Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 22, 2016

E. Berger
S. Nandakumar
M. Zanaty
Cisco Systems

March 21, 2016

Frame Marking RTP Header Extension
draft-ietf-avtext-framemarking-01

Abstract

This document describes a Frame Marking RTP header extension used to convey information about video frames that is critical for error recovery and packet forwarding in RTP middleboxes or network nodes. It is most useful when media is encrypted, and essential when the middlebox or node has no access to the media encryption keys. It is also useful for codec-agnostic processing of encrypted or unencrypted media, while it also supports extensions for codec-specific information.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Many widely deployed RTP topologies used in modern voice and video conferencing systems include a centralized component that acts as an RTP switch. It receives voice and video streams from each participant, which may be encrypted using SRTP [RFC3711], or extensions that provide participants with private media via end-to-end encryption that excludes the switch. The goal is to provide a set of streams back to the participants which enable them to render the right media content. In a simple video configuration, for example, the goal will be that each participant sees and hears just the active speaker. In that case, the goal of the switch is to receive the voice and video streams from each participant, determine the active speaker based on energy in the voice packets, possibly using the client-to-mixer audio level RTP header extension, and select the corresponding video stream for transmission to participants; see Figure 1.

In this document, an "RTP switch" is used as a common short term for the terms "switching RTP mixer", "source projecting middlebox", "source forwarding unit/middlebox" and "video switching MCU" as discussed in [I-D.ietf-avtcore-rtp-topologies-update].

```
+---+      +------------+      +---+
| A |<---->|            |<---->| B |
+---+      |            |      +---+
           |   RTP      |
+---+      |  Switch    |      +---+
| C |<---->|            |<---->| D |
+---+      +------------+      +---+
```

Figure 1: RTP switch

In order to properly support switching of video streams, the RTP switch typically needs some critical information about video frames in order to start and stop forwarding streams.

- o Because of inter-frame dependencies, it should ideally switch video streams at a point where the first frame from the new speaker can be decoded by recipients without prior frames, e.g switch on an intra-frame.
- o In many cases, the switch may need to drop frames in order to realize congestion control techniques, and needs to know which frames can be dropped with minimal impact to video quality.
- o Furthermore, it is highly desirable to do this in a way which is not specific to the video codec. Nearly all modern video codecs share common concepts around frame types.
- o It is also desirable to be able to do this for SRTP without requiring the video switch to decrypt the packets. SRTP will encrypt the RTP payload format contents and consequently this data is not usable for the switching function without decryption, which may not even be possible in the case of end-to-end encryption of private media.

A comprehensive discussion of SFU considerations around codec agnostic selective forwarding of RTP media is described in [I-D.draft-aboba-avtcore-sfu-rtp]

By providing meta-information about the RTP streams outside the encrypted media payload an RTP switch can do selective forwarding without decrypting the payload. This document provides a solution to this problem.

2. Solution

The solution uses RTP header extensions as defined in [RFC5285]. A subset of meta-information from the video stream is provided as an RTP header extension to allow an RTP switch to do generic selective forwarding of video streams encoded with potentially different video codecs.

2.1. Mandatory Extension

The following information are extracted from the media payload and sent in the Frame Marking RTP header extension.

- S: Start of Frame (1 bit) - MUST be 1 in the first packet in a frame within a layer; otherwise MUST be 0.
- E: End of Frame (1 bit) - MUST be 1 in the last packet in a frame within a layer; otherwise MUST be 0.
- I: Independent Frame (1 bit) - MUST be 1 for frames that can be decoded independent of prior frames, e.g. intra-frame, VPx keyframe, H.264 IDR [RFC6184], H.265 CRA/BLA; otherwise MUST be 0.
- D: Discardable Frame (1 bit) - MUST be 1 for frames that can be dropped, and still provide a decodable media stream; otherwise MUST be 0.
- B: Base Layer Sync (1 bit) - MUST be 1 if this frame only depends on the base layer; otherwise MUST be 0.
- TID: Temporal ID (3 bits) - The base temporal layer starts with 0, and increases with 1 for each higher temporal layer/sub-layer.
- LID: Layer ID (8 bits) - Identifies the spatial and quality layer encoded.

The layer information contained in TID and LID convey useful aspects of the layer structure that can be utilized in selective forwarding. Without further information about the layer structure, these identifiers can only be used for relative priority of layers. They convey a layer hierarchy with TID=0 and LID=0 identifying the base layer. Higher values of TID identify higher temporal layers with higher frame rates. Higher values of LID identify higher spatial or quality layers with higher resolutions and bitrates.

With further information, for example, possible future RTCP SDES items that convey full layer structure information, it may be possible to map these TIDs and LIDs to specific frame rates, resolutions and bitrates. Such additional layer information may be useful to forwarding decisions in the RTP switch, but is beyond the scope of this memo. The relative layer information is still useful for many selective forwarding decisions even without such additional layer information.

The Frame Marking RTP header extension is encoded using the one-byte header as described in [RFC5285] as shown below.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=1  |S|E|I|D|B| TID |   LID         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.2. Layer ID Mappings

2.2.1. H265 LID Mapping

The following shows H265-LayerID (6 bits) mapped to the generic LID field.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=1  |S|E|I|D|B| TID |0|0|  LayerID  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.2.2. VP9 LID Mapping

The following shows VP9 Layer encoding information (4 bits for spatial and quality) mapped to the generic LID field.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=1  |S|E|I|D|B| TID |0|0|0|0| RS| RQ|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.2.3. VP8 LID Mapping

The following shows the header extension for VP8 that contains no layer information.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=1  |S|E|I|D|B| TID |0|0|0|0|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.2.4. H264-SVC LID Mapping

The following shows H264-SVC Layer encoding information (3 bits for spatial and 4 bits quality) mapped to the generic LID field.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=1  |S|E|I|D|B| TID |0| DID |  QID  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.2.5. H264 (AVC) LID Mapping

The following shows the header extension for H264 (AVC) that contains no layer information.

```
 0                   1                   2
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=1  |S|E|I|D|B| TID |0|0|0|0|0|0|0|0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

2.3. Signaling information

The URI for declaring this header extension in an extmap attribute is "urn:ietf:params:rtp-hdrext:framemarking". It does not contain any extension attributes.

An example attribute line in SDP:

```
a=extmap:3 urn:ietf:params:rtp-hdrext:framemarking
```

2.4. Considerations on use

The header extension values MUST represent what is already in the RTP payload.

When a RTP switch needs to discard a received video frame due to congestion control considerations, it is RECOMMENDED that it preferably drop frames marked with the "discardable" bit.

When a RTP switch wants to forward a new video stream to a receiver, it is RECOMMENDED to select the new video stream from the first switching point (I bit set) and forward the same. A RTP switch can request a media source to generate a switching point for H.264 by sending Full Intra Request (RTCP FIR) as defined in [RFC5104], for example.

3. Security Considerations

In the Secure Real-Time Transport Protocol (SRTP) [RFC3711], RTP header extensions are authenticated but not encrypted. When header extensions are used some of the payload type information are exposed and is visible to middle boxes. The encrypted media data is not exposed, so this is not seen as a high risk exposure.

4. Acknowledgements

Many thanks to Bernard Aboba, Jonathan Lennox for their inputs.

5. IANA Considerations

This document defines a new extension URI to the RTP Compact HeaderExtensions sub-registry of the Real-Time Transport Protocol (RTP) Parameters registry, according to the following data:

Extension URI: urn:ietf:params:rtp-hdrext:framemarkinginfo
Description: Frame marking information for video streams
Contact: espeberg@cisco.com
Reference: RFC XXXX

Note to RFC Editor: please replace RFC XXXX with the number of this RFC.

6. References

6.1. Normative References

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

6.2. Informative References

[I-D.ietf-avtcore-rtp-topologies-update]
Westerlund, M. and S. Wenger, "RTP Topologies", draft-ietf-avtcore-rtp-topologies-update (work in progress), April 2013.

[I-D.draft-aboba-avtcore-sfu-rtp]
Aboba, B., "Codec-Independent Selective Forwarding", raft-aboba-avtcore-sfu-rtp-00 (work in progress), July 2015.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, DOI 10.17487/RFC3711, March 2004, <http://www.rfc-editor.org/info/rfc3711>.

[RFC5104] Wenger, S., Chandra, U., Westerlund, M., and B. Burman, "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)", RFC 5104, DOI 10.17487/RFC5104, February 2008, <http://www.rfc-editor.org/info/rfc5104>.

[RFC5285] Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, DOI 10.17487/RFC5285, July 2008, <http://www.rfc-editor.org/info/rfc5285>.

[RFC6184] Wang, Y., Even, R., Kristensen, T., and R. Jesup, "RTP Payload Format for H.264 Video", RFC 6184, DOI 10.17487/RFC6184, May 2011, <http://www.rfc-editor.org/info/rfc6184>.

Authors' Addresses

Espen Berger
Cisco Systems

Phone: +47 98228179
Email: espeberg@cisco.com

Suhas Nandakumar
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: snandaku@cisco.com

Mo Zanaty
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: mzanaty@cisco.com

M. Zanaty
E. Berger
S. Nandakumar
Cisco Systems

March 28, 2019

Frame Marking RTP Header Extension
draft-ietf-avtext-framemarking-09

Abstract

This document describes a Frame Marking RTP header extension used to convey information about video frames that is critical for error recovery and packet forwarding in RTP middleboxes or network nodes. It is most useful when media is encrypted, and essential when the middlebox or node has no access to the media decryption keys. It is also useful for codec-agnostic processing of encrypted or unencrypted media, while it also supports extensions for codec-specific information.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Many widely deployed RTP [RFC3550] topologies [RFC7667] used in modern voice and video conferencing systems include a centralized component that acts as an RTP switch. It receives voice and video streams from each participant, which may be encrypted using SRTP [RFC3711], or extensions that provide participants with private media [I-D.ietf-perc-private-media-framework] via end-to-end encryption where the switch has no access to media decryption keys. The goal is to provide a set of streams back to the participants which enable them to render the right media content. In a simple video configuration, for example, the goal will be that each participant sees and hears just the active speaker. In that case, the goal of the switch is to receive the voice and video streams from each participant, determine the active speaker based on energy in the voice packets, possibly using the client-to-mixer audio level RTP header extension [RFC6464], and select the corresponding video stream for transmission to participants; see Figure 1.

In this document, an "RTP switch" is used as a common short term for the terms "switching RTP mixer", "source projecting middlebox", "source forwarding unit/middlebox" and "video switching MCU" as discussed in [RFC7667].

```
+---+      +------------+      +---+
| A |<---->|            |<---->| B |
+---+      |            |      +---+
           |    RTP     |
+---+      |   Switch   |      +---+
| C |<---->|            |<---->| D |
+---+      +------------+      +---+
```

Figure 1: RTP switch

In order to properly support switching of video streams, the RTP switch typically needs some critical information about video frames in order to start and stop forwarding streams.

- o Because of inter-frame dependencies, it should ideally switch video streams at a point where the first frame from the new speaker can be decoded by recipients without prior frames, e.g switch on an intra-frame.
- o In many cases, the switch may need to drop frames in order to realize congestion control techniques, and needs to know which frames can be dropped with minimal impact to video quality.
- o Furthermore, it is highly desirable to do this in a payload format-agnostic way which is not specific to each different video codec. Most modern video codecs share common concepts around frame types and other critical information to make this codec-agnostic handling possible.
- o It is also desirable to be able to do this for SRTP without requiring the video switch to decrypt the packets. SRTP will encrypt the RTP payload format contents and consequently this data is not usable for the switching function without decryption, which may not even be possible in the case of end-to-end encryption of private media [I-D.ietf-perc-private-media-framework].

By providing meta-information about the RTP streams outside the encrypted media payload, an RTP switch can do codec-agnostic selective forwarding without decrypting the payload. This document specifies the necessary meta-information in an RTP header extension.

2. Key Words for Normative Requirements

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].

3. Frame Marking RTP Header Extension

This specification uses RTP header extensions as defined in [RFC8285]. A subset of meta-information from the video stream is provided as an RTP header extension to allow an RTP switch to do generic selective forwarding of video streams encoded with potentially different video codecs.

The Frame Marking RTP header extension is encoded using the one-byte header or two-byte header as described in [RFC8285]. The one-byte header format is used for examples in this memo. The two-byte header format is used when other two-byte header extensions are present in the same RTP packet, since mixing one-byte and two-byte extensions is not possible in the same RTP packet.

This extension is only specified for Source (not Redundancy) RTP Streams [RFC7656] that carry video payloads. It is not specified for audio payloads, nor is it specified for Redundancy RTP Streams. The (separate) specifications for Redundancy RTP Streams often include provisions for recovering any header extensions that were part of the original source packet. Such provisions SHALL be followed to recover the Frame Marking RTP header extension of the original source packet. Source packet frame markings may be useful when generating Redundancy RTP Streams; for example, the I and D bits can be used to generate extra or no redundancy, respectively, and redundancy schemes with source blocks can align source block boundaries with Independent frame boundaries as marked by the I bit.

A frame, in the context of this specification, is the set of RTP packets with the same RTP timestamp from a specific RTP synchronization source (SSRC).

3.1. Short Extension for Non-Scalable Streams

The following RTP header extension is RECOMMENDED for non-scalable streams. It MAY also be used for scalable streams if the sender has limited or no information about stream scalability. The ID is assigned per [RFC8285], and the length is encoded as L=0 which indicates 1 octet of data.

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=? |  L=0  |S|E|I|D|0 0 0 0|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The following information are extracted from the media payload and sent in the Frame Marking RTP header extension.

- o S: Start of Frame (1 bit) - MUST be 1 in the first packet in a frame; otherwise MUST be 0.
- o E: End of Frame (1 bit) - MUST be 1 in the last packet in a frame; otherwise MUST be 0. SHOULD match the RTP header marker bit in payload formats with such semantics for marking end of frame.
- o I: Independent Frame (1 bit) - MUST be 1 for frames that can be decoded independent of temporally prior frames, e.g. intra-frame, VPX keyframe, H.264 IDR [RFC6184], H.265 IDR/CRA/BLA/RAP [RFC7798]; otherwise MUST be 0.
- o D: Discardable Frame (1 bit) - MUST be 1 for frames the sender knows can be discarded, and still provide a decodable media stream; otherwise MUST be 0.
- o The remaining (4 bits) - are reserved for future use for non-scalable streams; they MUST be set to 0 upon transmission and ignored upon reception.

3.2. Long Extension for Scalable Streams

The following RTP header extension is RECOMMENDED for scalable streams. It MAY also be used for non-scalable streams, in which case TID, LID and TL0PICIDX MUST be 0 or omitted. The ID is assigned per [RFC8285], and the length is encoded as L=2 which indicates 3 octets of data when nothing is omitted, or L=1 for 2 octets when TL0PICIDX is omitted, or L=0 for 1 octet when both LID and TL0PICIDX are omitted.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=? |  L=2  |S|E|I|D|B| TID |   LID         |    TL0PICIDX  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
           or
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=? |  L=1  |S|E|I|D|B| TID |   LID         | (TL0PICIDX omitted)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
           or
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=? |  L=0  |S|E|I|D|B| TID | (LID and TL0PICIDX omitted)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The following information are extracted from the media payload and sent in the Frame Marking RTP header extension.

- S: Start of Frame (1 bit) - MUST be 1 in the first packet in a frame within a layer; otherwise MUST be 0.
- E: End of Frame (1 bit) - MUST be 1 in the last packet in a frame within a layer; otherwise MUST be 0. Note that the RTP header marker bit MAY be used to infer the last packet of the highest enhancement layer, in payload formats with such semantics.
- I: Independent Frame (1 bit) - MUST be 1 for frames that can be decoded independent of temporally prior frames, e.g. intra-frame, VPX keyframe, H.264 IDR [RFC6184], H.265 IDR/CRA/BLA/RAP [RFC7798]; otherwise MUST be 0. Note that this bit only signals temporal independence, so it can be 1 in spatial or quality enhancement layers that depend on temporally co-located layers but not temporally prior frames.
- D: Discardable Frame (1 bit) - MUST be 1 for frames the sender knows can be discarded, and still provide a decodable media stream; otherwise MUST be 0.
- B: Base Layer Sync (1 bit) - MUST be 1 if the sender knows this frame only depends on the base temporal layer; otherwise MUST be 0. If no scalability is used, this MUST be 0.
- TID: Temporal ID (3 bits) - The base temporal layer starts with 0, and increases with 1 for each higher temporal layer/sub-layer. If no scalability is used, this MUST be 0.
- LID: Layer ID (8 bits) - Identifies the spatial and quality layer encoded, starting with 0 and increasing with higher fidelity. If no scalability is used, this MUST be 0 or omitted to reduce length. When omitted, TL0PICIDX MUST also be omitted.
- TL0PICIDX: Temporal Layer 0 Picture Index (8 bits) - Running index of base temporal layer 0 frames when TID is 0. When TID is not 0, this indicates a dependency on the given index. If no scalability is used, or the running index is unknown, this MUST be omitted to

reduce length. Note that 0 is a valid running index value for TL0PICIDX.

The layer information contained in TID and LID convey useful aspects of the layer structure that can be utilized in selective forwarding. Without further information about the layer structure, these identifiers can only be used for relative priority of layers. They convey a layer hierarchy with TID=0 and LID=0 identifying the base layer. Higher values of TID identify higher temporal layers with higher frame rates. Higher values of LID identify higher spatial and/or quality layers with higher resolutions and/or bitrates.

With further information, for example, possible future RTCP SDES items that convey full layer structure information, it may be possible to map these TIDs and LIDs to specific frame rates, resolutions and bitrates. Such additional layer information may be useful for forwarding decisions in the RTP switch, but is beyond the scope of this memo. The relative layer information is still useful for many selective forwarding decisions even without such additional layer information.

3.2.1. Layer ID Mappings for Scalable Streams

3.2.1.1. H265 LID Mapping

The following shows the H265 [RFC7798] LayerID (6 bits) and TID (3 bits) from the NAL unit header mapped to the generic LID and TID fields.

The I bit MUST be 1 when the NAL unit type is 16-23 (inclusive), otherwise it MUST be 0.

The S and E bits MUST match the corresponding bits in PACI:PHES:TSCI payload structures.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=2  |S|E|I|D|B| TID |0|0|  LayerID  |    TL0PICIDX  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.2.1.2. H264-SVC LID Mapping

The following shows H264-SVC [RFC6190] Layer encoding information (3 bits for spatial/dependency layer, 4 bits for quality layer and 3 bits for temporal layer) mapped to the generic LID and TID fields.

The S, E, I and D bits MUST match the corresponding bits in PACSI payload structures.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=2  |S|E|I|D|B| TID |0| DID |  QID  |    TL0PICIDX  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.2.1.3. H264 (AVC) LID Mapping

The following shows the header extension for H264 (AVC) [RFC6184] that contains only temporal layer information.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=2  |S|E|I|D|B| TID |0|0|0|0|0|0|0|0|    TL0PICIDX  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.2.1.4. VP8 LID Mapping

The following shows the header extension for VP8 [RFC7741] that contains only temporal layer information.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID=2 |  L=2  |S|E|I|D|B| TID |0|0|0|0|0|0|0|0|    TL0PICIDX  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3.2.1.5. Future Codec LID Mapping

The RTP payload format specification for future video codecs SHOULD include a section describing the LID mapping and TID mapping for the codec. For example, the LID/TID mapping for the VP9 codec is described in the VP9 RTP Payload Format [I-D.ietf-payload-vp9].

3.3. Signaling Information

The URI for declaring this header extension in an extmap attribute is "urn:ietf:params:rtp-hdrext:framemarking". It does not contain any extension attributes.

An example attribute line in SDP:

