

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 28 April 2022

C. S. Perkins
University of Glasgow
25 October 2021

Sending RTP Control Protocol (RTCP) Feedback for Congestion Control in
Interactive Multimedia Conferences
draft-ietf-rmcat-rtp-cc-feedback-07

Abstract

This memo discusses the types of congestion control feedback that it is possible to send using the RTP Control Protocol (RTCP), and their suitability of use in implementing congestion control for unicast multimedia applications.

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 28 April 2022.

Copyright Notice

Copyright (c) 2021 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Possible Models for RTCP Feedback	2
3. What Feedback is Achievable With RTCP?	4
3.1. Scenario 1: Voice Telephony	4
3.2. Scenario 2: Point-to-Point Video Conference	7
4. Discussion and Conclusions	12
5. Security Considerations	12
6. IANA Considerations	12
7. Acknowledgements	13
8. Informative References	13
Author's Address	14

1. Introduction

The deployment of WebRTC systems [RFC8825] has resulted in high-quality video conferencing seeing extremely wide use. To ensure the stability of the network in the face of this use, WebRTC systems need to use some form of congestion control for their RTP-based media traffic. To develop such congestion control, it is necessary to understand the sort of congestion feedback that can be provided within the framework of RTP [RFC3550] and the RTP Control Protocol (RTCP). It then becomes possible to determine if this is sufficient for congestion control, or if some form of RTP extension is needed.

This memo considers the congestion feedback that can be sent using RTCP under the RTP/SAVPF profile [RFC5124] (the secure version of the RTP/AVPF profile [RFC4585]). This profile was chosen as it forms the basis for media transport in WebRTC [RFC8834] systems. Nothing in this memo is specific to the secure version of the profile, or to WebRTC, however.

2. Possible Models for RTCP Feedback

Several questions need to be answered when providing RTCP reception quality feedback for congestion control purposes. These include:

- * How often is feedback needed?
- * How much overhead is acceptable?
- * How much, and what, data does each report contain?

The key question is how often does the receiver need to send feedback on the reception quality it is experiencing, and hence the congestion state of the network? Traditional congestion control protocols, such as TCP, send acknowledgements with every packet (or, at least, every

couple of packets). That is straight-forward and low overhead when traffic is bidirectional and acknowledgements can be piggybacked onto return path data packets. It can also be acceptable, and can have reasonable overhead, to send separate acknowledgement packets when those packets are much smaller than data packets. It becomes a problem, however, when there is no return traffic on which to piggyback acknowledgements, and when acknowledgements are similar in size to data packets; this can be the case for some forms of media traffic, especially for voice over IP (VoIP) flows, but less so for video.

When considering multimedia traffic, it might make sense to consider less frequent feedback. For example, it might be possible to send a feedback packet once per video frame, or every few frames, or once per network round trip time (RTT). This could still give sufficiently frequent feedback for the congestion control loop to be stable and responsive while keeping the overhead reasonable when the feedback cannot be piggybacked onto returning data. In this case, it is important to note that RTCP can send much more detailed feedback than simple acknowledgements. For example, if it were useful, it could be possible to use an RTCP extended report (XR) packet [RFC3611] to send feedback once per RTT comprising a bitmap of lost and received packets, with reception times, over that RTT. As long as feedback is sent frequently enough that the control loop is stable, and the sender is kept informed when data leaves the network (to provide an equivalent to ACK clocking in TCP), it is not necessary to report on every packet at the instant it is received (indeed, it is unlikely that a video codec can react instantly to a rate change anyway, and there is little point in providing feedback more often than the codec can adapt).

The amount of overhead due to congestion control feedback that is considered acceptable has to be determined. RTCP feedback is sent in separate packets to RTP data, and this has some cost in terms of additional header overhead compared to protocols that piggyback feedback on return path data packets. The RTP standards have long said that a 5% overhead for RTCP traffic generally acceptable, while providing the ability to change this fraction. Is this still the case for congestion control feedback? Or is there a desire to either see more responsive feedback and congestion control, possibility with a higher overhead, or is lower overhead wanted, accepting that this might reduce responsiveness of the congestion control algorithm?

Finally, the details of how much, and what, data is to be sent in each report will affect the frequency and/or overhead of feedback. There is a fundamental trade-off that the more frequently feedback packets are sent, the less data can be included in each packet to keep the overhead constant. Does the congestion control need high rate but simple feedback (e.g., like TCP acknowledgements), or is it acceptable to send more complex feedback less often?

