Support for Reduced-Size RTCP, Opportunities and Consequences
draft-ietf-avt-rtcp-non-compound-06

Abstract

This memo discusses benefits and issues that arise when allowing RTCP packets to be transmitted with reduced size. The size can be reduced if the rules on how to create compound packets outlined in RFC3550 are removed or changed. Based on that analysis this memo defines certain changes to the rules to allow feedback messages to be sent as reduced-size RTCP packets under certain conditions when using the RTP AVPF profile (RFC 4585).
Table of Contents

1. Introduction ................................................. 3
2. Terminology ................................................. 3
3. Use Cases and Design Rationale ............................. 4
   3.1. RTCP Compound Packets (Background) .................... 4
   3.2. Use Cases for Reduced-Size RTCP ....................... 6
   3.3. Benefits of Reduced-Size RTCP ......................... 7
   3.4. Issues with Reduced-Size RTCP ......................... 8
       3.4.1. Middle Boxes .................................... 8
       3.4.2. Packet Validation ................................ 9
       3.4.3. Encryption/authentication ......................... 10
       3.4.4. RTP and RTCP Multiplex on the Same Port .......... 10
       3.4.5. Header Compression ................................ 10
4. Use of Reduced-size RTCP with AVPF ......................... 11
   4.1. Definition of Reduced-Size RTCP ....................... 11
   4.2. Algorithm Considerations ............................. 12
       4.2.1. Verification of Delivery ......................... 12
       4.2.2. Single vs Multiple RTCP in a Reduced-Size RTCP .. 13
       4.2.3. Enforcing Compound RTCP ......................... 13
       4.2.4. Immediate Mode ................................... 13
5. Signaling .................................................. 13
6. Security Considerations ................................. 14
7. IANA Considerations ..................................... 14
8. Acknowledgements ......................................... 14
9. References ............................................... 15
   9.1. Normative References .................................. 15
   9.2. Informative References ............................. 15
Authors’ Addresses ........................................ 16
Intellectual Property and Copyright Statements ............. 17
1. Introduction

In RTP [RFC3550] it is currently mandatory to send RTCP Control Protocol (RTCP) packets as compound packets containing at least a Sender Report (SR) or Receiver Report (RR), followed by a Source Description (SDES) packet containing at least the CNAME item. There are good reasons for this, as discussed below (see Section 3.1), however it does result in the minimal RTCP packets being quite large.

The RTP profile AVPF [RFC4585] specifies new RTCP packet types for feedback messages. Some of these feedback messages would benefit from being transmitted with minimal delay. AVPF does provide some mechanisms to support this, however for environments with low-bitrate links these messages can still consume a large amount of resources, and can introduce extra delay in the time it takes to completely send the compound packet in the network. It is therefore desirable to send just the feedback, without the other parts of a compound RTCP packet. This memo proposes such a mechanism, for this, and other, use cases, as discussed in Section 3.2.

The use of reduced-size RTCP is not without issues. This is discussed in Section 3.4. These issues need to be considered and are part of the motivation for this document.

Finally this document defines how AVPF is updated to allow for the transmission of reduced-size RTCP in a way that would not substantially affect the mechanisms that compound packets provide. The connection to AVPF is motivated by the fact that reduced-size RTCP is mainly beneficial for event driven feedback purposes and that the AVPF early and immediate modes make this possible.

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].

The naming convention for RTCP is often confusing. Below a list of RTCP terms and what they mean. See also section 6.1 in [RFC3550] and section 3.1 in [RFC4585] for details.

RTCP packet: Can be of different types, contains a fixed header part followed by structured elements depending on RTCP packet type.
Lower layer datagram: Can be interpreted as the UDP payload, it may however, depending on the transport be TCP or DCCP payload or something else. Synonymous to "underlying protocol" defined in 3 in [RFC3550].

Compound RTCP packet: A collection of two or more RTCP packets. A compound RTCP packet is transmitted in a lower layer datagram. It must contain at least an RTCP RR or SR packet and a SDES packet with the CNAME item. Often "compound" is left out, the interpretation of RTCP packet is therefore dependent on the context.

