University of Glasgow

School of Computing Science

60 YEARS OF COMPUTING AT GLASGOW

Congestion Feedback in RTCP

Colin Perkins

Presentation given to IETF RMCAT working group on 19 July 2017

Congestion Feedback in RTCP

- What is the overhead of sending congestion feedback in RTCP?
- This will depend on:
  - How often you want to send congestion feedback?
  - What information is included in congestion feedback packets?
  - How that information is formatted?
  - What other information must also be included in those packets?
  - The number and format of the media streams being sent

Considering two approaches:
1) draft-dt-rmcat-feedback-message-02
2) draft-holmer-rmcat-transport-wide-cc-extensions-01

Regular or reduced-size RTCP?
Are SR/RR packets included?
Are SDES packets included? What is the format of the CNAME?
How are stream associated?
How is cross-reporting handled?

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

- Congestion feedback can be sent in regular compound RTCP packets or reduced-size packets sent using RTP/AVPF early feedback
  - Send N_{nc} non-compound packets between every compound packet

draft-holmer-rmcat-transport-wide-cc-extensions-01

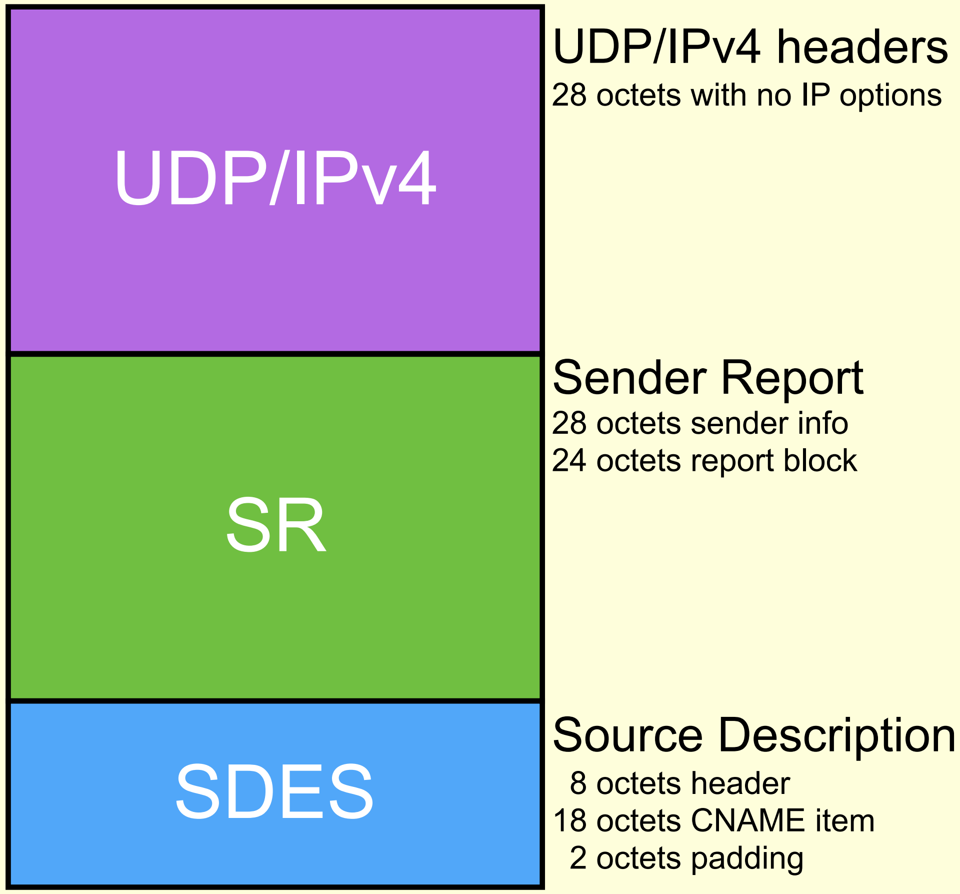
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|  FMT=15 |    PT=205     |           length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     SSRC of packet sender                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      SSRC of media source                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      base sequence number     |      packet status count      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 reference time                | fb pkt. count |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          packet chunk         |         packet chunk          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         packet chunk          |  recv delta   |  recv delta   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           recv delta          |  recv delta   | zero padding  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Single packet format used, whether sent in compound or reduced size packets
- Packet chunks are either a bit vector or RLE encoded
- Recv delta fields are one or two octets, depending on packet inter-arrival time
- RTCP packet size will vary with loss pattern and inter-arrival times

Requires RTP header extension for transport-wide sequence number if more than one media stream is sent

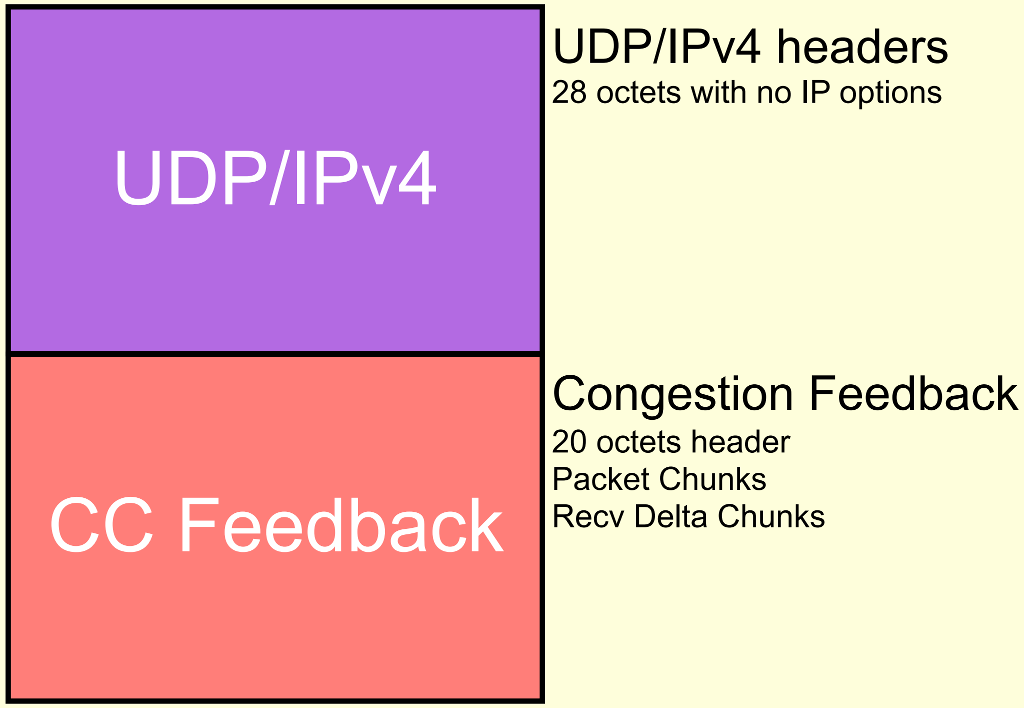
Scenario 1: VoIP – compound RTCP packets

- Compound RTCP packets contain:
  - Sender Report (SR)
  - Source Description (SDES) with CNAME item
- Packet size, S_c = 108 octets
  - UDP/IP + SR + SDES = 108 octets