3. What Feedback is Achievable With RTCP?

The following sections illustrate how the RTCP congestion control feedback report [RFC8888] can be used in different scenarios, and illustrate the overheads of this approach.

3.1. Scenario 1: Voice Telephony

In many ways, point-to-point voice telephony is the simplest scenario for congestion control, since there is only a single media stream to control. It's complicated, however, by severe bandwidth constraints on the feedback, to keep the overhead manageable.

Assume a two-party point-to-point voice-over-IP call, using RTP over UDP/IP. A rate adaptive speech codec, such as Opus, is used, encoded into RTP packets in frames of duration T_f seconds ($T_f = 20\text{ms}$ in many cases, but values up to 60ms are not uncommon). The congestion control algorithm requires feedback every N_r frames, i.e., every $N_r * T_f$ seconds, to ensure effective control. Both parties in the call send speech data or comfort noise with sufficient frequency that they are counted as senders for the purpose of the RTCP reporting interval calculation.

RTCP feedback packets can be full, compound, RTCP feedback packets, or non-compound RTCP packets. A compound RTCP packet is sent once for every N_{nc} non-compound RTCP packets.

Compound RTCP packets contain a Sender Report (SR) packet, a Source Description (SDS) packet, and an RTP Congestion Control Feedback (CCFB) packet [RFC8888]. Non-compound RTCP packets contain only the CCFB packet. Since each participant sends only a single RTP media stream, the extensions for RTCP report aggregation [RFC8108] and reporting group optimisation [RFC8861] are not used.

Within each compound RTCP packet, the SR packet will contain a sender information block (28 octets) and a single reception report block (24 octets), for a total of 52 octets. A minimal SDS packet will contain a header (4 octets) and a single chunk containing an SSRC (4 octets) and a CNAME item, and if the recommendations for choosing the CNAME [RFC7022] are followed, the CNAME item will comprise a 2 octet

header, 16 octets of data, and 2 octets of padding, for a total SDES packet size of 28 octets. The CCFB packets contains an RTCP header and SSRC (8 octets), a report timestamp (4 octets), the SSRC, beginning and ending sequence numbers (8 octets), and $2*Nr$ octets of reports, for a total of $20 + 2*Nr$ octets. The compound Secure RTCP packet will include 4 octets of trailer followed by an 80 bit (10 octet) authentication tag if HMAC-SHA1 authentication is used. If IPv4 is used, with no IP options, the UDP/IP header will be 28 octets in size. This gives a total compound RTCP packet size of $S_c = 142 + 2*Nr$ octets.

The non-compound RTCP packets will comprise just the CCFB packet, SRTCP trailer and authentication tag, and a UDP/IP header. It can be seen that these packets will be $S_{nc} = 62 + 2*Nr$ octets in size.

The RTCP reporting interval calculation ([RFC3550], Section 6.2) for a two-party session where both participants are senders, reduces to:

$$Trtcp = n * Srtcp / Brtcp$$

where $Srtcp = (S_c + N_{nc} * S_{nc}) / (1 + N_{nc})$ is the average RTCP packet size in octets, $Brtcp$ is the bandwidth allocated to RTCP in octets per second, and n is the number of participants in the RTP session (in this scenario, $n = 2$).

To ensure an RTCP report containing congestion control feedback is sent after every Nr frames of audio, it is necessary to set the RTCP reporting interval $Trtcp = Nr * T_f$, which when substituted into the previous gives $Nr * T_f = n * Srtcp / Brtcp$. Solving this to give the RTCP bandwidth, $Brtcp$, and expanding the definition of $Srtcp$ gives:

$$Brtcp = (n * (S_c + N_{nc} * S_{nc})) / (Nr * T_f * (1 + N_{nc})).$$

If we assume every report is a compound RTCP packet (i.e., $N_{nc} = 0$), the frame duration $T_f = 20ms$, and an RTCP report is sent for every second frame (i.e., 25 RTCP reports per second), this gives an RTCP feedback bandwidth, $Brtcp = 57kbps$. Increasing the frame duration, or reducing the frequency of reports, will reduce the RTCP bandwidth as shown in Table 1.

Tf (seconds)	Nr (frames)	rtcp_bw (kbps)
0.020	2	57.0
0.020	4	29.3
0.020	8	15.4
0.020	16	8.5
0.060	2	19.0
0.060	4	9.8
0.060	8	5.1
0.060	16	2.8

Table 1: RTCP bandwidth needed for VoIP feedback

The final row of Table 1 (60ms frames, report every 16 frames) sends RTCP reports once per second, giving an RTCP bandwidth overhead of 2.8kbps.