Minimal compound RTCP packet: A compound RTCP packet that contains the RTCP RR or SR packets and the SDES packet with the CNAME item with a specified ordering.

(Full) compound RTCP packet: A compound RTCP packet that conforms to the requirements on minimal compound RTCP packets and contains more RTCP packets.

Reduced-size RTCP packet: May contain one or more RTCP packets but does not follow the compound RTCP rules defined in section 6.1 in [RFC3550] and are thus neither a minimal or a full compound RTCP. See Section 4.1 for a full definition.

3. Use Cases and Design Rationale

3.1. RTCP Compound Packets (Background)

Section 6.1 in [RFC3550] specifies that an RTCP packet must be sent as a compound RTCP packet consisting of at least two individual RTCP packets, first an Sender Report (SR) or Receiver Report (RR), followed by additional packets including a mandatory SDES packet containing a CNAME Item for the transmitting source identifier (SSRC). Below is a short description what these RTCP packet types are used for.

1. The sender and receiver reports (see Section 6.4 of [RFC3550]) provides the RTP session participant with the Synchronisation Source (SSRC) Identifier of all RTP session participants. Having all participants send these packets periodically allows everyone to determine the current number of participants. This information is used in the transmission scheduling algorithm. Thus this is particularly important for new participants so that they quickly can establish a good estimate of the group size. Failure to do this would result in RTCP senders consuming too much bandwidth.
2. Before a new session participant has sent any RTP or RTCP packet, it can also avoid SSRC collisions with all the SSRCs it sees prior to that transmission. So the possibility to see a substantial amount of the participating sources minimizes the risk of any collision when selecting SSRC.

3. The sender and receiver reports contain some basic statistics usable for monitoring of the transport and thus enable adaptation. These reports become more useful if sent regularly as the receiver of a report can perform analysis to find trends between the individual reports. When used for media transmission adaptation the information become more useful the more frequently it is received, at least until one report per round-trip time (RTT) is achieved. Therefore there are, in most cases, no reason to not include the sender or receiver report in all RTCP packets.

4. The CNAME SDES item (See Section 6.5.1 of [RFC3550]) exists to allow receivers to determine which media flows that should be synchronized with each other, both within an RTP session and between different RTP sessions carrying different media types. Thus it is important to quickly receive this for each media sender in the session when joining an RTP session.

5. Sender Reports (SR) is used in combination with the above SDES CNAME mechanism to synchronize multiple RTP streams, such as audio and video. After having determined which media streams should be synchronized using the CNAME field, the receiver uses the Sender Report’s NTP and RTP timestamp fields to establish synchronization.

6. The CNAME SDES item also allows a session participant to detect SSRC collisions and separate them from routing loops. The 32-bit randomly selected SSRC has some probability of collisions. The CNAME is used as longer canonical identifier of an particular end-point instance that is bound to an SSRC. If that binding isn’t received and being current the receiver may not detect a SSRC collision, i.e. two different CNAMEs uses the same SSRC. It also can’t detect a RTP level routing loop resulting in that the same SSRC and CNAME arrives from multiple lower-layer source addresses.

Reviewing the above it is obvious that both SR/RR and the CNAME are very important for new session participants to be able to utilize any received media and to avoid flooding the network with RTCP reports. In addition, if not sent regularly the dynamic nature of the information provided would make it less useful.

The following sections will describe the cases when reduced-size RTCP
is beneficial and also show the possible issues that must be considered.

3.2. Use Cases for Reduced-Size RTCP

Below are listed a few use cases for reduced-size RTCP.

Control Plane Signaling: Open Mobile Alliance (OMA) Push-to-talk over Cellular (PoC) [OMA-PoC] makes use of reduced-size RTCP when transmitting certain events. The OMA POC service is primarily used over cellular links capable of IP transport, such as the GSM GPRS.