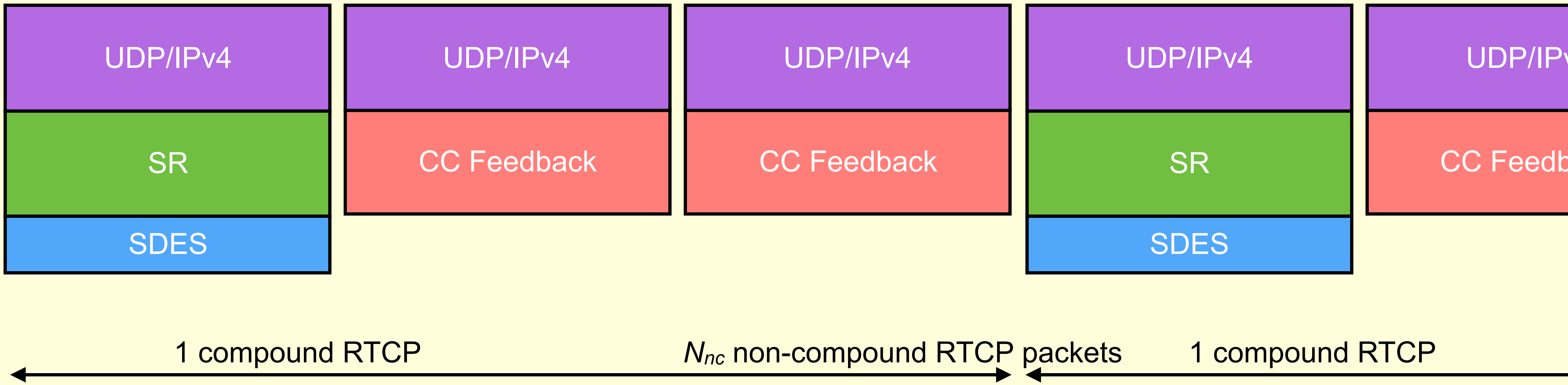


Scenario 1: VoIP – non-compound RTCP packets

- Non-compound RTCP packets contain:
  - Congestion control feedback (draft-holmer-rmcat-transport-wide-cc-extensions-01)
- Packet size, $S_{nc} = 48 + 2 + N_r$ octets
  - UDP/IP + RTCP congestion feedback header = 48 octets
  - Packet chunks: best case → no packets lost, single RLE packet chunk sufficient → 2 octets
  - Recv Delta chunks: bast case → deltas fit one octet format → N_r octets (one recv delta per packet sent)



Scenario 1: VoIP – average RTCP size



- Average RTCP packet size, $S_{rtcp} = (S_c + N_{nc} \times S_{nc}) / (1 + N_{nc})$ where $N_{nc} = 0$ if non-compound packets are not sent

Scenario 1: VoIP – RTCP bandwidth

- From RFC 3550: RTCP reporting interval, $T_{rtcp} = n \times S_{rtcp}/B_{rtcp}$ 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 $T_{rtcp} = N_r \times T_f$

  $\Rightarrow N_r \times T_f = n \times S_{rtcp}/B_{rtcp}$

  $\Rightarrow B_{rtcp} = (n \times (S_c + N_{nc} \times S_{nc})) / (N_r \times T_f \times (1 + N_{nc}))$

Scenario 1: VoIP – RTCP bandwidth requirements (1)

T_f (seconds)	N_r (frames)	B_{rtcp} (kbps)
20ms	2	42.2
20ms	4	21.1
20ms	8	10.6
20ms	16	5.3
60ms	2	14.1
60ms	4	7.0
60ms	8	3.5
60ms	16	1.8

Sending only compound RTCP packets

Note: this is best case; losses increase feedback packet size → increased B_{rtcp}

- 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 → 52kbps RTCP bandwidth
- Sending an RTCP report every 16th frame with 60ms frames → 2.4kbps RTCP bandwidth
  - This is 1 RTCP packet per second from each SSRC in the VoIP call

Scenario 1: VoIP – RTCP bandwidth requirements (2)

T_f (seconds)	N_r (frames)	B_{rtcp} (kbps)
20ms	2	31.3
20ms	4	15.8
20ms	8	8.1
20ms	16	4.2
60ms	2	10.4
60ms	4	5.3
60ms	8	2.7
60ms	16	1.4

Alternating compound and non-compound RTCP

- Required RTCP bandwidth is reduced if a non-compound packet is sent between compound packets
- Reduced header overheads – due to not sending SR/RR and SDES packets in some reports

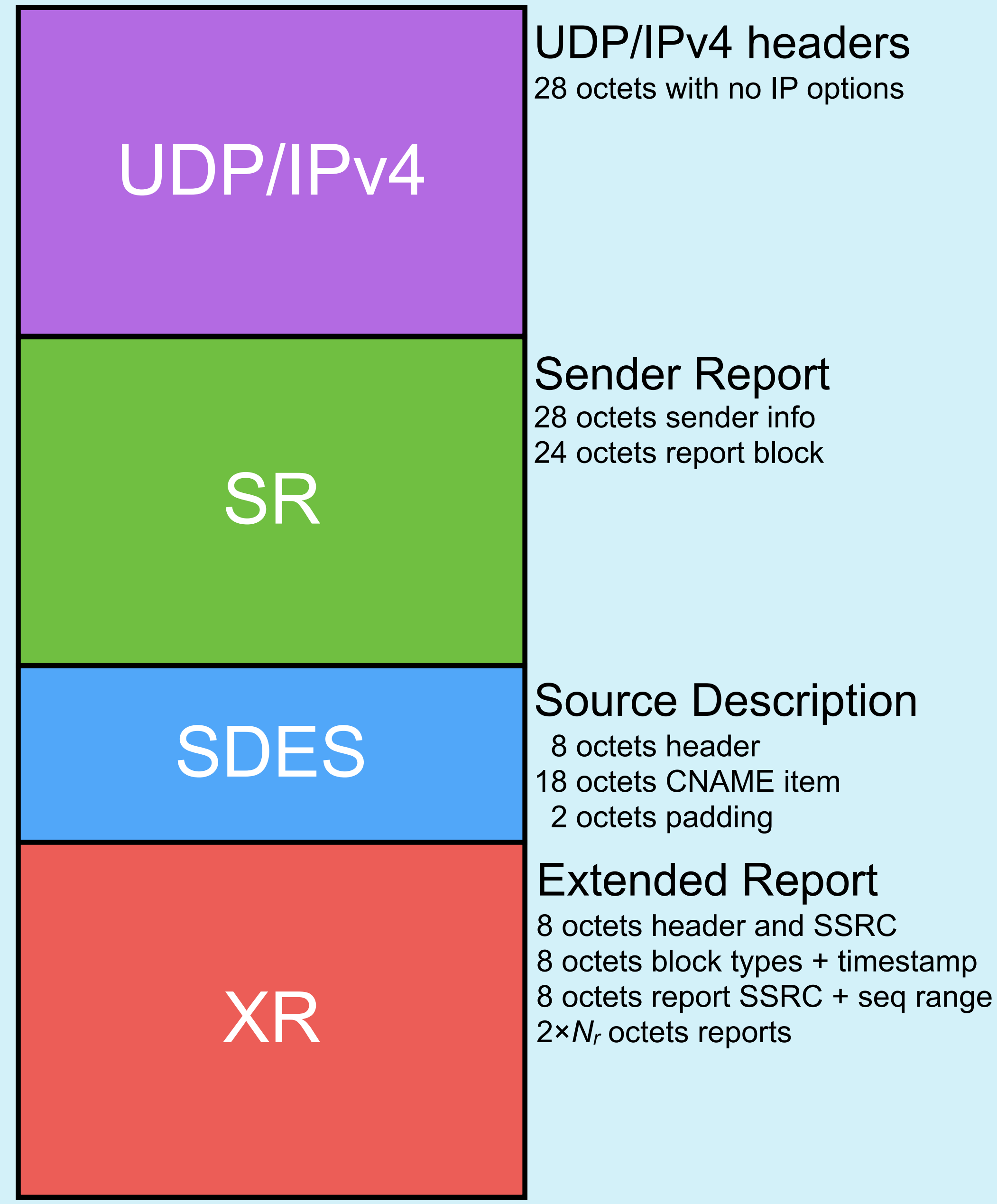
Note: this is best case; losses increase feedback packet size → increased B_{rtcp}