The overhead can be reduced by sending some reports in non-compound RTCP packets [RFC5506]. For example, if we alternate compound and non-compound RTCP packets, i.e., $N_{nc} = 1$, the calculation gives the results shown in Table 2.

Tf (seconds)	Nr (frames)	rtcp_bw (kbps)
0.020	2	41.4
0.020	4	21.5
0.020	8	11.5
0.020	16	6.5
0.060	2	13.8
0.060	4	7.2
0.060	8	3.8
0.060	16	2.2

Table 2: Required RTCP bandwidth for VoIP feedback (alternating compound and non-compound reports)

The RTCP bandwidth needed for 60ms frames, reporting every 16 frames (once per second), can be seen to drop to 2.2kbps. This calculation can be repeated for other patterns of compound and non-compound RTCP packets, feedback frequency, and frame duration, as needed.

Note: To achieve the RTCP transmission intervals above the RTP/SAVPF profile with $T_{rr_interval}=0$ is used, since even when using the reduced minimal transmission interval, the RTP/SAVP profile would only allow sending RTCP at most every 0.11s (every third frame of video). Using RTP/SAVPF with $T_{rr_interval}=0$ however is capable of fully utilizing the configured 5% RTCP bandwidth fraction.

3.2. Scenario 2: Point-to-Point Video Conference

Consider a point to point video call between two end systems. There will be four RTP flows in this scenario, two audio and two video, with all four flows being active for essentially all the time (the audio flows will likely use voice activity detection and comfort noise to reduce the packet rate during silent periods, but this does not cause the transmissions to stop).

Assume all four flows are sent in a single RTP session, each using a separate SSRC. The RTCP reports from the co-located audio and video SSRCs at each end point are aggregated [RFC8108], the optimisations in [RFC8861] are used, and RTCP congestion control feedback is sent [RFC8888].

When all members are senders, the RTCP reporting interval calculation in Section 6.2 and 6.3 of [RFC3550] and [RFC4585] reduces to:

$$\text{Trtcp} = n * \text{Srtcp} / \text{Brtcp}$$

where n is the number of members in the session, Srtcp is the average RTCP packet size in octets, and the Brtcp is the RTCP bandwidth in octets per second.

The average RTCP packet size, Srtcp , depends on the amount of feedback sent in each RTCP packet, on the number of members in the session, on the size of source description (RTCP SDES) information sent, and on the amount of congestion control feedback sent in each packet.

As a baseline, each RTCP packet will be a compound RTCP packet that contains an aggregate of a compound RTCP packet generated by the video SSRC and a compound RTCP packet generated by the audio SSRC. When the RTCP reporting group extensions are used, one of these SSRCs will be a reporting SSRC, to which the other SSRC will have delegated its reports. No non-compound RTCP packets are sent.

The aggregated compound RTCP packet from the non-reporting SSRC will contain an RTCP SR packet, an RTCP SDES packet, and an RTCP RGRS packet. The RTCP SR packet contains the 28 octet header and sender information, but no report blocks (since the reporting is delegated). The RTCP SDES packet will comprise a header (4 octets), originating SSRC (4 octets), a CNAME chunk, a terminating chunk, and any padding. If the CNAME follows [RFC7022] and [RFC8834] it will be 18 octets in size, and will need 1 octet of padding, making the SDES packet 28 octets in size. The RTCP RGRS packet will be 12 octets in size. This gives a total of $28 + 28 + 12 = 68$ octets.

The aggregated compound RTCP packet from the reporting SSRC will contain an RTCP SR packet, an RTCP SDES packet, and an RTCP congestion control feedback packet. The RTCP SR packet will contain two report blocks, one for each of the remote SSRCs (the report for the other local SSRC is suppressed by the reporting group extension), for a total of $28 + (2 * 24) = 76$ octets. The RTCP SDES packet will comprise a header (4 octets), originating SSRC (4 octets), a CNAME chunk, an RGRP chunk, a terminating chunk, and any padding. If the CNAME follows [RFC7022] and [RFC8834] it will be 18 octets in size.

The RGRP chunk similarly comprises 18 octets, and 3 octets of padding are needed, for a total of 48 octets. The RTCP congestion control feedback (CCFB) report comprises an 8 octet RTCP header and SSRC, a 4 octet report timestamp, and for each of the remote audio and video SSRCs, an 8 octet report header, and 2 octets per packet reported upon, and padding to a 4 octet boundary if needed; that is $8 + 4 + 8 + (2 * N_v) + 8 + (2 * N_a)$ where N_v is the number of video packets per report, and N_a is the number of audio packets per report.