Codec Control Signaling: An example that can be used with reduced-size RTCP is e.g TMMBR messages as specified in [RFC5104] which signal a request for a change in codec bitrate. The benefit of its use for these messages is in bad channel conditions as reduced-size RTCP are much more likely to be successfully transmitted than larger compound RTCP. This is critical as these messages are likely to occur when channel conditions are poor. Other examples of codec control usage for reduced-size RTCP are found in [MTSI-3GPP]

Feedback: An example of a feedback scenario that would benefit from reduced-size RTCP is Video streams with generic NACK. In cases where the RTT is shorter than the receiver buffer depth, generic NACK can be used to request retransmission of missing packets, thus improving playout quality considerably. If the generic NACK packets are transmitted as reduced-size RTCP, the bandwidth requirement for RTCP will be minimal, enabling more frequent feedback. Like in the Codec control case it is important that these packets can be transmitted with as little delay as possible. Another interesting use for reduced-size RTCP is in cases when regular feedback is needed, as described in Section 3.3.

Status Reports: One proposed idea is to transmit small measurement or status reports in reduced-size RTCP, and to be able to split the minimal compound RTCP and transmit the individual RTCP separately. The status reports can be used either by the endpoints or by other network monitoring boxes in the network. The benefit is that with some radio access technologies small packets are more robust to poor radio conditions than large packets. Additionally, with small (report) packets there is a smaller risk that the report packets will affect the channel that they report upon. Another benefit is that it is, with reduced-size RTCP, possible to allow e.g anonymous status reporting to be transmitted unencrypted. Something that may be beneficial for e.g network monitoring purposes.
3.3. Benefits of Reduced-Size RTCP

As mentioned in the introduction, most advantages of using reduced-size RTCP packets exists in cases when the available RTCP bitrate is limited. This because they can become substantially smaller than compound packets. A compound packet is forced to contain both an RR or an SR and the CNAME SDES item. The RR containing a report block for a single source is 32 bytes, an SR is 52 bytes. Both may be larger if they contain report blocks for multiple sources. The SDES packet containing a CNAME item will be 10 bytes plus the CNAME string length. Here it is reasonable that the CNAME string is at least 10 bytes to get a decent collision resistance. If the recommended form of user@host is used, then most strings will be longer than 20 characters. Thus a reduced-size RTCP can become at least 70-80 bytes smaller than the compound packet.

For low bitrate links the benefits of this reduction in size are as follows:

- For links where the packet loss rate grows with the packet size, smaller packets are less likely to be dropped. An example of such links are radio links. In the cellular world there exist links that are optimized to handle RTP packets sized for carrying compressed speech. This increases the capacity and coverage for voice services in a given wireless network. Minimal compound RTCP packets are commonly 2-3 times the size of a RTP packet carrying compressed speech. If the speech packet over such a bearer has a packet loss probability of p, then the RTCP packet will experience a loss probability of \(1-(1-p)^x\) where \(x\) is the number of fragments the compound packet will be split on the link layer, i.e. commonly into 2 or 3 fragments.

- Shorter serialization time, i.e the time it takes the link to transmit the packet. For slower links this time can be substantial. For example transmitting 120 bytes over an link interface capable of 30 kbps takes 32 milliseconds (ms) assuming uniform transmission rate.

In cases when reduced-size RTCP carry important and time sensitive feedback, both shorter serialization time and the lower loss probability are important to enable the best possible functionality. Having a packet loss rate that is much higher for the feedback packets compared to media packets hurts when trying to perform media adaptation, to for example handle the changed performance present at the cell border in a cellular system.

For high bitrate applications there is usually no problem to supply RTCP with sufficient bitrates. When using AVPF one can use the "trr-
int” parameter to restrict the regular reporting interval to approximately once per RTT or less often. As in most cases there is little reason to provide with regular reports of higher density than this. Any additional bandwidth can then be used for feedback messages. The benefit of reduced-size RTCP in this case is limited, but exists. One typical example is video using generic NACK in cases where the RTT is low. Using reduced-size RTCP would reduce the total amount of bits used for RTCP. This is primarily applicable if the number of reports is large. This would also result in lower processing delay and less complexity for the feedback packets as they do not need to query the RTCP database to construct the right messages.

As message size is generally a smaller issue at higher bitrates, it is also possible to transmit multiple RTCP in each lower layer datagram in these cases. The motivation behind reduced-size RTCP in this case is not size, rather it is to avoid the extra overhead caused by inclusion of the SR/RR and SDES CNAME items in each transmitted RTCP.

