IETF RMCAT Working Group   Z. Sarker
Internet-Draft   Ericsson AB
Intended status: Standards Track   C. Perkins
Expires: January 16, 2019   University of Glasgow
                                                            V. Singh
                                                        callstats.io
                                                         M. Ramalho
                                                       Cisco Systems
                                                       July 15, 2018

        RTP Control Protocol (RTCP) Feedback for Congestion Control
                draft-ietf-avtcore-cc-feedback-message-02

Abstract

   This document describes an RTCP feedback message intended to enable
   congestion control for interactive real-time traffic using RTP. The
   feedback message is designed for use with a sender-based congestion
   control algorithm, in which the receiver of an RTP flow sends RTCP
   feedback packets to the sender containing the information the sender
   needs to perform congestion control.

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 January 16, 2019.

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

   For interactive real-time traffic, such as video conferencing flows,
   the typical protocol choice is the Real-time Transport Protocol (RTP)
   running over the User Datagram Protocol (UDP). RTP does not provide
   any guarantee of Quality of Service (QoS), reliability, or timely
   delivery, and expects the underlying transport protocol to do so.
   UDP alone certainly does not meet that expectation. However, the RTP
   Control Protocol (RTCP) provides a mechanism by which the receiver of
   an RTP flow can periodically send transport and media quality metrics
   to the sender of that RTP flow. This information can be used by the
   sender to perform congestion control. In the absence of standardized
   messages for this purpose, designers of congestion control algorithms
   have developed proprietary RTCP messages that convey only those
   parameters needed for their respective designs. As a direct result,
   the different congestion control (i.e., rate adaptation) designs are
   not interoperable. To enable algorithm evolution as well as
   interoperability across designs (e.g., different rate adaptation
   algorithms), it is highly desirable to have generic congestion
   control feedback format.

   To help achieve interoperability for unicast RTP congestion control,
   this memo proposes a common RTCP feedback packet format that can be
   used by NADA [I-D.ietf-rmcat-nada], SCReAM [RFC8298], Google

Congestion Control [I-D.ietf-rmcat-gcc] and Shared Bottleneck
Detection [RFC8382], and hopefully also by future RTP congestion
control algorithms.

2. 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 [RFC2119].

In addition the terminology defined in [RFC3550], [RFC3551],
[RFC3611], [RFC4585], and [RFC5506] applies.

3. RTCP Feedback for Congestion Control

Based on an analysis of NADA [I-D.ietf-rmcat-nada], SCReAM [RFC8298],
Google Congestion Control [I-D.ietf-rmcat-gcc] and Shared Bottleneck
Detection [RFC8382], the following per-RTP packet congestion control
feedback information has been determined to be necessary:

o RTP sequence number: The receiver of an RTP flow needs to feedback
   the sequence numbers of the received RTP packets to the sender, so
   the sender can determine which packets were received and which
   were lost. Packet loss is used as an indication of congestion by
   many congestion control algorithms.

o Packet Arrival Time: The receiver of an RTP flow needs to feedback
   the arrival time of each RTP packet to the sender. Packet delay
   and/or delay variation (jitter) is used as a congestion signal by
   some congestion control algorithms.

o Packet Explicit Congestion Notification (ECN) Marking: If ECN
   [RFC3168], [RFC6679] is used, it is necessary to feedback the
   2-bit ECN mark in received RTP packets, indicating for each RTP
   packet whether it is marked not-ECT, ECT(0), ECT(1), or ECN-CE.
   If the path used by the RTP traffic is ECN capable the sender can
   use Congestion Experienced (ECN-CE) marking information as a
   congestion control signal.

Every RTP flow is identified by its Synchronization Source (SSRC)
identifier. Accordingly, the RTCP feedback format needs to group its
reports by SSRC, sending one report block per received SSRC.

As a practical matter, we note that host operating system (OS)
process interruptions can occur at inopportune times. Accordingly,
recording RTP packet send times at the sender, and the corresponding
RTP packet arrival times at the receiver, needs to be done with
deliberate care. This is because the time duration of host OS

interruptions can be significant relative to the precision desired in
the one-way delay estimates. Specifically, the send time needs to be
recorded at the last opportunity prior to transmitting the RTP packet
at the sender, and the arrival time at the receiver needs to be
recorded at the earliest available opportunity.

