information that describes recorded media and the CS to which they relate.

Pause and Resume during a Communication Session: Pause: The action of temporarily discontinuing the transmission and collection of RS media Resume: The action of recommencing the transmission and collection of RS media

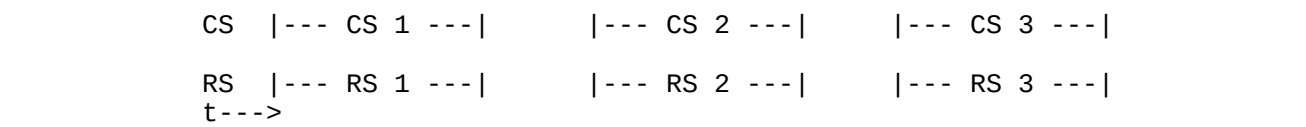
Most security-related terms in this document are to be understood in the sense defined in [RFC4949]; such terms include, but are not limited to, "authentication", "confidentiality", "encryption",

"identity", and "integrity".

4. Use Cases

Use Case 1: Full-time Recording: One Recording Session for each Communication Session.

For example, the diagram below shows the lifecycle of Communication Sessions (CS) and the relationship to the Recording Sessions (RS)



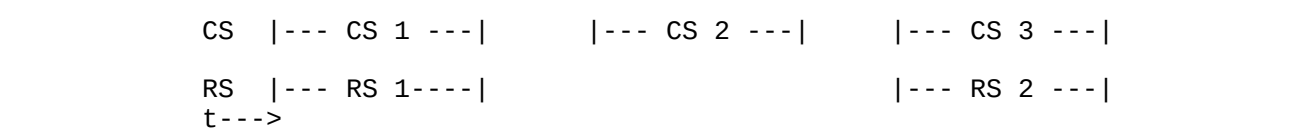
Record every CS for specific extension/person.

The need to record all calls is typically due to business process purposes (such as transaction confirmation or dispute resolution) or to ensure compliance with governmental regulations. Applications include enterprise, contact center, and financial trading floors.

Also commonly known as Total Recording.

Use Case 2: Selective Recording: Start a Recording Session when a Communication Session to be recorded is established.

In this example, Communication Sessions 1 and 3 are recorded but CS 2 is not.



Use Case 3: Start/Stop a Recording Session during a Communication Session.

The Recording Session starts during a Communication Session, either manually via a user-controlled mechanism (e.g. button on user's phone) or automatically via an application (e.g. a Contact Center customer service application) or business event. A Recording Session either ends during the Communication Session, or when the

Communication Session ends. One or more Recording Sessions may record each Communication Session.

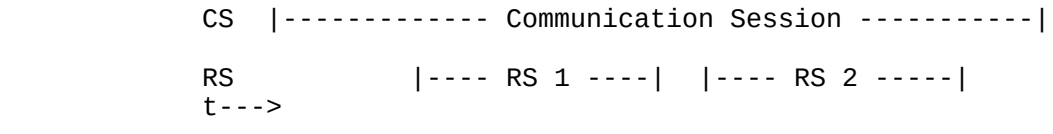


Figure 4

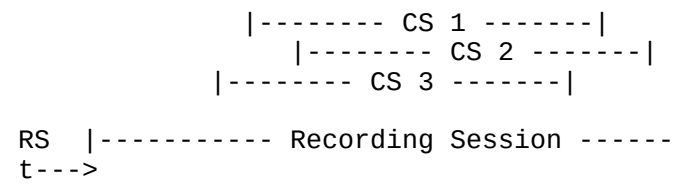
Use Case 4: Persistent Recording: A single Recording Session captures one or more Communication Sessions.

Figure 5

A Recording Session records continuously without interruption. Periods when there is no CS in progress must be reproduced upon playback (e.g. by recording silence during such periods or by not recording such periods but marking them by means of metadata for utilization on playback, etc.). Applications include financial trading desks and emergency (first-responder) service bureaus. The length of a Persistent Recording Session is independent from the length of the actual Communication Sessions. Persistent Recording Sessions avoid issues such as media clipping that can occur due to delays in Recording Session establishment.

The connection and attributes of media in the Recording Session are not dynamically signaled for each Communication Session before it can be recorded; however, codec re-negotiation is possible.