The complete compound RTCP packet contains the RTCP packets from both the reporting and non-reporting SSRCs, an SRTCP trailer and authentication tag, and a UDP/IPv4 header. The size of this RTCP packet is therefore: $262 + (2 * N_v) + (2 * N_a)$ octets. Since the aggregate RTCP packet contains reports from two SSRCs, the RTCP packet size is halved before use [RFC8108]. Accordingly, the size of the RTCP packets is:

$$Srtcp = (262 + (2 * N_v) + (2 * N_a)) / 2$$

How many RTP packets does the RTCP XR congestion control feedback packet included in these compound RTCP packets report on? That is, what are the values of N_v and N_a ? This depends on the RTCP reporting interval, $Trtcp$, the video bit rate and frame rate, R_f , the audio bit rate and framing interval, and whether the receiver chooses to send congestion control feedback in each RTCP packet it sends.

To simplify the calculation, assume it is desired to send one RTCP report for each frame of video received (i.e., $Trtcp = 1 / R_f$) and to include a congestion control feedback packet in each report. Assume that video has constant bit rate and frame rate, and that each frame of packet has to fit into a 1500 octet MTU. Further, assume that the audio takes negligible bandwidth, and that the audio framing interval can be varied within reasonable bounds, so that an integral number of audio frames align with video frame boundaries.

Table 3 shows the resulting values of N_v and N_a , the number of video and audio packets covered by each congestion control feedback report, for a range of data rates and video frame rates, assuming congestion control feedback is sent once per video frame. The table also shows the result of inverting the RTCP reporting interval calculation to find the corresponding RTCP bandwidth, $Brtcp$. The RTCP bandwidth is given in kbps and as a fraction of the data rate.

It can be seen that, for example, with a data rate of 1024 kbps and video sent at 30 frames-per-second, the RTCP congestion control feedback report sent for each video frame will include reports on 3 video packets and 2 audio packets. The RTCP bandwidth needed to sustain this reporting rate is 127.5kbps (12% of the data rate). This assumes an audio framing interval of 16.67ms, so that two audio packets are sent for each video frame.

Data Rate (kbps)	Video Frame Rate: Rf	Video Packets per Report: Nv	Audio Packets per Report: Na	Required RTCP bandwidth: Brtcp (kbps)
100	8	1	6	34.5 (34%)
200	16	1	3	67.5 (33%)
350	30	1	2	125.6 (35%)
700	30	2	2	126.6 (18%)
700	60	1	1	249.4 (35%)
1024	30	3	2	127.5 (12%)
1400	60	2	1	251.2 (17%)
2048	30	6	2	130.3 (6%)
2048	60	3	1	253.1 (12%)
4096	30	12	2	135.9 (3%)
4096	60	6	1	258.8 (6%)

Table 3: Required RTCP bandwidth, reporting on every frame

Use of reduced size RTCP [RFC5506] would allow the SR and SDES packets to be omitted from some reports. These "non-compound" (actually, compound but reduced size in this case) RTCP packets would contain an RTCP RGRS packet from the non-reporting SSRC, and an RTCP SDES RGRP packet and a congestion control feedback packet from the reporting SSRC. This will be $12 + 28 + 12 + 8 + 2*Nv + 8 + 2*Na$ octets, plus the SRTCP trailer and authentication tag, and a UDP/IP header. That is, the size of the non-compound packets would be $(110 + 2*Nv + 2*Na)/2$ octets. Repeating the analysis above, but alternating compound and non-compound reports gives results as shown in Table 4.

Data Rate (kbps)	Video Frame Rate: Rf	Video Packets per Report: Nv	Audio Packets per Report: Na	Required RTCP bandwidth: Brtcp (kbps)
100	8	1	6	24.1 (24%)
200	16	1	3	46.8 (23%)
350	30	1	2	86.7 (24%)
700	30	2	2	87.7 (12%)
700	60	1	1	171.6 (24%)
1024	30	3	2	88.6 (8%)
1400	60	2	1	173.4 (12%)
2048	30	6	2	91.4 (4%)
2048	60	3	1	175.3 (8%)
4096	30	12	2	97.0 (2%)
4096	60	6	1	180.9 (4%)

Table 4: Required RTCP bandwidth, reporting on every frame, with reduced-size reports

The use of reduced-size RTCP gives a noticeable reduction in the needed RTCP bandwidth, and can be combined with reporting every few frames rather than every frames. Overall, it is clear that the RTCP overhead can be reasonable across the range of data and frame rates, if RTCP is configured carefully.

