Video Frame Info
RTP Header Extension
draft-ietf-avtext-framemarking-03

Espen Berger, Suhas Nandakumar, Mo Zanaty
(Cisco)

AVTEXT WG
Note Well

- IPR declaration (ID#2876) from Vidyo, which may relate to scalability, perhaps more specifically to TL0PICIDX (but draw your own conclusions).
  - [https://datatracker.ietf.org/ipr/2876/](https://datatracker.ietf.org/ipr/2876/)

- Separated sections with/without scalability
  - 3.1 Extension for Non-Scalable Streams
  - 3.2 Extension for Scalable Streams

- Authors see no issues continuing to progress this draft given the licensing terms, but the WG needs to weigh in (again, draw your own conclusions).
Review: Main Motivation

Payload-Agnostic RTP Switch

• Payload may be encrypted
  – Avoid decryption cost to improve switch scale and latency

• Payload may be encrypted end-to-end
  – Impossible to decrypt / inspect payload without end-to-end keys

• Payload may be unknown format
  – Codec-agnostic switching can support any format, old or new
Review: More Motivations

Smarter RTP Switch
• Clean video switching at intra-frames
• Better recovery during packet loss
• Drop least important packets during congestion
• Drop scalable enhancement layers for constrained endpoints

Smarter Endpoints
• Better recovery during packet loss
Video Frame Info Extension in -02

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

• **S**: Start of Frame - MUST be 1 in the first packet in a frame within a layer.
• **E**: End of Frame - MUST be 1 in the last packet in a frame within a layer.
• **I**: Independent Frame - MUST be 1 for frames that can be decoded independent of prior frames, e.g. key/intra-frame; otherwise MUST be 0.
• **D**: Discardable Frame - MUST be 1 for frames that can be dropped, and still provide a decodable media stream; otherwise MUST be 0.
• **B**: Base Layer Sync - MUST be 1 if this frame only depends on the base layer; otherwise MUST be 0.
• **TID**: Temporal ID (3 bits) - The base temporal quality starts with 0, and increases with 1 for each temporal layer/sub-layer.
• **LID**: Layer ID (8 bits) - The spatial and quality layer ID defined by scalable codecs.
• **TL0PICIDX**: Temporal Base Layer 0 Picture Index (8 bits) - Running index of base temporal layer frames and dependencies on them.
Changes in -03

• Added a short variant for non-scalable streams
  – Omits all fields which only apply to scalability
  – Documented in new section 3.1
Editorial Changes in -03

• References to relevant codec payload formats were added to the Layer ID Mapping sections 3.2.1 - 3.2.1.5.

• References to drafts were updated to RFCs.
  – Except draft-aboba-avtcore-sfu-rtp which expired
  – Options:
    • Remove this reference
    • Keep it and progress that other draft
    • Merge relevant content from that draft
Open Issues

- Resolve IPR declaration actions, if any.
- Resolve references to expired drafts.
- Layer refresh indication needed for LRR draft?
- Add frame size? (mentioned on the list)
  - RID is a better solution
  - It can signal frame size more compactly (1 vs. 4 octets), will likely already be present (so really 0 octets), and more directly matches app needs (low/mid/high resolution which the app said it cares to distinguish)
  - It can signal all types of fidelities / constraints (frame rate, bitrate, etc.) in addition to frame size
  - Recommend RID instead of adding frame size
Next Steps

• Ready for WGLC after resolving open issues.

• Questions?

• Thank you!