Rapid Synchronisation of RTP Flows
draft-perkins-avt-rapid-rtp-sync-01.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on July 16, 2009.

Copyright Notice

Copyright (c) 2009 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 (http://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.
Abstract

This memo outlines how RTP multimedia sessions are synchronised, and discusses how rapidly such synchronisation can occur. We show that most RTP sessions can be synchronised immediately, but that the use of video switching MCUs or large multicast (SSM) groups can greatly increase the initial synchronisation delay. This increased delay can be unacceptable to some applications that use layered and/or multi-description codecs.

This memo changes to the RTCP timing rules to reduce the initial synchronisation delay for SSM sessions. A new RTP/AVPF feedback packet is defined to allow video switching MCUs to request rapid resynchronisation, and a new RTP header extension is defined to support rapid synchronisation for late joiners.

Table of Contents

1. Introduction ................................................. 3
2. Synchronisation of RTP Flows ............................. 3
   2.1. Initial Synchronisation Delay ....................... 4
       2.1.1. Unicast Sessions .............................. 4
       2.1.2. Source Specific Multicast (SSM) Sessions ...... 5
       2.1.3. Any Source Multicast (ASM) Sessions .......... 6
       2.1.4. Discussion ................................... 6
   2.2. Synchronisation for Late Joiners .................... 6
3. In-band Synchronisation ..................................... 7
4. Rapid Resynchronisation .................................... 9
5. Security Considerations ................................... 9
6. IANA Considerations ...................................... 9
7. Acknowledgements ......................................... 10
8. References ................................................ 10
   8.1. Normative References ............................... 10
   8.2. Informative References ............................. 10
Author’s Address ................................................. 11
1. Introduction

When using RTP to deliver multimedia content it’s often necessary to synchronise playout of audio and video components of a presentation. This is achieved using information contained in RTP Control Protocol (RTCP) Sender Report (SR) packets [1]. These are sent periodically, and the components of a multimedia session cannot be synchronised until an RTCP SR packet has been received for each flow. Recently, concern has been expressed that this initial synchronisation delay is problematic for some applications, for example those using layered or multiple description video coding. This memo reviews the operation of RTP synchronisation, describes the initial synchronisation delay that can be expected, and defines an enhancement to the Extended RTP Profile for RTCP-based Feedback (RTP/AVPF) [2], and a new RTP header extension, to provide faster synchronisation in some circumstances.

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 RFC 2119 [3].

2. Synchronisation of RTP Flows

RTP flows are synchronised by receivers based on information that is contained in RTCP SR packets generated by senders (specifically, the NTP and RTP timestamps). Each type of media (e.g. audio or video) is sent in a separate RTP session, and the receiver associates RTP flows to be synchronised by means of the canonical end-point identifier (CNAME) item included in the RTCP Source Description (SDES) packets generated by the sender. To ensure synchronisation, an RTP sender MUST therefore send periodic compound RTCP packets following Section 6 of RFC 3550 [1].

The timing of these periodic compound RTCP packets will depend on the number of members in each RTP session, the fraction of those that are sending data, the session bandwidth, the configured RTCP bandwidth fraction, and whether the session is multicast or unicast (see RFC 3550 Section 6.2 for details). In summary, RTCP control traffic is allocated a small fraction, generally 5%, of the session bandwidth, and of that fraction, one quarter is allocated to active RTP senders, while receivers use the remaining three quarters (these fractions can be configured via SDP [5]). Each member of an RTP session derives an RTCP reporting interval based on these fractions, whether the session is multicast or unicast, the number of members it has observed, and whether it is actively sending data or not. It then sends a compound RTCP packet on average once per reporting interval (the actual packet transmission time is randomised in the range [0.5 ... 1.5] times the reporting interval to avoid synchronisation of reports).
A minimum reporting interval of 5 seconds is RECOMMENDED, except that the delay before sending the initial report "MAY be set to half the minimum interval to allow quicker notification that the new participant is present" [1]. Also, for unicast sessions, "the delay before sending the initial compound RTCP packet MAY be zero" [1]. In addition, for unicast sessions, and for active senders in a multicast session, the fixed minimum reporting interval MAY be scaled to "360 divided by the session bandwidth in kilobits/second. This minimum is smaller than 5 seconds for bandwidths greater than 72 kb/s." [1]

2.1. Initial Synchronisation Delay

A multimedia session comprises a set of concurrent RTP sessions among a common group of participants, using one RTP session for each media type. For example, a videoconference (which is a multimedia session) might contain an audio RTP session and a video RTP session. To allow a receiver to synchronise the components of a multimedia session, a compound RTCP packet containing an RTCP SR packet and an RTCP SDES packet with a CNAME item MUST be sent to each of the RTP sessions in the multimedia session. A receiver cannot synchronise playout across the multimedia session until such RTCP packets have been received on all of the component RTP sessions. If there is no packet loss, this gives an expected initial synchronisation delay equal to the average time taken to receive the first RTCP packet in the RTP session with the longest RTCP reporting interval. This will vary between unicast and multicast RTP sessions.