In some cases, more than one concurrent Communication Session (on a single end-user apparatus, e.g. trading floor turret) is mixed into one Recording Session:



Use Case 5: Real-time Recording Controls.

For an active Recording Session, privacy or security reasons may demand not capturing a specific portion of a conversation. An example is for PCI (payment card industry) compliance where credit card info must be protected. One solution is to not record a caller speaking their credit card information.

An example of a real-time controls is Pause/Resume.

Use Case 6: IVR / Voice Portal Recording.

Self-service Interactive Voice Response applications may need to be recorded for application performance tuning or to meet compliance requirements.

Metadata about an IVR session recording must include session information and may include application context information (e.g. VoiceXML session variables, dialog names, etc.)

Use Case 7: Enterprise Mobility Recording.

Many agents and enterprise workers whose calls are to be recorded are not located on company premises.

Examples:

o Home-based agents or enterprise workers.

o Mobile phones of knowledge workers when they conduct work related (and legally required recording) calls. e.g. insurance agents, brokers, physicians.

Use Case 8: Geographically distributed or centralized recording.

Enterprises such as banks, insurance agencies, and retail stores may have many locations, possibly up to thousands of small sites. Frequently only phones and network infrastructure are installed in branches, without local recording services. In cases where calls inside or between branches must be recorded, a centralized recording system in data centers together with telephony infrastructure (e.g. PBX) may be deployed.

Use Case 9: Record complex call scenarios.

The following is an example of a scenario where one call that is recorded must be associated with a related call that also must be recorded.

o A Customer is in a conversation with a Customer Service Agent.

o Agent puts Customer on hold in order to consult with a Supervisor.

o Agent enters into a conversation with Supervisor.

o Agent disconnects from Supervisor, then reconnects with Customer.

o The Supervisor call must be associated with the original customer call.

Use case 10: High availability and continuous recording.

Specific deployment scenarios present different requirements for system availability, error handling, etc. including:

o An SRS must always be available at call setup time.

o No loss of media recording, including during failure of an SRS.

o The Communication Session must be terminated (or suitable notification given to parties) in the event of a recording failure.

Use Case 11: Record multi-channel, multi-media session.

Some applications require the recording of more than one media stream, possibly of different types. Media are synchronized, either at storage or at playback.

Speech analytics technologies (e.g. word spotting, emotion detection, speaker identification) may require speaker-separated recordings for optimum performance.

Multi-modal Contact Centers may include audio, video, IM or other

interaction modalities.

In trading floors environments, in order to minimize storage and recording system resources, it may be preferable to mix multiple concurrent calls (Communication Sessions) on different handsets/ speakers on the same turret into single recording session.

Use Case 12: Real-time media processing.

It must be possible for an SRS to support real-time media processing, such as speech analytics of trading floor interactions. Real-time analytics may be employed for automatic intervention (stopping interaction or alerting) if for example, a trader is not following regulations.

Speaker separation is required in order to reliably detect who is saying specific phrases.

5. Requirements

The following are requirements for SIP-based Media Recording:

o REQ-001 The mechanism MUST provide a means for using the SIP protocol for establishing, maintaining and terminating Recording Sessions between a Session Recording Client and a Session Recording Server.

o REQ-002 The mechanism MUST support the ability to record all CSs in their entirety.

o REQ-003 The mechanism MUST support the ability to record selected CSs in their entirety, according to policy.

o REQ-004 The mechanism MUST support the ability to record selected parts of selected CSs.

o REQ-005 The mechanism MUST support the ability to record a CS without loss of media of RS (for example, clipping media at the beginning of the CS) due to RS recording preparation and also, without impacting the quality or timing of the CS (for example, delaying the start of the CS while preparation for recording session). See Use Case 4 in Section 4 for more details.

o REQ-006 The mechanism MUST support the recording of IVR sessions.

o REQ-007 The mechanism MUST support the recording of RTP media types voice, DTMF (as defined by [RFC4733]), video, and text (as defined by

[RFC4103]).

o REQ-008 The mechanism MUST support the ability for an SRC to deliver mixed audio streams from multiple Communication Sessions to an SRS.

