AVTCORE J. Lennox
Internet-Draft Vidyo
Updates: 3550 (if approved) M. Westerlund
Intended status: Standards Track Ericsson
Expires: April 25, 2013 October 22, 2012

Real-Time Transport Protocol (RTP) Considerations for Endpoints Sending Multiple Media Streams
draft-lennox-avtcore-rtp-multi-stream-01

Abstract

This document expands and clarifies the behavior of the Real-Time Transport Protocol (RTP) endpoints when they are sending multiple media streams in a single RTP session. In particular, issues involving Real-Time Transport Control Protocol (RTCP) messages are described.

Table of Contents

1. Introduction

At the time The Real-Time Tranport Protocol (RTP) [RFC3550] was originally written, and for quite some time after, endpoints in RTP sessions typically only transmitted a single media stream per RTP session, where separate RTP sessions were typically used for each distinct media type.

Recently, however, a number of scenarios have emerged (discussed further in Section 3) in which endpoints wish to send multiple RTP media streams, distinguished by distinct RTP synchronization source (SSRC) identifiers, in a single RTP session. Although RTP's initial design did consider such scenarios, the specification was not consistently written with such use cases in mind. The specifications are thus somewhat unclear.

The purpose of this document is to expand and clarify [RFC3550]'s language for these use cases. The authors believe this does not result in any major normative changes to the RTP specification, however this document defines how the RTP specification shall be interpreted. In these cases, this document updates RFC3550.

The document starts with terminology and some use cases where multiple sources will occur. This is followed by some case studies to try to identify issues that exist and need considerations. This is followed by RTP and RTCP recommendations to resolve issues. Next are security considerations and remaining open issues.

2. Terminology

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 RFC 2119 [RFC2119] and indicate requirement levels for compliant implementations.

3. Use Cases For Multi-Stream Endpoints

This section discusses several use cases that have motivated the development of endpoints that send multiple streams in a single RTP session.

3.1. Multiple-Capturer Endpoints

The most straightforward motivation for an endpoint to send multiple media streams in a session is the scenario where an endpoint has

multiple capture devices of the same media type and characteristics. For example, telepresence endpoints, of the type described by the CLUE Telepresence Framework [I-D.ietf-clue-framework] is designed, often have multiple cameras or microphones covering various areas of a room.

3.2. Multi-Media Sessions

Recent work has been done in RTP [I-D.ietf-avtcore-multi-media-rtp-session] and SDP [I-D.ietf-mmusic-sdp-bundle-negotiation] to update RTP's historical assumption that media streams of different media types would always be sent on different RTP sessions. In this work, a single endpoint's audio and video media streams (for example) are instead sent in a single RTP session.

3.3. Multi-Stream Mixers

There are several RTP topologies which can involve a central box which itself generates multiple media streams in a session.

One example is a mixer providing centralized compositing for a multi-capturer scenario like the one described in Section 3.1. In this case, the centralized node is behaving much like a multi-capturer endpoint, generating several similar and related sources.

More complicated is the Source Projecting Mixer, see Section 3.6 [I-D.westerlund-avtcore-rtp-topologies-update], which is a central box that receives media streams from several endpoints, and then selectively forwards modified versions of some of the streams toward the other endpoints it is connected to. Toward one destination, a separate media source appears in the session for every other source connected to the mixer, "projected" from the original streams, but at any given time many of them may appear to be inactive (and thus receivers, not senders, in RTP). This box is an RTP mixer, not an RTP translator, in that it terminates RTCP reporting about the mixed streams, and it can re-write SSRCs, timestamps, and sequence numbers, as well as the contents of the RTP payloads, and can turn sources on and off at will without appearing to be generating packet loss. Each projected stream will typically preserve its original RTCP source description (SDES) information.

4. Issue Cases

This section tries to illustrate a few cases that have been determined to cause issues.

4.1. Cascaded Multi-party Conference with Source Projecting Mixers

This issue case tries to illustrate the effect of having multiple SSRCs sent by an endpoint, by considering the traffic between two source-projecting mixers in a large multi-party conference.

For concreteness, consider a 200-person conference, where 16 sources are viewed at any given time. Assuming participants are distributed evenly among the mixers, each mixer would have 100 sources "behind" (projected through) it, of which at any given time eight are active senders. Thus, the RTP session between the mixers consists of two endpoints, but 200 sources.

The RTCP bandwidth implications of this scenario are discussed further in Section 8.4.

(TBD: Other examples?)

5. Multi-Stream Endpoint RTP Media Recommendations

While an endpoint MUST (of course) stay within its share of the available session bandwidth, as determined by signalling and congestion control, this need not be applied independently or uniformly to each media stream. In particular, session bandwidth MAY be reallocated among an endpoint's media streams, for example by varying the bandwidth use of a variable-rate codec, or changing the codec used by the media stream, up to the constraints of the session's negotiated (or declared) codecs. This includes enabling or disabling media streams as more or less bandwidth becomes available.