Scenario 1: VoIP – compound RTCP packets

- Using draft-dt-rmcat-feedback-message-02

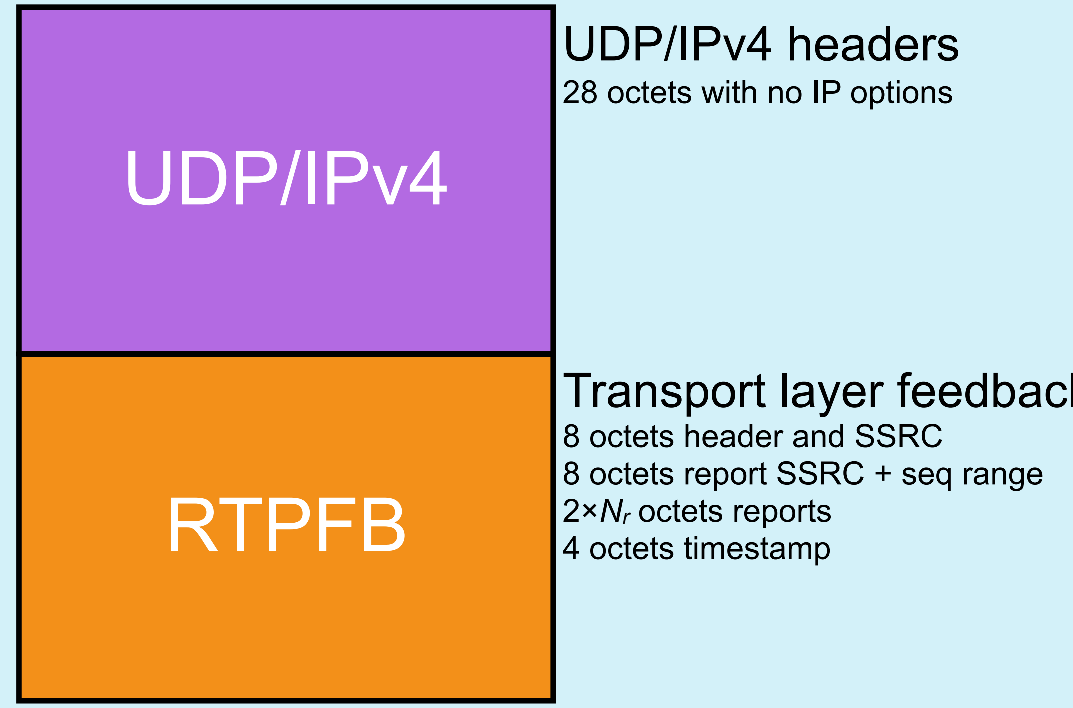
- Compound RTCP packets contain:
  - Sender Report (SR)
  - Source Description (SDES) with CNAME item
  - Extended Report (XR) with congestion control feedback (draft-dt-rmcat-feedback-message-01)

- Packet size, $S_c = 132 + 2 \times N_r$ octets



Scenario 1: VoIP – non-compound RTCP packets

- Non-compound RTCP packets contain:
  - RTP/AVPF transport layer feedback packet
    (draft-dt-rmcat-feedback-message-01)
- Packet size, $S_{nc} = 48 + 2 \times N_r$ octets



Scenario 1: VoIP – RTCP bandwidth requirements (1)

T_f (seconds)	N_r (frames)	B_{rtcp} (kbps)
20ms	2	53.1
20ms	4	27.3
20ms	8	14.5
20ms	16	8.0
60ms	2	17.7
60ms	4	9.1
60ms	8	4.8
60ms	16	2.7

Sending only compound RTCP packets

Note: B_{rtcp} independent of loss rate

- 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 → 53kbps RTCP bandwidth
- Sending an RTCP report every 16th frame with 60ms frames → 2.7kbps RTCP bandwidth
  - This is 1 RTCP packet per second from each SSRC in the VoIP call

Scenario 1: VoIP – RTCP bandwidth requirements (2)

T_f (seconds)	N_r (frames)	B_{rtcp} (kbps)
20ms	2	36.7
20ms	4	19.1
20ms	8	10.4
20ms	16	6.0
60ms	2	12.2
60ms	4	6.4
60ms	8	3.5
60ms	16	2.0

Alternating compound and non-compound RTCP

- Required RTCP bandwidth is reduced if a non-compound packet is sent between compound packets
- Reduced header overheads – due to not sending SR/RR and SDES packets in some reports

Note: B_{rtcp} independent of loss rate

Scenario 1: VoIP – Comparison

T_f (seconds)	N_r (frames)	B_{rtcp} (kbps)	B_{rtcp} (kbps)
20ms	2	31.3	36.7
20ms	4	15.8	19.1
20ms	8	8.1	10.4
20ms	16	4.2	6.0
60ms	2	10.4	12.2
60ms	4	5.3	6.4
60ms	8	2.7	3.5
60ms	16	1.4	2.0

In best case, draft-holmer-rmcat-transport-wide-cc-extensions-01 more efficient compared to draft-dt-rmcat-feedback-message-02 – but doesn't report on ECN (saving primarily due to not including congestion reports in compound packets – still more efficient if congestion reports *are* sent in compound packets, but only just)

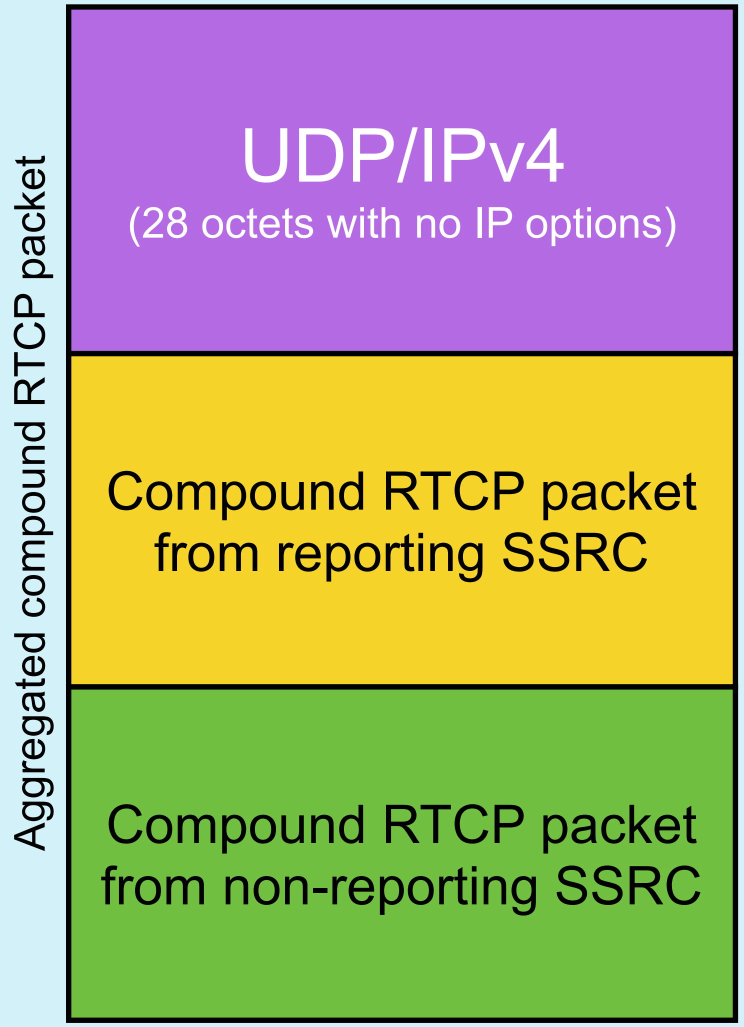
With more complex loss patterns, still saves compared to draft-dt-rmcat-feedback-message-02, but benefit is less