```
a=extmap:3 urn:ietf:params:rtp-hdrext:framemarking
```

3.4. Usage Considerations

The header extension values MUST represent what is already in the RTP payload.

When an RTP switch needs to discard a received video frame due to congestion control considerations, it is RECOMMENDED that it preferably drop frames marked with the D (Discardable) bit set, or the highest values of TID and LID, which indicate the highest temporal and spatial/quality enhancement layers, since those typically have fewer dependenices on them than lower layers.

When an RTP switch wants to forward a new video stream to a receiver, it is RECOMMENDED to select the new video stream from the first switching point with the I (Independent) bit set in all spatial layers and forward the same. An RTP switch can request a media source to generate a switching point by sending Full Intra Request (RTCP FIR) as defined in [RFC5104], for example.

3.4.1. Relation to Layer Refresh Request (LRR)

Receivers can use the Layer Refresh Request (LRR) [I-D.ietf-avtext-lrr] RTCP feedback message to upgrade to a higher layer in scalable encodings. The TID/LID values and formats used in LRR messages MUST correspond to the same values and formats specified in Section 3.2.

Because frame marking can only be used with temporally-nested streams, temporal-layer LRR refreshes are unnecessary for frame-marked streams. Other refreshes can be detected based on the I bit being set for the specific spatial layers.

3.4.2. Scalability Structures

The LID and TID information is most useful for fixed scalability structures, such as nested hierarchical temporal layering structures, where each temporal layer only references lower temporal layers or the base temporal layer. The LID and TID information is less useful, or even not useful at all, for complex, irregular scalability structures that do not conform to common, fixed patterns of inter-layer dependencies and referencing structures. Therefore it is RECOMMENDED to use LID and TID information for RTP switch forwarding decisions only in the case of temporally nested scalability structures, and it is NOT RECOMMENDED for other (more complex or irregular) scalability structures.

4. Security Considerations

In the Secure Real-Time Transport Protocol (SRTP) [RFC3711], RTP header extensions are authenticated but usually not encrypted. When header extensions are used some of the payload type information are exposed and visible to middle boxes. The encrypted media data is not exposed, so this is not seen as a high risk exposure.

5. Acknowledgements

Many thanks to Bernard Aboba, Jonathan Lennox, and Stephan Wenger for their inputs.

6. IANA Considerations

This document defines a new extension URI to the RTP Compact HeaderExtensions sub-registry of the Real-Time Transport Protocol (RTP) Parameters registry, according to the following data:

Extension URI: urn:ietf:params:rtp-hdrext:framemarkinginfo
Description: Frame marking information for video streams
Contact: mzanaty@cisco.com
Reference: RFC XXXX

Note to RFC Editor: please replace RFC XXXX with the number of this RFC.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC6184] Wang, Y., Even, R., Kristensen, T., and R. Jesup, "RTP Payload Format for H.264 Video", RFC 6184, DOI 10.17487/RFC6184, May 2011, <https://www.rfc-editor.org/info/rfc6184>.

[RFC6190] Wenger, S., Wang, Y., Schierl, T., and A. Eleftheriadis, "RTP Payload Format for Scalable Video Coding", RFC 6190, DOI 10.17487/RFC6190, May 2011, <https://www.rfc-editor.org/info/rfc6190>.

[RFC7741] Westin, P., Lundin, H., Glover, M., Uberti, J., and F. Galligan, "RTP Payload Format for VP8 Video", RFC 7741, DOI 10.17487/RFC7741, March 2016, <https://www.rfc-editor.org/info/rfc7741>.

[RFC7798] Wang, Y., Sanchez, Y., Schierl, T., Wenger, S., and M. Hannuksela, "RTP Payload Format for High Efficiency Video Coding (HEVC)", RFC 7798, DOI 10.17487/RFC7798, March 2016, <https://www.rfc-editor.org/info/rfc7798>.

[RFC8285] Singer, D., Desineni, H., and R. Even, Ed., "A General Mechanism for RTP Header Extensions", RFC 8285, DOI 10.17487/RFC8285, October 2017, <https://www.rfc-editor.org/info/rfc8285>.

7.2. Informative References

[I-D.ietf-avtext-lrr] Lennox, J., Hong, D., Uberti, J., Holmer, S., and M. Flodman, "The Layer Refresh Request (LRR) RTCP Feedback Message", draft-ietf-avtext-lrr-07 (work in progress), July 2017.

[I-D.ietf-payload-vp9] Uberti, J., Holmer, S., Flodman, M., Lennox, J., and D. Hong, "RTP Payload Format for VP9 Video", draft-ietf-payload-vp9-06 (work in progress), July 2018.

[I-D.ietf-perc-private-media-framework] Jones, P., Benham, D., and C. Groves, "A Solution Framework for Private Media in Privacy Enhanced RTP Conferencing", draft-ietf-perc-private-media-framework-09 (work in progress), February 2019.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <https://www.rfc-editor.org/info/rfc3550>.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, DOI 10.17487/RFC3711, March 2004, <https://www.rfc-editor.org/info/rfc3711>.

[RFC5104] Wenger, S., Chandra, U., Westerlund, M., and B. Burman, "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)", RFC 5104, DOI 10.17487/RFC5104, February 2008, <https://www.rfc-editor.org/info/rfc5104>.

[RFC6464] Lennox, J., Ed., Ivov, E., and E. Marocco, "A Real-time Transport Protocol (RTP) Header Extension for Client-to-Mixer Audio Level Indication", RFC 6464, DOI 10.17487/RFC6464, December 2011, <https://www.rfc-editor.org/info/rfc6464>.

[RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, Ed., "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <https://www.rfc-editor.org/info/rfc7656>.

[RFC7667] Westerlund, M. and S. Wenger, "RTP Topologies", RFC 7667, DOI 10.17487/RFC7667, November 2015, <https://www.rfc-editor.org/info/rfc7667>.

Authors' Addresses

Mo Zanaty
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: mzanaty@cisco.com

Espen Berger
Cisco Systems

Phone: +47 98228179
Email: espeberg@cisco.com

Suhas Nandakumar
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: snandaku@cisco.com

Payload Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 22, 2016

J. Lennox
D. Hong
Vidyo
J. Uberti
S. Holmer
M. Flodman
Google

March 21, 2016

The Layer Refresh Request (LRR) RTCP Feedback Message
draft-ietf-avtext-lrr-02

Abstract

This memo describes the RTCP Payload-Specific Feedback Message "Layer Refresh Request" (LRR), which can be used to request a state refresh of one or more substreams of a layered media stream. It also defines its use with several RTP payloads for scalable media formats.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This memo describes an RTCP [RFC3550] Payload-Specific Feedback Message [RFC4585] "Layer Refresh Request" (LRR). It is designed to allow a receiver of a layered media stream to request that one or more of its substreams be refreshed, such that it can then be decoded by an endpoint which previously was not receiving those layers, without requiring that the entire stream be refreshed (as it would be if the receiver sent a Full Intra Request (FIR) [RFC5104] (see also [I-D.wenger-avtext-avpf-ccm-layered]).

The feedback message is applicable both to temporally and spatially scaled streams, and to both single-stream and multi-stream scalability modes.

2. Conventions, Definitions and Acronyms

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].

2.1. Terminology

A "Layer Refresh Point" is a point in a scalable stream after which a decoder, which previously had been able to decode only some (possibly none) of the available layers of stream, is able to decode a greater number of the layers.

For spatial (or quality) layers, layer refresh typically requires that a spatial layer be encoded in a way that references only lower-layer subpictures of the current picture, not any earlier pictures of that spatial layer. Additionally, the encoder must promise that no earlier pictures of that spatial layer will be used as reference in the future.

In a layer refresh, however, other layers than the ones requested for refresh may still maintain dependency on earlier content of the stream. This is the difference between a layer refresh and a Full Intra Request [RFC5104]. This minimizes the coding overhead of refresh to only those parts of the stream that actually need to be refreshed at any given time.

An illustration of spatial layer refresh of an enhancement layer is shown below.

```
     ... <--  S1  <--  S1       S1  <--  S1  <-- ...
               |        |        |        |
              \/       \/       \/       \/
     ... <--  S0  <--  S0  <--  S0  <--  S0  <-- ...

               1        2        3        4
```

In this illustration, frame 3 is a layer refresh point for spatial layer S1; a decoder which had previously only been decoding spatial layer S0 would be able to decode layer S1 starting at frame 3.

Figure 1

An illustration of spatial layer refresh of a base layer is shown below.

```
     ... <--  S1  <--  S1  <--  S1  <--  S1  <-- ...
               |        |        |        |
              \/       \/       \/       \/
     ... <--  S0  <--  S0       S0  <--  S0  <-- ...

               1        2        3        4
```

In this illustration, frame 3 is a layer refresh point for spatial layer S0; a decoder which had previously not been decoding the stream at all could decode layer S0 starting at frame 3.

Figure 2

For temporal layers, layer refresh requires that the layer be "temporally nested", i.e. use as reference only earlier frames of a lower temporal layer, not any earlier frames of this temporal layer, and also promise that no future frames of this temporal layer will reference frames of this temporal layer before the refresh point. In many cases, the temporal structure of the stream will mean that all frames are temporally nested, in which case decoders will have no need to send LRR messages for the stream.

An illustration of temporal layer refresh is shown below.

```
        ...  <----- T1  <------ T1          T1  <------ ...
                   /           /           /
                 |_          |_          |_
     ... <--  T0  <------ T0  <------ T0  <------ T0  <--- ...

               1     2     3     4     5     6     7
```

In this illustration, frame 6 is a layer refresh point for temporal layer T1; a decoder which had previously only been decoding temporal layer T0 would be able to decode layer T1 starting at frame 6.

Figure 3

An illustration of an inherently temporally nested stream is shown below.

```
                    T1          T1          T1
                   /           /           /
                 |_          |_          |_
  ... <--  T0  <------ T0  <------ T0  <------ T0  <--- ...

           1     2     3     4     5     6     7
```

In this illustration, the stream is temporally nested in its ordinary structure; a decoder receiving layer T0 can begin decoding layer T1 at any point.

Figure 4

3. Layer Refresh Request

A layer refresh frame can be requested by sending a Layer Refresh Request (LRR), which is an RTCP payload-specific feedback message [RFC4585] asking the encoder to encode a frame which makes it possible to upgrade to a higher layer. The LRR contains one or two tuples, indicating the layer the decoder wants to upgrade to, and (optionally) the currently highest layer the decoder can decode.

The specific format of the tuples, and the mechanism by which a receiver recognizes a refresh frame, is codec-dependent. Usage for several codecs is discussed in Section 4.

LRR follows the model of the Full Intra Request (FIR) [RFC5104](Section 3.5.1) for its retransmission, reliability, and use in multipoint conferences.

The LRR message is identified by RTCP packet type value PT=PSFB and FMT=TBD. The FCI field MUST contain one or more LRR entries. Each entry applies to a different media sender, identified by its SSRC.

3.1. Message Format

The Feedback Control Information (FCI) for the Layer Refresh Request consists of one or more FCI entries, the content of which is depicted in Figure 5. The length of the LRR feedback message MUST be set to 2+3*N, where N is the number of FCI entries.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              SSRC                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Seq nr.       |C| Payload Type| Reserved                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Target Layer Index            | Current Layer Index (opt)     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5

SSRC (32 bits) The SSRC value of the media sender that is requested to send a layer refresh point.

Seq nr. (8 bits) Command sequence number. The sequence number space is unique for each pairing of the SSRC of command source and the SSRC of the command target. The sequence number SHALL be increased by 1 modulo 256 for each new command. A repetition SHALL NOT increase the sequence number. The initial value is arbitrary.

C (1 bit) A flag bit indicating whether the "Current Layer Index" field is present in the FCI. If this bit is false, the sender of the LRR message is requesting refresh of all layers up to and including the target layer.

Payload Type (7 bits) The RTP payload type for which the LRR is being requested. This gives the context in which the target layer index is to be interpreted.

Reserved (16 bits) All bits SHALL be set to 0 by the sender and SHALL be ignored on reception.

Target Layer Index (16 bits) The target layer for which the receiver wishes a refresh point. Its format is dependent on the payload type field.

Current Layer Index (16 bits) If C is 1, the current layer being decoded by the receiver. This message is not requesting refresh of layers at or below this layer. If C is 0, this field SHALL be set to 0 by the sender and SHALL be ignored on reception.

3.2. Semantics

Within the common packet header for feedback messages (as defined in section 6.1 of [RFC4585]), the "SSRC of packet sender" field indicates the source of the request, and the "SSRC of media source" is not used and SHALL be set to 0. The SSRCs of the media senders to which the LRR command applies are in the corresponding FCI entries. A LRR message MAY contain requests to multiple media senders, using one FCI entry per target media sender.

Upon reception of LRR, the encoder MUST send a decoder refresh point (see section Section 2.1) as soon as possible.

The sender MUST consider congestion control as outlined in section 5 of [RFC5104], which MAY restrict its ability to send a layer refresh point quickly.

4. Usage with specific codecs

In order for LRR to be used with a scalable codec, the format of the target layer and current target layer fields needs to be specified for that codec’s RTP packetization. New RTP packetization specifications for scalable codecs SHOULD define how this is done. (The VP9 payload [I-D.ietf-payload-vp9], for instance, has done so.) This section defines the layer index fields for use with several existing scalable codecs.

4.1. H264 SVC

H.264 SVC [RFC6190] defines temporal, dependency (spatial), and quality scalability modes.