2.1.1. Unicast Sessions

For unicast multimedia sessions, senders SHOULD transmit an initial compound RTCP packet (containing an RTCP SR packet and an RTCP SDES packet with a CNAME item) immediately on joining each RTP session in the multimedia session. The individual RTP sessions are considered to be joined once any in-band signalling for NAT traversal (e.g. [6]) and/or security keying (e.g. [7],[8]) has concluded, and the media path is open. This implies that the initial RTCP packet is sent in parallel with the first data packet following the guidance in RFC 3550 that "the delay before sending the initial compound RTCP packet MAY be zero" and, in the absence of any packet loss, flows can be synchronised immediately.

Note that NAT pinholes, firewall holes, quality-of-service, and media security keys have should been negotiated as part of the signalling, whether in-band or out-of-band, before the first RTCP packet is sent. This should ensure that any middleboxes are ready to accept traffic, and reduce the likelihood that the initial RTCP packet will be lost.
2.1.2. Source Specific Multicast (SSM) Sessions

For multicast sessions, the delay before sending the initial RTCP packet, and hence the synchronisation delay, varies with the session bandwidth and the number of members in the session. For a multicast multimedia session, the average synchronisation delay will depend on the slowest of the component RTP sessions; this will generally be the session with the lowest bandwidth (assuming all the RTP sessions have the same number of members).

When sending to a multicast group, the reduced minimum RTCP reporting interval of 360 seconds divided by the session bandwidth in kilobits per second [1] should be used when synchronisation latency is likely to be an issue. Also, as usual, the reporting interval is halved for the first RTCP packet. Depending on the session bandwidth and the number of members, this gives the following average synchronisation delays:

<table>
<thead>
<tr>
<th>Session Bandwidth</th>
<th>2 Receivers</th>
<th>3 Receivers</th>
<th>4 Receivers</th>
<th>5 Receivers</th>
<th>10 Receivers</th>
<th>100 Receivers</th>
<th>1000 Receivers</th>
<th>10000 Receivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 kbps</td>
<td>2.73</td>
<td>4.10</td>
<td>5.47</td>
<td>5.47</td>
<td>5.47</td>
<td>5.47</td>
<td>5.47</td>
<td>5.47</td>
</tr>
<tr>
<td>16 kbps</td>
<td>2.50</td>
<td>2.50</td>
<td>2.73</td>
<td>2.73</td>
<td>2.73</td>
<td>2.73</td>
<td>2.73</td>
<td>2.73</td>
</tr>
<tr>
<td>32 kbps</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>64 kbps</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>128 kbps</td>
<td>1.41</td>
<td>1.41</td>
<td>1.41</td>
<td>1.41</td>
<td>1.41</td>
<td>1.41</td>
<td>1.41</td>
<td>1.41</td>
</tr>
<tr>
<td>256 kbps</td>
<td>0.70</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>512 kbps</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>1 Mbps</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>2 Mbps</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>4 Mbps</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Figure 1: Average RTCP Reporting Interval (seconds)

These numbers assume a single-source multicast channel with a single active sender, which the rules in RFC 3550 section 6.3 give a fixed fraction of the RTCP bandwidth irrespective of the number of receivers. It can be seen that they are sufficient for lip-synchronisation without excessive delay, but might be viewed as having too much latency for synchronising parts of a layered video stream.

The RTCP interval is randomised in the usual manner, so the minimum synchronisation delay will be half these intervals, and the maximum delay will be 1.5 times these intervals. Note also that these RTCP intervals are calculated assuming perfect knowledge of the number of members in the session. In practice, an implementation will have
only limited knowledge of the size of the session when joining, and will likely send its initial report early compared to these values, following the RTCP reconsideration rules.

2.1.3. Any Source Multicast (ASM) Sessions

(tbd)

For ASM sessions, the fraction of members that are senders plays an important role, and imply more variation in average RTCP reporting interval.

2.1.4. Discussion

For unicast sessions, the existing RTCP SR-based mechanism allows for immediate synchronisation, provided the initial RTCP packet is not lost.

For SSM sessions, the initial synchronisation delay is sufficient for lip-synchronisation, but may be larger than desired for some layered codecs. The rationale for not sending immediate RTCP packets for multicast groups is to avoid implosion of requests when large numbers of members simultaneously join the group ("flash crowd"). This is not an issue for SSM senders, since there can be at most one sender, so it might be desirable to allow SSM senders to send an immediate RTCP SR on joining a session (as is currently allowed for unicast sessions, which also don’t suffer from the implosion problem). SSM receivers using unicast feedback would not be allowed to send immediate RTCP. This would be a change to RFC 3550, if accepted.