3.1. RTCP Congestion Control Feedback Report

Congestion control feedback can be sent as part of a regular
scheduled RTCP report, or in an RTP/AVPF early feedback packet. If
sent as early feedback, congestion control feedback MAY be sent in a
non-compound RTCP packet [RFC5506] if the RTP/AVPF profile [RFC4585]
or the RTP/SAVPF profile [RFC5124] is used.

Irrespective of how it is transported, the congestion control
feedback is sent as a Transport Layer Feedback Message (RTCP packet
type 205). The format of this RTCP packet is shown in Figure 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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |V=2|P| FMT=CCFB |   PT = 205    |          length               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 SSRC of RTCP packet sender                    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  SSRC of 1st RTP Stream                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        begin_seq              |         end_seq+1             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |L|ECN|  Arrival time offset    | ...                          .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   .                                                              .
   .                                                              .
   .                                                              .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  SSRC of nth RTP Stream                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |        begin_seq              |         end_seq+1             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |L|ECN|  Arrival time offset    | ...                          |
   .                                                              .
   .                                                              .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  Report Timestamp (32bits)                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
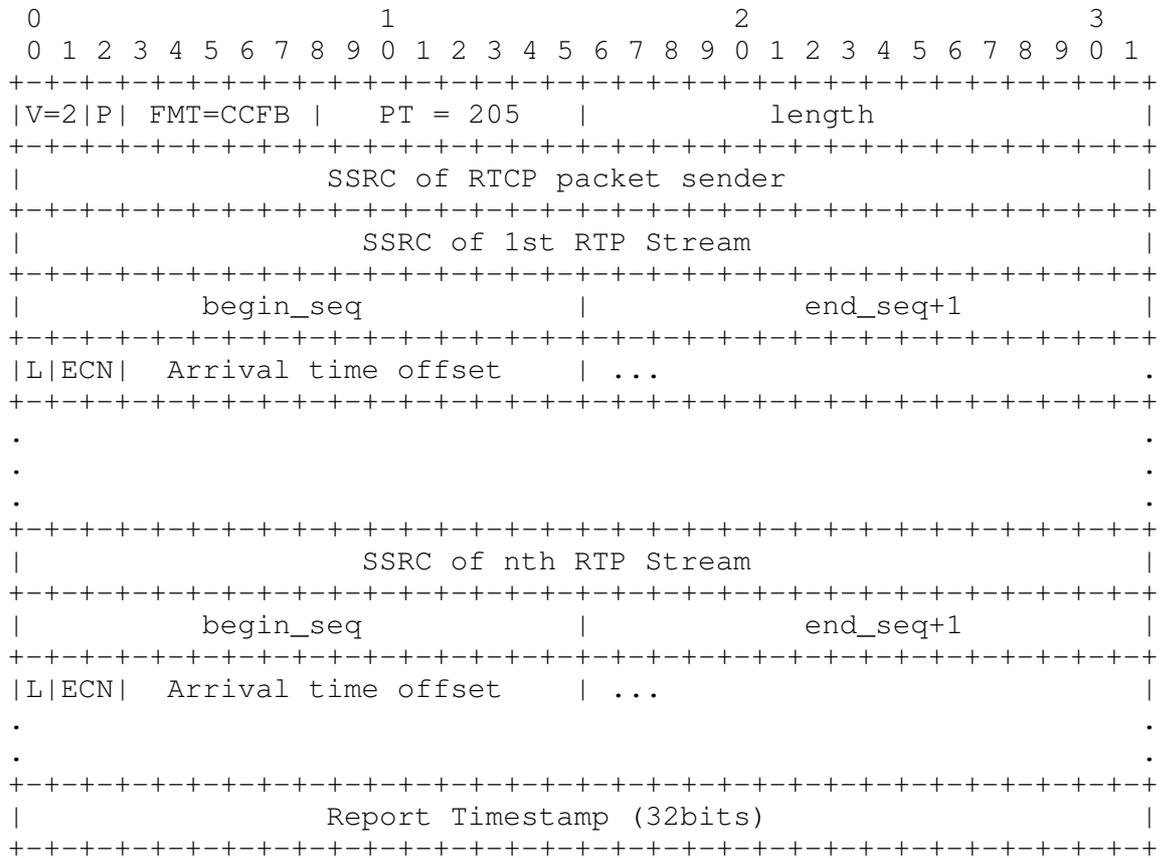
     Figure 1: RTCP Congestion Control Feedback Packet Format

   The first eight octets comprise a standard RTCP header, with PT=205
   and FMT=CCFB indicating that this is a congestion control feedback
   packet, and with the SSRC set to that of the sender of the RTCP
   packet. (NOTE TO RFC EDITOR: please replace CCFB here and in the
   above diagram with the IANA assigned RTCP feedback packet type, and
   remove this note)

   Section 6.1 of [RFC4585] requires the RTCP header to be followed by
   the SSRC of the RTP flow being reported upon. Accordingly, the RTCP
   header is followed by a report block for each SSRC from which RTP
   packets have been received, followed by a Report Timestamp.