6. Multi-Stream Endpoint RTCP Recommendations

This section contains a number of different RTCP clarifications or recommendations that enables more efficient and simpler behavior without loss of functionality.

The Real-Time Transport Control Protocol (RTCP) is defined in Section 6 of [RFC3550], but it is largely documented in terms of "participants". In many cases, the specification's recommendations for "participants" should be interpreted as applying to individual media streams, rather than to endpoints. This section describes several concrete cases where this applies.

6.1. RTCP Reporting Requirement

For each of an endpoint's media media streams, whether or not it is currently sending media, SR/RR and SDES packets MUST be sent at least once per RTCP report interval. (For discussion of the content of SR or RR packets' reception statistic reports, see Section 8.)

6.2. Initial Reporting Interval

When a new media stream is added to a unicast session, the sentence in [RFC3550]'s Section 6.2 applies: "For unicast sessions ... the delay before sending the initial compound RTCP packet MAY be zero." This applies to individual media sources as well. Thus, endpoints MAY send an initial RTCP packet for an SSRC immediately upon adding it to a unicast session.

This allowance also applies, as written, when initially joining a unicast session. However, in this case some caution should be excersied if the end-point or mixer has a large number of sources (SSRCs) as this can create a significant burst. How big an issue this depends on the number of source to send initial SR or RR and Session Description CNAME items for in relation to the RTCP bandwidth. TBD: Maybe some recommendation here?

6.3. Compound RTCP Packets

Section 6.1 gives the following advice to RTP translators and mixers:

> It is RECOMMENDED that translators and mixers combine individual RTCP packets from the multiple sources they are forwarding into one compound packet whenever feasible in order to amortize the packet overhead (see Section 7). An example RTCP compound packet as might be produced by a mixer is shown in Fig. 1. If the overall length of a compound packet would exceed the MTU of the network path, it SHOULD be segmented into multiple shorter compound packets to be transmitted in separate packets of the underlying protocol. This does not impair the RTCP bandwidth estimation because each compound packet represents at least one distinct participant. Note that each of the compound packets MUST begin with an SR or RR packet.

Note: To avoid confusion, an RTCP packet is an individual item, such as a Sender Report (SR), Receiver Report (RR), Source Description (SDES), Goodbye (BYE), Application Defined (APP), Feedback [RFC4585] or Extended Report (XR) [RFC3611] packet. A compound packet is the combination of two or more such RTCP packets where the first packet must be an SR or an RR packet, and which contains a SDES packet containing an CNAME item. Thus the above results in compound RTCP

packets that contain multiple SR or RR packets from different sources as well as any of the other packet types. There are no restrictions on the order the packets may occur within the compound packet, except the regular compound rule, i.e. starting with an SR or RR.

This advice applies to multi-media-stream endpoints as well, with the same restrictions and considerations. (Note, however, that the last sentence does not apply to AVPF [RFC4585] or SAVPF [RFC5124] feedback packets if Reduced-Size RTCP [RFC5506] is in use.)

Due to RTCP's randomization of reporting times, there is a fair bit of tolerance in precisely when an endpoint schedules RTCP to be sent. Thus, one potential way of implementing this recommendation would be to randomize all of an endpoint's sources together, with a single randomization schedule, so an MTU's worth of RTCP all comes out simultaneously.

TBD: Multiplexing RTCP packets from multiple different sources may require some adjustment to the calculaton of RTCP's avg_rtcp_size, as the RTCP group interval is proportional to avg_rtcp_size times the group size.

7. RTCP Bandwidth Considerations When Sources have Greatly-Differing Bandwidths

it is possible for an RTP session to carry sources of greatly differing bandwidths. One example is the scenario of [I-D.ietf-avtcore-multi-media-rtp-session], when audio and video are sent in the same session. However, this can occur even within a single media type, for example a video session carrying both 5 fps QCIF and 60 fps 1080p HD video, or an audio session carrying both G.729 and L24/48000/6 audio.

TBD: recommend how RTCP bandwidths should be chosen in these scenarios. Likely, these recommendations will be different for sessions using AVPF-based profiles (where the trr-int parameter is available) than for those using AVP.

8. Grouping of RTCP Reception Statistics and Other Feedback

As required by [RFC3550], an endpoint MUST send reception reports about every active media stream it is receiving, from at least one local source.

However, a naive application of the RTP specification's rules could be quite inefficient. In particular, if a session has N media

sources (active and inactive), and has S senders in each reporting interval, there will either be N*S report blocks per reporting interval, or (per the round-robinning recommendations of [RFC3550] Section 6.1) reception sources would be unnecessarily round-robinned. In a session where most media sources become senders reasonably frequently, this results in quadratically many reception report blocks in the conference, or reporting delays proportional to the number of session members.