Independently of the link type there are additional benefits with sending feedback in small reduced-size RTCP. Applications that use RTCP AVPF in early or immediate mode to send frequent event driven feedback. Under these circumstances, the risk that the RTCP bandwidth becomes too high during periods of heavy feedback signaling is reduced.

In cases when regular feedback is needed, such as the profile under development for TCP friendly rate control (TFRC) for RTP [I-D.ietf-avt-tfrc-profile], the size of compound RTCP can result in very high bandwidth requirements if the round trip time is short. For this particular application reduced-size RTCP gives a very substantial improvement.

### 3.4. Issues with Reduced-Size RTCP

This section describes the known issues with reduced-size RTCP and also a brief analysis.

#### 3.4.1. Middle Boxes

Middle boxes in the network may discard RTCP that do not follow the rules outlined in section 6.1 of RFC3550. Newer report types may be interpreted as unknown by the middle box. For instance if the payload type number is 207 instead of 200 or 201 it may be treated as unknown. The effect of this might for instance be that compound RTCP would get through while the reduced-size RTCP would be lost.
3.4.2. Packet Validation

A reduced-size RTCP packet will be discarded by the packet validation code in Appendix A of [RFC3550]

Weakened Packet Validation: The packet validation code needs to be rewritten to accept reduced-size RTCP. This in particular affects section 9.1 in [RFC3550] in the sense that the header verification must take into account that the payload type numbers for the (first) RTCP in the lower layer datagram may differ from 200 or 201 (SR or RR). One potential effect of this change is much weaker validation that received packets actually are RTCP, and not packets of some other type being wrongly delivered. Thus some consideration should be done to ensure the best possible validation is available. For example restricting reduced-size RTCP to contain only some specific RTCP packet types, that is preferably signalled on a per-session basis. However, the application of a security mechanisms for source authentication on the packets will provide much stronger protection.

Old RTP Receivers: Any RTCP receiver without updated packet validation code will discard the reduced-size RTCP which means that the receiver will not see e.g the contained feedback messages. The effect of this depends on the type of feedback message and the role of the receiver. For example this may cause complete function loss in the case of attempting to use a reduced size NACK message (see Section 6.2.1 of [RFC4585]) to non updated media sender in a session using the retransmission scheme defined by [RFC4588]. This type of discarding would also effect the feedback suppression defined in AVPF. The result would be a partitioning of the receivers within the session between old ones only seeing the compound RTCP feedback messages and the newer ones seeing both. Where the old ones may send feedback messages for events already reported on in reduced-size RTCP.

Bandwidth Considerations: The discarding of reduced-size RTCP would effect the RTCP transmission calculation in the following way: the avg_rtcp_size value would become larger than for RTP receivers that exclude the reduced-size RTCP in this calculation (assuming that reduced-size RTCP are smaller than compound ones). Therefore these senders would under-utilize the available bitrate and send with a longer interval than updated receivers. For most sessions this should not be an issue. However for sessions with a large portion of reduced-size RTCP may result in that the updated receivers time out non-updated senders prematurely. This is not
likely to occur as the time between RTCP transmission needs to become 5 times that used by the reduced-sized senders when sending compound.

Computation of avg_rtcp_size: Long intervals between compound RTCP and many reduced-size RTCP in between may lead to a computation of a value for avg_rtcp_size that varies greatly over time. Investigation shows that although it varies this is not enough of a problem to warrant further changes or complexities to the RTCP scheduling algorithm.

3.4.3. Encryption/authentication

SRTP presents a problem for reduced-size RTCP. Section 3.4 in [RFC3711] states "SRTCP MUST be given packets according to that requirement in the sense that the first part MUST be a sender report or a receiver report".

Upon examination of how SRTP process packets it becomes obvious that SRTP has no real dependency on that the first packet is either an SR or an RR packet. What is needed is the common RTCP packet header, which is present in all the packet types, with a source SSRC. The conclusion is therefore that it is possible to use reduced-size RTCP with SRTP.

