AVTCORE Working Group
Internet-Draft
Intended status: Standards Track
Expires: 16 August 2025

Y. He
Qualcomm
C. Herglotz
FAU
E. Francois
InterDigital

12 February 2025

RTP Control Protocol (RTCP) Messages for Temporal-Spatial Resolution draft-ietf-avtcore-rtcp-green-metadata-05

Abstract

This specification describes an RTCP feedback message format for the ISO/IEC International Standard 23001-11, known as Energy Efficient Media Consumption (Green metadata), developed by the ISO/IEC JTC 1/SC 29/ WG 3 MPEG System. The RTCP payload format specified in this specification enables receivers to provide feedback to the senders and thus allows for short-term adaptation and feedback-based energy efficient mechanisms to be implemented. The payload format has broad applicability in real-time video communication services.

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 https://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 16 August 2025.

Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document.

Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction

ISO/IEC 23001-11 specification, Energy Efficient Media Consumption (green metadata) [GreenMetadata], specifies metadata that facilitates reduction of energy usage during media consumption. Two main types of metadata are defined in the specification. The first type consists of metadata generated by a video encoder which provides information about the decoding complexity of the delivered bitstream and about the quality of the decoded content. This first type of metadata is conveyed via the supplemental enhancement information (SEI) message mechanism specified in the video coding standard ITU-T Recommendation H.264 and ISO/IEC 14496-10 [AVC], H.265 and ISO/IEC 23008-5 [HEVC], H.266 and ISO/IEC 23090-3 [VVC].

The second type consists of metadata generated by a decoder as feedback conveyed to the encoder to adapt the decoder energy consumption. This specification focuses on this second type of metadata which is conveyed as extension of RTCP feedback messages [RFC4585]. The feedback in the second type of metadata specified in ISO/IEC 23001-11 [GreenMetadata] includes decoder operations reduction request, coding tools configuration request and temporal and spatial scaling request. This specification defines new RTCP payload format for the temporal and spatial resolution request and notification feedback message.

2. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Abbreviations

AVPF: The extended RTP profile for RTCP-based feedback

FCI: Feedback Control Information [RFC4585]

FMT: Feedback Message Type [RFC4585]

PSFB: Payload-specific FB message [RFC4585]

TSRR: Temporal-Spatial Resolution Request

TSRN: Temporal-Spatial Resolution Notification

CCM: Codec Control Messages [RFC5104]

4. Format of RTCP Feedback Messages

This document extends the RTCP feedback messages defined in the RTP/AVPF [RFC4585] and [RFC5104] by defining a temporal-spatial resolution feedback message. The message can be used by the receiver to inform the sender of the desirable coding temporal resolution (frame rate) and spatial resolution of the bitstream delivered, and by the sender to indicate the coding temporal and spatial resolution it will use henceforth.

RTCP temporal-spatial resolution feedback message follows a similar message format as RTCP Temporal-Spatial Trade-off Request and Notification [RFC5104]. The message may be sent in a regular full compound RTCP packet or in an early RTCP packet, as per the RTP/AVPF rules.

This specification specifies two additional payload-specific feedback messages: Temporal-Spatial Resolution Request (TSRR) and Temporal-Spatial Resolution Notification (TSRN)

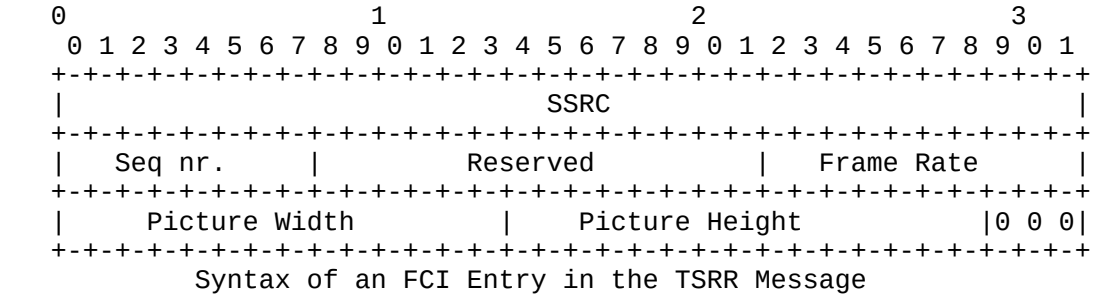
4.1. Temporal-Spatial Resolution Request

The TSRR feedback message is identified by RTCP packet type value PT=PSFB and FMT=12.

