Overheads of Congestion Feedback in RTCP

draft-ietf-rmcat-rtp-cc-feedback-07

Colin Perkins
RTCP Congestion Control Feedback Overhead

- The draft outlines the overhead of using RTCP congestion control feedback packets, as described in RFC 8888

- Assumes a modern RTCP implementation:
  - RTP/AVPF or RTP/SAVPF profile
  - Non-compound RTCP packets
  - RTCP XR
  - RFC 7022 format for SDES CNAME items
  - RFC 8108 aggregated RTCP feedback
  - RFC 8861 reporting groups
Scenario 1: VoIP

• Two-party point-to-point VoIP call

• Speech frames sent every $T_f$ seconds; both participants sending
• Want to send congestion feedback every $N_r$ frames

• Desire RTCP reporting interval = $T_f \times N_r$ seconds

• RTCP packets can be regular compound packets or non-compound packets sent using RTP/AVPF early feedback
  • Send $N_{nc}$ non-compound packets between every compound packet
Scenario 1: VoIP – compound RTCP packets

- Compound SRTCP packets contain:
  - Sender Report (SR)
  - Source Description (SDES) with CNAME item
  - Congestion control feedback
  - Packet size, $S_c = 142 + 2 \times N_r$ octets
Scenario 1: VoIP – non-compound RTCP packets

- Non-compound SRTCP packets contain:
  - Congestion control feedback

- Packet size, $S_{nc} = 62 + 2 \times N_{r}$ octets

- UDP/IPv4 headers: 28 octets with no IP options
- Congestion Control Feedback:
  - 8 octets header and SSRC
  - 4 octets timestamp
  - 8 octets report SSRC + seq range
  - $2 \times N_{r}$ octets reports
- Authentication Tag:
  - 4 octets SRTCP index
  - 10 octets AUTH tag
Scenario 1: VoIP – average RTCP size

- Average RTCP packet size, \( S_{\text{rtcp}} = \left( S_c + N_{nc} \times S_{nc} \right) / (1 + N_{nc}) \)
  where \( N_{nc} = 0 \) if non-compound packets are not sent
Scenario 1: VoIP – RTCP bandwidth

- From RFC 3550: RTCP reporting interval, $T_{rtcp} = n \times \frac{S_{rtcp}}{B_{rtcp}}$
  where:
  - $n$ is the number of participants ($n = 2$ in this scenario)
  - $S_{rtcp} = (S_c + N_{nc} \times S_{nc}) / (1 + N_{nc})$ is the average RTCP packet size in octets
  - $B_{rtcp}$ is the bandwidth allocated to RTCP in octets per second

- To report every $N_r$ frames, we want to set $T_{rtcp} = N_r \times T_f$
  \[
  \Rightarrow N_r \times T_f = n \times \frac{S_{rtcp}}{B_{rtcp}}
  \]
  \[
  \Rightarrow B_{rtcp} = \left( n \times (S_c + N_{nc} \times S_{nc}) \right) / \left( N_r \times T_f \times (1 + N_{nc}) \right)
  \]
Scenario 1: VoIP – RTCP bandwidth requirements (1)

<table>
<thead>
<tr>
<th>$T_f$ (seconds)</th>
<th>$N_r$ (frames)</th>
<th>$B_{rtcp}$ (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20ms</td>
<td>2</td>
<td>57.0</td>
</tr>
<tr>
<td>20ms</td>
<td>4</td>
<td>29.3</td>
</tr>
<tr>
<td>20ms</td>
<td>8</td>
<td>15.4</td>
</tr>
<tr>
<td>20ms</td>
<td>16</td>
<td>8.5</td>
</tr>
<tr>
<td>60ms</td>
<td>2</td>
<td>19.0</td>
</tr>
<tr>
<td>60ms</td>
<td>4</td>
<td>9.8</td>
</tr>
<tr>
<td>60ms</td>
<td>8</td>
<td>5.1</td>
</tr>
<tr>
<td>60ms</td>
<td>16</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Sending only compound RTCP packets