3.4.4. RTP and RTCP Multiplex on the Same Port

In applications which multiplex RTP and RTCP on the same port, as defined in [I-D.ietf-avt-rtp-and-rtcp-mux], care must be taken to ensure that the de-multiplexing is done properly even though RTCP are reduced size. The downside of reduced size RTCP is that more values representing RTCP packets exist, reducing the available RTP payload type space. However, section 4 in [I-D.ietf-avt-rtp-and-rtcp-mux] already requires the corresponding RTP payload type range not be used when performing this multiplexing.

3.4.5. Header Compression

Two issues are related to header compression, possible changes are left for future work:

- Payload type number identification: The RoHC header compression algorithm [RFC3095] needs to create different compression contexts for RTP and RTCP for optimum performance. If RTP and RTCP are multiplexed on the same port the classification may be based on payload type numbers. The classification algorithm must here acknowledge the fact that the payload type number for (the first) RTCP may differ from 200 or 201.
4. Use of Reduced-size RTCP with AVPF

Based on the above analysis it seems feasible to allow transmission of reduced-size RTCP under some restrictions:

- First of all it is important that compound RTCP are transmitted at regular intervals to ensure that the mechanisms maintained by the compound packets, like feedback reporting works. The tracking of session size and number of participants warrants mentioning again as this ensures that the RTCP bandwidth remain bounded independent of the number of session participants.

- Second, as the compound RTCP are also used to establish and maintain synchronization between media, any newly joining participant in a session would need to receive compound RTCP from the media sender(s).

This implies that the regular transmission of compound RTCP MUST be maintained throughout an RTP session. Reduced-size RTCP should be restricted to be used as extra RTCP (e.g feedback) sent in cases when a regular compound RTCP packet would not otherwise have been sent.

The usage of reduced-size RTCP SHALL only be done in RTP sessions operating in AVPF [RFC4585] Early or Immediate mode. Reduced-size RTCP SHALL NOT be sent until at least one compound RTCP has been sent. In Immediate mode all feedback messages MAY be sent as reduced-size RTCP. In early mode a feedback message scheduled for transmission as an Early RTCP, i.e not a Regular RTCP, MAY be sent as reduced-size RTCP. All RTCP that are scheduled for transmission as Regular RTCP SHALL be sent as compound RTCP as indicated by AVPF [RFC4585].

4.1. Definition of Reduced-Size RTCP

A reduced-size RTCP packet is an RTCP packet with the following properties that makes it deviate from the compound RTCP packet definition given in section 6.1 in [RFC3550]:

- Contains one or more RTCP packet(s)

- Any RTCP packet type allowed
Internet-Draft      Reduced size RTCP in RTP profile           July 2008

o MUST NOT be used for regular (scheduled) RTCP report purposes

o MUST NOT be used with the RTP/AVP profile [RFC3551].

4.2. Algorithm Considerations

4.2.1. Verification of Delivery

If an application is to use reduced-size RTCP it is important to verify that the reduced-size RTCP packets actually reach the session participants. As outlined above in Section 3.4.1 and Section 3.4.2 packets may be discarded along the path or in the end-point.

A few verification rules are RECOMMENDED to ensure robust RTCP transmission and reception and to solve the identified issues when reduced-size RTCP is used:

o The end-point issue can be solved by introducing signaling that informs if all session participants are capable of reduced-size RTCP. See Section 5.

o The middle box issue is more difficult and here one will be required to use heuristics to determine if the reduced-size RTCP are delivered or not. However in many cases the feedback messages sent using reduced-size RTCP will result in either explicit or implicit indications that they have been received. An example of is the RTP retransmission [RFC4588] that results from a NACK message [RFC4585]. Another example is the Temporary Maximum Media Bitrate Notification message resulting from a Temporary Maximum Media Bitrate Request [RFC5104]. A third example is the presence of a Decoder Refresh Point [RFC5104] in the video media stream resulting from the Full Intra Request sent.

o An algorithm to detect consistent failure of delivery of reduced-size RTCP must be used by any application using it. The details of this algorithm is application dependent and therefore outside the scope of this document.

o A method to detect if reduced-size RTCP are discarded is to send a single SR packet in a lower layer datagram, then check that the timestamp is echoed back in the corresponding RR packet. This verification method is not completely safe however as it SR is still one of the expected packet types.