   Each report block begins with the SSRC of the received RTP Stream on
   which it is reporting. Following this, the report block contains a
   16-bit packet metric block for each RTP packet with sequence number,
   seq, in the range begin_seq <= seq < end_seq+1 (calculated using
   arithmetic modulo 65535 to account for possible sequence number wrap-
   around). If the number of 16-bit packet metric blocks included in
   the report block is not a multiple of two, then 16 bits of zero
   padding MUST be added after the last packet metric block, to align
   the end of the packet metric blocks with the next 32 bit boundary.
   Each report block MUST NOT include more than 16384 packet metric
   blocks (i.e., it MUST NOT report on more than one quarter of the
   sequence number space in a single report).

   The contents of each 16-bit packet metric block comprises the L, ECN,
   and ATO fields are as follows:

   o L (1 bit): is a boolean to indicate if the packet was received. 0
      represents that the packet was not yet received and all the
      subsequent bits (ECN and ATO) are also set to 0. 1 represent the
      packet was received and the subsequent bits in the block need to
      be parsed.

   o ECN (2 bits): is the echoed ECN mark of the packet. These are set
      to 00 if not received, or if ECN is not used.

   o Arrival time offset (ATO, 13 bits): is the arrival time of the RTP
      packet at the receiver. It is measured as an offset from the time
      at which the RTCP congestion control feedback report packet is
      sent. The arrival time offset is calculated by subtracting the
      reception time of the RTP packet denoted by this 16 bit packet
      metric block from the Report Timestamp (RTS) field of the RTCP
      congestion control feedback report packet in which the packet
      metric report block is contained. The arrival time offset is
      measured in units of 1/1024 seconds (this unit is chosen to give
      exact offsets from the RTS field). If the measured value is
      greater than 8189/1024 seconds (the value that would be coded as

0x1FFD), the value 0x1FFE MUST be reported to indicate an over-
range positive measurement. If the measurement is unavailable,
the value 0x1FFF MUST be reported.

The RTCP congestion control feedback report packet concludes with the
Report Timestamp field (RTS, 32 bits). This represents the time
instant when the report packet was generated. The value of RTS field
is derived from the same wallclock used to generate the NTP timestamp
field in RTCP Sender Report (SR) packets. It is formatted as the
middle 32 bits of an NTP format timestamp, as described in Section 4
of [RFC3550].

RTCP congestion control feedback packets SHOULD include a report
block for each SSRC that is being congestion controlled. The
sequence number ranges reported on in consecutive reports for an SSRC
SHOULD be consecutive and SHOULD NOT overlap (i.e., begin_seq for a
report is expected to be one greater, modulo 65535, than end_seq of
the previous report for that SSRC). If overlapping reports are sent,
the information in the later report updates that in any previous
reports for packets included in both reports (although note that such
updated information will likely arrive too late to affect congestion
control decisions at the sender). Reports that cover RTP sequence
number ranges that are more than 16384 (i.e., one quarter of the
sequence number space) ahead of the last end_seq received from an
SSRC, or behind the last begin_seq received from an SSRC, modulo
65535 to account for wrap-around, SHOULD be ignored. An exception to
this occurs if sender has sent RTP packets using more than one
quarter of the sequence number space since it last received an RTCP
congestion control feedback packet, then a report on recently sent
RTP packets ought to be accepted, to allow recovery from report
packet loss.

If no packets are received from an SSRC in a reporting interval, a
report block MAY be sent with begin_seq and end_seq+1 both set to the
highest sequence number previously received from that SSRC (or, the
report can simply to omitted). The corresponding SR/RR packet will
have a non-increased extended highest sequence number received field
that will inform the sender that no packets have been received, but
it can ease processing to have that information available in the
congestion control feedback reports too.

4. Feedback Frequency and Overhead

There is a trade-off between speed and accuracy of reporting, and the
overhead of the reports. [I-D.ietf-rmcat-rtp-cc-feedback] discusses
this trade-off, suggests desirable RTCP feedback rates, and provides
guidance on how to configure the RTCP bandwidth fraction, etc., to
make appropriate use of the reporting block described in this memo.