The FCI field MUST contain one or more TSRR FCI entries.

4.1.1. Message format

The content of the FCI entry for the Temporal-Spatial Resolution Request is depicted in Figure 1.



Syntax of an FCI Entry in the TSRR Message

Figure 1

SSRC (32 bits): The Synchronization Source (SSRC) of the media sender that is requested to apply the frame rate and picture resolution.

Seq nr. (8 bits): Request sequence number. The sequence number space is unique for pairing of the SSRC of request source and the SSRC of the request 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.

Reserved (14 bits): All bits SHALL be set to 0 by the sender and SHALL be ignored on reception.

Frame Rate (10 bits): frames_per_second. This field specifies the frame rate as defined in clause 6.3 of [GreenMetadata]. An integer value between 1 and 1023 that indicates the coding frame rate that is requested. The value of Frame Rate equal to 0 is invalid.

Picture Width (14 bits): pic_width_in_luma_samples. This field specifies the picture width as defined in clause 6.3 of [GreenMetadata]. An integer value between 1 and 16383 that indicates the coding picture width in the units of luma samples that is requested. The value of Picture Width equal to 0 is invalid.

Picture Height (14 bits): pic_height_in_luma_samples. This specifies the picture height as defined in clause 6.3 of [GreenMetadata]. An integer value between 1 and 16383 that indicates the coding picture height in the units of luma samples that is requested. The value of Picture Height equal to 0 is invalid.

4.1.2. Semantics

A decoder can suggest a temporal-spatial resolution by sending a TSRR message to an encoder. If the encoder is capable of adjusting its temporal-spatial resolution, it SHOULD take into account the received TSRR message for future coding of pictures. The temporal and spatial resolutuions in a TSRR message SHALL be less than or equal to the temporal and spatial resolutions negotiated via SDP.

The reaction to the reception of more than one TSRR message by a media sender from different media receivers is left open to the implementation. The selected Frame Rate, Picture Width and Picture Height SHALL be communicated to the media receivers by means of the TSRN message (see Section 4.2).

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 TSRR applies are in the corresponding FCI entries.

A TSRR message MAY contain requests to multiple media senders, using one FCI entry per target media sender.

4.1.3. Timing Rules

The timing follows the rules outlined in section 3 of [RFC4585]. This request message is not time critical and SHOULD be sent using regular RTCP timing. Only if it is known that the user interface requires quick feedback, the message MAY be sent with early or immediate feedback timing.

4.1.4. Handling of Message in Mixers and Translators

A mixer or media translator that encodes content sent to the session participant issuing the TSRR SHALL consider the request to determine if it can fulfill it by changing its own encoding parameters. A media translator unable to fulfill the request MAY forward the request unaltered towards the media sender. A mixer encoding for multiple session participants will need to consider the joint needs of these participants before generating a TSRR on its own behalf towards the media sender.

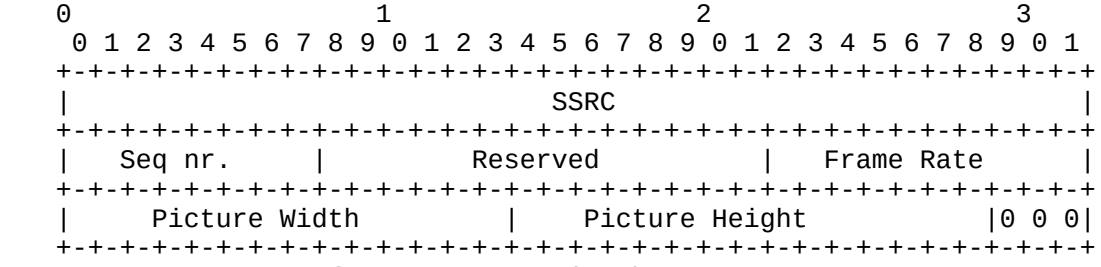
4.2. Temporal-Spatial Resolution Notification (TSRN)

The TSRN message is identified by RTCP packet type value PT=PSFB and FMT=13.

The FCI field SHALL contain one or more TSRN FCI entries.

4.2.1. Message format

The content of the FCI entry for the Temporal-Spatial Resolution Notification is depicted in Figure 2.