Note: A mixed audio stream is where several related Communication Sessions are carried in a single Recording Session. A mixed media stream is typically produced by a mixer function. The RS MAY be informed about the composition of the mixed streams through session metadata.

o REQ-009: The mechanism MUST support the ability for an SRC to deliver mixed audio streams from different parties of a given Communication Session to an SRS.

o REQ-010 The mechanism MUST support the ability to deliver to the SRS multiple media streams for a given CS.

o REQ-011 The mechanism MUST support the ability to pause and resume the transmission and collection of RS media.

o REQ-012 The mechanism MUST include a means for providing the SRS with metadata describing CSs that are being recorded, including the media being used and the identifiers of parties involved.

o REQ-013 The mechanism MUST include a means for the SRS to be able to correlate RS media with CS participant media.

o REQ-014 Metadata format must be agnostic of the transport protocol.

o REQ-015: The mechanism MUST support a means to stop the recording.

o REQ-016: The mechanism MUST support a means for a recording-aware UA involved in a CS to request at session establishment time that the CS should be recorded or should not be recorded, the honoring of such a request being dependent on policy.

o REQ-017: The mechanism MUST support a means for a recording-aware UA involved in a CS to request during a session that the recording of the CS should be started, paused, resumed or stopped, the honoring of such a request being dependent on policy. Such recording-aware UA MUST be notified about outcome of such requests.

o REQ-018 The mechanism MUST NOT prevent the application of tones or announcements during recording or at the start of a CS to support notification to participants that the call is being recorded or may be recorded.

o REQ-019 The mechanism MUST provide a means of indicating to recording-aware UAs whether recording is taking place, for appropriate rendering at the user interface.

o REQ-020 The mechanism MUST provide a way for metadata to be conveyed to the SRS incrementally during the CS.

o REQ-021 The mechanism MUST NOT prevent high availability deployments.

o REQ-022 The mechanism MUST provide means for facilitating synchronization of the recorded media streams and metadata.

o REQ-023 The mechanism MUST provide means for facilitating synchronization among the recorded media streams.

o REQ-024 The mechanism MUST provide means to relate recording and recording controls such as start/stop/pause/resume to the wall clock time.

o REQ-025 The mechanism MUST provide means for an SRS to authenticate the SRC on RS initiation.

o REQ-026 The mechanism MUST provide means for an SRC to authenticate the SRS on RS initiation.

o REQ-027 The mechanism MUST include a means for ensuring that the integrity of the metadata sent from SRC to SRS is an accurate representation of the original CS metadata.

o REQ-028 The mechanism MUST include a means for ensuring that the integrity of the media sent from SRC to SRS is an accurate representation of the original CS media.

o REQ-029 The mechanism MUST include a means for ensuring the confidentiality of the Metadata sent from SRC to SRS.

o REQ-030 The mechanism MUST provide a means to support RS confidentiality.

o REQ-031 The mechanism MUST support the ability to deliver to the SRS multiple media streams of the same media type (e.g. audio, video). For example in the case of delivering unmixed audio for each participant in the CS.

6. Privacy Considerations

Respecting the privacy rights and wishes of users engaged in a call is of paramount importance. In many jurisdictions participants have a right to know that the session is being recorded or might be recorded, and have a right to opt out, either by terminating the call or by demanding that the call not be recorded. Therefore this document contains requirements for being able to notify users that a call is being recorded and for users to be able to request that a call not be recorded. Use cases where users participating in a call are not informed that the call is or might be recorded are outside the scope of this document. In particular, lawful intercept is outside the scope of this document.

Requirements for participant notification of recording vary widely by jurisdiction. In a given deployment, not all users will be authorized to stop the recording of a CS (although any user can terminate its participation in a CS). Typically users within the domain that is carrying out the recording will be subject to policies of that domain concerning whether CSs are recorded. For example, in a call centre, agents will be subject to policies of the call centre and may or may not have the right to prevent the recording of a CS or part of a CS. Users calling into the call centre, on the other hand, will typically have to ask the agent not to record the CS. If the agent is unable to prevent recording, or if the caller does not trust the agent, the only option generally is to terminate the CS.

