Using RTCP Feedback for Unicast Multimedia Congestion Control

draft-ietf-rmcat-rtp-cc-feedback-01

Colin Perkins
Motivation

- Transport protocol provides a feedback loop

- Dynamics of congestion control depend on rate of feedback, and type of information returned

- RTCP provides a feedback channel for RTP-based applications – what sort of feedback can it provide?
Questions to ask regarding congestion feedback:

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

How often can feedback be sent in RTCP?

- Per-packet – probably not
- Per-video frame – yes, with reasonable assumptions – details follow
- Per-RTT – yes in many cases, provided RTT is not too low
- Conclusion: if configured correctly, RTCP can support congestion control, provided an appropriate feedback packet is defined
RTCP Feedback

- Four types of feedback can be used:
  - Regular RTP reports [RFC 3550]
  - RTP/AVPF feedback [RFC 4584]
    - Avoid UDP/IP header overhead per report
  - Reporting groups [draft-ietf-avtcore-rtp-multi-stream-optimisation-12]
    - Avoid sending unnecessary reports from co-located SSRCs

- Support in WebRTC:
  - RTCP reporting groups are OPTIONAL in draft-ietf-rtcweb-rtp-usage-26, the others are required (expect aggregated RTCP reports are not widely implemented at the sender)
  - The RTCP XR report defined in draft-dt-rmcat-feedback-message-00 is assumed to be used
Example Scenario

- Point-to-point video conference
- Two parties, each sending audio and video
- Media bundled onto single 5-tuple → 4 SSRCs
  - 1 audio SSRC, 1 video SSRC, for each party

- Can we send a congestion report for every video frame using RTCP?
Aggregation and Reporting Groups

- Aggregate feedback → each RTCP packet is a compound packet, comprising a compound RTCP packet generated by the audio SSRC and a compound RTCP packet generated by the video SSRC.

- RTCP reporting groups are used:
  - One SSRC is designated as the reporting SSRC.
  - The other SSRC delegates its reports to that SSRC.
  - The reports are aggregated, so it doesn’t matter which is chosen as reporting SSRC (slides assume video SSRC is reporting SSRC).
Non-reporting SSRC

- Compound RTCP packet from video SSRC
- Compound RTCP packet from audio SSRC
- SR (empty)
  - 28 octets
- SDES CNAME
  - 28 octets
- RGRS
  - 12 octets
Reporting SSRC

Compound RTCP packet from video SSRC

Aggregated compound RTCP packet

SR + 2 report blocks
76 octets

Compound RTCP packet from audio SSRC

SDES CNAME+ RGRP
48 octets

XR RC2F
32 + 2*N_a + 2*N_v

28 SR header and sender info
24 Report block for remote audio
24 Report block for remote video

4 SDES header
4 Originating SSRC
18 CNAME
18 RGRP
1 Terminator
3 Padding

8 XR header
12 RC2F audio report header
2 * number of audio packets
12 RC2F video report header
2 * number of video packets
Report Size: Overall

<table>
<thead>
<tr>
<th>UDP/IP</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound RTCP</td>
<td>76 SR</td>
</tr>
<tr>
<td>packet from video SSRC</td>
<td>48 SDES</td>
</tr>
<tr>
<td></td>
<td>32 + 2<em>N_a + 2</em>N_v XR</td>
</tr>
<tr>
<td></td>
<td>= 156 + 2<em>N_a + 2</em>N_v octets</td>
</tr>
</tbody>
</table>

| Compound RTCP packet from audio SSRC | 28 SR |
|                                      | 28 SDES |
|                                      | 12 RGRS |
|                                      | = 68 octets |

Total size = 28 + 156 + 2*N_a + 2*N_v + 68 + 14
= 266 + 2*N_a + 2*N_v octets
What are $N_a$ and $N_v$?

Assume:

- $\text{video\_bit\_rate\_bps} \approx \text{session\_bw}$
- $\text{video\_packets\_per\_second} = \frac{\text{video\_bit\_rate\_bps}}{8} / \text{mtu}$
- $\text{audio\_packets\_per\_second} = 50$

- $N_v = \text{ceil}(\text{video\_packets\_per\_second} / \text{fps})$
- $N_a = \text{ceil}(\text{audio\_packets\_per\_second} / \text{fps})$

(These assumptions are not realistic)
RTCP Reporting Interval

- Reporting interval for RTP/AVPF with $T_{rr\_interval} = 0$ is $rtcp\_interval = \text{avg}_\text{rtc}_\text{p}\_size \times n / rtcp\_bw$

- For our scenario:
  - $n = 4$
  - $\text{avg}_\text{packet}\_size = 266 + 2*Na + 2*N_v$
  - $rtcp\_bw = 5\%$ of $session\_bw$
  - Because of aggregation, $\text{avg}_\text{packet}\_size$ is halved

[draft-ietf-avtcore-rtp-multi-stream-11]
Can we report per frame?

If session bandwidth > ~2.5Mbps we can report on each frame of 30fps video

\[ session\_bw = 2.5 \times 1024 \times 1024/8 \]
\[ fps = 30 \]
\[ mtu = 1500 \]
\[ Na = \text{ceil}(50/fps) \]
\[ Nv = \text{ceil}((session\_bw/mtu)/fps) \]
\[ \text{avg\_rtcp\_packet\_size} = (42 + 68 + 156 + (2 \times Nv) + (2 \times Na))/2 \]
\[ n = 4 \]
\[ \text{rtcp\_interval} = (\text{avg\_rtcp\_packet\_size} \times n) / (session\_bw \times 0.05) \]

\[ session\_bw = 327,680 \]
\[ fps = 30 \]
\[ mtu = 1,500 \]
\[ Na = 2 \]
\[ Nv = 8 \]
\[ \text{avg\_rtcp\_packet\_size} = 143 \]
\[ n = 4 \]
\[ \text{rtcp\_interval} = 0.03949121094 \]
Summary and Next Steps

• With the RTCP congestion feedback format, and standard RTCP features, we can report on every frame of 30fps video if video bandwidth > 2.5Mbps

• Obvious ways to optimise this, without changing the congestion report format
  • RGRP extensions have high overhead

• Analysis is very preliminary – ongoing