Syntax of an FCI Entry in the TSRN Message

Figure 2

SSRC (32 bits): The Synchronization Source (SSRC) of the source of the TSRR that resulted in this notification.

Seq nr. (8 bits): The sequence number value from the TSRR that is being acknowledged.

Reserved (14 bits): All bits SHALL be set to 0 by the sender and SHALL be ignored on reception.

Frame Rate (10 bits): The frame rate the media sender is using henceforth.

Picture Width (14 bits): The coding picture width the media sender is using henceforth.

Picture Height (14 bits): The coding picture height the media sender is using henceforth.

It is to note that the returned value (Frame Rate, Picture Width, Picture Height) may differ from the requested one, for example, in cases where a media encoder cannot change its frame rate or picture resolution, or when the requested temporal and spatial resolutions are larger than the temporal and spatial resolutions negotiated via SDP, or when pre-recorded content is used.

4.2.2. Semantics

This feedback message is used to acknowledge the reception of a TSRR. For each TSRR received targeted at the session participant, a TSRN FCI entry SHALL be sent in a TSRN feedback message. A single TSRN message MAY acknowledge multiple requests using multiple FCI entries. The Frame Rate, Picture Width and Picture Height value included SHALL be the same in all FCI entries of the TSRN message. Including an FCI for each requestor allows each requesting entity to determine that the media sender received the request. The notification SHALL also be sent in response to TSRR repetitions received. If the request receiver has received TSRR with several different sequence numbers from a single requestor, it SHALL only respond to the request with the highest (modulo 256) sequence number. Note that the highest sequence number may be a smaller integer value due to the wrapping of the field. Appendix A.1 of [RFC3550] has an algorithm for keeping track of the highest received sequence number for RTP packets; it could be adapted for this usage.

The TSRN SHALL include the Temporal-Spatial Resolution Frame Rate, Picture Width and Picture Height that will be used as a result of the request. This is not necessarily the same Frame Rate, Picture Width and Picture Height as requested, as the media sender may need to aggregate requests from several requesting session participants. It may also have some other policies or rules that limit the selection.

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 Notification, and the "SSRC of media source" is not used and SHALL be set to 0. The SSRCs of the requesting entities to which the Notification applies are in the corresponding FCI entries.

4.2.3. Timing Rules

The timing follows the rules outlined in section 3 of [RFC4585]. This acknowledgement message is not extremely time critical and SHOULD be sent using regular RTCP timing.

4.2.4. Handling of TSRN in Mixers and Translators

A mixer or translator that acts upon a TSRR SHALL also send the corresponding TSRN. In cases where it needs to forward a TSRR itself, the notification message MAY need to be delayed until the TSRR has been responded to.

5. Security Considerations

The defined messages have certain properties that have security implications. These must be addressed and taken into account by users of this protocol.

Spoofed or maliciously created feedback messages of the type defined in this specification can have the following implications:

* severely reduced picture resolution due to false TSRR messages that sets the picture width and height to a very low value;

* severely reduced frame rate due to false TSRR messages that sets the frame rate to a very low value.

* severely increasedd picture resolution due to false TSRR messages that sets the picture width and height to a value that is larger than the value negotiated via SDP;

* severely increased frame rate due to false TSRR messages that sets the frame rate to a value that is larger than the value negotiated via SDP.

To prevent these attacks, there is a need to apply authentication and integrity protection of the feedback messages. This can be accomplished against threats external to the current RTP session using the RTP profile that combines Secure RTP [SRTP] and AVPF into SAVPF [SAVPF]. In the mixer cases, separate security contexts and filtering can be applied between the mixer and the participants, thus protecting other users on the mixer from a misbehaving participant.

6. SDP Definitions

The capability of handling messages defined in this specification MAY be exchanged at a higher layer such as SDP. This specification follows all the rules defined in AVPF [RFC4585] and CCM [RFC5104] for an "rtcp-fb" attribute relating to the payload type in a session description.

