Network Working Group
Internet-Draft
Updates: 5104 (if approved)
Intended status: Standards Track
Expires: May 21, 2017
S. Wenger
J. Lennox
Vidyo, Inc.
B. Burman
M. Westerlund
Ericsson
November 17, 2016

Using Codec Control Messages in the RTP Audio-Visual Profile with Feedback with Layered Codecs
draft-ietf-avtext-avpf-ccm-layered-03

Abstract

This document updates RFC5104 by fixing a shortcoming in the specification language of the Codec Control Message Full Intra Request (FIR) as defined in RFC5104 when using it 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.

Table of Contents

1. Introduction and Problem Statement

The Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF) [RFC4585] and Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF) [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, a number of options exist how to transport those layers in RTP. With reference to A Taxonomy of Semantics and Mechanisms for Real-Time Transport Protocol (RTP) Sources [RFC7656]):

single layers or groups of layers may be sent in their own RTP streams in MRST or MRMT mode;

using media-codec specific multiplexing mechanisms, multiple layers may be sent in a single RTP stream in SRST mode.

The dependency relationship between layers in a truly layered, pyramid-shaped bitstream 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. Such a layering structure is assumed henceforth; for more exotic layering structures please see Section 5.

Implementation experience has shown that the Full Intra Request command as defined in [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 [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 [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 [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. Require a media sender to send a Decoder Refresh Point after the media sender has received a Full Intra Request over an RTCP stream associated with any of the RTP streams over which a part of the layered bitstream is transported;

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 bitstream is in use for which the above rules apply.

While, clearly, the reaction to FIR for layered codecs in [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 [RFC4585], [RFC5104]. 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 remainder of this section replaces the definition of Decoder Refresh Point in section 2.2 of [RFC5104] in its entirety.

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 that is conveyed in-band. In [H.264], for example, a decoder refresh point contains those parameter set Network Adaptation Layer (NAL) units that generate parameter sets necessary for the decoding of the following slice/data partition NAL units. (That is assuming the parameter sets have not been conveyed out of band.)

When a layered codec is in use, the above definition--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

A media receiver or middlebox may decide to send a FIR command based on the guidance provided in Section 4.3.1 of [RFC5104]. When sending the FIR command, it MUST target the RTP stream that carries the base layer of the layered bitstream, and this is done by setting the Feedback Control Information (FCI, and in particular the SSRC field therein) to refer to the SSRC of 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. and 5. of [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 bitstreams (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 the use of a layered bitstream. In general, it is expected that implementers know when layered bitstreams (in its commonly understood sense: with inter-layer prediction between pyramided-arranged layers) are in use and when not, and can therefore implement the above updates to RFC 5104 correctly. However, there are scenarios in which layered codecs are employed creating non-pyramid shaped bitstreams. Those scenarios may be viewed as somewhat exotic today but clearly are supported by certain video coding syntaxes, such as H.264/SVC. When blindly applying the above rules to those non-pyramid-arranged layering structures, suboptimal system behavior would result. Nothing would break, and there would not be an interoperability 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 bitstream 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 a layered bitstream. 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 within a single scalable bitstream (as defined in the video coding standard), but without inter-layer prediction. In such a scenario, it is potentially, though not necessarily, counter-productive to send a decoder refresh point on all RTP streams using that payload format and SSRC. It is beyond the scope of this document to discuss optimized reactions to FIRs received on RTP streams carrying such exotic bitstreams.

One good indication of the likely use of pyramid-shaped 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 Session Description Protocol (SDP) Grouping Framework [RFC5888] and the layer dependency RFC 5583 [RFC5583].

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

Between them, AVPF [RFC4585] and Codec Control Messages [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 section 6.3.1 of [RFC4585]. The prudent response to a PLI message received for an enhancement layer is to "repair" that enhancement layer and all dependent enhancement layers through appropriate source-coding specific means. However, the reference layer(s) used by the enhancement layer for which the PLI was received does not require repair. 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. Thus, there is nothing that needs to be specified herein.

6.2. Slice Loss Indication (SLI)

SLI is defined in section 6.3.2 of [RFC4585]. The authors' current understanding is that the prudent response to a SLI message received for an enhancement layer is to "repair" the affected spatial area of that enhancement layer and all dependent enhancement layers through appropriate source-coding specific means. As in PLI, the reference layers used by the enhancement layer for which the SLI was received

do not need to be repaired. Again, as in PLI, the encoder can determine by itself what constitutes a dependent enhancement layer and does not need help from the system stack in doing so. Thus, 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 section 6.3.3 of [RFC4585]. 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 indicating a missing reference picture forces the encoder to appropriately handle that missing reference picture in the layer indicated, and all dependent layers. Thus, 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 section 4.3.2 and 4.3.3 of [RFC5104], respectively. The TSTR request communicates 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. Instead, it is suggested that an implementation be liberal in

accepting TSTR messages, and upon receipt responding in TSTN indicating "no change". Further, it is suggested that new implementations do not send TSTR messages except when operating in SRST mode as defined in [RFC7656]. Finally implementers are encouraged to 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 section 4.3.4 of [RFC5104]. 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 AVPF [RFC4585] (as updated by Support for Reduced-Size Real-Time Transport Control Protocol (RTCP): Opportunities and Consequences [RFC5506]) and Codec Control Messages [RFC5104] apply. The clarified response to FIR does not introduce additional security considerations.

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.

[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.

[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>.

[RFC5888] Camarillo, G. and H. Schulzrinne, "The Session Description Protocol (SDP) Grouping Framework", RFC 5888, DOI 10.17487/RFC5888, June 2010, <http://www.rfc-editor.org/info/rfc5888>.

[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.

draft-wenger-avtext-avpf-ccm-layered-00-00: initial version

draft-ietf-avtext-avpf-ccm-layered-00: resubmit as avtext WG draft per IETF95 and list confirmation by Rachel 4/25/2016

draft-ietf-avtext-avpf-ccm-layered-00: In section "Identifying the use of Layered Codecs (Informative)", removed last sentence that could be misread that the explicit signaling of simulcasting in conjunction with payload formats supporting layered coding implies no layering.

draft-ietf-avtext-avpf-ccm-layered-01: clarifications in section 5.

draft-ietf-avtext-avpf-ccm-layered-02: addressing WGLC comments, mostly editorial; see reflector discussions 09/2016

draft-ietf-avtext-avpf-ccm-layered-03: addressing AD writeup comments, editorial

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 2
SE- 164 80 Kista
Sweden

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