Specifications for RTP congestion control algorithms can also provide guidance.

It is a general understanding that the congestion control algorithms will work better with more frequent feedback - per packet feedback. However, RTCP bandwidth and transmission rules put some upper limits on how frequently the RTCP feedback messages can be send from the RTP receiver to the RTP sender. It has been shown [I-D.ietf-rmcat-rtp-cc-feedback] that in most cases a per frame feedback is a reasonable assumption on how frequent the RTCP feedback messages can be transmitted. It has also been noted that even if a higher frequency of feedback is desired it is not viable if the feedback messages starts to compete against the RTP traffic on the feedback path during congestion period. Analyzing the feedback interval requirement [feedback-requirements] it can be seen that the candidate algorithms can perform with a feedback interval range of 50-200ms. A value within this range need to be negotiated at session setup.

5. SDP Signalling

A new "ack" feedback parameter, "ccfb", is defined to indicate the use of the RTP Congestion Control feedback packet format defined in Section 3. The ABNF definition of this SDP parameter extension is:

```
rtcp-fb-ack-param = <See Section 4.2 of [RFC4584]>
rtcp-fb-ack-param =/ ccfb-par
ccfb-par          = SP "ccfb"
```

The offer/answer rules for these SDP feedback parameters are specified in the RTP/AVPF profile [RFC4585].

6. Design Rationale

The primary function of RTCP SR/RR packets is to report statistics on the reception of RTP packets. The reception report blocks sent in these packets contain information about observed jitter, fractional packet loss, and cumulative packet loss. It was intended that this information could be used to support congestion control algorithms, but experience has shown that it is not sufficient for that purpose. An efficient congestion control algorithm requires more fine grained information on per packet reception quality than is provided by SR/RR packets to react effectively.

The Codec Control Messages for the RTP/AVPF profile [RFC5104] include a Temporary Maximum Media Bit Rate (TMMBR) message. This is used to convey a temporary maximum bit rate limitation from a receiver of RTP packets to their sender. Even though it was not designed to replace

congestion control, TMMBR has been used as a means to do receiver
based congestion control where the session bandwidth is high enough
to send frequent TMMBR messages, especially when used with non-
compound RTCP packets [RFC5506]. This approach requires the receiver
of the RTP packets to monitor their reception, determine the level of
congestion, and recommend a maximum bit rate suitable for current
available bandwidth on the path; it also assumes that the RTP sender
can/will respect that bit rate. This is the opposite of the sender
based congestion control approach suggested in this memo, so TMMBR
cannot be used to convey the information needed for a sender based
congestion control. TMMBR could, however, be viewed a complementary
mechanism that can inform the sender of the receiver's current view
of acceptable maximum bit rate.

A number of RTCP eXtended Report (XR) blocks have previously been
defined to report details of packet loss, arrival times [RFC3611],
delay [RFC6843], and ECN marking [RFC6679]. It is possible to
combine several such XR blocks to report the detailed loss, arrival
time, and ECN marking marking information needed for effective
sender-based congestion control. However, the result has high
overhead both in terms of bandwidth and complexity, due to the need
to stack multiple reports.

Considering these issues, we believe it appropriate to design a new
RTCP feedback mechanism to convey information for sender based
congestion control algorithms. The new congestion control feedback
RTCP packet described in Section 3 provides such a mechanism.

7. Acknowledgements

   This document is an outcome of RMCAT design team discussion. We
   would like to thank all participants specially Sergio Mena, Xiaoquing
   Zhu, Stefan Holmer, David, Ingemar Johansson, Randell Jesup, Ingemar
   Johansson, and Magnus Westerlund for their valuable contribution to
   the discussions and to the document.

8. IANA Considerations

   The IANA is requested to register one new RTP/AVPF Transport-Layer
   Feedback Message in the table for FMT values for RTPFB Payload Types
   [RFC4585] as defined in Section 3.1:

        Name: CCFB
        Long name: RTP Congestion Control Feedback
        Value: (to be assigned by IANA)
        Reference: (RFC number of this document, when published)

The IANA is also requested to register one new SDP "rtcp-fb"
attribute "ack" parameter, "ccfb", in the SDP ("ack" and "nack"
Attribute Values) registry:

        Value name: ccfb
        Long name: Congestion Control Feedback
        Usable with: ack
        Reference: (RFC number of this document, when published)

9. Security Considerations

   The security considerations of the RTP specification [RFC3550], the
   applicable RTP profile (e.g., [RFC3551], [RFC3711], or [RFC4585]),
   and the RTP congestion control algorithm that is in use (e.g.,
   [I-D.ietf-rmcat-nada], [RFC8298], [I-D.ietf-rmcat-gcc], or [RFC8382])
   apply.