6.1. Extension of the rtcp-fb Attribute

This specification defines a new parameter "tsrr" to the "ccm" feedback value defined in CCM [RFC5104] to indicate support of the Temporal-Spatial Resolution Request/Notification (TSRR/TSRN). All the rules described in [RFC4585] for rtcp-fb attribute relating to payload type and to multiple rtcp-fb attributes in a session description also apply to the new feedback messages defined in this specification.

rtcp-fb-ccm-param =/ SP "tsrr" ; Temporal-Spatial Resolution

6.2. Examples

Example 1: The following SDP describes a point-to-point video call with H.266, with the originator of the call declaring its capability to support the FIR and TSRR/TSRN codec control messages. The SDP is carried in a high-level signaling protocol like SIP.

```
v=0
o=alice 3203093520 3203093520 IN IP4 host.example.com
s=Point-to-Point call
c=IN IP4 192.0.2.124
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 51372 RTP/AVPF 98
a=rtpmap:98 H266/90000
a=rtcp-fb:98 ccm tsrr
a=rtcp-fb:98 ccm fir
```

In the above example, when the sender receives a TSRR message from the remote party it is capable of adjusting the trade-off as indicated in the RTCP TSRN feedback message.

Example 2: The following example describes the Offer/Answer implications for the codec control messages. The offerer wishes to support "tsrr", "fir" and "tmmbr". The offered SDP is

```
-------------> Offer
```

```
v=0
o=alice 3203093520 3203093520 IN IP4 host.example.com
s=Offer/Answer
c=IN IP4 192.0.2.124
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 51372 RTP/AVPF 98
a=rtpmap:98 H266/90000
a=rtcp-fb:98 ccm tsrr
a=rtcp-fb:98 ccm fir
a=rtcp-fb:* ccm tmmbr smaxpr=120
```

The answerer wishes to support only the FIR and TSRR/TSRN messages and the answerer SDP is

```
<---------------- Answer

v=0
o=alice 3203093520 3203093524 IN IP4 otherhost.example.com
s=Offer/Answer
c=IN IP4 192.0.2.37
m=audio 47190 RTP/AVP 0
a=rtpmap:0 PCMU/8000
m=video 53273 RTP/AVPF 98
a=rtpmap:98 H266/90000
a=rtcp-fb:98 ccm tsrr
a=rtcp-fb:98 ccm fir
```

7. IANA Considerations

Registration of new FMT Values for the Temporal-Spatial Resolution Request and Temporal-Spatial Resolution Notification is to be performed in the "FMT Values for RTPFB Payload Types" table [RFC4585] as defined in Section 4.1 and Section 4.2.

8. References

8.1. Normative References

[GreenMetadata]
"ISO/IEC FDIS 23001-11, Information technology - MPEG Systems Technologies - Part 11: Energy-Efficient Media Consumption (green metadata)", 2023, <https://www.iso.org/standard/83674.html>.

[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>.

[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>.

[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, <https://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, <https://www.rfc-editor.org/info/rfc5104>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

8.2. Informative References

[AVC] "Advanced video coding, ITU-T Recommendation H.264", 2021, <https://www.itu.int/rec/T-REC-H.264>.

[HEVC] "High efficiency video coding, ITU-T Recommendation H.265", 2021, <https://www.itu.int/rec/T-REC-H.265>.

[SAVPF] Ott, J. and E. Carrara, ""Extended Secure RTP Profile for RTCP-based Feedback (RTP/SAVPF)"", 2008, <https://datatracker.ietf.org/doc/pdf/rfc5124>.

[SRTP] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol(SRTP)", 2004, <https://datatracker.ietf.org/doc/pdf/rfc3711>.

[VVC] "Versatile Video Coding, ITU-T Recommendation H.266", 2022, <http://www.itu.int/rec/T-REC-H.266>.

Appendix A. Change History

To RFC Editor: PLEASE REMOVE ThIS SECTION BEFORE PUBLICATION

draft-ietf-avtcore-rtcp-green-metadata-00 ....initial version

draft-ietf-avtcore-rtcp-green-metadata-01 ....title and editorial changes

draft-ietf-avtcore-rtcp-green-metadata-02 ....editorial changes

draft-ietf-avtcore-rtcp-green-metadata-03 ....no changes

draft-ietf-avtcore-rtcp-green-metadata-04 ....no changes

draft-ietf-avtcore-rtcp-green-metadata-05 ....editorial changes

Authors' Addresses

Yong He
Qualcomm
5775 Morehouse Drive
San Diego, 92121
United States of America
Email: yonghe@qti.qualcomm.com

Christian Herglotz
FAU
Schlossplatz 4
91054 Erlangen
Germany
Email: christian.herglotz@fau.de

Edouard Francois
InterDigital
975 Avenue des Champs Blancs
35576 Cesson-Sevigne
France
Email: edouard.francois@interdigital.com