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

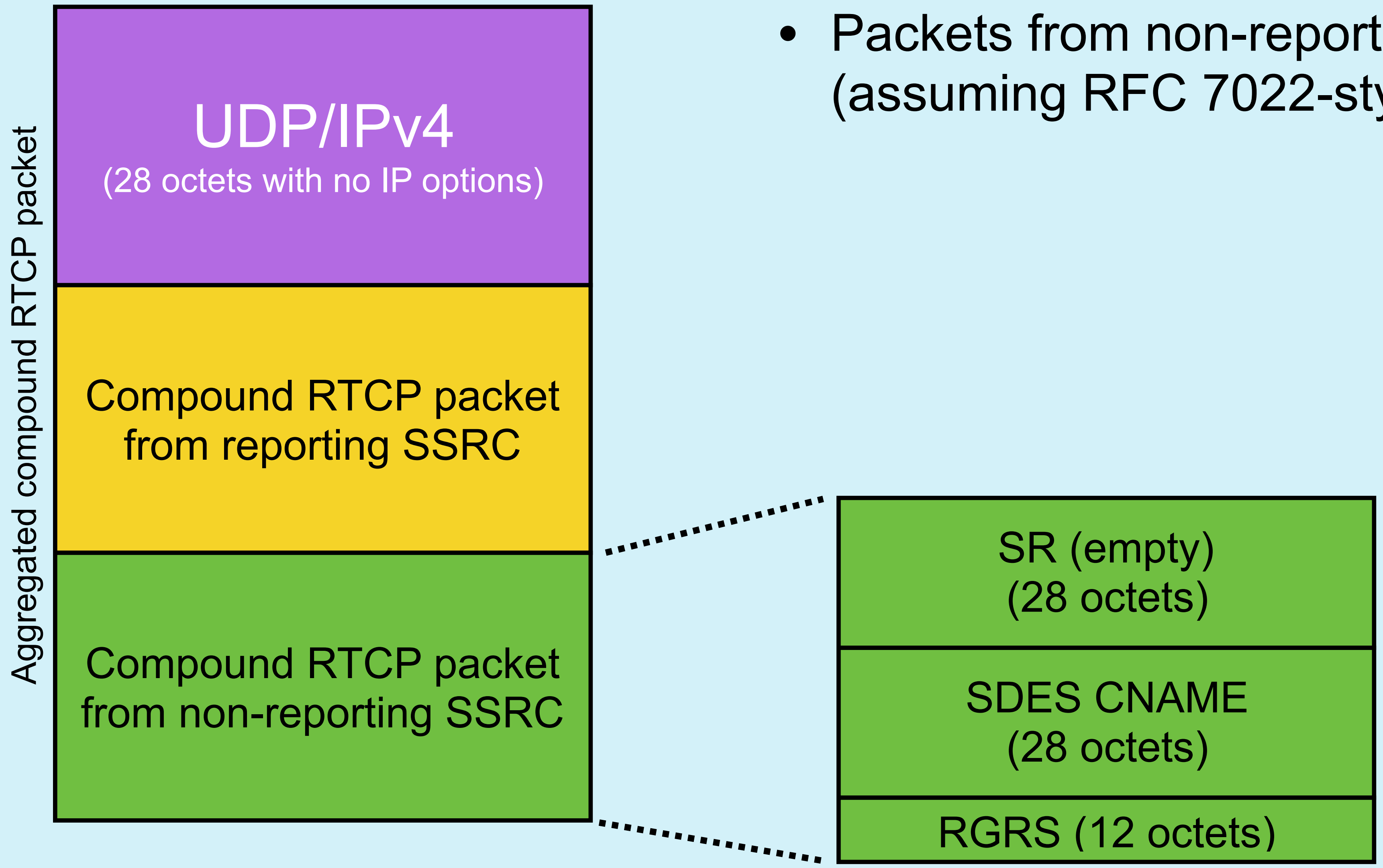
- Video frame interval = T_f (i.e., frame rate = $1/T_f$ frames per second)
- Desire RTCP reporting interval = $N_r \times T_f$
  - If $N_r = 1$, report every frame
  - If $N_r = 2$, report every other frame
  - …
- Packets can be sent as compound or reduced size (non-compound) RTCP packets

Scenario 2: Video conference – compound packets



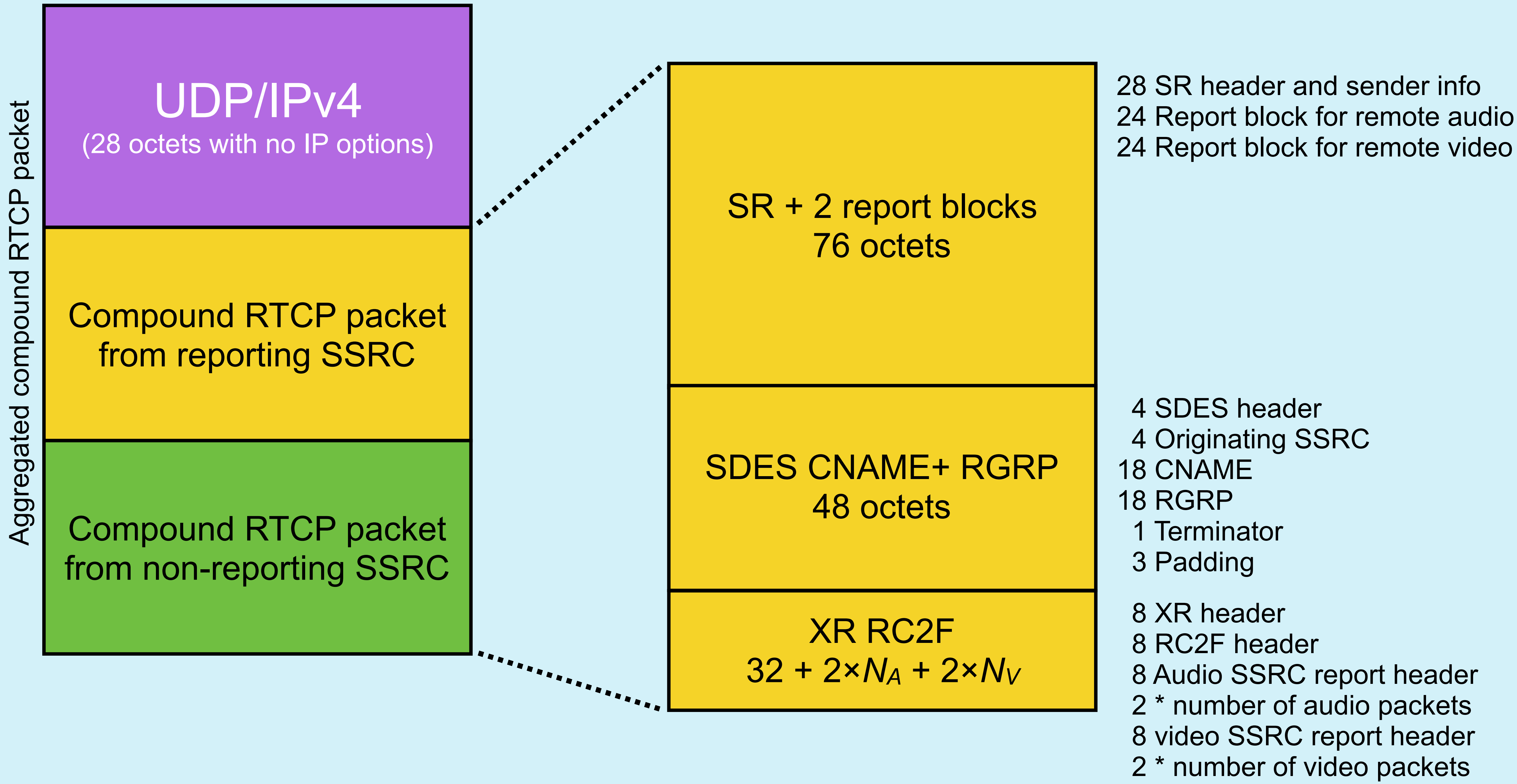
- Two SSRC → need to aggregate feedback into a single RTCP packet
  - Each packet is an aggregation of a compound RTCP packet from the audio SSRC and a compound RTCP packet from the video SSRC
- Assume RTCP reporting groups are used:
  - One SSRC is designated as the reporting SSRC
  - The other SSRC 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



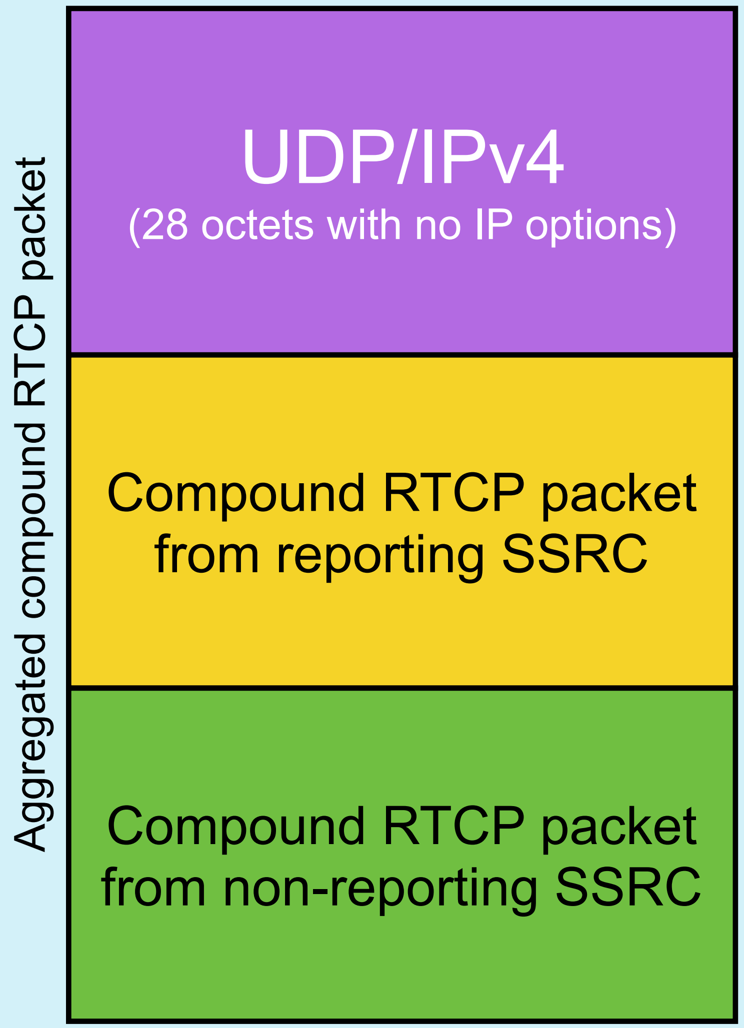
- Packets from non-reporting SSRC are 68 octets (assuming RFC 7022-style CNAME)

