DISPATCH K. Rehor, Ed.
Internet-Draft Cisco Systems
Intended status: Informational R. Jain
Expires: April 1, 2011 IPC Systems
L. Portman
NICE Systems
A. Hutton
Siemens Enterprise Communications
September 28, 2010

Requirements for SIP-based Media Recording (SIPREC)
draft-ietf-siprec-req-02

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 done 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 from an end-point that originates media (or that has access to it) to a recording device. This is being referred to as SIP-based Media Recording.

Status of this Memo

Copyright Notice

Table of Contents

1. 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.

2. 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. The recorded media content must be an exact copy of the actual conversation (i.e. clipping and loss of media are unacceptable). A new call attempt would be automatically rejected if the recording device becomes temporarily unavailable. An existing call would be dropped in the same situation. In contrast, support call centers typically only record a subset of the calls, and calls must not fail regardless of the availability of the recording device.

Furthermore, the scale and cost burdens vary widely, in all markets, where the different needs for solution capabilities such as media injection, transcoding, and security-related needs do not conform well to a one-size-fits-all model. 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." 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, 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. (The requirements for session metadata delivery are specified separately [draft-ram-siprec-metadata-00]).

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. In addition, lawful intercept is outside the scope of this document.

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 a logical function that typically archives media for extended durations of time and provides interfaces for search and retrieval of the archived media. An SRS is typically implemented as a multi-port device that is capable of receiving media from several sources simultaneously. An SRS is typically also 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 itself or another entity working on its behalf (such as a SIP Application Server) may act as the source of the recording metadata.

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

```
+----------------------------------------------------------------+
|                  (Session Recording Client)                    |
|   +-------------+                           +-----------+      |
|   |             |   Communication Session   |           |      |
|   |      A      |<------------------------->|     B     |      |
|   |             |                           |           |      |
|   +-------------+                           +-----------+      |
+----------------------------------------------------------------+
```

```
                                      |
                                      | Recording
                                      | Session
                                      |
                                      v
                               +-------------+
                               |             |
                               |   Session   |
                               |  Recording  |
                               |   Server    |
                               |             |
                               +-------------+
```

Figure 1

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

SIPREC: The set of SIP extensions that supports recording of Communication Sessions.

Pause and Resume during a Communication Session: Pause: The action of temporarily discontinuing the recording of media during a CS. Resume: The action of recommencing the recording of media for a CS following a pause.

4. Example Deployment Architectures

A recording system deployment consists of the Recording Client and Recording Server. Recording Control is bi-directional; Recording Media and Communication Session (CS) Metadata are sent from the SRC

to SRS.

```
      +-------------+                          +--------------+
      |             |       CS Metadata        |              |
      |   Session   |------------------------->|   Session    |
      |  Recording  |                          |  Recording   |
      |   Client    |    Recording Control     |   Server     |
```

```
      |    (SRC)     |<------------------------->|     (SRS)     |
      |              |                           |               |
      |              |      Recorded Media       |               |
      |              |-------------------------->|               |
      |              |                           |               |
      +--------------+                           +---------------+
```

Figure 2

5. Use Cases

Use Case 1: Full-time Recording: One (or more, in the case of redundant recording) 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)

```
CS   |--- CS 1 ---|      |--- CS 2 ---|     |--- CS 3 ---|

RS   |--- RS 1 ---|      |--- RS 2 ---|     |--- RS 3 ---|
```

Figure 3

Record every call 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.

```
CS  |--- CS 1 ---|      |--- CS 2 ---|     |--- CS 3 ---|

RS  |--- RS 1----|                         |--- RS 2 ---|
```

Figure 4

Use Case 3: Dynamic Recording: 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 per Communication Session:

```
CS  |------------- Communication Session -----------|

RS           |---- RS 1 ----|  |---- RS 2 -----|
```

Figure 5

Also known as Mid-session or Mid-call Recording.

Use Case 4: Persistent Recording: A single Recording Session captures one or more Communication Sessions, in sequence (Fig. 6) or in parallel (Fig. 7). The recording session is a single RTP stream, therefore consists of a single offer/answer exchange. There may be mid-session RE-INVITE offer/answer exchanges for codec changes or for moving the RTP streams to handle failure scenarios.

```
          |--- CS 1 ---|      |--- CS 2 ---|     |--- CS 3 ---|

RS  |---------------------- Recording Session ---------------------|
```

Figure 6

A Recording Session records continuously without interruption. Applications include financial trading desks and emergency (first-responder) service bureaus. The length of a Persistent Recording Sessions 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. CS details and CS metadata will still be signaled, but can be post-correlated to the recorded media. There will still need to be a means of correlating the recorded media connection/packets to the Communication Session, however this may be on a permanent filter-type basis, such as based on a SIP AoR of an agent that is always recorded.

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:

```
              |-------- CS 1 -------|
                 |-------- CS 2 -------|
            |-------- CS 3 -------|

RS  |----------- Recording Session --------------|
```

Figure 7

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 (DTMF or ASR) 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 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. i.e. insurance agents, brokers, physicians.

Use Case 8: Geographically distributed or centralized recording.

Global banks with multiple branches up to thousands of small sites.

o Only phones and network infrastructure in branches, no recording services.

o Internal calls inside or between branches must be recorded.

o Centralized recording system in data centers together with telephony infrastructure (e.g. PBX).

Use Case 9: Record complex call scenarios.

