                 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

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

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

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

   This Internet-Draft will expire on September 22, 2016.

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 |
     +---+      +-----------+      +---+
```
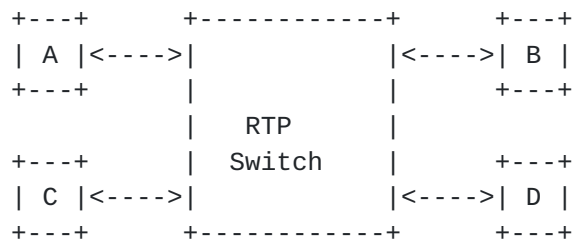

                       Figure 1: RTP switch

   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.

   o S: Start of Frame (1 bit) - MUST be 1 in the first packet in a
      frame within a layer; otherwise MUST be 0.
   o E: End of Frame (1 bit) - MUST be 1 in the last packet in a frame
      within a layer; otherwise MUST be 0.
   o 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.
   o 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.
   o B: Base Layer Sync (1 bit) - MUST be 1 if this frame only depends
      on the base layer; otherwise MUST be 0.
   o TID: Temporal ID (3 bits) - The base temporal layer starts with 0,
      and increases with 1 for each higher temporal layer/sub-layer.
   o 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