Scenario 2: Video conference – compound packets



- Packets from reporting SSRC are $156 + 2 \times N_A + 2 \times N_V$ octets

Scenario 2: Video conference – compound packets



- 28
- 156 + 2×N_A + 2×N_V octets
- 68 octets
- Total = 252 + 2×N_A + 2×N_V octets
- Since this reports on two SSRCs, it is halved before use: S_c = (252 + 2×N_A + 2×N_V)/2

Scenario 2: Video conference – B_{rtcp} calculation

- Assume:
  - Constant rate media
  - Video frames equal size
  - Audio at 50 packets per second (20ms frames)
  - MTU around 1500 octets
- RTCP bandwidth calculation as for scenario 1:

$$B_{rtcp} = (n \times (S_c + N_{nc} \times S_{nc})) / (N_r \times T_f \times (1 + N_{nc}))$$

with
$S_c = (252 + 2 \times N_A + 2 \times N_V)/2$
$N_{nc} = 0$
T_f based on chosen video frame rate
$N_r = 1$ (report on every frame)

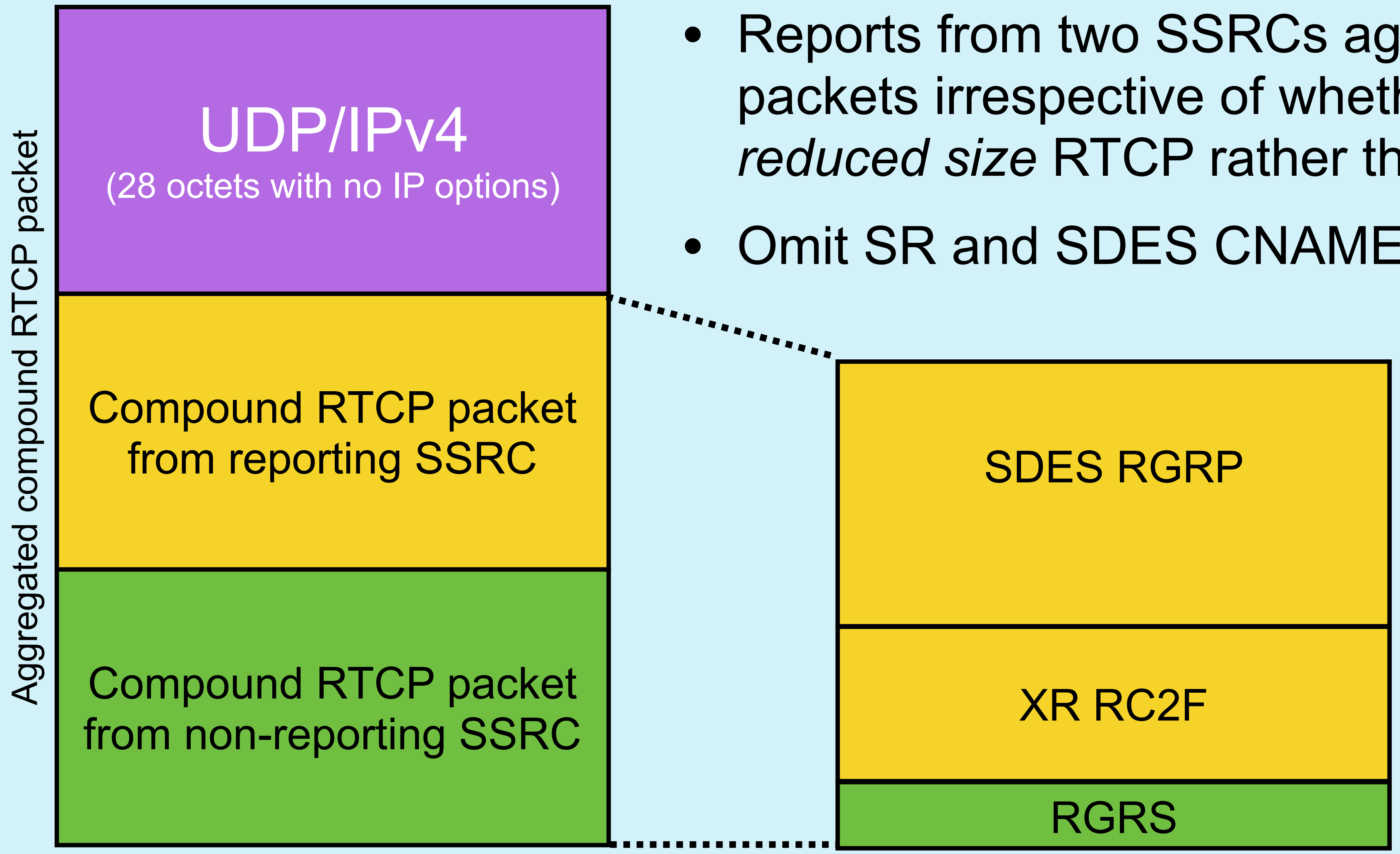
Scenario 2: Video conference – required RTCP bandwidth

Media Rate (kbps)	Video Frame Rate ($1/T_f$)	Video packets per report: N_v	Audio packets per report: N_a	Required RTCP bandwidth, B_{rtcp} in kbps (and as % of media rate)
100	8	1	6	33.3 (33%)
200	16	1	3	65.0 (33%)
350	30	1	2	120.1 (35%)
700	30	2	2	121.9 (17%)
700	60	1	1	240.0 (34%)
1024	30	3	2	122.8 (12%)
1400	60	2	1	241.8 (17%)
2048	30	6	2	125.6 (6%)
2048	60	3	1	243.8 (12%)
4096	30	12	2	131.3 (3%)
4096	60	6	1	249.4 (6%)

Sending only compound RTCP packets

B_{rtcp} scales linearly with N_r (i.e., reporting every 2nd frame halves the required RTCP bandwidth)

Scenario 2: Video conference – reduced size packets



- Reports from two SSRCs aggregated into compound packets irrespective of whether RFC5506 is used – *reduced size* RTCP rather than *non-compound* RTCP
- Omit SR and SDES CNAME from aggregated packet

- Gives $S_{nc} = (96 + 2 \times N_v + 2 \times N_a)/2$
- Repeat calculation with $N_{nc} = 1$ indicating that we alternate regular and reduced size RTCP

Scenario 2: Video conference – required RTCP bandwidth

Media Rate (kbps)	Video Frame Rate ($1/T_f$)	Video packets per report: N_v	Audio packets per report: N_a	Required RTCP bandwidth, B_{rtcp} in kbps (and as % of media rate)
100	8	1	6	23.5 (23%)
200	16	1	3	45.5 (23%)
350	30	1	2	84.4 (24%)
700	30	2	2	85.3 (12%)
700	60	1	1	166.9 (24%)
1024	30	3	2	86.2 (8%)
1400	60	2	1	168.8 (12%)
2048	30	6	2	89.1 (4%)
2048	60	3	1	170.6 (8%)
4096	30	12	2	94.7 (2%)
4096	60	6	1	176.3 (4%)