If the verification fails it is strongly RECOMMENDED that only compound RTCP according to the rules outlined in RFC3550 is transmitted.
4.2.2. Single vs Multiple RTCP in a Reduced-Size RTCP

The result of the definition in Section 4.1 may be that the resulting size of reduced-size RTCP can become larger than a normal compound RTCP. For applications that use access types that are sensitive to packet size (see Paragraph 2 in Section 3.3) it is strongly RECOMMENDED that the use of reduced-size RTCP is limited to the transmission of single RTCP in each lower layer datagram. The methods to determine the need for this is outside the scope of this draft.

In general, as the benefit with large sized reduced-size RTCP packets is very limited, it is strongly RECOMMENDED to transmit large reduced-size RTCP packets as compound RTCP packets instead.

4.2.3. Enforcing Compound RTCP

As discussed earlier it is important that the transmission of compound RTCP occurs at regular intervals. However, this will occur as long as the RTCP senders follow the AVPF scheduling algorithm defined in Section 3.5 in [RFC4585]. This as all regular RTCP MUST be full compound RTCP. Note that also in immediate mode is there a requirement on sending regular RTCP.

4.2.4. Immediate Mode

Section 3.3 in RFC4585 gives the option to use AVPF Immediate mode as long as the groupsize is below a certain limit. As transmission using reduced-size RTCP may reduce the bandwidth demand it opens up for a more liberal use of immediate mode.

5. Signaling

This document defines the "a=rtcp-rsize" SDP [RFC4566] attribute to indicate if the session participant is capable of supporting reduced-size RTCP for applications that uses SDP for configuration of RTP sessions. It is required that a participant that proposes the use of reduced-size RTCP itself supports the reception of reduced-size RTCP.

An offering client that wish to use reduced-size RTCP MUST include the attribute "a=rtcp-rsize" in the SDP offer. If "a=rtcp-rsize" is present in the offer SDP, the answerer that supports reduced-size RTCP and wish to use it SHALL include the "a=rtcp-rsize" attribute in the answer.

In declarative usage such as RTSP [RFC2326] and SAP [RFC2974] of SDP the presence of the attribute indicates that the session participant...
MAY use reduced size RTCP packets in its RTCP transmissions.

6. Security Considerations

The security considerations of RTP [RFC3550] and AVPF [RFC4585] will apply also to reduced-size RTCP. The reduction in validation strength for received packets on the RTCP port may result in a higher degree of acceptance of spurious data as real RTCP. This vulnerability can mostly be addressed by usage of any security mechanism that provide authentication, one example such mechanism is SRTP [RFC3711].

7. IANA Considerations

Following the guidelines in [RFC4566], the IANA is requested to register one new SDP attribute:

- Contact name, email address and telephone number: Authors of RFCXXXX

- Attribute-name: rtcp-rsize

- Long-form attribute name: Reduced-size RTCP

- Type of attribute: media-level

- Subject to charset: no

This attribute defines the support for reduced-size RTCP, i.e the possibility to transmit RTCP that does not conform to the rules for compound RTCP defined in RFC3550. It is a property attribute, which does not take a value.

Note to RFC Editor: please replace "RFC XXXX" above with the RFC number of this memo, and remove this note.

8. Acknowledgements

The authors would like to thank all the people who gave feedback on this document.

This document also contain some text copied from [RFC3550], [RFC4585] and [RFC3711]. We take the opportunity to thank the authors of said documents.
9. References

9.1. Normative References


9.2. Informative References


Authors’ Addresses

Ingemar Johansson
Ericsson AB
Laboratoriegrand 11
SE-971 28 Lulea
SWEDEN

Phone: +46 73 0783289
Email: ingemar.s.johansson@ericsson.com

Magnus Westerlund
Ericsson AB
Faeroegatan 6
SE-164 80 Stockholm
SWEDEN

Phone: +46 8 7190000
Email: magnus.westerlund@ericsson.com