   A receiver that intentionally generates inaccurate RTCP congestion
   control feedback reports might be able trick the sender into sending
   at a greater rate than the path can support, thereby congesting the
   path. This will negatively impact the quality of experience of that
   receiver. Since RTP is an unreliable transport, a sender can
   intentionally leave a gap in the RTP sequence number space without
   causing harm, to check that the receiver is correctly reporting
   losses.

   An on-path attacker that can modify RTCP congestion control feedback
   packets can change the reports to trick the sender into sending at
   either an excessively high or excessively low rate, leading to denial
   of service. The secure RTCP profile [RFC3711] can be used to
   authenticate RTCP packets to protect against this attack.

10. References

10.1. Normative References

   [I-D.ietf-rmcat-rtp-cc-feedback]
              Perkins, C., "RTP Control Protocol (RTCP) Feedback for
              Congestion Control in Interactive Multimedia Conferences",
              draft-ietf-rmcat-rtp-cc-feedback-03 (work in progress),
              November 2016.

   [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>.

   [RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition
              of Explicit Congestion Notification (ECN) to IP",
              RFC 3168, DOI 10.17487/RFC3168, September 2001,
              <https://www.rfc-editor.org/info/rfc3168>.

   [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>.

   [RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and
              Video Conferences with Minimal Control", STD 65, RFC 3551,
              DOI 10.17487/RFC3551, July 2003, <https://www.rfc-
              editor.org/info/rfc3551>.

   [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>.

   [RFC6679] Westerlund, M., Johansson, I., Perkins, C., O'Hanlon, P.,
              and K. Carlberg, "Explicit Congestion Notification (ECN)
              for RTP over UDP", RFC 6679, DOI 10.17487/RFC6679, August
              2012, <https://www.rfc-editor.org/info/rfc6679>.

10.2. Informative References

   [feedback-requirements]
              "RMCAT Feedback Requirements",
              <://www.ietf.org/proceedings/95/slides/slides-95-rmcat-
              1.pdf>.

   [I-D.ietf-rmcat-gcc]
              Holmer, S., Lundin, H., Carlucci, G., Cicco, L., and S.
              Mascolo, "A Google Congestion Control Algorithm for Real-
              Time Communication", draft-ietf-rmcat-gcc-02 (work in
              progress), July 2016.

   [I-D.ietf-rmcat-nada]
              Zhu, X., Pan, R., Ramalho, M., Cruz, S., Jones, P., Fu,
              J., and S. D'Aronco, "NADA: A Unified Congestion Control
              Scheme for Real-Time Media", draft-ietf-rmcat-nada-04
              (work in progress), March 2017.

   [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>.

   [RFC6843] Clark, A., Gross, K., and Q. Wu, "RTP Control Protocol
              (RTCP) Extended Report (XR) Block for Delay Metric
              Reporting", RFC 6843, DOI 10.17487/RFC6843, January 2013,
              <https://www.rfc-editor.org/info/rfc6843>.

   [RFC8298] Johansson, I. and Z. Sarker, "Self-Clocked Rate Adaptation
              for Multimedia", RFC 8298, DOI 10.17487/RFC8298, December
              2017, <https://www.rfc-editor.org/info/rfc8298>.

   [RFC8382] Hayes, D., Ed., Ferlin, S., Welzl, M., and K. Hiorth,
              "Shared Bottleneck Detection for Coupled Congestion
              Control for RTP Media", RFC 8382, DOI 10.17487/RFC8382,
              June 2018, <https://www.rfc-editor.org/info/rfc8382>.

Authors' Addresses

   Zaheduzzaman Sarker
   Ericsson AB
   Luleae
   Sweden

   Phone: +46107173743
   Email: zaheduzzaman.sarker@ericsson.com

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

Email: csp@csperkins.org

Varun Singh
CALLSTATS I/O Oy
Annankatu 31-33 C 42
Helsinki 00100
Finland

Email: varun.singh@iki.fi
URI: http://www.callstats.io/

Michael A. Ramalho
Cisco Systems, Inc.
6310 Watercrest Way Unit 203
Lakewood Ranch, FL 34202
USA

Phone: +1 919 476 2038
Email: mramalho@cisco.com