Record a call that is associated with another call.

Example:

o Customer in conversation with Agent

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

o Agent in conversation with Supervisor.

o Agent disconnects from Supervisor, 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) 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 is 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 save 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.

Recorder must support real-time media processing, such as speech analytics.

Recording and real-time analytics of trading floor interactions (including video and instant messaging). Real time analytics is required for automatic intervention (stopping interaction or alert) if for example, trader is not following regulations.

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

6. Requirements

The following are requirements for SIP-based Media Recording:

o REQ-000 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-001 The mechanism MUST support the ability to record all CSs in their entirety.

o REQ-002 The mechanism MUST support the ability to record selected

CSs in their entirety, according to policy.

o REQ-003 The mechanism MUST support the ability to record selected parts of selected CSs.

o REQ-004 The mechanism MUST support the ability to record a CS without an intentional loss of media (for example, clipping media at the beginning of the CS) and without impacting the quality or timing of the CS (for example, delaying the start of the CS while preparation for recording takes place). See Use Case 4 in Section 5.

o REQ-005 The mechanism MUST support establishing Recording Sessions from the SRC to the SRS. This requirement typically applies when the decision about whether a session should be recorded or not resides in the SRC.

o REQ-006 The mechanism MUST support establishing Recording Sessions from the SRS to the SRC (SRS initiates recording). This requirement typically applies when the decision about whether a session should be recorded or not resides in the SRS.

o REQ-007 The mechanism MUST support the recording of IVR sessions.

o REQ-008 The mechanism SHOULD support SRS failover and migration of Recording Sessions to a working SRS without disconnecting the

Communication Sessions.

o REQ-009 A request for a new Recording Session MUST be rejected by the SRS if service is unavailable (e.g. system overload, disk full, etc.)

o REQ-010 A request for a new Recording Session MUST be redirected to an available SRS.

o REQ-012 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 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-012bis: 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-013 The mechanism MUST support the ability to deliver multiple

media streams for a given Communication Session over separate Recording Sessions to the SRS.

o REQ-014 The mechanism MUST support the ability to deliver multiple media streams for a given Communication Session over a single Recording Session to the SRS.

o REQ-015 The mechanism MUST support the ability to pause and resume the Recording Session from the SRC.

o REQ-016 The mechanism MUST support the ability to pause and resume the Recording Session from the SRS.

o REQ-017 The mechanism MUST provide the SRS with metadata describing CSs that are being recorded, including the media being used and the identities of parties involved.

o REQ-018 The mechanism MUST provide the SRS with the means to correlate RS media with CS participant media described in metadata.

o REQ-019 The mechanism MUST support the ability to transport the metadata in the same SIP dialog as the Recording Session.

o REQ-020 The mechanism MUST support the ability to transport the metadata outside of the Recording Session SIP dialog.

o REQ-021 Metadata format must be agnostic of the transport protocol.

o REQ-022: The mechanism MUST support a means to cancel and discard a Recording Session during a Communication Session.

o REQ-023 The mechanism MUST support a means for a SIP UA involved in a CS to request, prior to the start of recording, that the CS not be recorded

o REQ-024 The mechanism MUST provide a means of indicating to the end users of a Communication Session that the session in which they are participating is being recorded.

Examples include: inject tones into the Communication Session from the SRC, play a message at the beginning of a session, a visual indicator on a display, etc.

o REQ-028 The mechanism MUST NOT prevent high availability deployments.

o REQ-029 The mechanism MUST support a means of providing security (confidentiality, integrity and authentication) for the SIPREC.

o REQ-030 The mechanism MUST provide a means for the Recording Session identifier so that the Recording Session itself is labeled as a SIP session that is established for the purpose of recording.

o REQ-032 If the Communication Session is encrypted, the Recording Session MUST use different keys.

o REQ-033 The mechanism SHALL support means to relate Recording Session(s) with Communication Session(s).

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, eavesdropping protection, and non-repudiation 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 SIPREC needs the ability to be authenticated and protected from eavesdropping and non-repudiation. Requirements are detailed in the requirements section.

8. IANA Considerations

This document has no IANA actions.

9. Acknowledgements

Thanks to Dan Wing, Alan Johnson, Vijay Gurbani, and Cullen Jennings for their help with this document, and to all the members of the DISPATCH WG mailing list for providing valuable input to this work.

10. Contributors

In addition to the editors, the following people provided substantial technical and writing contributions to this document, listed alphabetically:

Hadriel Kaplan
Acme Packet
71 Third Ave.
Burlington, MA 01803
USA
hkaplan@acmepacket.com

Henry Lum
Genesys, Alcatel-Lucent
1380 Rodick Road, Suite 200
Markham, Ontario L3R 4G5
Canada
henry.lum@genesyslab.com

Martin Palmer
BT Global Services
Annandale House,1 Hanworth Road,
Sunbury on Thames Middlesex TW16 5DJ
UK
martin.4.palmer@bt.com

Dave Smith
Genesys, Alcatel-Lucent
2001 Junipero Serra Blvd, Daly City, CA 94014
USA
dsmith@genesyslab.com

11. Normative References

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

[RFC2804] IAB and IESG, "IETF Policy on Wiretapping", RFC 2804, May 2000.

[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.

Authors' Addresses

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

Email: krehor@cisco.com

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

Email: rajnish.jain@ipc.com

Leon Portman
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