- Chart gives the required RTCP bandwidth, $B_{rtcp}$, to send a report after every $N_r$ frames with frames being sent every $T_f$ seconds
  - Total RTCP bandwidth for the session: each participant gets half of this
  - Compound packets only: $N_{nc} = 0$

- Sending an RTCP report every 2nd frame with 20ms frames $\rightarrow$ 57kbps RTCP bandwidth
- Sending an RTCP report every 16th frame with 60ms frames $\rightarrow$ 2.8kbps RTCP bandwidth
  - This is 1 RTCP packet per second from each SSRC in the VoIP call
Scenario 1: VoIP – RTCP bandwidth requirements (2)

- RTCP bandwidth reduced by sending non-compound packet between compound packets
- Reduced header overheads, due to not sending SR/RR and SDES packets in some reports
- Can further lower overhead by sending compound packets less often

<table>
<thead>
<tr>
<th>$T_f$ (seconds)</th>
<th>$N_r$ (frames)</th>
<th>$B_{rtcp}$ (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20ms</td>
<td>2</td>
<td>41.4</td>
</tr>
<tr>
<td>20ms</td>
<td>4</td>
<td>21.5</td>
</tr>
<tr>
<td>20ms</td>
<td>8</td>
<td>11.5</td>
</tr>
<tr>
<td>20ms</td>
<td>16</td>
<td>6.5</td>
</tr>
<tr>
<td>60ms</td>
<td>2</td>
<td>13.8</td>
</tr>
<tr>
<td>60ms</td>
<td>4</td>
<td>7.2</td>
</tr>
<tr>
<td>60ms</td>
<td>8</td>
<td>3.8</td>
</tr>
<tr>
<td>60ms</td>
<td>16</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Alternating compound and non-compound RTCP
Scenario 2: Video conference

- 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

- Video frame interval = $T_f$ (i.e., frame rate = $1/T_f$ frames per second)
- Desire RTCP reporting interval = $N_r \times T_f$
  - If $N_r = 1$, report every frame
  - If $N_r = 2$, report every other frame
  - ...
- Packets can be sent as compound or reduced size (non-compound) RTCP packets
Scenario 2: Video conference – compound packets

- Two SSRCs → need to aggregate feedback into a single RTCP packet
  - Each packet is an aggregation of a compound RTCP packet from the audio SSRC and a compound RTCP packet from 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
Scenario 2: Video conference – compound packets

- Packets from non-reporting SSRC are 68 octets (assuming RFC 7022-style CNAME)
Scenario 2: Video conference – compound packets

- Packets from reporting SSRC are $152 + 2 \times N_A + 2 \times N_V$ octets
Scenario 2: Video conference – compound packets

- **UDP/IPv4** (28 octets with no IP options)
  - 28 octets
  - $152 + 2 \times N_A + 2 \times N_V$ octets
  - 68 octets
  - 14 octets
  - Total = $262 + 2 \times N_A + 2 \times N_V$ octets

Since this reports on two SSRCs, it is halved before use:

$$S_c = \frac{(262 + 2 \times N_A + 2 \times N_V)}{2}$$
Scenario 2: Video conference – $B_{rtcp}$ calculation

- Assume:
  - Constant rate media
  - Equal size video frames; audio framing aligned to video
  - MTU around 1500 octets
- RTCP bandwidth calculation as for scenario 1:

\[
B_{rtcp} = \frac{n \times (S_c + N_{nc} \times S_{nc})}{N_r \times T_f \times (1 + N_{nc})}
\]

with

- $S_c = \frac{(262 + 2 \times N_A + 2 \times N_V)}{2}$
- $N_{nc} = 0$ (sending only compound RTCP)
- $T_f$ based on chosen video frame rate
- $N_r = 1$ (report on every frame)
### Scenario 2: Video conference – required RTCP bandwidth