4. Discussion and Conclusions

RTCP as it is currently specified cannot be used to send per-packet congestion feedback with reasonable overhead.

RTCP can, however, be used to send congestion feedback on each frame of video sent, provided the session bandwidth exceeds a couple of megabits per second (the exact rate depending on the number of session participants, the RTCP bandwidth fraction, and what RTCP extensions are enabled, and how much detail of feedback is needed). For lower rate sessions, the overhead of reporting on every frame becomes high, but can be reduced to something reasonable by sending reports once per N frames (e.g., every second frame), or by sending non-compound RTCP reports in between the regular reports.

If it is desired to use RTCP in something close to it's current form for congestion feedback in WebRTC, the multimedia congestion control algorithm needs be designed to work with feedback sent every few frames, since that fits within the limitations of RTCP. The provided feedback will be more detailed than just an acknowledgement, however, and will provide a loss bitmap, relative arrival time, and received ECN marks, for each packet sent. This will allow congestion control that is effective, if slowly responsive, to be implemented.

The format described in [RFC8888] seems sufficient for the needs of congestion control feedback. There is little point optimising this format: the main overhead comes from the UDP/IP headers and the other RTCP packets included in the compound packets, and can be lowered by using the [RFC5506] extensions and sending reports less frequently.

Further study of the scenarios of interest is needed, to ensure that the analysis presented is applicable to other media topologies, and to sessions with different data rates and sizes of membership.

5. Security Considerations

An attacker that can modify or spoof RTCP congestion control feedback packets can manipulate the sender behaviour to cause denial of service. This can be prevented by authentication and integrity protection of RTCP packets, for example using the secure RTP profile [RFC3711][RFC5124], or by other means as discussed in [RFC7201].

6. IANA Considerations

There are no actions for IANA.

7. Acknowledgements

Thanks to Magnus Westerlund and the members of the RMCAT feedback design team for their feedback.

8. Informative References

- [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>>.
- [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, <<https://www.rfc-editor.org/info/rfc3611>>.
- [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>>.
- [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>>.
- [RFC5124] Ott, J. and E. Carrara, "Extended Secure RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/SAVPF)", RFC 5124, DOI 10.17487/RFC5124, February 2008, <<https://www.rfc-editor.org/info/rfc5124>>.
- [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, <<https://www.rfc-editor.org/info/rfc5506>>.
- [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, <<https://www.rfc-editor.org/info/rfc7022>>.
- [RFC7201] Westerlund, M. and C. Perkins, "Options for Securing RTP Sessions", RFC 7201, DOI 10.17487/RFC7201, April 2014, <<https://www.rfc-editor.org/info/rfc7201>>.

- [RFC8108] Lennox, J., Westerlund, M., Wu, Q., and C. Perkins, "Sending Multiple RTP Streams in a Single RTP Session", RFC 8108, DOI 10.17487/RFC8108, March 2017, <<https://www.rfc-editor.org/info/rfc8108>>.
- [RFC8825] Alvestrand, H., "Overview: Real-Time Protocols for Browser-Based Applications", RFC 8825, DOI 10.17487/RFC8825, January 2021, <<https://www.rfc-editor.org/info/rfc8825>>.
- [RFC8834] Perkins, C., Westerlund, M., and J. Ott, "Media Transport and Use of RTP in WebRTC", RFC 8834, DOI 10.17487/RFC8834, January 2021, <<https://www.rfc-editor.org/info/rfc8834>>.
- [RFC8861] Lennox, J., Westerlund, M., Wu, Q., and C. Perkins, "Sending Multiple RTP Streams in a Single RTP Session: Grouping RTP Control Protocol (RTCP) Reception Statistics and Other Feedback", RFC 8861, DOI 10.17487/RFC8861, January 2021, <<https://www.rfc-editor.org/info/rfc8861>>.
- [RFC8888] Sarker, Z., Perkins, C., Singh, V., and M. Ramalho, "RTP Control Protocol (RTCP) Feedback for Congestion Control", RFC 8888, DOI 10.17487/RFC8888, January 2021, <<https://www.rfc-editor.org/info/rfc8888>>.

Author's Address

Colin Perkins
University of Glasgow
School of Computing Science
Glasgow
G12 8QQ
United Kingdom

Email: msp@csperkins.org