EV-VBR RTP payload format proposal overview

AVT WG @ IETF 70

Ari Lakaniemi
Nokia Research Center
December 3, 2007
Nokia may own IPR related to this draft. Nokia may submit a written IPR declaration, according to BCP79, pertaining to this draft, as soon as our experts have evaluated the situation.
Presentation overview

• Codec overview
• Proposed format overview
• Open issues
• Next steps
Codec overview

Scalable/layered wideband speech/audio codec (to be) specified in ITU-T SG 16

- Core codec sampling rate 16 kHz
- Core layer + 4 enhancement layers provide bit-rates 8, 12, 16, 24, 32 kbit/s
- Interoperable coding mode with the AMR-WB codec (at 12.65 kbit/s)
- The work on core codec to be completed early 2008

Work in progress (also) for extension options

- Super-wideband (SWB) providing wider audio bandwidth
- Stereo option
- To be finalized around mid-2008
Proposed payload format overview

Flexible packet/payload structure, enabling
• any number of frames per packet
• any (contiguous) subset of layers (of a frame) per packet

Enables usage with
• Single RTP session – i.e. all layers carried within a single RTP session
  • Enables low IP/UDP/RTP overhead...
  • ... but requires media-aware network elements to enable in-network scaling
• Multiple RTP sessions – i.e. subsets of layers in their own RTP sessions
  • May result in relatively high IP/UDP/RTP overhead...
  • ... but facilitates simple scaling by also media-agnostic network element

Progressive CRC checksum
• Enable dropping parts of payload WITHOUT affecting the payload checksum → simple scalability

Codec bit-rate/configuration control
• T.b.d → Currently one of the open issues
Proposed payload structure details

Overall packet structure: payload header (i.e. CRC) followed by encoded data

<table>
<thead>
<tr>
<th>RTP header</th>
<th>CRC</th>
<th>TX block(1)</th>
<th>TX block(2)</th>
<th>TX block(n)</th>
</tr>
</thead>
</table>

TX block format – the 1st (primary) TX block

<table>
<thead>
<tr>
<th>Layer ID</th>
<th>NF</th>
<th>Encoded EV-VBR data</th>
</tr>
</thead>
</table>

TX block format – a subsequent (secondary) TX block

<table>
<thead>
<tr>
<th>Layer ID</th>
<th>NF</th>
<th>Encoded EV-VBR data</th>
<th>Tail</th>
</tr>
</thead>
</table>

Data fields shortly:
- CRC (i.e. payload header): checksum over primary TX block
- Layer ID: layer configuration ID specifying the encoded data carried in this TX block
- NF: Number of frames in this TX block
- Tail: extra bits to force the progressive CRC checksum at this TX block to the desired value
Proposed payload structure examples

Examples on allocating two encoded frames into RTP packets

- $L_{x,y}$ denotes Layer x of frame y; Colors indicate TX blocks

- All layers in single packet, separate TX blocks for each layer

  \[
  \begin{array}{cccccccccccc}
  \text{RTP} & L_{1,1} & L_{1,2} & L_{2,1} & L_{2,2} & L_{3,1} & L_{3,2} & L_{4,1} & L_{4,2} & L_{5,1} & L_{5,2} \\
  \end{array}
  \]

- All layers in single packet, separate TX blocks for core and enhancement layers

  \[
  \begin{array}{cccccccccccc}
  \text{RTP} & L_{1,1} & L_{1,2} & L_{2,1} & L_{2,2} & L_{3,1} & L_{3,2} & L_{4,1} & L_{4,2} & L_{5,1} & L_{5,2} \\
  \end{array}
  \]

- Core and enhancement layers separately, two separate TX blocks for enh. layers

  \[
  \begin{array}{cccccccccccc}
  \text{RTP}_1 & L_{1,1} & L_{1,2} & & & & & & & & \\
  \text{RTP}_2 & L_{2,1} & L_{2,2} & L_{3,1} & L_{3,2} & L_{4,1} & L_{4,2} & L_{5,1} & L_{5,2} \\
  \end{array}
  \]

- All layers in separate packets

  \[
  \begin{array}{cccccccccccc}
  \text{RTP}_1 & L_{1,1} & L_{1,2} & & & & & & & & \\
  \text{RTP}_2 & L_{2,1} & L_{2,2} & & & & & & & & \\
  \text{RTP}_3 & L_{3,1} & L_{3,2} & & & & & & & & \\
  \text{RTP}_4 & L_{4,1} & L_{4,2} & & & & & & & & \\
  \text{RTP}_5 & L_{5,1} & L_{5,2} & & & & & & & & \\
  \end{array}
  \]
Open issue 1: Cross-layer/cross-session time synchronization

Possible solutions include

• Re-use RTCP based cross-session time sync mechanism (used e.g. for lip-sync)
  • Pros: Well-known and proven mechanism
  Does not require sending additional data
  • Cons: No cross-layer/cross-session sync until first RTCP SRs received (on all layers/sessions)

• Pre-synchronize the RTP timestamps in the transmitting end
  • Pros: Simple mechanism that does not require sending additional data
  Not dependent on protocol/profile behavior (e.g. timing of RTCP packets)
  • Cons: Payload specific solution (but also allows usage of RTCP based sync)

• Include synchronization data elements in the payload (e.g. a cross-layer timing reference in all/selected payloads)
  • Pros: Not dependent on protocol/profile behavior
  • Cons: Requires sending small amount of additional data within the payload
  Payload specific solution
Open issue 2: Codec bit-rate/configuration control

Current draft proposes to use RTCP-APP packet for bit-rate/configuration control
  • Based on initial feedback does not seem appropriate

Other possibilities include
  • In-band signaling
    • Pros: Enables fast feedback loop, can be tailored for the EV-VBR codec
    • Cons: Not (well) in line with the RTP framework
  • New payload specific message (RTCP packet type, AVPF FB packet)
    • Pros: Can be tailored for the codec
    • Cons: Codec specific solution
  • New payload independent (RTCP packet type, AVPF FB packet)
    • Pros: Re-use of control message(s) for several payloads
    • Cons: Covering codecs with different requirements probably a challenging task
      (→ possibly a complex/sub-optimal solution?)
  • Re-use an existing mechanism, e.g. TMMBR message of CCM
    • Pros: Existing generic solution
    • Cons: Requires usage of (S)AVPF, enables only control of bit-rate
Open issue 3: Layer configuration signaling

Layer configuration signaling in session set-up

- Capability description quite trivial, but offer/answer usage may need further considerations
- Current solution allows answer to modify the offered layer configuration (to a subset of offered layers)
  - May result in strange things in multi-session configurations
    → Should this be limited to single-session configurations?
  - Is this desirable? Is this needed? Is this useful?
- Maybe separate media parameters for sending and receiving preferences/capabilities?
Next steps

Follow-up ITU-T SG16 work on the codec and reflect possible changes in the draft

Resolve & incorporate current open issues into the draft

Accommodate SWB & stereo options into payload format once more detailed (and final) information is available
  • Note that G.729.1 codec and EV-VBR codec will share these features
    → Common format or separate payload formats for both core codecs?

Adoption as an AVT WG item sometime in the near future?