For ASM session... (tbd)

In all cases, it is possible that the initial RTCP SR packet is lost. In this case, the receiver will not be able to synchronise the media until the reporting interval has passed, and the next RTCP SR packet is sent. This is undesirable. Section 4 defines a new RTP/AVPF transport layer feedback message to request an RTCP SR be generated, allowing rapid resynchronisation in the case of packet loss.

2.2. Synchronisation for Late Joiners

Synchronisation between RTP sessions is potentially slower for late joiners, than for participants present at the start of the session. The reasons for this are two-fold:

1. Many of the optimisations that allow rapid transmission of RTCP SR packets apply only at the start of a session. This implies that a new participant may have to wait a complete RTCP reporting
interval for each session before receiving the necessary data to synchronise media streams. This might potentially take several seconds, depending on the configured session bandwidth and the number of participants.

2. Additional synchronisation delay comes from the nature of the RTCP timing rules. Packets are generated on average once per reporting interval but with the exact transmission times being randomised +/- 50% to avoid synchronisation of reports. This is important to avoid network congestion in multicast sessions, but does mean that the timing of RTCP SR reports for different RTP sessions aren’t synchronised. Accordingly, a receiver must estimate the skew on the NTP-format clock in order to align RTP timestamps across sessions. This estimation is an essential part of an RTP synchronisation implementation, and can be done exactly given sufficient reports. Collecting sufficient RTCP SR data to perform this estimation, however, may require several reports, further increasing the synchronisation delay.

These delays are likely an issue for tuning in to an ongoing multicast RTP session, or for video switching MCUs.

3. In-band Synchronisation

The RTP header extension mechanism defined in [4] can be adopted to carry an OPTIONAL NTP format wall clock timestamp in RTP data packets. If such a timestamp is included, it MUST correspond to the same time instant as the RTP timestamp in the packet’s header, and MUST be derived from the same clock used to generate the NTP format timestamps included in RTCP SR packets. The format of such a header extension is show below.
Note: it’s unfortunate that three octets of padding are needed to align the header extension. It may be worth defining a variant that sends only the lower 56 bits of the NTP format timestamp, to reduce the overheads (assuming the top 8 bits can be inferred), although this is incompatible with the equivalent ISMA mechanism.

An NTP format wall clock timestamp may be included on any RTP packets the sender chooses, but is expected to be most useful:

1. When sent on the RTP packets corresponding to a video random access point, and on the associated audio packets, to allow rapid synchronisation for late joiners and in video switching scenarios.

2. When used with a layered, multi-description, or multi-view codec, to provide exact synchronisation between layers, descriptions, or views without requiring receivers to estimate clock skew between wall and media clocks.

In all cases, irrespective of whether in-band NTP format timestamps are included or not, regular RTCP SR packets MUST be sent to provide backwards compatibility with receivers that synchronise RTP flows according to RFC 3550 [1].

(tbd: signalling for this header extension)
4. Rapid Resynchronisation

The general format of an RTP/AVPF transport layer feedback message is shown below.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|   FMT   | PT=RTPFB=205  |          length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  SSRC of packet sender                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  SSRC of media source                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:            Feedback Control Information (FCI)                 :
:                                                               :
```

A new feedback message type, RTCP-SR-REQ, is defined with FMT = XXX. (the next available FMT is 5?) This MAY be sent to indicate that a receiver is unable to synchronise media streams, and desires that the media source send an RTCP SR packet as soon as possible (within the constraints of RTCP the early feedback rules). On receipt of this, the media source SHOULD generate an RTCP SR packet as soon as possible within the RTCP early feedback rules. That RTCP SR packet MAY be sent as a non-compound RTCP packet, if this has been negotiated.

The Feedback Control Information (FCI) part of the packet is empty. The SSRC of packet sender indicates the member that is unable to synchronise media streams, while the SSRC of media source indicates the sender of the media it is unable to synchronise. The length MUST equal 2.

(tbd: discuss what happens if the feedback target is not co-located with the sender)

5. Security Considerations

The security considerations of the RTP specification [1] and RTP/AVPF profile [2] apply. No additional security considerations apply due to the RTP/AVPF rapid resynchronisation mechanism defined in Section 4.

6. IANA Considerations

(tbd - this needs to register the new RTP/AVPF transport layer
7. Acknowledgements

This memo has benefitted from discussions with numerous members of the IETF AVT working group, including Magnus Westerlund, Thomas Schierl, and Jonathan Lennox. The mechanism in Section 3 was suggested by Dave Singer, matching a similar mechanism specified by ISMA.

8. References

8.1. Normative References


8.2. Informative References


Author’s Address

Colin Perkins
University of Glasgow
Department of Computing Science
Sir Alwyn Williams Building
Lilybank Gardens
Glasgow  G12 8QQ
UK

Email: csp@csperkins.org