<table>
<thead>
<tr>
<th>Media Rate (kbps)</th>
<th>Video Frame Rate ($1/T_f$)</th>
<th>Video packets per report: $N_v$</th>
<th>Audio packets per report: $N_a$</th>
<th>Required RTCP bandwidth, $B_{rtcp}$ in kbps (and as % of media rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>34.5 (34%)</td>
</tr>
<tr>
<td>200</td>
<td>16</td>
<td>1</td>
<td>3</td>
<td>67.5 (33%)</td>
</tr>
<tr>
<td>350</td>
<td>30</td>
<td>1</td>
<td>2</td>
<td>125.6 (35%)</td>
</tr>
<tr>
<td>700</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>126.6 (18%)</td>
</tr>
<tr>
<td>700</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>249.4 (35%)</td>
</tr>
<tr>
<td>1024</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>127.5 (12%)</td>
</tr>
<tr>
<td>1400</td>
<td>60</td>
<td>2</td>
<td>1</td>
<td>251.2 (17%)</td>
</tr>
<tr>
<td>2048</td>
<td>30</td>
<td>6</td>
<td>2</td>
<td>130.3 (6%)</td>
</tr>
<tr>
<td>2048</td>
<td>60</td>
<td>3</td>
<td>1</td>
<td>253.1 (12%)</td>
</tr>
<tr>
<td>4096</td>
<td>30</td>
<td>12</td>
<td>2</td>
<td>135.9 (3%)</td>
</tr>
<tr>
<td>4096</td>
<td>60</td>
<td>6</td>
<td>1</td>
<td>258.8 (6%)</td>
</tr>
</tbody>
</table>

Sending only compound RTCP

$B_{rtcp}$ scales linearly with $N_r$

→ reporting every 2nd frame halves bandwidth
Scenario 2: Video conference – reduced size packets

- Aggregate reports from two SSRCs – reduced size RTCP rather than non-compound RTCP – allowing to omit SR and SDES CNAME from aggregated packet

\[ S_{nc} = \frac{110 + 2 \times N_v + 2 \times N_a}{2} \]

- Gives \( S_{nc} \)
- Repeat calculation with \( N_{nc} = 1 \) indicating that we alternate regular and reduced size RTCP
### Scenario 2: Video conference – required RTCP bandwidth

<table>
<thead>
<tr>
<th>Media Rate (kbps)</th>
<th>Video Frame Rate (1/Tf)</th>
<th>Video packets per report: Nv</th>
<th>Audio packets per report: Na</th>
<th>Required RTCP bandwidth, B&lt;sub&gt;rtcp&lt;/sub&gt; in kbps (and as % of media rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>24.1 (24%)</td>
</tr>
<tr>
<td>200</td>
<td>16</td>
<td>1</td>
<td>3</td>
<td>46.8 (23%)</td>
</tr>
<tr>
<td>350</td>
<td>30</td>
<td>1</td>
<td>2</td>
<td>86.7 (24%)</td>
</tr>
<tr>
<td>700</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>87.7 (12%)</td>
</tr>
<tr>
<td>700</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>171.6 (24%)</td>
</tr>
<tr>
<td>1024</td>
<td>30</td>
<td>3</td>
<td>2</td>
<td>88.6 (8%)</td>
</tr>
<tr>
<td>1400</td>
<td>60</td>
<td>2</td>
<td>1</td>
<td>173.4 (12%)</td>
</tr>
<tr>
<td>2048</td>
<td>30</td>
<td>6</td>
<td>2</td>
<td>91.4 (4%)</td>
</tr>
<tr>
<td>2048</td>
<td>60</td>
<td>3</td>
<td>1</td>
<td>175.3 (8%)</td>
</tr>
<tr>
<td>4096</td>
<td>30</td>
<td>12</td>
<td>2</td>
<td>97.0 (2%)</td>
</tr>
<tr>
<td>4096</td>
<td>60</td>
<td>6</td>
<td>1</td>
<td>180.9 (4%)</td>
</tr>
</tbody>
</table>

Alternating regular and reduced-size RTCP packets

$B_{rtcp}$ scales linearly with $N_r$

→ reporting every 2nd frame halves bandwidth

Overheads significantly reduced
Changes in -07

• Brings the calculation up-to-date with published RTCP congestion control feedback format [RFC 8888] and corrects some minor errors

• Removes placeholders for multi-party and screen sharing sessions – too many variables to easily characterise
Next Steps

• Draft illustrates factors that influence RTCP congestion control feedback overhead, to illustrate how the format can be used and configured

• Is it useful enough to publish? If so, it’s ready