```
+---------------+---------------+
|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R| DID |  QID  | TID |RES      |
+---------------+---------------+
```

Figure 6

Figure 6 shows the format of the layer index field for H.264 SVC streams. This is designed to follow the same layout as the third and fourth bytes of the H.264 SVC NAL unit extension, which carry the stream’s layer information. The "R" and "RES" fields MUST be set to 0 on transmission and ignored on reception. See [RFC6190] Section 1.1.3 for details on the DID, QID, and TID fields.

A dependency or quality layer refresh of a given layer in H.264 SVC can be identified by the "I" bit (idr_flag) in the extended NAL unit header, present in NAL unit types 14 (prefix NAL unit) and 20 (coded scalable slice). Layer refresh of the base layer can also be identified by its NAL unit type of its coded slices, which is "5" rather than "1". A dependency or quality layer refresh is complete once this bit has been seen on all the appropriate layers (in decoding order) above the current layer index (if any, or beginning from the base layer if not) through the target layer index.

Note that as the "I" bit in a PACSI header is set if the corresponding bit is set in any of the aggregated NAL units it describes; thus, it is not sufficient to identify layer refresh when NAL units of multiple dependency or quality layers are aggregated.

In H.264 SVC, temporal layer refresh information can be determined from various Supplemental Encoding Information (SEI) messages in the bitstream.

Whether an H.264 SVC stream is scalably nested can be determined from the Scalability Information SEI message’s temporal_id_nesting flag. If this flag is set in a stream’s currently applicable Scalability Information SEI, receivers SHOULD NOT send temporal LRR messages for that stream, as every frame is implicitly a temporal layer refresh point. (The Scalability Information SEI message may also be available in the signaling negotiation of H.264 SVC, as the sprop-scalability-info parameter.)

If a stream’s temporal_id_nesting flag is not set, the Temporal Level Switching Point SEI message identifies temporal layer switching points. A temporal layer refresh is satisfied when this SEI message is present in a frame with the target layer index, if the message’s delta_frame_num refers to a frame with the requested current layer index. (Alternately, temporal layer refresh can also be satisfied by a complete state refresh, such as an IDR.) Senders which support receiving LRR for non-temporally-nested streams MUST insert Temporal Level Switching Point SEI messages as appropriate.

4.2. VP8

The VP8 RTP payload format [I-D.ietf-payload-vp8] defines temporal scalability modes. It does not support spatial scalability.

```
+---------------+---------------+
|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|TID| RES                       |
+---------------+---------------+
```

Figure 7

Figure 7 shows the format of the layer index field for VP8 streams. The "RES" fields MUST be set to 0 on transmission and be ignored on reception. See [I-D.ietf-payload-vp8] Section 4.2 for details on the TID field.

A VP8 layer refresh point can be identified by the presence of the "Y" bit in the VP8 payload header. When this bit is set, this and all subsequent frames depend only on the current base temporal layer. On receipt of an LRR for a VP8 stream, A sender which supports LRR MUST encode the stream so it can set the Y bit in a packet whose temporal layer is at or below the target layer index.

Note that in VP8, not every layer switch point can be identified by the Y bit, since the Y bit implies layer switch of all layers, not just the layer in which it is sent. Thus the use of LRR with VP8 can result in some inefficiency in transmision. However, this is not expected to be a major issue for temporal structures in normal use.

4.3. H265

The initial version of the H.265 payload format [I-D.ietf-payload-rtp-h265] defines temporal scalability, with protocol elements reserved for spatial or other scalability modes (which are expected to be defined in a future version of the specification).

```
+---------------+---------------+
|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RES         |  LayerId  | TID |
+-------------+-----------------+
```

Figure 8

Figure 8 shows the format of the layer index field for H.265 streams. This is designed to follow the same layout as the first and second bytes of the H.265 NAL unit header, which carry the stream’s layer information. The "RES" field MUST be set to 0 on transmission and ignored on reception. See [I-D.ietf-payload-rtp-h265] Section 1.1.4 for details on the LayerId and TID fields.

H.265 streams signal whether they are temporally nested, using the vps_temporal_id_nesting_flag in the Video Parameter Set (VPS), and the sps_temporal_id_nesting_flag in the Sequence Parameter Set (SPS). If this flag is set in a stream’s currently applicable VPS or SPS, receivers SHOULD NOT send temporal LRR messages for that stream, as every frame is implicitly a temporal layer refresh point.

If a stream’s sps_temporal_id_nesting_flag is not set, the NAL unit types 2 to 5 inclusively identify temporal layer switching points. A layer refresh to any higher target temporal layer is satisfied when a NAL unit type of 4 or 5 with TID equal to 1 more than current TID is seen. Alternatively, layer refresh to a target temporal layer can be incrementally satisfied with NAL unit type of 2 or 3. In this case, given current TID = TO and target TID = TN, layer refresh to TN is satisfied when NAL unit type of 2 or 3 is seen for TID = T1, then TID = T2, all the way up to TID = TN. During this incremental process, layer refresh to TN can be completely satisfied as soon as a NAL unit type of 2 or 3 is seen.

Of course, temporal layer refresh can also be satisfied whenever any Intra Random Access Point (IRAP) NAL unit type (with values 16-23, inclusively) is seen. An IRAP picture is similar to an IDR picture in H.264 (NAL unit type of 5 in H.264) where decoding of the picture can start without any older pictures.

In the (future) H.265 payloads that support spatial scalability, a spatial layer refresh of a specific layer can be identified by NAL units with the requested layer ID and NAL unit types between 16 and 21 inclusive. A dependency or quality layer refresh is complete once NAL units of this type have been seen on all the appropriate layers (in decoding order) above the current layer index (if any, or beginning from the base layer if not) through the target layer index.

5. Usage with different scalability transmission mechanisms

Several different mechanisms are defined for how scalable streams can be transmitted in RTP. The RTP Taxonomy [RFC7656] Section 3.7 defines three mechanisms: Single RTP Stream on a Single Media Transport (SRST), Multiple RTP Streams on a Single Media Transport (MRST), and Multiple RTP Streams on Multiple Media Transports (MRMT).

The LRR message is applicable to all these mechanisms. For MRST and MRMT mechanisms, the "media source" field of the LRR FCI is set to the SSRC of the RTP stream containing the layer indicated by the Current Layer Index (if "C" is 1), or the stream containing the base encoded stream (if "C" is 0). For MRMT, it is sent on the RTP session on which this stream is sent. On receipt, the sender MUST refresh all the layers requested in the stream, simultaneously in decode order.

6. Security Considerations

All the security considerations of FIR feedback packets [RFC5104] apply to LRR feedback packets as well. Additionally, media senders receiving LRR feedback packets MUST validate that the payload types and layer indices they are receiving are valid for the stream they are currently sending, and discard the requests if not.

7. SDP Definitions

Section 7 of [RFC5104] defines SDP procedures for indicating and negotiating support for codec control messages (CCM) in SDP. This document extends this with a new codec control command, "lrr", which indicates support of the Layer Refresh Request (LRR).

Figure 9 gives a formal Augmented Backus-Naur Form (ABNF) [RFC5234] showing this grammar extension, extending the grammar defined in [RFC5104].

```
rtcp-fb-ccm-param =/ SP "lrr"    ; Layer Refresh Request
```

Figure 9: Syntax of the "lrr" ccm

The Offer-Answer considerations defined in [RFC5104] Section 7.2 apply.

8. IANA Considerations

This document defines a new entry to the "Codec Control Messages" subregistry of the "Session Description Protocol (SDP) Parameters" registry, according to the following data:

Value name: lrr

Long name: Layer Refresh Request Command

Usable with: ccm

Reference: RFC XXXX

This document also defines a new entry to the "FMT Values for PSFB Payload Types" subregistry of the "Real-Time Transport Protocol (RTP) Parameters" registry, according to the following data:

Name: LRR

Long Name: Layer Refresh Request Command

Value: TBD

Reference: RFC XXXX

9. References

9.1. Normative References

[I-D.ietf-payload-rtp-h265]
Wang, Y., Sanchez, Y., Schierl, T., Wenger, S., and M. Hannuksela, "RTP Payload Format for H.265/HEVC Video", draft-ietf-payload-rtp-h265-15 (work in progress), November 2015.

[I-D.ietf-payload-vp8]
Westin, P., Lundin, H., Glover, M., Uberti, J., and F. Galligan, "RTP Payload Format for VP8 Video", draft-ietf-payload-vp8-17 (work in progress), September 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC4585] Ott, J., Wenger, S., Sato, N., Burmeister, C., and J. Rey, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 4585, DOI 10.17487/RFC4585, July 2006, <http://www.rfc-editor.org/info/rfc4585>.

[RFC5104] Wenger, S., Chandra, U., Westerlund, M., and B. Burman, "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)", RFC 5104, DOI 10.17487/RFC5104, February 2008, <http://www.rfc-editor.org/info/rfc5104>.

[RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, DOI 10.17487/RFC5234, January 2008, <http://www.rfc-editor.org/info/rfc5234>.

[RFC6190] Wenger, S., Wang, Y., Schierl, T., and A. Eleftheriadis, "RTP Payload Format for Scalable Video Coding", RFC 6190, DOI 10.17487/RFC6190, May 2011, <http://www.rfc-editor.org/info/rfc6190>.

9.2. Informative References

[I-D.ietf-payload-vp9]
Uberti, J., Holmer, S., Flodman, M., Lennox, J., and D. Hong, "RTP Payload Format for VP9 Video", draft-ietf-payload-vp9-01 (work in progress), October 2015.

[I-D.wenger-avtext-avpf-ccm-layered]
Wenger, S., Lennox, J., Burman, B., and M. Westerlund, "Using Codec Control Messages in the RTP Audio-Visual Profile with Feedback with Layered Codecs", draft-wenger-avtext-avpf-ccm-layered-00 (work in progress), December 2015.

[RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, Ed., "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <http://www.rfc-editor.org/info/rfc7656>.

Authors' Addresses

Jonathan Lennox
Vidyo, Inc.
433 Hackensack Avenue
Seventh Floor
Hackensack, NJ 07601
US

Email: jonathan@vidyo.com

Danny Hong
Vidyo, Inc.
433 Hackensack Avenue
Seventh Floor
Hackensack, NJ 07601
US

Email: danny@vidyo.com

Justin Uberti
Google, Inc.
747 6th Street South
Kirkland, WA 98033
USA

Email: justin@uberti.name

Stefan Holmer
Google, Inc.
Kungsbron 2
Stockholm 111 22
Sweden

Email: holmer@google.com

Magnus Flodman
Google, Inc.
Kungsbron 2
Stockholm 111 22
Sweden

Email: mflodman@google.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 4, 2016

A. Roach
Mozilla
S. Nandakumar
Cisco Systems
P. Thatcher
Google

March 03, 2016

RTP Stream Identifier (RID) Source Description (SDES)
draft-ietf-avtext-rid-01

Abstract

This document defines and registers an RTCP SDES item, RID, for identification of RTP streams associated with Encoded Streams and Dependent Streams.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

RTP sessions frequently consist of multiple streams, each of which is identified at any given time by its SSRC; however, the SSRC associated with a stream is not guaranteed to be stable over its lifetime. Within a session, these streams can be tagged with a number of identifiers, including CNAMEs and MSIDs [I-D.ietf-mmusic-msid]. Unfortunately, none of these have the proper ordinality to refer to an individual stream; all such identifiers can appear in more than one stream at a time. While approaches that use unique Payload Types (PTs) per stream have been used in some applications, this is a semantic overloading of that field, and one for which its size is inadequate: in moderately complex systems that use PT to uniquely identify every potential combination of codec configuration and unique stream, it is possible to simply run out of values.

To address this situation, we define a new RTCP SDES identifier that uniquely identifies a single stream. A key motivator for defining this identifier is the ability to differentiate among different encodings of a single Source Stream that are sent simultaneously (i.e., simulcast). This need for unique identification extends to Dependent Streams (i.e., layers used by a layered codec).

At the same time, when Redundancy RTP Streams are in use, we also need an identifier that connects such streams to the RTP stream for which they are providing redundancy. To that end, when this new identifier is in use, it appears (and contains the same value) in both in the Redundancy RTP Stream as well as the stream it is correcting.

For lack of a better term, we have elected to call this term "RID," which loosely stands for "RTP stream IDentifier." It should be noted that this isn't an overly-precise use of the term "RTP Stream," due to the lack of an existing well-defined term for the construct we are attempting to identify. See Section 3 for a formal definition of the exact scope of a RID.

The use of RIDs in SDP is described in [I-D.ietf-mmusic-rid].

Finally, to accommodate the potential need to identify Redundancy RTP streams independently of the stream for which they are defining redundancy, we specify a second identifier, RRID (Redundancy RTP Stream IDentifier).

2. Key Words for Requirements

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]

3. Terminology

In this document, the terms "Source Stream", "Encoded Stream," "RTP Stream", "Source RTP Stream", "Dependent Stream", "Received RTP Stream", and "Redundancy RTP Stream" are used as defined in [RFC7656].

For Encoded Streams, the RID refers to the "Source RTP Stream" as defined by [RFC7656] Section 2.1.10. For Dependent Streams, it refers to the RTP Stream that, like the Source RTP Stream of an Encoded Stream, is the RTP Stream that is not a Redundancy RTP Stream. For conciseness, we define the term "RID RTP Stream" to refer to this construct.

For clarity, when RID is used, Redundancy RTP Streams that can be used to repair Received RTP Streams will use the same RID value as the Received RTP Stream they are intended to be combined with. If applications want to identify individual redundancy streams, they can add an RRID to them instead of or in addition to the RID.

4. Usage of RID in RTP and RTCP

The RTP fixed header includes the payload type number and the SSRC values of the RTP stream. RTP defines how you de-multiplex streams within an RTP session; however, in some use cases, applications need further identifiers in order to effectively map the individual RID RTP Streams to their equivalent payload configurations in the SDP.

This specification defines two new RTCP SDES items [RFC3550]. The first item is 'RID', which is used to carry RID identifiers within RTCP SDES packets. This makes it possible for a receiver to associate received RTP packets (identifying the RID RTP Stream) with a media description having the format constraint specified. The second is 'RRID', which can be used to carry a unique identifier to designate Redundancy RTP Streams independently of the stream for which they provide redundancy.

This specification also uses the RTP header extension for RTCP SDES items [I-D.ietf-avtext-sdes-hdr-ext] to allow carrying RID and RRID information in RTP packets. This allowes correlation at stream startup, or after stream changes where the use of RTCP may not be sufficiently responsive.

4.1. RTCP 'RID' SDES Extension

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      RID=TBD  |     length    | rid                         ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The rid payload is UTF-8 encoded and is not null-terminated.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

4.2. RTCP 'RRID' SDES Extension

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     RRID=TBD  |     length    | rrid                        ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The rrid payload is UTF-8 encoded and is not null-terminated.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

4.3. RTP 'RID' and 'RRID' Header Extensions

Because recipients of RTP packets will typically need to know which streams they correspond to immediately upon receipt, this specification also defines a means of carrying RID and RRID identifiers in RTP extension headers, using the technique described in [I-D.ietf-avtext-sdes-hdr-ext].

As described in that document, the header extension element can be encoded using either the one-byte or two-byte header, and the identification-tag payload is UTF-8 encoded, as in SDP.

As the identification-tag is included in an RTP header extension, there should be some consideration about the packet expansion caused by the identification-tag. To avoid Maximum Transmission Unit (MTU) issues for the RTP packets, the header extension's size needs to be taken into account when the encoding media. Note that set of header extensions included in the packet needs to be padded to the next 32-bit boundary [RFC5285].

It is RECOMMENDED that the identification-tag is kept short. In many cases, a one-byte tag will be sufficient; it is RECOMMENDED that implementations use the shortest identifier that fits their purposes.

5. IANA Considerations

5.1. New RID SDES item

RFC EDITOR NOTE: Please replace RFCXXXX with the RFC number of this document.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

This document adds the MID SDES item to the IANA "RTCP SDES item types" registry as follows:

```
Value:          TBD
Abbrev.:        RID
Name:           RTP Stream Identifier
Reference:      RFCXXXX
```

5.2. New RRID SDES item

RFC EDITOR NOTE: Please replace RFCXXXX with the RFC number of this document.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

This document adds the MID SDES item to the IANA "RTCP SDES item types" registry as follows:

```
Value:          TBD
Abbrev.:        RRID
Name:           Redundancy RTP Stream Identifier
Reference:      RFCXXXX
```

6. Security Considerations

The actual identifiers used for RIDs and RRIDs are expected to be opaque. As such, they are not expected to contain information that would be sensitive, were it observed by third-parties.

7. Acknowledgements

Many thanks for review and input from Cullen Jennings, Magnus Westerlund, Colin Perkins, Peter Thatcher, Jonathan Lennox, and Paul Kyzivat.

8. References

8.1. Normative References

[I-D.ietf-avtext-sdes-hdr-ext]
Westerlund, M., Burman, B., Even, R., and M. Zanaty, "RTP Header Extension for RTCP Source Description Items", draft-ietf-avtext-sdes-hdr-ext-05 (work in progress), March 2016.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC5285] Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, DOI 10.17487/RFC5285, July 2008, <http://www.rfc-editor.org/info/rfc5285>.

[RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, Ed., "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <http://www.rfc-editor.org/info/rfc7656>.

8.2. Informative References

[I-D.ietf-mmusic-msid]
Alvestrand, H., "WebRTC MediaStream Identification in the Session Description Protocol", draft-ietf-mmusic-msid-11 (work in progress), October 2015.

[I-D.ietf-mmusic-rid]
Thatcher, P., Zanaty, M., Nandakumar, S., Burman, B., Roach, A., and B. Campen, "RTP Payload Format Constraints", draft-ietf-mmusic-rid-04 (work in progress), February 2016.

Authors' Addresses

Adam Roach
Mozilla

Email: adam@nostrum.com

Suhas Nandakumar
Cisco Systems

Email: snandaku@cisco.com

Peter Thatcher
Google

Email: pthatcher@google.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 9, 2017

A. Roach
Mozilla
S. Nandakumar
Cisco Systems
P. Thatcher
Google

October 06, 2016

RTP Stream Identifier Source Description (SDES)
draft-ietf-avtext-rid-09

Abstract

This document defines and registers two new RTCP Stream Identifier Source Description (SDES) items. One, named RtpStreamId, is used for unique identification of RTP streams. The other, RepairedRtpStreamId, can be used to identify which stream a redundancy RTP stream is to be used to repair.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

RTP sessions frequently consist of multiple streams, each of which is identified at any given time by its SSRC; however, the SSRC associated with a stream is not guaranteed to be stable over its lifetime. Within a session, these streams can be tagged with a number of identifiers, including CNAMEs and MSIDs [I-D.ietf-mmusic-msid]. Unfortunately, none of these have the proper ordinality to refer to an individual stream; all such identifiers can appear in more than one stream at a time. While approaches that use unique Payload Types (PTs) per stream have been used in some applications, this is a semantic overloading of that field, and one for which its size is inadequate: in moderately complex systems that use PT to uniquely identify every potential combination of codec configuration and unique stream, it is possible to simply run out of values.

To address this situation, we define a new RTCP Stream Identifier Source Description (SDES) identifier, RtpStreamId, that uniquely identifies a single RTP stream. A key motivator for defining this identifier is the ability to differentiate among different encodings of a single Source Stream that are sent simultaneously (i.e., simulcast). This need for unique identification extends to dependent streams (e.g., where layers used by a layered codec are transmitted on separate streams).

At the same time, when redundancy RTP streams are in use, we also need an identifier that connects such streams to the RTP stream for which they are providing redundancy. For this purpose, we define an additional SDES identifier, RepairedRtpStreamId. This identifier can appear only in packets associated with a redundancy RTP stream. They carry the same value as the RtpStreamId of the RTP stream that the redundant RTP stream is correcting.

2. Terminology

In this document, the terms "source stream", "RTP stream", "source RTP stream", "dependent stream", "received RTP stream", and "redundancy RTP stream" are used as defined in [RFC7656].

The following acronyms are also used:

- CNAME: Canonical End-Point Identifier, defined in [RFC3550]
- MID: Media Identification, defined in [I-D.ietf-mmusic-sdp-bundle-negotiation]
- MSID: Media Stream Identifier, defined in [I-D.ietf-mmusic-msid]
- RTCP: Real-time Transport Control Protocol, defined in [RFC3550]
- RTP: Real-time Transport Protocol, defined in [RFC3550]
- SDES: Source Description, defined in [RFC3550]
- SSRC: Synchronization Source, defined in [RFC3550]

3. Usage of RtpStreamId and RepairedRtpStreamId in RTP and RTCP

The RTP fixed header includes the payload type number and the SSRC values of the RTP stream. RTP defines how you de-multiplex streams within an RTP session; however, in some use cases, applications need further identifiers in order to effectively map the individual RTP Streams to their equivalent payload configurations in the SDP.

This specification defines two new RTCP SDES items [RFC3550]. The first item is 'RtpStreamId', which is used to carry RTP stream identifiers within RTCP SDES packets. This makes it possible for a receiver to associate received RTP packets (identifying the RTP stream) with a media description having the format constraint specified. The second is 'RepairedRtpStreamId', which can be used in redundancy RTP streams to indicate the RTP stream repaired by a redundancy RTP stream.

To be clear: the value carried in a RepairedRtpStreamId will always match the RtpStreamId value from another RTP stream in the same session. For example, if a source RTP stream is identified by RtpStreamId "A", then any redundancy RTP stream that repairs that source RTP stream will contain a RepairedRtpStreamId of "A" (if this mechanism is being used to perform such correlation). These redundant RTP streams may also contain their own unique RtpStreamId.

This specification also uses the RTP header extension for RTCP SDES items [I-D.ietf-avtext-sdes-hdr-ext] to allow carrying RtpStreamId and RepairedRtpStreamId values in RTP packets. This allows correlation at stream startup, or after stream changes where the use of RTCP may not be sufficiently responsive. This speed of response is necessary since, in many cases, the stream cannot be properly processed until it can be identified.

RtpStreamId and RepairedRtpStreamId values are scoped by source identifier (e.g., CNAME) and by media session. When the media is multiplexed using the BUNDLE extension [I-D.ietf-mmusic-sdp-bundle-negotiation], these values are further scoped by their associated MID values. For example: an RtpStreamId of "1" may be present in the stream identified with a CNAME of "1234@example.com", and may also be present in a stream with a CNAME of "5678@example.org", and these would refer to different streams. Similarly, an RtpStreamId of "1" may be present with an MID of "A", and again with a MID of "B", and also refer to two different streams.

Note that the RepairedRtpStreamId mechanism is limited to indicating one repaired stream per redundancy stream. If systems require correlation for schemes in which a redundancy stream contains information used to repair more than one stream, they will have to use a more complex mechanism than the one defined in this specification.

As with all SDES items, RtpStreamId and RepairedRtpStreamId are limited to a total of 255 octets in length. RtpStreamId and RepairedStreamId are constrained to contain only alphanumeric characters. For avoidance of doubt, the only allowed byte values for these IDs are decimal 48 through 57, 65 through 90, and 97 through 122.

3.1. RTCP 'RtpStreamId' SDES Extension

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|RtpStreamId=TBD|     length    | RtpStreamId                 ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The RtpStreamId payload is ASCII encoded and is not null-terminated.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

3.2. RTCP 'RepairedRtpStreamId' SDES Extension

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Repaired...=TBD|     length    | RepairRtpStreamId           ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The RepairedRtpStreamId payload is ASCII encoded and is not null-terminated.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

3.3. RTP 'RtpStreamId' and 'RepairedRtpStreamId' Header Extensions

Because recipients of RTP packets will typically need to know which streams they correspond to immediately upon receipt, this specification also defines a means of carrying RtpStreamId and RepairedRtpStreamId identifiers in RTP extension headers, using the technique described in [I-D.ietf-avtext-sdes-hdr-ext].

As described in that document, the header extension element can be encoded using either the one-byte or two-byte header, and the identification-tag payload is ASCII-encoded.

As the identifier is included in an RTP header extension, there should be some consideration given to the packet expansion caused by the identifier. To avoid Maximum Transmission Unit (MTU) issues for the RTP packets, the header extension's size needs to be taken into account when encoding media. Note that the set of header extensions included in the packet needs to be padded to the next 32-bit boundary [RFC5285].

In many cases, a one-byte identifier will be sufficient to distinguish streams in a session; implementations are strongly encouraged to use the shortest identifier that fits their purposes. Implementors are warned, in particular, not to include any information in the identifier that is derived from potentially user-identifying information, such as user ID or IP address. To avoid identification of specific implementations based on their pattern of tag generation, implementations are encouraged to use a simple scheme that starts with the ASCII digit "1", and increments by one for each subsequent identifier.

4. IANA Considerations

4.1. New RtpStreamId SDES item

RFC EDITOR NOTE: Please replace RFCXXXX with the RFC number of this document.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

This document adds the RtpStreamId SDES item to the IANA "RTP SDES item types" registry as follows:

Value:	TBD
Abbrev.:	RtpStreamId
Name:	RTP Stream Identifier
Reference:	RFCXXXX

4.2. New RepairRtpStreamId SDES item

RFC EDITOR NOTE: Please replace RFCXXXX with the RFC number of this document.

RFC EDITOR NOTE: Please replace TBD with the assigned SDES identifier value.

This document adds the RepairedRtpStreamId SDES item to the IANA "RTP SDES item types" registry as follows:

Value:	TBD
Abbrev.:	RepairedRtpStreamId
Name:	Repaired RTP Stream Identifier
Reference:	RFCXXXX

4.3. New RtpStreamId Header Extension URI

RFC EDITOR NOTE: Please replace RFCXXXX with the RFC number of this document.

This document defines a new extension URI in the RTP SDES Compact Header Extensions sub-registry of the RTP Compact Header Extensions registry sub-registry, as follows

Extension URI: urn:ietf:params:rtp-hdrext:sdes:rtp-stream-id
Description: RTP Stream Identifier Contact: adam@nostrum.com
Reference: RFCXXXX

4.4. New RepairRtpStreamId Header Extension URI

RFC EDITOR NOTE: Please replace RFCXXXX with the RFC number of this document.

This document defines a new extension URI in the RTP SDES Compact Header Extensions sub-registry of the RTP Compact Header Extensions registry sub-registry, as follows

Extension URI: urn:ietf:params:rtp-hdrext:sdes:repaired-rtp-sream-id
Description: RTP Repaired Stream Identifier Contact: adam@nostrum.com
Reference: RFCXXXX

5. Security Considerations

Although the identifiers defined in this document are limited to be strictly alphanumeric, SDES items have the potential to carry any string. As a consequence, there exists a risk that it might carry privacy-sensitive information. Implementations need to take care when generating identifiers so that they do not contain information that can identify the user or allow for long term tracking of the device. Following the generation recommendations in Section 3.3 will result in non-instance-specific labels, with only minor fingerprinting possibilities in the total number of used RtpStreamIds and RepairedRtpStreamIds.

Even if the SDES items are generated to convey as little information as possible, implementors are strongly encouraged to encrypt SDES items - both in RTCP and RTP header extensions - so as to preserve privacy against third parties.

As the SDES items are used for identification of the RTP streams for different application purposes, it is important that the intended values are received. An attacker, either a third party or malicious RTP middlebox, that removes, or changes the values for these SDES items, can severely impact the application. The impact can include failure to decode or display the media content of the RTP stream. It can also result in incorrectly attributing media content to identifiers of the media source, such as incorrectly identifying the speaker. To prevent this from occurring due to third party attacks, integrity and source authentication is needed.

Options for Securing RTP Sessions [RFC7201] discusses options for how encryption, integrity and source authentication can be accomplished.

6. Acknowledgements

Many thanks for review and input from Cullen Jennings, Magnus Westerlund, Colin Perkins, Jonathan Lennox, and Paul Kyzivat. Magnus Westerlund provided substantially all of the Security Considerations section.

7. References

7.1. Normative References

[I-D.ietf-avtext-sdes-hdr-ext]
Westerlund, M., Burman, B., Even, R., and M. Zanaty, "RTP Header Extension for RTCP Source Description Items", draft-ietf-avtext-sdes-hdr-ext-07 (work in progress), June 2016.

[I-D.ietf-mmusic-sdp-bundle-negotiation]
Holmberg, C., Alvestrand, H., and C. Jennings, "Negotiating Media Multiplexing Using the Session Description Protocol (SDP)", draft-ietf-mmusic-sdp-bundle-negotiation-32 (work in progress), August 2016.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC5285] Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, DOI 10.17487/RFC5285, July 2008, <http://www.rfc-editor.org/info/rfc5285>.

[RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, Ed., "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <http://www.rfc-editor.org/info/rfc7656>.

7.2. Informative References

[I-D.ietf-mmusic-msid]
Alvestrand, H., "WebRTC MediaStream Identification in the Session Description Protocol", draft-ietf-mmusic-msid-15 (work in progress), July 2016.

[RFC7201] Westerlund, M. and C. Perkins, "Options for Securing RTP Sessions", RFC 7201, DOI 10.17487/RFC7201, April 2014, <http://www.rfc-editor.org/info/rfc7201>.

Authors' Addresses

Adam Roach
Mozilla

Email: adam@nostrum.com

Suhas Nandakumar
Cisco Systems

Email: snandaku@cisco.com

Peter Thatcher
Google

Email: pthatcher@google.com

Network Working Group M. Westerlund
Internet-Draft B. Burman
Intended status: Standards Track Ericsson
Expires: September 3, 2016 R. Even
Huawei Technologies
M. Zanaty
Cisco Systems
March 2, 2016

RTP Header Extension for RTCP Source Description Items
draft-ietf-avtext-sdes-hdr-ext-05

Abstract

Source Description (SDES) items are normally transported in RTP control protocol (RTCP). In some cases it can be beneficial to speed up the delivery of these items. Mainly when a new source (SSRC) joins an RTP session and the receivers needs this source’s identity, relation to other sources, or its synchronization context, all of which may be fully or partially identified using SDES items. To enable this optimization, this document specifies a new RTP header extension that can carry SDES items.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

This specification defines an RTP header extension [RFC3550][RFC5285] that can carry RTCP source description (SDES) items. By including selected SDES items in a header extension the determination of relationship and synchronization context for new RTP streams (SSRCs) in an RTP session can be optimized. Which relationship and what information depends on the SDES items carried. This becomes a complement to using only RTCP for SDES Item delivery.

It is important to note that not all SDES items are appropriate to transmit using RTP header extensions. Some SDES items performs binding or identifies synchronization context with strict timeliness requirements, while many other SDES items do not have such requirements. In addition, security and privacy concerns for the SDES item information need to be considered. For example, the Name and Location SDES items are highly sensitive from a privacy perspective and should not be transported over the network without strong security. No use case has identified where this information is required at the same time as the first RTP packets arrive. A few seconds delay before such information is available to the receiver appears acceptable. Therefore only appropriate SDES items will be registered for use with this header extension, such as CNAME.

First, some requirements language and terminology are defined. The following section motivates why this header extension is sometimes required or at least provides a significant improvement compared to waiting for regular RTCP packet transmissions of the information. This is followed by a specification of the header extension and usage recommendations. Next, a sub-space of the header-extension URN is defined to be used for existing and future SDES items, and then the appropriate existing SDES items are registered.

2. Definitions

2.1. 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].

2.2. Terminology

This document uses terminology defined in "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources" [RFC7656]. In particular the following definitions:

Media Source

RTP Stream

Media Encoder

Participant

3. Motivation

Source Description (SDES) items are associated with a particular SSRC and thus RTP stream. The source description items provide various meta data associated with the SSRC. How important it is to have this data no later than when receiving the first RTP packets depends on the item itself. The CNAME item is one item that is commonly needed either at reception of the first RTP packet for this SSRC, or at least by the time the first media can be played out. If it is not available, the synchronization context cannot be determined and thus any related streams cannot be correctly synchronized. Thus, this is a valuable example for having this information early when a new RTP stream is received.

The main reason for new SSRCs in an RTP session is when media sources are added. This can be either because an end-point is adding a new actual media source, or additional participants in a multi-party session are added to the session. Another reason for a new SSRC can be an SSRC collision that forces both colliding parties to select new SSRCs.

For the case of rapid media synchronization, one may use the RTP header extension for Rapid Synchronization of RTP Flows [RFC6051]. This header extension carries the clock information present in the RTCP sender report (SR) packets. It however assumes that the CNAME binding is known, which can be provided via signaling [RFC5576] in some cases, but not all. Thus an RTP header extension for carrying SDES items like CNAME is a powerful combination to enable rapid synchronization in all cases.

The Rapid Synchronization of RTP Flows specification does provide an analysis of the initial synchronization delay for different sessions depending on number of receivers as well as on session bandwidth (Section 2.1 of [RFC6051]). These results are applicable also for other SDES items that have a similar time dependency until the information can be sent using RTCP. These figures can be used to determine the benefit of reducing the initial delay before information is available for some use cases.

That document also discusses the case of late joiners, and defines an RTCP Feedback format to request synchronization information, which is another potential use case for SDES items in RTP header extension. It would for example be natural to include CNAME SDES item with the header extension containing the NTP formatted reference clock to ensure synchronization.

There is an another SDES item that can benefit from timely delivery, and an RTP header extension SDES item was therefore defined for it:

MID: This is a media description identifier that matches the value of the Session Description Protocol (SDP) [RFC4566] a=mid attribute, to associate RTP streams multiplexed on the same transport with their respective SDP media description as described in [I-D.ietf-mmusic-sdp-bundle-negotiation].

4. Specification

This section first specifies the SDES item RTP header extension format, followed by some usage considerations.

4.1. SDES Item Header Extension

An RTP header extension scheme allowing for multiple extensions is defined in "A General Mechanism for RTP Header Extensions" [RFC5285]. That specification defines both short and long item headers. The short headers (One-byte) are restricted to 1 to 16 bytes of data, while the long format (Two-byte) supports a data length of 0 to 255 bytes. Thus the RTP header extension formats are capable of supporting any SDES item from a data length perspective.

The ID field, independent of short or long format, identifies both the type of RTP header extension and, in the case of the SDES item header extension, the type of SDES item. The mapping is done in signaling by identifying the header extension and SDES item type using a URN, which is defined in the IANA consideration (Section 5) for the known SDES items appropriate to use.

4.1.1. One-Byte Format

The one-byte header format for an SDES item extension element consists of the one-byte header (defined in Section 4.2 of [RFC5285]), which consists of a 4-bit ID followed by a 4-bit length field (len) that identifies the number of data bytes (len value +1) following the header. The data part consists of len+1 bytes of UTF-8 text. The type of text and its mapping to the SDES item type is determined by the ID field value.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID   |  len  | SDES Item text value ...                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1

4.1.2. Two-Byte Format

The two-byte header format for an SDES item extension element consists of the two-byte header (defined in Section 4.3 of [RFC5285]), which consists of an 8-bit ID followed by an 8-bit length field (len) that identifies the number of data bytes following the header. The data part consists of len bytes of UTF-8 text. The type of text and its mapping to the SDES item type is determined by the ID field value.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      ID       |      len      |  SDES Item text value ...     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2

4.2. Usage of the SDES Item Header Extension

This section discusses various usage considerations; which form of header extension to use, the packet expansion, and when to send SDES items in header extension.

4.2.1. One or Two Byte Headers

The RTP header extensions for SDES items MAY use either the one-byte or two-byte header formats, depending on the text value size for the used SDES items and the requirement from any other header extensions used. The one-byte header SHOULD be used when all non SDES item header extensions supports the one-byte format and all SDES item text values contain at most 16 bytes. Note that the RTP header extension specification does not allow mixing one-byte and two-byte headers for the same RTP stream (SSRC), so if the value size of any of the SDES items value requires the two-byte header, then all other header extensions MUST also use the two-byte header format.

For example using CNAMEs that are generated according to "Guidelines for Choosing RTP Control Protocol (RTCP) Canonical Names (CNAMEs)" [RFC7022], using short term persistent values, and if 96-bit random values prior to base64 encoding are sufficient, then they will fit into the one-byte header format.

An RTP middlebox needs to take care choosing between one-byte headers and two-byte headers when creating the first packets for an outgoing stream (SSRC) with header extensions. First of all it needs to consider all the header extensions that may potentially be used. Secondly, it needs to know the size of the SDES items that are going to be included, and use two bytes headers if any are longer than 16 bytes. An RTP middlebox that forwards a stream, i.e. not mixing it or combing it with other streams, may be able to base its choice on the header size in incoming streams. This is assuming that the middlebox does not modify the stream or add additional header extensions to the stream it sends, in which case it needs to make its own decision.

4.2.2. MTU and Packet Expansion

The RTP packet size will clearly increase when a header extension is included. How much depends on the type of header extensions and their data content. The SDES items can vary in size. There are also some use-cases that require transmitting multiple SDES items in the same packet to ensure that all relevant data reaches the receiver. An example of that is when both CNAME, a MID, and the rapid time synchronization extension from RFC 6051 are needed. Such a combination is quite likely to result in at least 16+3+8 bytes of data plus the headers, which will be another 7 bytes for one-byte headers, plus two bytes of header padding to make the complete header extension word aligned, thus in total 36 bytes.

If the packet expansion cannot be taken into account when producing the RTP payload, it can cause an issue. An RTP payload that is created to meet a particular IP level Maximum Transmission Unit (MTU), taking the addition of IP/UDP/RTP headers but not RTP header extensions into account, could exceed the MTU when the header extensions are present, thus resulting in IP fragmentation. IP fragmentation is known to negatively impact the loss rate due to middleboxes unwilling or not capable of dealing with IP fragments, as well as increasing the target surface for other types of packet losses.

As this is a real issue, the media encoder and payload packetizer should be flexible and be capable of handling dynamically varying payload size restrictions to counter the packet expansion caused by header extensions. If that is not possible, some reasonable worst case packet expansion should be calculated and used to reduce the RTP payload size of all RTP packets the sender transmits.

4.2.3. Transmission Considerations

The general recommendation is to only send header extensions when needed. This is especially true for SDES items that can be sent in periodic repetitions of RTCP throughout the whole session. Thus, the different usages (Section 4.2.4) have different recommendations. First some general considerations for getting the header extensions delivered to the receiver:

1. The probability for packet loss and burst loss determine how many repetitions of the header extensions will be required to reach a targeted delivery probability, and if burst loss is likely, what distribution would be needed to avoid getting all repetitions of the header extensions lost in a single burst.

2. If a set of packets are all needed to enable decoding, there is commonly no reason for including the header extension in all of these packets, as they share fate. Instead, at most one instance of the header extension per independently decodable set of media data would be a more efficient use of the bandwidth.

3. How early the SDES item information is needed, from the first received RTP data or only after some set of packets are received, can guide if the header extension(s) should be in all of the first N packets or be included only once per set of packets, for example once per video frame.

4. The use of RTP level robustness mechanisms, such as RTP retransmission [RFC4588], or Forward Error Correction, e.g., [RFC5109] may treat packets differently from a robustness perspective, and SDES header extensions should be added to packets that get a treatment corresponding to the relative importance of receiving the information.

As a summary, the number of header extension transmissions should be tailored to a desired probability of delivery taking the receiver population size into account. For the very basic case, N repetitions of the header extensions should be sufficient, but may not be optimal. N is selected so that the header extension target delivery probability reaches 1-P^N, where P is the probability of packet loss. For point to point or small receiver populations, it might also be possible to use feedback, such as RTCP, to determine when the information in the header extensions has reached all receivers and stop further repetitions. Feedback that can be used includes the RTCP XR Loss RLE report block [RFC3611], which will indicate successful delivery of particular packets. If the RTP/AVPF Transport Layer Feedback Messages for generic NACK [RFC4585] is used, it can indicate the failure to deliver an RTP packet with the header extension, thus indicating the need for further repetitions. The normal RTCP report blocks can also provide an indicator of successful delivery, if no losses are indicated for a reporting interval covering the RTP packets with the header extension. Note that loss of an RTCP packet reporting on an interval where RTP header extension packets were sent, does not necessarily mean that the RTP header extension packets themselves were lost.

4.2.4. Different Usages

4.2.4.1. New SSRC

A new SSRC joins an RTP session. As this SSRC is completely new for everyone, the goal is to ensure, with high probability, that all RTP session participants receives the information in the header extension. Thus, header extension transmission strategies that allow some margins in the delivery probability should be considered.

4.2.4.2. Late Joiner

In a multi-party RTP session where one or a small number of receivers join a session where the majority of receivers already have all necessary information, the use of header extensions to deliver relevant information should be tailored to reach the new receivers. The trigger to send header extensions can for example either be RTCP from new receiver(s) or an explicit request like the Rapid Resynchronization Request defined in [RFC6051]. In centralized topologies where an RTP middlebox is present, it can be responsible for transmitting the known information, possibly stored, to the new session participant only, and not repeat it to all the session participants.

4.2.4.3. Information Change

If the SDES information is tightly coupled with the RTP data, and the SDES information needs to be updated, then the use of the RTP header extension is superior to RTCP. Using the RTP header extension ensures that the information is updated on reception of the related RTP media, ensuring synchronization between the two. Continued use of the old SDES information can lead to undesired effects in the application. Thus, header extension transmission strategies with high probability of delivery should be chosen.

4.2.5. SDES Items in RTCP

The RTP header extension information, i.e. SDES items, can and will be sent also in RTCP. Therefore, it is worth making some reflections on this interaction. As an alternative to the header extension, it is possible to schedule a non-regular RTCP packet transmission containing important SDES items, if one uses an RTP/AVPF-based RTP profile. Depending on which mode one’s RTCP feedback transmitter is working on, extra RTCP packets may be sent as immediate or early packets, enabling more timely SDES information delivery.

There are however two aspects that differ between using RTP header extensions and any non-regular transmission of RTCP packets. First, as the RTCP packet is a separate packet, there is no direct relation and also no fate sharing between the relevant media data and the SDES information. The order of arrival for the packets will matter. With a header-extension, the SDES items can be ensured to arrive if the media data to play out arrives. Secondly, it is difficult to determine if an RTCP packet is actually delivered. This, as the RTCP packets lack both sequence number and a mechanism providing feedback on the RTCP packets themselves.

4.2.6. Update Flaps

The SDES item may arrive both in RTCP and in RTP header extensions, potentially causing the value to flap back and forth at the time of updating. There are at least two reasons for these flaps. The first one is packet reordering, where a pre-update RTP or RTCP packet with an SDES item is delivered to the receiver after the first RTP/RTCP packet with the updated value. The second reason is the different code-paths for RTP and RTCP in implementations. An update to the sender’s SDES item parameter can take a different time to propagate to the receiver than the corresponding media data. For example, an RTCP packet with the SDES item included that may have been generated prior to the update can still reside in a buffer and be sent unmodified. The update of the item’s value can at the same time cause RTP packets to be sent including the header extension, prior to the RTCP packet being sent.

However, most of these issues can be avoided by the receiver performing some checks before updating the receiver’s stored value. To handle flaps caused by reordering, only SDES items received in RTP packets with a higher extended sequence number than the last change shall be applied, i.e. discard items that can be determined to be older than the current one. For compound RTCP packets, which will contain an Sender Report (SR) packet (assuming an active RTP sender), the receiver can use the RTCP SR Timestamp field to determine at what approximate time it was transmitted. If the timestamp is earlier than the last received RTP packet with a header extension carrying an SDES item, and especially if carrying a previously used value, the SDES item in the RTCP SDES packet can be ignored. Note that media processing and transmission pacing can easily cause the RTP header timestamp field as well as the RTCP SR timestamp field to not match with the actual transmission time.

5. IANA Considerations

This section makes the following requests to IANA:

- o Create a new sub-registry reserved for RTCP SDES items with the URN sub-space "urn:ietf:params:rtp-hdrext:sdes:" in the RTP Compact Header Extensions registry.

- o Register the SDES items appropriate for use with the RTP header extension defined in this document.

RFC-editor note: Please replace all occurrences of RFCXXXX with the RFC number this specification receives when published.

5.1. Registration of an SDES Base URN

IANA is requested to register the below entry in the RTP Compact Header Extensions registry:

```
Extension URI: urn:ietf:params:rtp-hdrext:sdes
Description:   Reserved as base URN for RTCP SDES items that are also
               defined as RTP Compact header extensions.
Contact:       Authors of [RFCXXXX]
Reference:     [RFCXXXX]
```

The reason to register a base URN for an SDES sub-space is that the name represents an RTCP Source Description item, where a specification is strongly recommended [RFC3550].

5.2. Creation of an SDES Sub-Registry

IANA is requested to create a sub-registry to the RTP Compact Header Extensions registry, with the same basic requirements, structure and layout as the RTP Compact Header Extensions registry.

- o Registry name: RTP SDES Compact Header Extensions

- o Specification: RFCXXXX and RFCs updating RFCXXXX

- o Information required: Same as for RTP Header Extensions [RFC5285] registry

- o Review process: Same as for RTP Header Extensions [RFC5285] registry, with the following requirements added to the expert review:

   1. Any registration using an Extension URI that starts with "urn:ietf:params:rtp-hdrext:sdes:" (Section 5.1) MUST also have a registered Source Description item in the "RTP SDES item types" registry.

2. A security and privacy consideration for the SDES item MUST be provided with the registration.

3. Information MUST be provided on why this SDES item requires timely delivery, motivating it to be transported in a header extension rather than as RTCP only.

- o Size and format of entries: Same as for RTP Header Extensions [RFC5285] registry.

- o Initial assignments: See Section 5.3 below.

5.3. Registration of SDES Items

It is requested that the following SDES item is registered in the newly formed RTP SDES Compact Header Extensions registry:

```
Extension URI: urn:ietf:params:rtp-hdrext:sdes:cname
Description:   Source Description: Canonical End-Point Identifier
               (SDES CNAME)
Contact:       Authors of [RFCXXXX]
Reference:     [RFCXXXX]
```

6. Security Considerations

Source Description items may contain data that are sensitive from a security perspective. There are SDES items that are or may be sensitive from a user privacy perspective, like CNAME, NAME, EMAIL, PHONE, LOC and H323-CADDR. Some may contain sensitive information, like NOTE and PRIV, while others may be sensitive from profiling implementations for vulnerability or other reasons, like TOOL. The CNAME sensitivity can vary depending on how it is generated and what persistence it has. A short term CNAME identifier generated using a random number generator [RFC7022] may have minimal security implications, while a CNAME of the form user@host has privacy concerns, and a CNAME generated from a MAC address has long term tracking potentials.

In RTP sessions where any type of confidentiality protection is enabled for RTCP, the SDES item header extensions MUST also be protected. This implies that to provide confidentiality, users of SRTP need to implement and use encrypted header extensions per [RFC6904]. The security level that is applied to RTCP packets carrying SDES items SHOULD also be applied to SDES items carried as RTP header extensions. If the security level is chosen to be different for an SDES item in RTCP and RTP header extension, it is important to motivate the exception, and to consider the security properties as the worst in each aspect for the different configurations.

As the SDES items are used by the RTP based application to establish relationships between RTP streams or between an RTP stream and information about the originating participant, there SHOULD be strong integrity protection and source authentication of the header extensions. If not, an attacker can modify the SDES item value to create erroneous relationship bindings in the receiving application.

7. Acknowledgements

The authors likes to thank the following individuals for feedback and suggestions; Colin Perkins, Ben Campbell.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC5285] Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, DOI 10.17487/RFC5285, July 2008, <http://www.rfc-editor.org/info/rfc5285>.

[RFC6904] Lennox, J., "Encryption of Header Extensions in the Secure Real-time Transport Protocol (SRTP)", RFC 6904, DOI 10.17487/RFC6904, April 2013, <http://www.rfc-editor.org/info/rfc6904>.

8.2. Informative References

[I-D.ietf-mmusic-sdp-bundle-negotiation] Holmberg, C., Alvestrand, H., and C. Jennings, "Negotiating Media Multiplexing Using the Session Description Protocol (SDP)", draft-ietf-mmusic-sdp-bundle-negotiation-27 (work in progress), February 2016.

[RFC3611] Friedman, T., Ed., Caceres, R., Ed., and A. Clark, Ed., "RTP Control Protocol Extended Reports (RTCP XR)", RFC 3611, DOI 10.17487/RFC3611, November 2003, <http://www.rfc-editor.org/info/rfc3611>.

[RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, DOI 10.17487/RFC4566, July 2006, <http://www.rfc-editor.org/info/rfc4566>.

[RFC4585] Ott, J., Wenger, S., Sato, N., Burmeister, C., and J. Rey, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 4585, DOI 10.17487/RFC4585, July 2006, <http://www.rfc-editor.org/info/rfc4585>.

[RFC4588] Rey, J., Leon, D., Miyazaki, A., Varsa, V., and R. Hakenberg, "RTP Retransmission Payload Format", RFC 4588, DOI 10.17487/RFC4588, July 2006, <http://www.rfc-editor.org/info/rfc4588>.

[RFC5109] Li, A., Ed., "RTP Payload Format for Generic Forward Error Correction", RFC 5109, DOI 10.17487/RFC5109, December 2007, <http://www.rfc-editor.org/info/rfc5109>.

[RFC5576] Lennox, J., Ott, J., and T. Schierl, "Source-Specific Media Attributes in the Session Description Protocol (SDP)", RFC 5576, DOI 10.17487/RFC5576, June 2009, <http://www.rfc-editor.org/info/rfc5576>.

[RFC6051] Perkins, C. and T. Schierl, "Rapid Synchronisation of RTP Flows", RFC 6051, DOI 10.17487/RFC6051, November 2010, <http://www.rfc-editor.org/info/rfc6051>.

[RFC7022] Begen, A., Perkins, C., Wing, D., and E. Rescorla, "Guidelines for Choosing RTP Control Protocol (RTCP) Canonical Names (CNAMEs)", RFC 7022, DOI 10.17487/RFC7022, September 2013, <http://www.rfc-editor.org/info/rfc7022>.

[RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, Ed., "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <http://www.rfc-editor.org/info/rfc7656>.

Authors' Addresses

Magnus Westerlund
Ericsson
Farogatan 6
SE-164 80 Stockholm
Sweden

Phone: +46 10 714 82 87
Email: magnus.westerlund@ericsson.com

Bo Burman
Ericsson
Kistavagen 25
Stockholm 16480
Sweden

Email: bo.burman@ericsson.com

Roni Even
Huawei Technologies
Tel Aviv
Israel

Email: roni.even@mail01.huawei.com

Mo Zanaty
Cisco Systems
7100 Kit Creek
RTP, NC 27709
USA

Email: mzanaty@cisco.com

AVTEXT Working Group
INTERNET-DRAFT
Intended Status: Standards Track
Expires: September 22, 2016

J. Xia
R. Even
R. Huang
Huawei
L. Deng
China Mobile

March 21, 2016

RTP/RTCP extension for RTP Splicing Notification
draft-ietf-avtext-splicing-notification-05

Abstract

Content splicing is a process that replaces the content of a main multimedia stream with other multimedia content, and delivers the substitutive multimedia content to the receivers for a period of time. The splicer is designed to handle RTP splicing and needs to know when to start and end the splicing.

This memo defines two RTP/RTCP extensions to indicate the splicing related information to the splicer: an RTP header extension that conveys the information in-band and an RTCP packet that conveys the information out-of-band.

Status of this Memo

Copyright and License Notice

Table of Contents

1 Introduction

Splicing is a process that replaces some multimedia content with other multimedia content and delivers the substitutive multimedia content to the receivers for a period of time. In some predictable splicing cases, e.g., advertisement insertion, the splicing duration needs to be inside of the specific, pre-designated time slot. Certain timing information about when to start and end the splicing must be first acquired by the splicer in order to start the splicing. This document refers to this information as the Splicing Interval.

[SCTE35] provides a method that encapsulates the Splicing Interval inside the MPEG2-TS layer in cable TV systems. When transported in RTP, an middle box designed as the splicer to decode the RTP packets and search for the Splicing Interval inside the payloads is required. The middle box is either a translator or a mixer as described in [RFC6828]. The need for such processing increases the workload of the middle box and limits the number of RTP sessions the middle box can support.

The document defines an RTP header extension [RFC5285] used by the main RTP sender to provide the Splicing Interval by including it in the RTP packets.

However, the Splicing Interval conveyed in the RTP header extension might not reach the splicer successfully. Any splicing un-aware middlebox on the path between the RTP sender might strip this RTP header extension.

To increase robustness against such case, the document also defines a complementary RTCP packet type to carry the same Splicing Interval to the splicer.

1.1 Terminology

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].

In addition, we define following terminologies:

Main RTP sender:

The sender of RTP packets carrying the main RTP stream.

Splicer:

An intermediary node that inserts substitutive content into a main

RTP stream. The splicer sends substitutive content to the RTP receiver instead of the main content during splicing. It is also responsible for processing RTCP traffic between the RTP sender and the RTP receiver.

Splicing-In Point

A virtual point in the RTP stream, suitable for substitutive content entry, typically in the boundary between two independently decodable frames.

Splicing-Out Point

A virtual point in the RTP stream, suitable for substitutive content exit, typically in the boundary between two independently decodable frames.

Splicing Interval:

The NTP-format timestamps, representing the main RTP sender wallclock time, for the Splicing-In point and Splicing-Out point per [RFC6828] allowing the splicer to know when to start and end the RTP splicing.

Substitutive RTP Sender:

The sender of RTP packets carrying the RTP stream that will replace the content in the main RTP stream.

2 Overview of RTP Splicing Notification

A splicer is designed to handle splicing on the RTP layer at the reserved time slots set by the main RTP sender. This implies that the splicer must first know the Splicing Interval from the main RTP sender before it can start splicing. The splicer can be a mixer as described in [RFC6828].

When a new splicing is forthcoming, the main RTP sender needs to send the Splicing Interval to the splicer. The Splicing Interval SHOULD be sent more than once to mitigate the possible packet loss. To enable the splicer to get the substitutive content before the splicing starts, the main RTP sender MUST send the Splicing Interval far ahead. For example, the main RTP sender can estimate when to send the Splicing Interval based on the round-trip time (RTT) following the mechanisms in section 6.4.1 of [RFC3550] when the splicer sends RTCP RR to the main sender.

The substitutive sender also needs to learn the Splicing Interval from the main RTP sender in advance, and thus estimates when to transfer the substitutive content to the splicer. The Splicing Interval could be transmitted from the main RTP sender to the substitutive content using some out-of-band mechanisms, for example, a proprietary mechanism to exchange the Splicing Interval, or the substitutive sender is implemented together with the main RTP sender inside a single device. To ensure the Splicing Interval is valid for both the main RTP sender and the substitutive RTP sender, the two senders MUST share a common reference clock so that the splicer can achieve accurate splicing. The requirements for the common reference clock (e.g. resolution, skew) depend on the codec used by the media content.

In this document, the main RTP sender uses a pair of NTP-format timestamps, to indicate when to start and end the splicing to the splicer: the timestamp of the first substitutive RTP packet at the splicing in point, and the timestamp of the first main RTP packet at the splicing out point.

When the substitutive RTP sender gets the Splicing Interval, it must prepare the substitutive stream. The main and the substitutive content providers MUST ensure that the RTP timestamp of the first substitutive RTP packet that would be presented to the receivers corresponds to the same time instant as the former NTP-format timestamp in the Splicing Interval. To enable the splicer to know the first substitutive RTP packet it needs to send, the substitutive RTP sender MUST send the substitutive RTP packet ahead of the Splicing In point, allowing the splicer to find out the timestamp of this first RTP packet in the substitutive RTP stream, e.g., using a prior RTCP SR (Sender Report) message.

When the splicing will end, the main content provider and the substitutive content provider MUST ensure the RTP timestamp of the first main RTP packet that would be presented on the receivers corresponds to the same time instant as the latter NTP-format timestamp in the Splicing Interval.

3 Conveying Splicing Interval in RTP/RTCP extensions

This memo defines two backwards compatible RTP extensions to convey the Splicing Interval to the splicer: an RTP header extension and an RTCP splicing notification message.

3.1 RTP Header Extension

The RTP header extension mechanism defined in [RFC5285] can be adapted to carry the Splicing Interval consisting of a pair of NTP-format timestamps. Two variants are defined for this header extension: one-byte header variant and two-byte header variant.

For one-byte header variant, it carries the 7 octets splicing-out NTP-format timestamp (lower 24-bit part of the Seconds of a NTP-format timestamp and the 32 bits of the Fraction of a NTP-format timestamp as defined in [RFC5905]), followed by the 8 octets splicing-in NTP-format timestamp (64-bit NTP-format timestamp as defined in [RFC5905]). The top 8 bits of the splicing-out NTP timestamp are inferred from the top 8 bits of the splicing-in NTP timestamp. This is unambiguous, under the assumption that the splicing-out time is after the splicing-in time, and the splicing interval is less than 2^25 seconds. Therefore, if the value of 7 octets splicing-out NTP-format timestamp is smaller than the value of 7 lower octets splicing-in NTP-format, it implies a wrap of the 56-bit splicing-out NTP-format timestamp which means the top 8-bit value of the 64-bit splicing-out is equal to the top 8-bit value of splicing-in NTP Timestamp plus 0x01. Otherwise, the top 8 bits of splicing-out NTP timestamp is equal to the top 8 bits of splicing-in NTP Timestamp.

The format is shown in Figures 1.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+E
|   ID  | L=15  | OUT NTP timestamp format - Seconds (bit 8-31) |x
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+t
|       OUT NTP timestamp format - Fraction (bit 0-31)          |e
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+n
|       IN  NTP timestamp format - Seconds (bit 0-31)           |s
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+i
|       IN  NTP timestamp format - Fraction (bit 0-31)          |o
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+n
```

Figure 1: Splicing Interval
Using the One-Byte Header Format

For two-byte header variant, it carries the 6 octets splicing-out NTP-format timestamp (lower 16-bit part of the Seconds of a NTP-format timestamp and the 32 bits of the Fraction of a NTP-format timestamp as defined in [RFC5905]), followed by the 8 octets splicing-in NTP-format timestamp (64-bit NTP-format timestamp as defined in [RFC5905]). Also, the top 16 bits of the splicing-out NTP timestamp are inferred from the top 16 bits of the splicing-in NTP timestamp, under the assumption that the splicing-out time is after the splicing-in time, and the splicing interval is less than 2^16 seconds. Similar to the one-byte header, if the value of 6 octets splicing-out NTP-format timestamp is smaller than the value of 6 lower octets splicing-in NTP-format, it implies a wrap of the 48-bit splicing-out NTP-format timestamp which means the top 16-bit value of the 64-bit splicing-out is equal to the top 16-bit value of splicing-in NTP Timestamp plus 0x01. Otherwise, the top 16 bits of splicing-out NTP timestamp is equal to the top 16 bits of splicing-in NTP Timestamp.

The format is shown in Figures 2.

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+E
 |   ID          |    L=14       | OUT NTP timestamp  - Seconds  |x
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+t
 |       OUT NTP timestamp format - Fraction (bit 0-31)          |e
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+n
 |       IN  NTP timestamp format - Seconds (bit 0-31)           |s
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+i
 |       IN  NTP timestamp format - Fraction (bit 0-31)          |o
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+n
```

Figure 2: Splicing Interval
Using the Two-Byte Header Format

Since the inclusion of an RTP header extension will reduce the efficiency of RTP header compression, it is RECOMMENDED that the main sender inserts the RTP header extensions into only a number of RTP packets, instead of all the RTP packets, prior to the splicing in.

After the splicer intercepts the RTP header extension and derives the Splicing Interval, it will generate its own stream and SHOULD NOT include the RTP header extension in outgoing packets to reduce header overhead.

3.2 RTCP Splicing Notification Message

In addition to the RTP header extension, the main RTP sender includes the Splicing Interval in an RTCP splicing notification message. Whether or not the timestamps are included in the RTP header extension, the main RTP sender MUST send the RTCP splicing notification message. This provide robustness in the case where a middlebox strips RTP header extensions. The main RTP sender MUST make sure the splicing information contained in the RTCP splicing notification message consistent with the information included in the RTP header extensions.

The RTCP splicing notification message is a new RTCP packet type. It has a fixed header followed by a pair of NTP-format timestamps:

```
     0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |V=2|P|reserved |    PT=TBA     |             length            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                              SSRC                             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             IN NTP Timestamp (most significant word)          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             IN NTP Timestamp (least significant word)         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             OUT NTP Timestamp (most significant word)         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             OUT NTP Timestamp (least significant word)        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: RTCP Splicing Notification Message

The RSI packet includes the following fields:

Length: 16 bits

As defined in [RFC3550], the length of the RTCP packet in 32-bit words minus one, including the header and any padding.

SSRC: 32 bits

The SSRC of the Main RTP Sender.

Timestamp: 64 bits

Indicates the wallclock time when this splicing starts and ends. The full-resolution NTP-format timestamp is used, which is a 64-bit, unsigned, fixed-point number with the integer part in the first 32 bits and the fractional part in the last 32 bits. This format is same as the NTP timestamp field in the RTCP Sender Report (Section 6.4.1 of [RFC3550]).

The RTCP splicing notification message can be included in the RTCP compound packet together with RTCP SR generated at the main RTP sender, and hence follows the compound RTCP rules defined in Section 6.1 in [RFC3550].

If the use of non-compound RTCP [RFC5506] was previously negotiated between the sender and the splicer, the RTCP splicing notification message may be sent as non-compound RTCP packets. In some cases that the mapping from RTP timestamp to NTP timestamp changes, e.g., clock drift happening before the splicing event, it may be required to send RTCP SR or even updated Splicing Interval information timely to update the timestamp mapping for accurate splicing.

When the splicer intercepts the RTCP splicing notification message, it SHOULD NOT forward the message to the down-stream receivers in order to reduce RTCP bandwidth consumption. And if the splicer wishes to prevent the downstream receivers from detecting splicing, it MUST NOT forward the message.