Since traffic is received by endpoints, however, rather than by media sources, there is not actually any need for this quadratic expansion. All that is needed is for each endpoint to report all the remote sources it is receiving.

Thus, this document defines a new RTCP mechanism, Reporting Groups, to indicate sources which originate from the same endpoint, and which therefore would have identical reecption reports.

8.1. Semantics and Behavior of Reporting Groups

An RTCP Reporting Group indicates that a set of sources originate from a single entity in an RTP session, and therefore all the sources in the group's view of the network is identical. Typically, a Reporting Group corresponds to a physical entity in the network.

If reporting groups are in use, an endpoint MUST NOT send reception reports from one source in a reporting group about another one in the same group ("self-reports"). Similarly, an endpoint MUST NOT send reception reports about a remote media source from more than one sources in a reporting group ("cross-reports"). Instead, it MUST pick one of its local media sources as the "reporting" source for each remote media source, and use it to send reception reports for that remote source; all its other media sources MUST NOT send any reception reports for that remote media source.

An endpoint MAY choose different local media sources as the reporting source for different remote media sources (for example, it could choose to send reports about remote audio sources from a local audio source, and reports about remote video sources from a local video source), or it MAY choose a single local source for all its reports. This reporting source MUST also be the source for any AVPF Feedback [RFC4585] or Extended Report (XR) [RFC3611] packets about the corresponding remote sources as well. If a reporting source leaves the session (i.e., if it sends a BYE, or leaves the group without sending BYE under the rules of [RFC3550] section 6.3.7), another reporting source MUST be chosen, if the sources it was reporting on are still in the session.

If AVPF feedback is in use, a reporting source MAY send immediate or early feedback at any point when any member of the reporting group could validly do so.

An endpoint SHOULD NOT create single-source reporting groups, unless it is anticipated that the group might have additional sources added to it in the future.

8.2. RTCP Source Description (SDES) item for Reporting Groups

A new Source Description (SDES) item, "RGRP", indicates that a sources is a member of a specified reporting group. Syntactically, its format is the same as the RTCP CNAME [RFC6222], and MUST be chosen with the same global-uniqueness and privacy considerations as CNAME.

Every source which belongs to a reporting group MUST include an RGRP SDES item in an SDES packet, alongside its CNAME, in every compound RTCP packet in which it sends an RR or SR packet. (I.e., in every RTCP packet it sends, unless Reduced-Size RTCP [RFC5506] is in use.)

8.3. SDP signaling for Reporting Groups

TBD

8.4. Bandwidth Benefits of RTCP Reporting Groups

To understand the benefits of RTCP reporting groups, consider the scenario described in Section 4.1. This scenario describes an environment in which the two endpoints in a session each have a hundred sources, of which eight each are sending within any given reporting interval.

For ease of analysis, we can make the simplifying approximation that the duration of the RTCP reporting interval is equal to the total size of the RTCP packets sent during an RTCP interval, divided by the RTCP bandwidth. (This will be approximately true in scenarios where the bandwidth is not so high that the minimum RTCP interval is reached.) For further simplification, we can assume RTCP senders are following the recommendations of Section 6.3; thus, the per-packet transport-layer overhead will be small relative to the RTCP data. Thus, only the actual RTCP data itself need be considered.

In a report interval in this scenario, there will, as a baseline, be 200 SDES packets, 184 RR packets, and 16 SR packets. This amounts to approximately 6.5 kB of RTCP per report interval, assuming 16-byte CNAMEs and no other SDES information.

Using naive everyone-reports-on-every-sender feedback rules, each of the 184 receivers will send 16 report blocks, and each of the 16 senders will send 15. This amounts to approximately 76 kB of report block traffic per interval; 92% of RTCP traffic consists of report blocks.

If reporting groups are used, however, there is only 0.4 kB of reports per interval, with no loss of useful information. Additionally, there will be (assuming 16-byte RGRPs as well) an additional 3.2 kB per cycle of RGRP SDES items. Put another way, the naive case's reporting interval is approximately 7.5 times longer than if reporting groups are in use.

8.5. Consequences of RTCP Reporting Groups

The RTCP traffic generated by receivers using RTCP Reporting Groups might appear, to observers unaware of these semantics, to be generated by receivers who are experiencing a network disconnection, as the non-reporting sources appear not to be receiving a given sender at all.

This could be a potentially critical problem for such a sender useing RTCP for congestion control, as such a sender might think that it is sending so much traffic that it is causing complete congestion collapse.

However, such an interpretation of the session statistics would require a fairly sophisticated RTCP analysis. Any receiver of RTCP statistics which is just interested in information about itself needs to be prepared that any given reception report might not contain information about a specific media source, because reception reports in large conferences can be round-robined.