Privacy considerations also extend to what happens to a recording once it has been created. Typical issues are who can access the recording (e.g., receive a copy of the recording, view the metadata, play back the media, etc.), for what purpose the recording can be used (e.g., for training purposes, for quality control purposes, etc.) and for how long the recording is to be retained before deletion. These are typically policies of the domain that makes the recording, rather than policies of individual users involved in a recorded CS, whether those users be in the same domain or in a different domain. Taking the call centre example again, agents might be made aware of call centre policy regarding retention and use of recordings as part of their employment contract, and callers from outside the call centre might be given some information about policy when notified that a CS will be recorded (e.g., through an announcement that says that calls may be recorded for quality purposes).

This document does not specify any requirements for a user engaged in a CS to be able to dictate policy for what happens to a recording, or for such information to be conveyed from an SRC to an SRS. It is assumed that the SRS has access to policy applicable to its

environment and can ensure that recordings are stored and used in accordance with that policy.

7. Security Considerations

Session recording has substantial security implications, for the SIP UA's being recorded, the SRC, and the SRS.

For the SIP UA's involved in the Communication Session, the requirements in this draft enable the UA to identify that a Communication Session is being recorded and for the UA to request that a given Communication Session is not subject to recording.

Since humans don't typically look at or know about protocol signaling such as SIP, and indeed the SIP session might have originated through a PSTN Gateway without any ability to pass on in-signaling indications of recording, users can be notified of recording in the media itself through voice announcements, a visual indicator on the endpoint, or other means.

With regards to security implications of the protocol(s), clearly there is a need for authentication, authorization and eavesdropping protection for the solution. The SRC needs to know the SRS it is communicating with is legitimate, and vice-versa, even if they are in different domains. Both the signaling and media for the Recording Session need the ability to be authenticated and protected from eavesdropping. Requirements are detailed in the requirements section.

Communication Sessions and Recording Sessions can require different security levels both for signaling and media, depending on deployment configurations. For some environments, for example, SRS and SRC will be collocated in a secure network region and therefore the RS will not require the same protection level as a CS that extends over a public network, for example. For other environments, the SRS can be located in a public cloud, for example, and the RS will require a higher protection level than the CS. For these reasons, there is not a direct relationship between the security level of Communication Sessions and the security level of Recording Sessions.

A malicious or corrupt SRC can tamper with media and metadata relating to a CS before sending to an SRS. Also CS media and signaling can be tampered with in the network prior to reaching an SRC, unless proper means are provided to ensure integrity protection during transmission on the CS. Means for ensuring the correctness of media and metadata emitted by an SRC are outside the scope of this work. Other organizational and technical controls will need to be

used to prevent tampering.

8. IANA Considerations

This document has no IANA actions.

9. Acknowledgements

Thanks to Dan Wing, Alan Johnson, Vijay Gurbani, Cullen Jennings, Hadriel Kaplan, Henry Lum, Dave Smith, Martin Palmer, Alissa Cooper, Deepanshu Gautam, Paul Kyzivat, Parthasarathi R, Ram Mohan R, and Charles Eckel for their significant contributions and assistance with this document and Working Group, and to all the members of the DISPATCH WG and SIPREC WG mailing lists for providing valuable input to this work.

10. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[RFC4103] Hellstrom, G. and P. Jones, "RTP Payload for Text Conversation", RFC 4103, June 2005.

[RFC4733] Schulzrinne, H. and T. Taylor, "RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals", RFC 4733, December 2006.

[RFC4949] Shirey, R., "Internet Security Glossary, Version 2", RFC 4949, August 2007.

Authors' Addresses

Ken Rehor (editor)
Cisco Systems
170 West Tasman Dr.
Mail Stop SJC30/2/
San Jose, CA 95134
USA

Email: krehor@cisco.com

Leon Portman (editor)
NICE Systems
8 Hapnina
Ra'anana 43017
Israel

Email: leon.portman@nice.com

Andrew Hutton
Siemens Enterprise Communications

Email: andrew.hutton@siemens-enterprise.com
URI: http://www.siemens-enterprise.com

Rajnish Jain
IPC Systems
777 Commerce Drive
Fairfield, CT 06825
USA

Email: rajnish.jain@ipc.com