4 Reducing Splicing Latency

When splicing starts or ends, the splicer outputs the multimedia content from another sender to the receivers. Given that the receivers must first acquire certain information ([RFC6285] refers to this information as Reference Information) to start processing the multimedia data, either the main RTP sender or the substitutive sender SHOULD provide the Reference Information together with its multimedia content to reduce the delay caused by acquiring the Reference Information. The methods by which the Reference Information is distributed to the receivers is out of scope of this memo.

Another latency element is synchronization caused delay. The receivers must receive enough synchronization metadata prior to synchronizing the separate components of the multimedia streams when splicing starts or ends. Either the main RTP sender or the substitutive sender SHOULD send the synchronization metadata early enough so that the receivers can play out the multimedia in a synchronized fashion. The main RTP sender and the substitutive sender can be coordinated by some proprietary out-of-band mechanisms to decide when and whom to send the metadata. If both send the information, the splicer SHOULD pick one based on the current situation, e.g., choosing media sender when synchronizing the main media content while choosing the information from the substitutive sender when synchronizing the spliced content. The mechanisms defined in [RFC6051] are RECOMMENDED to be adopted to reduce the possible synchronization delay.

5 Failure Cases

This section examines the implications of losing RTCP splicing notification message and the other failure case, e.g., the RTP header extension is stripped on the path.

Given that there may be a splicing un-aware middlebox on the path between the main RTP sender and the splicer, the main and the substitutive RTP senders can use one heuristic to verify whether or not the Splicing Interval reaches the splicer.

The splicer can be implemented to have its own SSRC, and send RTCP reception reports to the senders of the main and substitutive RTP streams. This allows the senders to detect problems on the path to the splicer. Alternatively, it is possible to implement the splicer such that it has no SSRC, and does not send RTCP reports; this prevents the senders from being able to monitor the quality to the path to the splicer.

If the splicer has an SSRC and sends its own RTCP reports, it can choose not to pass RTCP reports it receives from the receivers to the senders. This will stop the senders from being able to monitor the quality of the paths from the splicer to the receivers.

A splicer that has an SSRC can choose to pass RTCP reception reports from the receivers back to the senders, after modifications to account for the splicing. This will allow the senders the monitor the quality of the paths from the splicer to the receivers. A splicer that does not have its own SSRC has to forward and translation RTCP reports from the receiver, otherwise the senders will not see any receivers in the RTP session.

If the splicer is implemented following [RFC6828], it will have its own SSRC and will send its own RTCP reports, and will forward translated RTCP reports from the receivers.

Upon the detection of a failure, the splicer can communicate with the main sender and the substitutive sender in some out of band signaling ways to fall back to the payload specific mechanisms it supports, e.g., MPEG-TS splicing solution defined in [SCTE35], or just abandon the splicing.

6 Session Description Protocol (SDP) Signaling

This document defines the URI for declaring this header extension in an extmap attribute to be "urn:ietf:params:rtp-hdrext:splicing-interval".

This document extends the standard semantics defined in SDP Grouping Framework [RFC5888] with a new semantic: SPLICE to represent the relationship between the main RTP stream and the substitutive RTP stream. Only 2 m-lines are allowed in the SPLICE group. The main RTP stream is the one with the extended extmap attribute, and the other one is substitutive stream. A single m-line MUST NOT be included in different SPLICE groups at the same time. The main RTP sender provides the information about both main and substitutive sources.

The extended SDP attribute specified in this document is applicable for offer/answer content [RFC3264] and do not affect any rules when negotiating offer and answer. When used with multiple m-lines, substitutive RTP MUST be applied only to the RTP packets whose SDP m-line is in the same group with the substitutive stream using SPLICE and has the extended splicing extmap attribute. This semantic is also applicable for BUNDLE cases.

The following examples show how SDP signaling could be used for splicing in different cases.

6.1 Declarative SDP

```
v=0
o=xia 1122334455 1122334466 IN IP4 splicing.example.com
s=RTP Splicing Example
t=0 0
a=group:SPLICE 1 2
m=video 30000 RTP/AVP 100
i=Main RTP Stream
c=IN IP4 233.252.0.1/127
a=rtpmap:100 MP2T/90000
a=extmap:1 urn:ietf:params:rtp-hdrext:splicing-interval
a=mid:1
m=video 30002 RTP/AVP 100
i=Substitutive RTP Stream
c=IN IP4 233.252.0.2/127
a=sendonly
a=rtpmap:100 MP2T/90000
a=mid:2
```

Figure 3: Example SDP for a single-channel splicing scenario

The splicer receiving the SDP message above receives one MPEG2-TS stream (payload 100) from the main RTP sender (with multicast destination address of 233.252.0.1) on port 30000, and/or receives another MPEG2-TS stream from the substitutive RTP sender (with multicast destination address of 233.252.0.2) on port 30002. But at a particular point in time, the splicer only selects one stream and outputs the content from the chosen stream to the downstream receivers.

6.2 Offer/Answer without BUNDLE

SDP Offer - from main RTP sender

```
v=0
o=xia 1122334455 1122334466 IN IP4 splicing.example.com
s=RTP Splicing Example
t=0 0
a=group:SPLICE 1 2
m=video 30000 RTP/AVP 31 100
i=Main RTP Stream
c=IN IP4 splicing.example.com
a=rtpmap:31 H261/90000
a=rtpmap:100 MP2T/90000
a=extmap:1 urn:ietf:params:rtp-hdrext:splicing-interval
a=sendonly
a=mid:1
m=video 40000 RTP/AVP 31 100
i=Substitutive RTP Stream
c=IN IP4 substitutive.example.com
a=rtpmap:31 H261/90000
a=rtpmap:100 MP2T/90000
a=sendonly
a=mid:2
```

SDP Answer - from splicer

```
v=0
o=xia 1122334455 1122334466 IN IP4 splicer.example.com
s=RTP Splicing Example
t=0 0
a=group:SPLICE 1 2
m=video 30000 RTP/AVP 100
i=Main RTP Stream
c=IN IP4 splicer.example.com
a=rtpmap:100 MP2T/90000
a=extmap:1 urn:ietf:params:rtp-hdrext:splicing-interval
a=recvonly
a=mid:1
m=video 40000 RTP/AVP 100
i=Substitutive RTP Stream
c=IN IP4 splicer.example.com
a=rtpmap:100 MP2T/90000
a=recvonly
a=mid:2
```

6.3 Offer/Answer with BUNDLE: All Media are spliced

In this example, the bundled audio and video media have their own substitutive media for splicing:

1. An Offer, in which the offerer assigns a unique address and a substitutive media to each bundled "m="line for splicing within the BUNDLE group.

2. An answer, in which the answerer selects its own BUNDLE address, and leave the substitutive media untouched.

SDP Offer - from main RTP sender

```
v=0
o=alice 1122334455 1122334466 IN IP4 splicing.example.com
s=RTP Splicing Example
c=IN IP4 splicing.example.com
t=0 0
a=group:SPLICE foo 1
a=group:SPLICE bar 2
a=group:BUNDLE foo bar
m=audio 10000 RTP/AVP 0 8 97
a=mid:foo
b=AS:200
a=rtpmap:0 PCMU/8000
a=rtpmap:8 PCMA/8000
a=rtpmap:97 iLBC/8000
a=extmap:1 urn:ietf:params:rtp-hdrext:splicing-interval
a=sendonly
m=video 10002 RTP/AVP 31 32
a=mid:bar
b=AS:1000
a=rtpmap:31 H261/90000
a=rtpmap:32 MPV/90000
a=extmap:2 urn:ietf:params:rtp-hdrext:splicing-interval
a=sendonly
m=audio 20000 RTP/AVP 0 8 97
i=Substitutive audio RTP Stream
c=IN IP4 substitive.example.com
a=rtpmap:0 PCMU/8000
a=rtpmap:8 PCMA/8000
a=rtpmap:97 iLBC/8000
a=sendonly
a=mid:1
m=video 20002 RTP/AVP 31 32
i=Substitutive video RTP Stream
```

```
      c=IN IP4 substitive.example.com
      a=rtpmap:31 H261/90000
      a=rtpmap:32 MPV/90000
      a=mid:2
      a=sendonly
```

SDP Answer - from the splicer

```
      v=0
      o=bob 2808844564 2808844564 IN IP4 splicer.example.com
      s=RTP Splicing Example
      c=IN IP4 splicer.example.com
      t=0 0
      a=group:SPLICE foo 1
      a=group:SPLICE bar 2
      a=group:BUNDLE foo bar
      m=audio 30000 RTP/AVP 0
      a=mid:foo
      b=AS:200
      a=rtpmap:0 PCMU/8000
      a=extmap:1 urn:ietf:params:rtp-hdrext:splicing-interval
      a=recvonly
      m=video 30000 RTP/AVP 32
      a=mid:bar
      b=AS:1000
      a=rtpmap:32 MPV/90000
      a=extmap:2 urn:ietf:params:rtp-hdrext:splicing-interval
      a=recvonly
      m=audio 30002 RTP/AVP 0
      i=Substitutive audio RTP Stream
      c=IN IP4 splicer.example.com
      a=rtpmap:0 PCMU/8000
      a=recvonly
      a=mid:1
      m=video 30004 RTP/AVP 32
      i=Substitutive video RTP Stream
      c=IN IP4 splicer.example.com
      a=rtpmap:32 MPV/90000
      a=mid:2
      a=recvonly
```

6.4 Offer/Answer with BUNDLE: a Subset of Media are Spliced

In this example, the substitutive media only applies for video when splicing:

1. An Offer, in which the offerer assigns a unique address to each bundled "m="line within the BUNDLE group, and assigns a substitutive

media to the bundled video "m=" line for splicing.

2. An answer, in which the answerer selects its own BUNDLE address, and leave the substitutive media untouched.

SDP Offer - from the main RTP sender:

```
v=0
o=alice 1122334455 1122334466 IN IP4 splicing.example.com
s=RTP Splicing Example
c=IN IP4 splicing.example.com
t=0 0
a=group:SPLICE bar 2
a=group:BUNDLE foo bar
m=audio 10000 RTP/AVP 0 8 97
a=mid:foo
b=AS:200
a=rtpmap:0 PCMU/8000
a=rtpmap:8 PCMA/8000
a=rtpmap:97 iLBC/8000
a=sendonly
m=video 10002 RTP/AVP 31 32
a=mid:bar
b=AS:1000
a=rtpmap:31 H261/90000
a=rtpmap:32 MPV/90000
a=extmap:2 urn:ietf:params:rtp-hdrext:splicing-interval
a=sendonly
m=video 20000 RTP/AVP 31 32
i=Substitutive video RTP Stream
c=IN IP4 substitutive.example.com
a=rtpmap:31 H261/90000
a=rtpmap:32 MPV/90000
a=mid:2
a=sendonly
```

SDP Answer - from the splicer:

```
v=0
o=bob 2808844564 2808844564 IN IP4 splicer.example.com
s=RTP Splicing Example
c=IN IP4 splicer.example.com
t=0 0
a=group:SPLICE bar 2
a=group:BUNDLE foo bar
m=audio 30000 RTP/AVP 0
a=mid:foo
b=AS:200
```

```
a=rtpmap:0 PCMU/8000
a=recvonly
m=video 30000 RTP/AVP 32
a=mid:bar
b=AS:1000
a=rtpmap:32 MPV/90000
a=extmap:2 urn:ietf:params:rtp-hdrext:splicing-interval
a=recvonly
m=video 30004 RTP/AVP 32
i=Substitutive video RTP Stream
c=IN IP4 splicer.example.com
a=rtpmap:32 MPV/90000
a=mid:2
a=recvonly
```

7 Security Considerations

The security considerations of the RTP specification [RFC3550] and the general mechanism for RTP header extensions [RFC5285] apply. The splicer can either be a mixer or a translator, and has all the security considerations on these two standard RTP intermediaries. However, the splicer replaces some content with other content in RTP packet, thus breaking any RTP-level end-to-end security, such as source authentication and integrity protection.

End to end source authentication is not possible with any known existing splicing solution. A new solution can theoretically be developed that enables identifying the participating entities and what each provides, i.e., the different media sources, main and substituting, and the splicer providing the RTP-level integration of the media payloads in a common timeline and synchronization context.

Since splicer works as a trusted entity, any RTP-level or outside security mechanism, such IPsec[RFC4301] or Datagram Transport Layer Security [RFC6347], will use a security association between the splicer and the receiver. When using the Secure Real-Time Transport Protocol (SRTP) [RFC3711], the splicer could be provisioned with the same security association as the main RTP sender.

If there is a concern about the confidentiality of the splicing time information, header extension encryption [RFC6904] SHOULD be used. However, the malicious endpoint may get the splicing time information by other means, e.g., inferring from the communication between the main and substitutive content sources. To avoid the insertion of invalid substitutive content, the splicer MUST have some mechanisms to authenticate the substitutive stream source.

For cases that the splicing time information is changed by a malicious endpoint, the splicing may fail since it will not be available at the right time for the substitutive media to arrive, which may also break an undetectable splicing. To mitigate this effect, the splicer SHOULD NOT forward the splicing time information RTP header extension defined in Section 4.1 to the receivers. And it MUST NOT forward this header extension when considering an undetectable splicing.

8 IANA Considerations

8.1 RTCP Control Packet Types

Based on the guidelines suggested in [RFC5226], a new RTCP packet format has been registered with the RTCP Control Packet Type (PT) Registry:

Name: SNM

Long name: Splicing Notification Message

Value: TBA

Reference: This document

8.2 RTP Compact Header Extensions

The IANA has also registered a new RTP Compact Header Extension [RFC5285], according to the following:

Extension URI: urn:ietf:params:rtp-hdrext:splicing-interval

Description: Splicing Interval

Contact: Jinwei Xia <xiajinwei@huawei.com>

Reference: This document

8.3 SDP Grouping Semantic Extension

This document request IANA to register the new SDP grouping semantic extension called "SPLICE".

Semantics: Splice

Token:SPLICE

Reference: This document

Contact: Jinwei Xia <xiajinwei@huawei.com>

9 Acknowledgement

The authors would like to thank the following individuals who help to review this document and provide very valuable comments: Colin Perkins, Bo Burman, Stephen Botzko, Ben Campbell.

10 References

10.1 Normative References

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

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC3264] Rosenberg, J., and H. Schulzrinne, "An Offer/Answer Model with the Session Description Protocol (SDP)", RFC 3264, June 2002.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC5285] Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, July 2008.

[RFC5888] Camarillo, G. and H. Schulzrinne, "The Session Description Protocol (SDP) Grouping Framework", RFC 5888, June 2010.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

[RFC6051] Perkins, C. and T. Schierl, "Rapid Synchronisation of RTP Flows", RFC 6051, November 2010.

[RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", RFC 6347, January 2012.

10.2 Informative References

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5506] Johansson, I. and M. Westerlund, "Support for Reduced-Size Real-Time Transport Control Protocol (RTCP): Opportunities and Consequences", RFC 5506, April 2009.

[RFC6285] Ver Steeg, B., Begen, A., Van Caenegem, T., and Z. Vax, "Unicast-Based Rapid Acquisition of Multicast RTP Sessions", RFC 6285, June 2011.

[RFC6904] Lennox, J.,"Encryption of Header Extensions in the Secure Real-Time Transport Protocol (SRTP)", April 2013.

[SCTE35] Society of Cable Telecommunications Engineers (SCTE), "Digital Program Insertion Cueing Message for Cable", 2011.

[RFC6828] Xia, J., "Content Splicing for RTP Sessions", RFC 6828, January 2013.