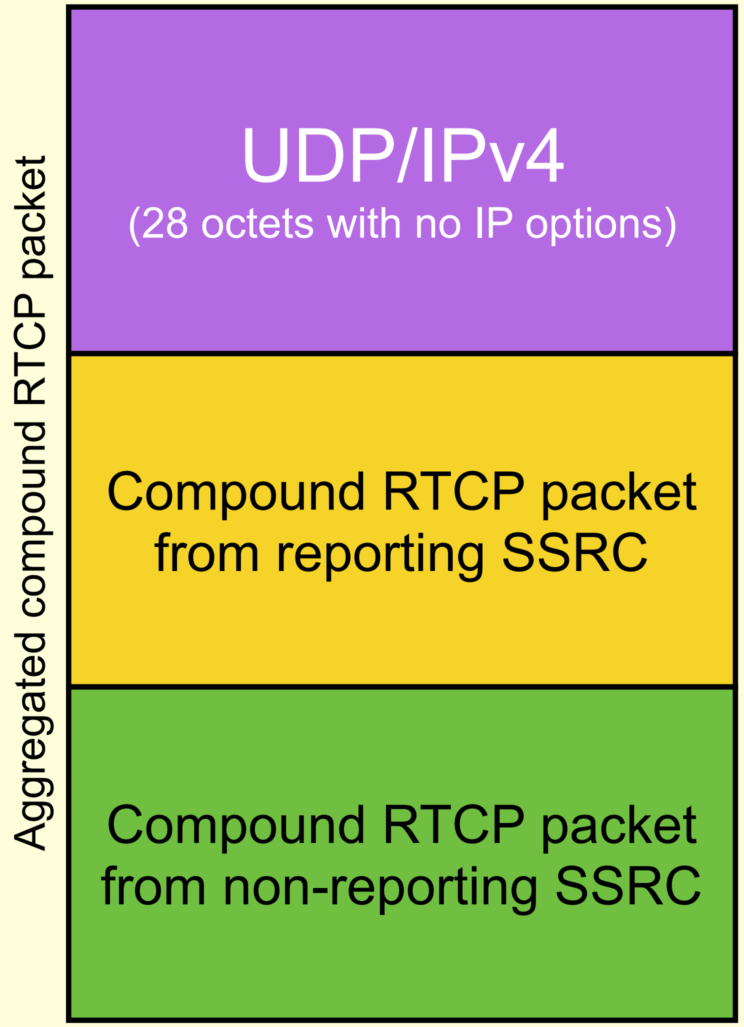
Alternating regular and reduced-size RTCP packets

B_{rtcp} scales linearly with N_r (i.e., reporting every 2nd frame halves the required RTCP bandwidth)

Scenario 2: Video conference

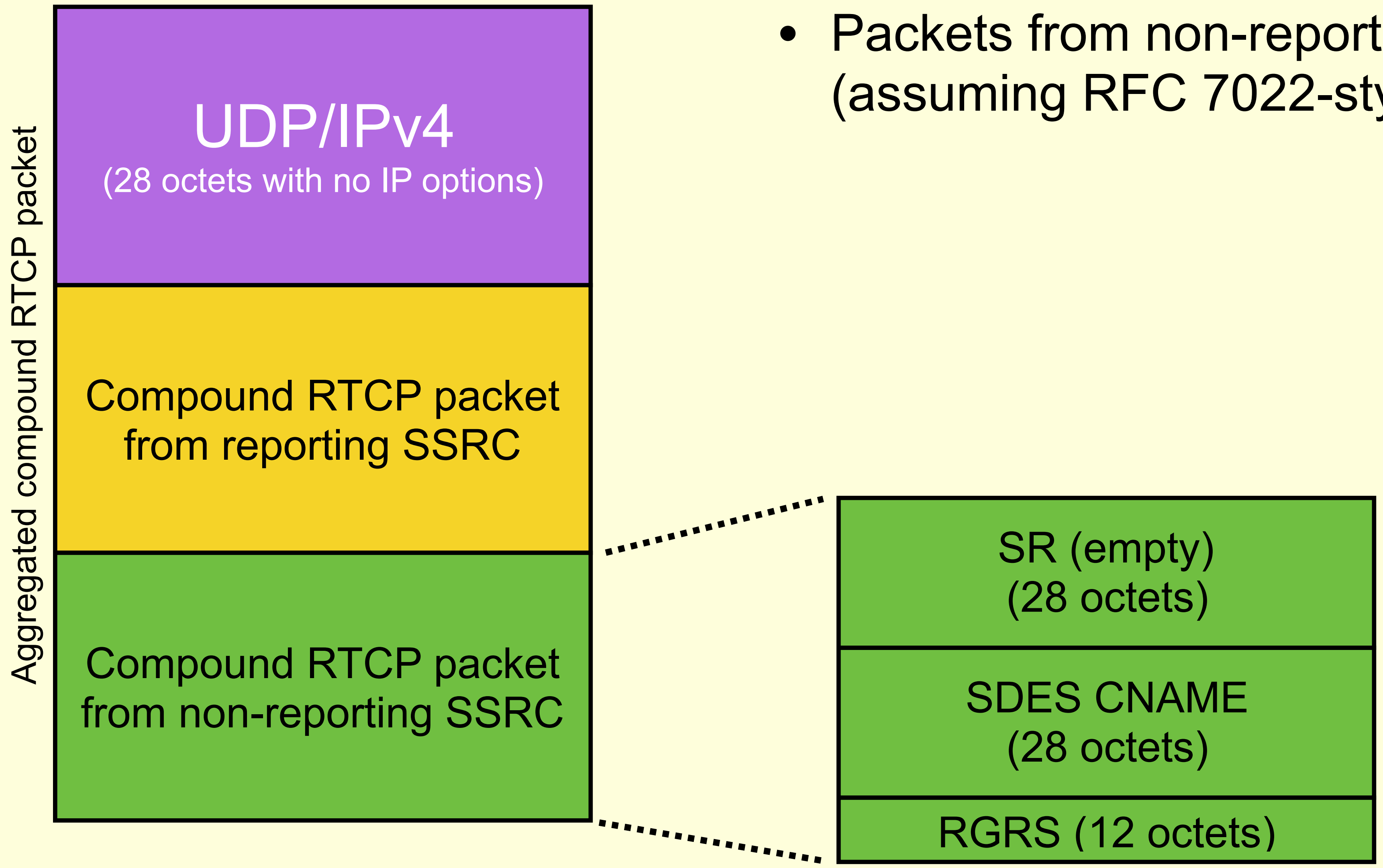
- Audio and video send regular RTCP reports as a compound packet
- Separate non-compound transport-level RTCP congestion feedback packets sent
  - RTP packets require extra header extension → 8 octets per packet