Thus, it is unclear to what extent such backward compatibility issues would actually cause trouble in practice.

9. Security Considerations

In the secure RTP protocol (SRTP) [RFC3711], the cryptographic context of a compound SRTCP packet is the SSRC of the sender of the first RTCP (sub-)packet. This could matter in some cases, especially for keying mechanisms such as Mikey [RFC3830] which use per-SSRC keying.

Other than that, the standard security considerations of RTP apply; sending multiple media streams from a single endpoint does not appear to have different security consequences than sending the same number

of streams.

10. Open Issues

At this stage this document contains a number of open issues. The below list tries to summarize the issues:

1. Further clarifications on how to handle the RTCP scheduler when sending multiple sources in one compound packet.
2. How should the use of reporting groups be signaled in SDP?
3. How should the RTCP avg_rtcp_size be calculated when RTCP packets are routinely multiplexed among multiple RTCP senders?
4. Do we need to provide a recommendation for unicast session joiners with many sources to not use 0 initial minimal interval from bit-rate burst perspective?

11. IANA Considerations

This document adds an additional SDES type to the IANA "RTCP SDES Item Types" Registry, as follows:

Value	Abbrev	Name	Reference
TBD	RGRP	Reporting Group	[RFCXXXX]

Figure 1: Initial Contents of IANA Source Attribute Registry

(Note to the RFC-Editor: please replace "TBD" with the IANA-assigned value, and "XXXX" with the number of this document, prior to publication as an RFC.)

12. References

12.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[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, July 2006.

[RFC5124] Ott, J. and E. Carrara, "Extended Secure RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/SAVPF)", RFC 5124, February 2008.

[RFC5506] Johansson, I. and M. Westerlund, "Support for Reduced-Size Real-Time Transport Control Protocol (RTCP): Opportunities and Consequences", RFC 5506, April 2009.

[RFC6222] Begen, A., Perkins, C., and D. Wing, "Guidelines for Choosing RTP Control Protocol (RTCP) Canonical Names (CNAMEs)", RFC 6222, April 2011.

12.2. Informative References

[I-D.ietf-avtcore-multi-media-rtp-session]
Westerlund, M., Perkins, C., and J. Lennox, "Multiple Media Types in an RTP Session", draft-ietf-avtcore-multi-media-rtp-session-01 (work in progress), October 2012.

[I-D.ietf-clue-framework]
Romanow, A., Duckworth, M., Pepperell, A., and B. Baldino, "Framework for Telepresence Multi-Streams", draft-ietf-clue-framework-07 (work in progress), October 2012.

[I-D.ietf-mmusic-sdp-bundle-negotiation]
Holmberg, C. and H. Alvestrand, "Multiplexing Negotiation Using Session Description Protocol (SDP) Port Numbers", draft-ietf-mmusic-sdp-bundle-negotiation-01 (work in progress), August 2012.

[I-D.westerlund-avtcore-rtp-topologies-update]
Westerlund, M. and S. Wenger, "RTP Topologies", draft-westerlund-avtcore-rtp-topologies-update-01 (work in progress), October 2012.

[RFC3611] Friedman, T., Caceres, R., and A. Clark, "RTP Control Protocol Extended Reports (RTCP XR)", RFC 3611, November 2003.

[RFC3830] Arkko, J., Carrara, E., Lindholm, F., Naslund, M., and K. Norrman, "MIKEY: Multimedia Internet KEYing", RFC 3830, August 2004.

Appendix A. Changes From Earlier Versions

Note to the RFC-Editor: please remove this section prior to publication as an RFC.

A.1. Changes From Draft -00

- Added the Reporting Group semantic to explicitly indicate which sources come from a single endpoint, rather than leaving it implicit.
- Specified that Reporting Group semantics (as they now are) apply to AVPF and XR, as well as to RR/SR report blocks.
- Added a description of the cascaded source-projecting mixer, along with a calculation of its RTCP overhead if reporting groups are not in use.
- Gave some guidance on how the flexibility of RTCP randomization allows some freedom in RTCP multiplexing.
- Clarified the language of several of the recommendations.
- Added an open issue discussing how avg_rtcp_size should be calculated for multiplexed RTCP.
- Added an open issue discussing RTCP bandwidths should be chosen for sessions where source bandwidths greatly differ.

Authors' Addresses

Jonathan Lennox
Vidyo, Inc.
433 Hackensack Avenue
Seventh Floor
Hackensack, NJ 07601
US

Email: jonathan@vidyo.com

Magnus Westerlund
Ericsson
Farogatan 6
SE-164 80 Kista
Sweden

Phone: +46 10 714 82 87
Email: magnus.westerlund@ericsson.com