Authors' Addresses

Jinwei Xia
Huawei

Email: xiajinwei@huawei.com

Roni Even
Huawei

Email: ron.even.tlv@gmail.com

Rachel Huang
Huawei

Email: rachel.huang@huawei.com

Lingli Deng
China Mobile

Email: denglingli@chinamobile.com

Network Working Group
Internet-Draft
Updates: 5104 (if approved)
Intended status: Standards Track
Expires: June 11, 2016

S. Wenger
J. Lennox
Vidyo, Inc.
B. Burman
M. Westerlund
Ericsson

December 09, 2015

Using Codec Control Messages in the RTP Audio-Visual Profile with Feedback with Layered Codecs

draft-wenger-avtext-avpf-ccm-layered-00

Abstract

This document fixes a shortcoming in the specification language of the Codec Control Message Full Intra Request (FIR) as defined in RFC5104 when using with layered codecs. In particular, a Decoder Refresh Point needs to be sent by a media sender when a FIR is received on any layer of the layered bitstream, regardless on whether those layers are being sent in a single or in multiple RTP flows. The other payload-specific feedback messages defined in RFC 5104 and RFC 4585 as updated by RFC 5506 have also been analyzed, and no corresponding shortcomings have been found.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction and Problem Statement

RFC 4585 [RFC4585] and RFC 5104 [RFC5104] specify a number of payload-specific feedback messages which a media receiver can use to inform a media sender of certain conditions, or make certain requests. The feedback messages are being sent as RTCP receiver reports, and RFC 4585 specifies timing rules that make the use of those messages practical for time-sensitive codec control.

Since the time those RFCs were developed, layered codecs have gained in popularity and deployment. Layered codecs use multiple sub-bitstreams called layers to represent the content in different fidelities. Depending on the media codec and its RTP payload format in use, single layers or groups of layers may be sent in their own

RTP streams (in MRST or MRMT mode as defined in RFC 7656 [RFC7656]), or multiplexed (using media-codec specific multiplexing mechanisms) in a single RTP stream (SRST mode as defined in RFC 7656 [RFC7656]). The dependency relationship between layers forms a directed graph, with the base layer at the root. Enhancement layers depend on the base layer and potentially on other enhancement layers, and the target layer and all layers it depends on have to be decoded jointly in order to re-create the uncompressed media signal at the fidelity of the target layer.

Implementation experience has shown that the Full Intra Request command as defined in RFC 5104 [RFC5104] is underspecified when used with layered codecs and when more than one RTP stream is used to transport the layers of a layered bitstream at a given fidelity. In particular, from the RFC 5104 [RFC5104] specification language it is not clear whether an FIR received for only a single RTP stream of multiple RTP streams covering the same layered bitstream necessarily triggers the sending of a Decoder Refresh Point (as defined in RFC 5104 [RFC5104] section 2.2) for all layers, or only for the layer which is transported in the RTP stream which the FIR request is associated with.

This document fixes this shortcoming by:

a. Updating the definition of the Decoder Refresh Point (as defined in RFC 5104 [RFC5104] section 2.2) to cover layered codecs, in line with the corresponding definitions used in a popular layered codec format, namely H.264/SVC [H.264]. Specifically, a decoder refresh point, in conjunction with layered codecs, resets the state of the whole decoder, which implies that it includes hard or gradual single-layer decoder refresh for all layers;

b. Requiring that, when a media sender receives a Full Intra Request over the RTCP stream associated with any of the RTP streams over which a part of the layered bitstream is transported, to send a Decoder Refresh Point;

c. Require that a media receiver sends the FIR on the RTCP stream associated with the base layer (the option of receiving FIR on enhancement layer-associated RTCP stream as specified in point b) above is kept for backward compatibility); and

d. Providing guidance on how to detect that a layered codec is in use for which the above rules apply.

While, clearly, the reaction to FIR for layered codecs in RFC 5104 [RFC5104] and companion documents is underspecified, it appears that this is not the case for any of the other payload-specific codec control messages defined in any of RFC 4585 [RFC4585], RFC 5104 [RFC5104], or RFC 5506 [RFC5506]. A brief summary of the analysis that led to this conclusion is also included in this document.

2. 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].

3. Updated definition of Decoder Refresh Point

The text below updates the definition of Decoder Refresh Point in section 2.2 of RFC 5104 [RFC5104].

Decoder Refresh Point: A bit string, packetized in one or more RTP packets, that completely resets the decoder to a known state.

Examples for "hard" single layer decoder refresh points are Intra pictures in H.261 [H.261], H.263 [H.263], MPEG-1 [MPEG-1], MPEG-2 [MPEG-2], and MPEG-4 [MPEG-4]; Instantaneous Decoder Refresh (IDR) pictures in H.264 [H.264], and H.265 [H.265]; and Keyframes in VP8 [RFC6386] and VP9 [I-D.grange-vp9-bitstream]. "Gradual" decoder refresh points may also be used; see for example H.264 [H.264]. While both "hard" and "gradual" decoder refresh points are acceptable in the scope of this specification, in most cases the user experience will benefit from using a "hard" decoder refresh point.

A decoder refresh point also contains all header information above the syntactical level of the picture layer (or equivalent, depending on the video compression standard) that is conveyed in-band. In H.264 [H.264], for example, a decoder refresh point contains parameter set Network Adaptation Layer (NAL) units that generate parameter sets necessary for the decoding of the following slice/data partition NAL units (and that are not conveyed out of band).

When a layered codec is in use, the above definition (and, in particular, the requirement to COMPLETELY reset the decoder to a known state) implies that the decoder refresh point includes hard or gradual single layer decoder refresh points for all layers.

4. Full Intra Request for Layered Codecs

When a media receiver or middlebox has decided to send a FIR command (based on the guidance provided in Section 4.3.1 of RFC 5104 [RFC5104], it MUST do so in the RTCP stream related to the forward RTP stream that carries the base layer of the layered bitstream, and the Feedback Control Information (FCI, and in particular the SSRC field therein) MUST also refer to the forward RTP stream that carries the base layer.

When a Full Intra Request Command is received by the designated media sender in the RTCP stream associated with any of the RTP streams in which any layer of a layered bitstream are sent, the designated media sender MUST send a Decoder Refresh Point (Section 3) as defined above at its earliest opportunity. The requirements related to congestion control on the forward RTP streams as specified in sections 3.5.1.5 of RFC 5104 [RFC5104] apply for the RTP streams both in isolation and combined.

Note: the requirement to react to FIR commands associated with enhancement layers is included for robustness and backward compatibility reasons.

5. Identifying the use of Layered Codecs (Informative)

The above modifications to RFC 5104 unambiguously define how to deal with FIR when layered bitstreams are in use. However, it is surprisingly difficult to identify this situation. In general, it is expected that implementers know when layered coding (in its commonly understood sense: with inter-layer prediction between pyramided-arranged layers) is in use and when not, and can therefore implement the above updates to RFC 5104 correctly. However, there are use cases of the use of layered codecs that may be viewed as somewhat exotic today but clearly are supported by the video coding syntax, in which the above rules would lead to suboptimal system behavior. Nothing would break, and there would not be an interop failure, but the user experience may suffer through the sending or receiving of Decoder Refresh Points at times or on parts of the bitstream that are unnecessary from a user experience viewpoint. Therefore, this informative section is included that provides the current understanding of when a layered codec is in use and when not.

The key observation made here is that the RTP payload format negotiated for the RTP streams, in isolation, is not necessarily an indicator for the use of layering. Some layered codecs (including H.264/SVC) can form decodable bitstreams including only (one or more) enhancement layers, without the base layer, effectively creating simulcastable sub-bitstreams in a scalable bitstream that does not take advantage of inter-layer prediction. In such a scenario, it is potentially (though not necessarily) unnecessary--or even counter-productive--to send a decoder refresh point on all RTP streams using that payload format and SSRC.

One good indication of the likely use of layering with interlayer prediction is when the various RTP streams are "bound" together on the signaling level. In an SDP environment, this would be the case if they are marked as being dependent from each other using the grouping framework RFC 4588 [RFC4588] and the layer dependency RFC 5583 [RFC5583]. Conversely, one good indication of the use of simulcast is when simulcasting is explicitly being signaled, for example through the use of [I-D.ietf-mmusic-sdp-simulcast], except when simulcast stream identifiers are explicitly defined as dependent according to [I-D.ietf-mmusic-rid].

6. Layered Codecs and non-FIR codec control messages (Informative)

Between them, RFC 4585 [RFC4585] (as updated by RFC 5506 [RFC5506]) and RFC 5104 [RFC5104] define a total of seven Payload-specific Feedback messages. For the FIR command message, guidance has been provided above. In this section, some information is provided with respect to the remaining six codec control messages.

6.1. Picture Loss Indication (PLI)

PLI is defined in RFC 4585 [RFC4585] section 6.3.1. The prudent response to a PLI message received for an enhancement layer is to "repair" (through whatever source-coding specific means) that enhancement layer and all dependent enhancement layers, but not the reference layer(s) used by the enhancement layer for which the PLI was received. The encoder can figure out by itself what constitutes a dependent enhancement layer and does not need help from the system stack in doing so. Insofar, there is nothing that needs to be specified herein.

6.2. Slice Loss Indication (SLI)

SLI is defined in RFC 4585 [RFC4585] section 6.3.2. The authors' current understanding is that the prudent response to a SLI message received for an enhancement layer is to "repair" (through whatever source-coding specific means) the affected spatial area of that enhancement layer and all dependent enhancement layers, but not the reference layers used by the enhancement layer for which the SLI was received. The encoder can figure out by itself what constitutes a dependent enhancement layer and does not need help from the system stack in doing so. Insofar, there is nothing that needs to be specified herein. SLI has seen very little implementation and, as far as it is known, none in conjunction with layered systems.

6.3. Reference Picture Selection Indication (RPSI)

RPSI is defined in RFC 4585 [RFC4585] section 6.3.3. While a technical equivalent of RPSI has been in use with non-layered systems for many years, no implementations are known in conjunction of layered codecs. The authors’ current understanding is that the reception of an RPSI message on any layer forces the encoder to "repair" the bitstream on that layer and all dependent layers without the need of any system-provided guidance. Insofar, RPSI should work without further need for specification language.

6.4. Temporal-Spatial Trade-off Request and Notification (TSTR/TSTN)

TSTN/TSTR are defined in RFC 5104 [RFC5104] section 4.3.2 and 4.3.3, respectively. The TSTR request allows to communicate (typically user-interface-obtained) guidance of the preferred trade-off between spatial quality and frame rate. A technical equivalent of TSTN/TSTR has seen deployment for many years in non-scalable systems.

The Temporal-Spatial Trade-off request and notification messages include an SSRC target, which (similarly to FIR) may refer to an RTP stream carrying a base layer, an enhancement layer, or multiple layers. Therefore, the authors’ current understanding is that the semantics of the message applies to the layers present in the targeted RTP stream.

It is noted that per-layer TSTR/TSTN is a mechanism that is, in some ways, counterproductive in a system using layered codecs. Given a sufficiently complex layered bitstream layout, a sending system has flexibility in adjusting the spatio/temporal quality balance by adding and removing temporal, spatial, or quality enhancement layers. At present it is unclear whether an allowed (or even recommended) option to the reception of a TSTR is to adjust the bit allocation within the layer(s) present in the addressed RTP stream, or to adjust the layering structure accordingly--which can involve more than just the addressed RTP stream.

Until there is a sufficient critical mass of implementation practice, it is probably prudent for an implementer not to assume either of the two options (or any middleground that may exist between the two), be liberal in accepting TSTR messages, perhaps responding in TSTN indicating "no change," not sending TSTR messages except when operating in SRST mode as defined in RFC 7656 [RFC7656], and contribute to the IETF documentation of any implementation requirements that make per-layer TSTR/TSTN useful.

6.5. H.271 Video Back Channel Message (VBCM)

VBCM is defined in RFC 5104 [RFC5104] section 4.3.4. What was said above for RPSI (Section 6.3) applies here as well.

7. Acknowledgements

The authors want to thank Mo Zanaty for useful discussions.

8. IANA Considerations

This memo includes no request to IANA.

9. Security Considerations

The security considerations of RFC 4585 [RFC4585] (as updated by RFC 5506 [RFC5506]) and RFC 5104 [RFC5104] apply. The clarified response to FIR does not require any updates.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4585] Ott, J., Wenger, S., Sato, N., Burmeister, C., and J. Rey, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 4585, DOI 10.17487/RFC4585, July 2006, <http://www.rfc-editor.org/info/rfc4585>.

[RFC5104] Wenger, S., Chandra, U., Westerlund, M., and B. Burman, "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)", RFC 5104, DOI 10.17487/RFC5104, February 2008, <http://www.rfc-editor.org/info/rfc5104>.

[RFC5506] Johansson, I. and M. Westerlund, "Support for Reduced-Size Real-Time Transport Control Protocol (RTCP): Opportunities and Consequences", RFC 5506, DOI 10.17487/RFC5506, April 2009, <http://www.rfc-editor.org/info/rfc5506>.

10.2. Informative References

[H.261] ITU-T, "ITU-T Rec. H.261: Video codec for audiovisual services at p x 64 kbit/s", 1993, <http://handle.itu.int/11.1002/1000/1088>.

[H.263] ITU-T, "ITU-T Rec. H.263: Video coding for low bit rate communication", 2005, <http://handle.itu.int/11.1002/1000/7497>.

[H.264] ITU-T, "ITU-T Rec. H.264: Advanced video coding for generic audiovisual services", 2014, <http://handle.itu.int/11.1002/1000/12063>.

[H.265] ITU-T, "ITU-T Rec. H.265: High efficiency video coding", 2015, <http://handle.itu.int/11.1002/1000/12455>.

[I-D.grange-vp9-bitstream]
Grange, A. and H. Alvestrand, "A VP9 Bitstream Overview", draft-grange-vp9-bitstream-00 (work in progress), February 2013.

[I-D.ietf-mmusic-rid]
Thatcher, P., Zanaty, M., Nandakumar, S., Burman, B., Roach, A., and B. Campen, "RTP Payload Format Constraints", draft-ietf-mmusic-rid-00 (work in progress), November 2015.

[I-D.ietf-mmusic-sdp-simulcast]
Burman, B., Westerlund, M., Nandakumar, S., and M. Zanaty, "Using Simulcast in SDP and RTP Sessions", draft-ietf-mmusic-sdp-simulcast-03 (work in progress), October 2015.

[MPEG-1] ISO/IEC, "ISO/IEC 11172-2:1993 Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 2: Video", 1993.

[MPEG-2] ISO/IEC, "ISO/IEC 13818-2:2013 Information technology -- Generic coding of moving pictures and associated audio information -- Part 2: Video", 2013.

[MPEG-4] ISO/IEC, "ISO/IEC 14496-2:2004 Information technology -- Coding of audio-visual objects -- Part 2: Visual", 2004.

[RFC4588] Rey, J., Leon, D., Miyazaki, A., Varsa, V., and R. Hakenberg, "RTP Retransmission Payload Format", RFC 4588, DOI 10.17487/RFC4588, July 2006, <http://www.rfc-editor.org/info/rfc4588>.

[RFC5583] Schierl, T. and S. Wenger, "Signaling Media Decoding Dependency in the Session Description Protocol (SDP)", RFC 5583, DOI 10.17487/RFC5583, July 2009, <http://www.rfc-editor.org/info/rfc5583>.

[RFC6386] Bankoski, J., Koleszar, J., Quillio, L., Salonen, J., Wilkins, P., and Y. Xu, "VP8 Data Format and Decoding Guide", RFC 6386, DOI 10.17487/RFC6386, November 2011, <http://www.rfc-editor.org/info/rfc6386>.

[RFC7656] Lennox, J., Gross, K., Nandakumar, S., Salgueiro, G., and B. Burman, Ed., "A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources", RFC 7656, DOI 10.17487/RFC7656, November 2015, <http://www.rfc-editor.org/info/rfc7656>.

Appendix A. Change Log

NOTE TO RFC EDITOR: Please remove this section prior to publication.

-00: initial version

Authors' Addresses

Stephan Wenger
Vidyo, Inc.

Email: stewe@stewe.org

Jonathan Lennox
Vidyo, Inc.

Email: jonathan@vidyo.com

Bo Burman
Ericsson
Kistavagen 25
SE - 164 80 Kista
Sweden

Email: bo.burman@ericsson.com

Magnus Westerlund
Ericsson
Farogatan 6
SE- 164 80 Kista
Sweden

Phone: +46107148287
Email: magnus.westerlund@ericsson.com