Scenario 2: Video conference – compound packets



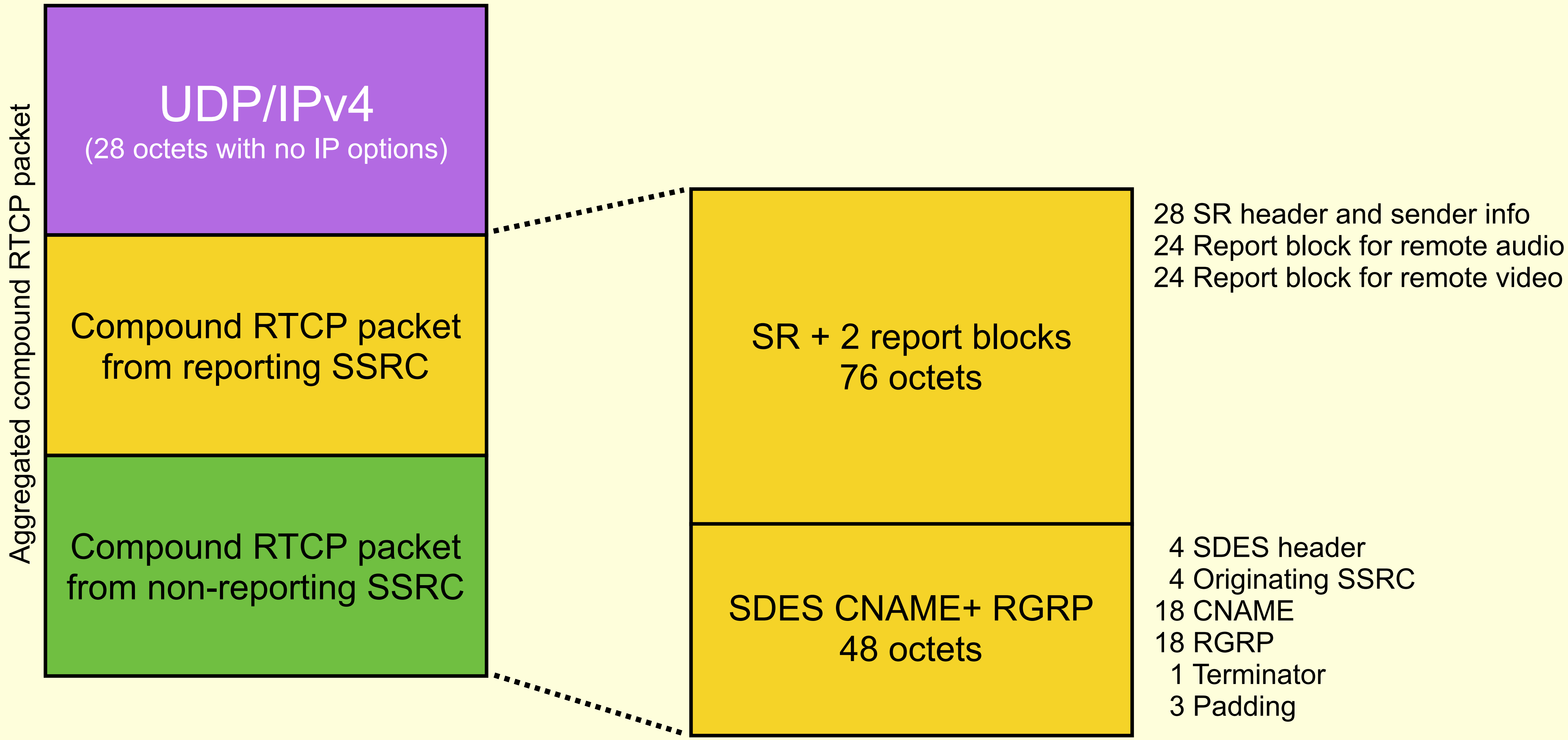
- Two SSRC → need to aggregate feedback into a single RTCP packet
  - Each packet is an aggregation of a compound RTCP packet from the audio SSRC and a compound RTCP packet from the video SSRC
- Assumes RTCP reporting groups are used:
  - One SSRC is designated as the reporting SSRC
  - The other SSRC 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)

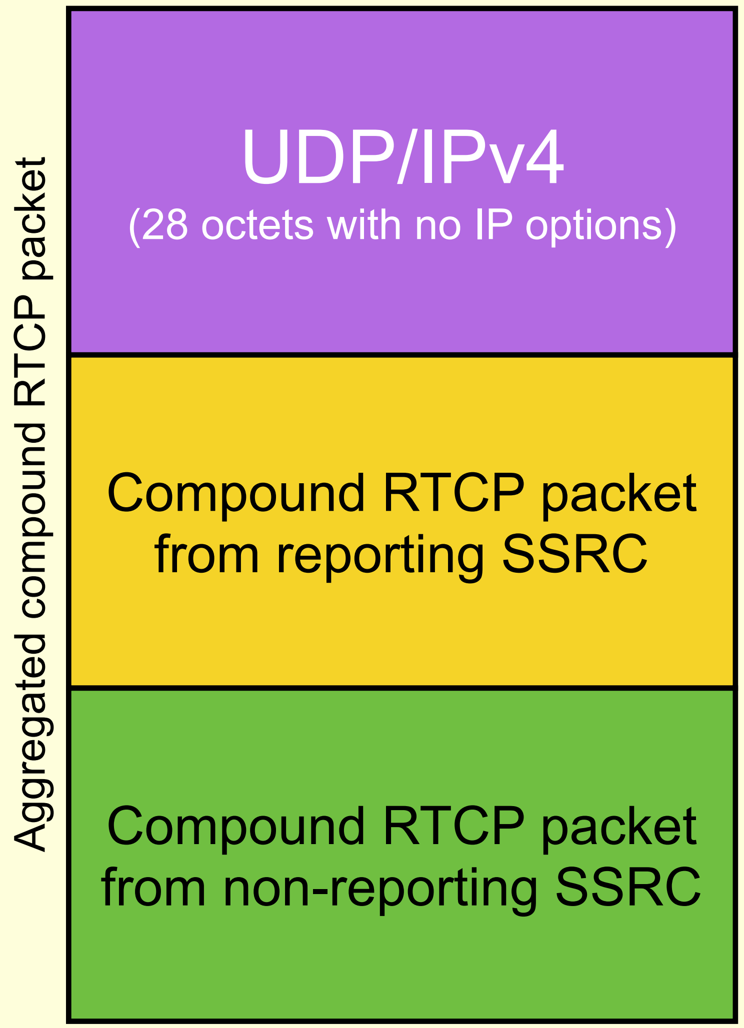
Scenario 2: Video conference – compound packets



- Packets from reporting SSRC are 124 octets

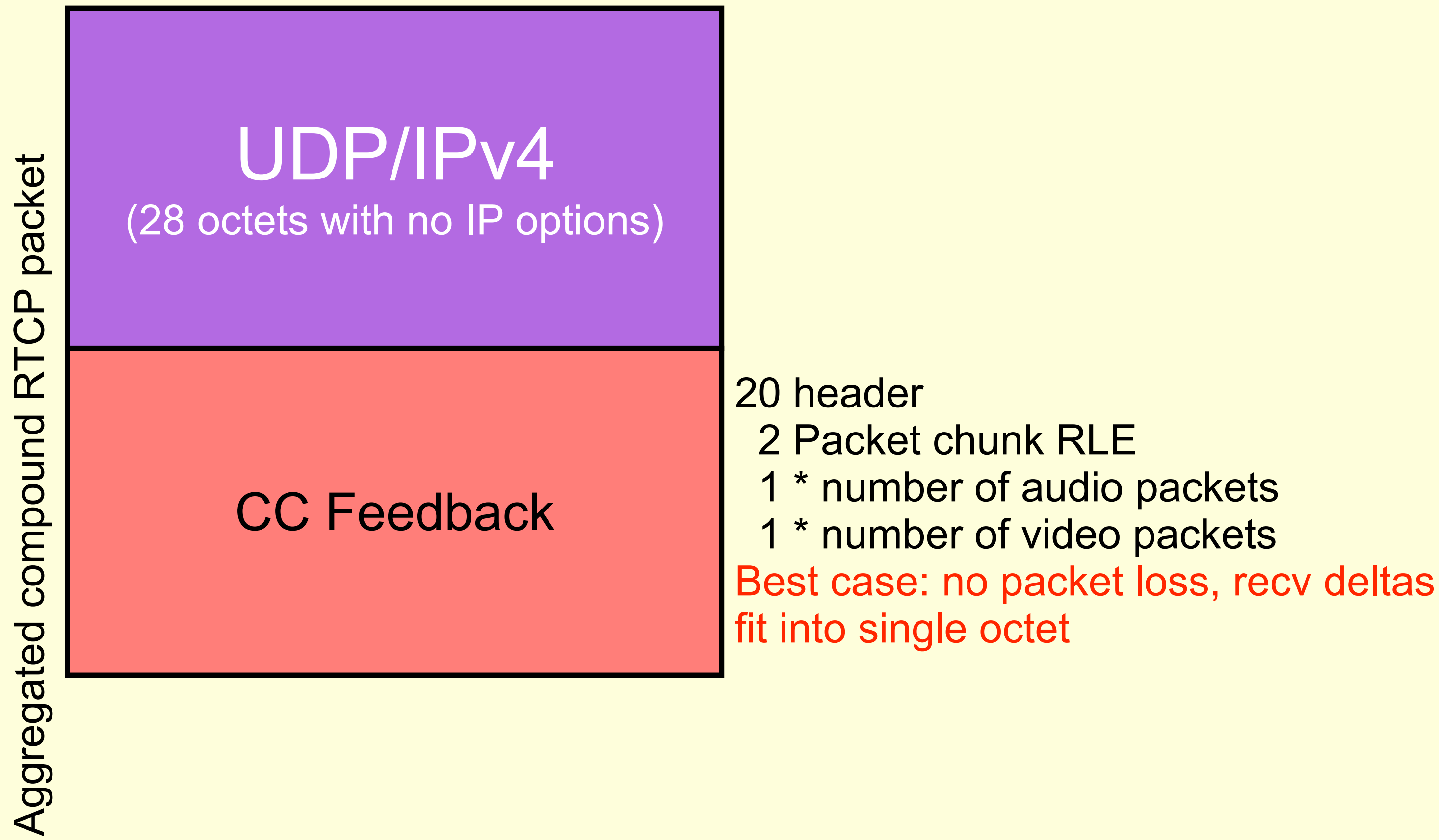
28

Scenario 2: Video conference – compound packets



- 28
- 124 octets
- 68 octets
- Total = 220 octets
- Since this reports on two SSRCs, it is halved before use: S_c = 220/2 = 110

Scenario 2: Video conference – reduced-size packets



- Reduced size packets: $S_{nc} = 50 + N_a + N_v$ octets (best case)
- A single CC Feedback packet is sent, reporting on all sources

30

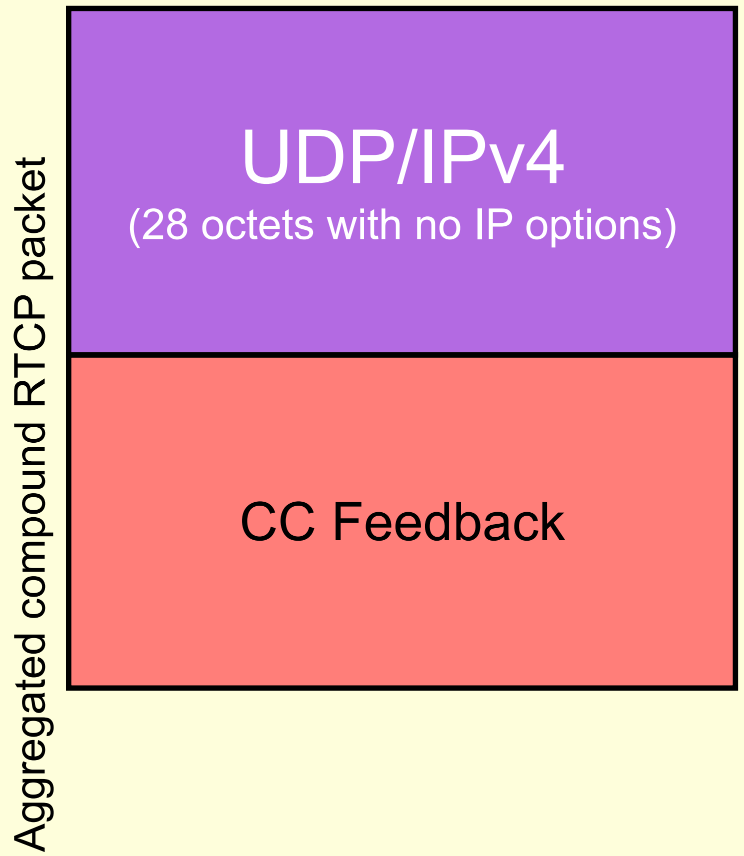
Scenario 2: Video conference – required RTCP bandwidth

Media Rate (kbps)	Video Frame Rate ($1/T_f$)	Video packets per report: N_v	Audio packets per report: N_a	Required RTCP bandwidth, B_{rtcp} in kbps (and as % of media rate)
100	8	1	6	20.9 (21%)
200	16	1	3	41.0 (21%)
350	30	1	2	76.4 (22%)
700	30	2	2	76.9 (11%)
700	60	1	1	151.9 (22%)
1024	30	3	2	77.3 (8%)
1400	60	2	1	152.8 (11%)
2048	30	6	2	78.8 (4%)
2048	60	3	1	153.8 (8%)
4096	30	12	2	81.6 (2%)
4096	60	6	1	156.6 (4%)

Best case – packet loss or timing variation will affect B_{rtcp}

Excludes overhead of RTP header extension

Scenario 2: Video conference – reduced-size packets



20 header
2 Packet chunk RLE
$\lceil Na + Nv / 7 \rceil$ octets of packet chunks
2 * (Na + Nv) octets of recv deltas
Worst case: unpredictable packet loss → send bit vector of loss; timing variation large enough to need 2 octet recv delta

- Reduced size packets: $S_{nc} = 50 + \lceil Na + Nv / 7 \rceil + 2 \times (Na + Nv)$ octets (worst case)
- A single CC Feedback packet is sent, reporting on all sources

32

Scenario 2: Video conference – required RTCP bandwidth

Media Rate (kbps)	Video Frame Rate ($1/T_f$)	Video packets per report: N_v	Audio packets per report: N_a	Required RTCP bandwidth, B_{rtcp} in kbps (and as % of media rate)
100	8	1	6	21.9 (22%)
200	16	1	3	42.3 (21%)
350	30	1	2	78.3 (22%)
700	30	2	2	79.2 (11%)
700	60	1	1	154.7 (22%)
1024	30	3	2	80.2 (8%)
1400	60	2	1	156.6 (11%)
2048	30	6	2	83.4 (4%)
2048	60	3	1	158.4 (8%)
4096	30	12	2	89.1 (2%)
4096	60	6	1	165.1 (4%)

Worst case – required RTCP bandwidth increases

Excludes overhead of RTP header extension

Scenario 2: Video conference – required RTCP bandwidth

Media Rate (kbps)	Video Frame Rate ($1/T_f$)	Video packets per report: N_v	Audio packets per report: N_a	RTCP Bandwidth (worst case)	RTCP Bandwidth (best case)	RTCP Bandwidth
100	8	1	6	21.9 (22%)	20.9 (21%)	23.5 (23%)
200	16	1	3	42.3 (21%)	41.0 (21%)	45.5 (23%)
350	30	1	2	78.3 (22%)	76.4 (22%)	84.4 (24%)
700	30	2	2	79.2 (11%)	76.9 (11%)	85.3 (12%)
700	60	1	1	154.7 (22%)	151.9 (22%)	166.9 (24%)
1024	30	3	2	80.2 (8%)	77.3 (8%)	86.2 (8%)
1400	60	2	1	156.6 (11%)	152.8 (11%)	168.8 (12%)
2048	30	6	2	83.4 (4%)	78.8 (4%)	89.1 (4%)
2048	60	3	1	158.4 (8%)	153.8 (8%)	170.6 (8%)
4096	30	12	2	89.1 (2%)	81.6 (2%)	94.7 (2%)
4096	60	6	1	165.1 (4%)	156.6 (4%)	176.3 (4%)

Results for draft-holmer-rmcat-transport-wide-cc-extensions-01 exclude overhead of RTP header extension

Scenario 2: Video conference – required RTCP bandwidth

Media Rate (kbps)	Video Frame Rate ($1/T_f$)	Video packets per report: N_v	Audio packets per report: N_a	RTCP Bandwidth (worst case) + overhead	RTCP Bandwidth (best case) + overhead	RTCP Bandwidth
100	8	1	6	25.4 (25%)	24.4 (24%)	23.5 (23%)
200	16	1	3	46.3 (23%)	45.0 (23%)	45.5 (23%)
350	30	1	2	83.9 (24%)	82.0 (23%)	84.4 (24%)
700	30	2	2	79.2 (11%)	84.4 (12%)	85.3 (12%)
700	60	1	1	162.2 (23%)	159.4 (23%)	166.9 (24%)
1024	30	3	2	89.5 (9%)	86.7 (8%)	86.2 (8%)
1400	60	2	1	167.8 (12%)	164.1 (12%)	168.8 (12%)
2048	30	6	2	98.4 (5%)	93.8 (5%)	89.1 (4%)
2048	60	3	1	173.4 (8%)	168.8 (8%)	170.6 (8%)
4096	30	12	2	115.3 (3%)	107.8 (3%)	94.7 (2%)
4096	60	6	1	190.3 (5%)	182.8 (4%)	176.3 (4%)

Adding RTP header extension overhead: $8 \times (N_a + N_v)$ octets per RTCP reporting interval

$= (8 \times (N_a + N_v)) / T_f$ octets per second (since reporting interval = frame rate)

Conclusions

- Compared performance of draft-dt-rmcat-feedback-message-02 and draft-holmer-rmcat-transport-wide-cc-extensions-01
  - draft-holmer-rmcat-transport-wide-cc-extensions-01 needs less RTCP bandwidth for VoIP
    - Doesn't include ECN feedback
  - draft-dt-rmcat-feedback-message-02 generally needs less RTCP bandwidth to report on video once RTP header extension overhead taken into account
    - But RTCP bandwidth needed is highly dependent on data rate and packet loss patterns
    - Compatible with standard RTP use of SSRCs
    - Conveys ECN feedback