VVC/H.266 Payload Format

Shuai Zhao and Stephan Wenger, Tencent

VVC/H.266 Introduction

- Hybrid Video Codec, similar to H.264, H.265
- Standardized in JVET (ITU-T / MPEG), finalization version 1 7/2020
- Trades complexity against coding efficiency
- Complex picture segmentation mechanisms to support MTU size matching, parallelization, application driven tiling
- Native support temporal/spatial/SNR scalability, akin AV1
- Parameter Set concept – option for out-of-band transmission of data structures pertaining to more than a single picture
- NAL unit concept – NAL unit header defined by JVET co-serves as basic payload header (for most packets in naïve implementations)

VVC / H.266 Payload Format Design Principles

- Start from RFC 7798 (HEVC payload format)
  - Deployed, generally works, though lots of options not exercised
- Adapt to changes in the NAL unit header
  - Need to update current draft the latest after Jan 2020 JVET meeting
- Remove as much complexity as we can
  - Reflecting implementation practice of RFC 7798, and earlier payload formats for NAL-unit based video codecs
- Deal with transport plane first, signaling later, “required sections” last
- Target 2020 publication date, shortly after publication of VVC

Open Issues

1. Support for MRST/MRMT modes; see RFC 7656 Section 3.5
   - This relates to scalability, and how scalable layers are transported
   - SRST: Single RTP stream on a Single media Transport
   - MRST: Multiple RTP streams on a Single media Transport
   - MRMT: Multiple RTP streams on Multiple media Transports
2. Need for PAyload Content Information (PACI) packets akin RFC 7798
   - Payload Header extension mechanism
   - RFC 7798 includes one instantiation: Temporal Scalability Control Information
3. Interleaving Support
   - RFC 7798 allowed for interleaving of packets, and DOND-based signaling), mostly for error resilience. Should that be removed for cleaner IP situation?
4. Adopt as AVTCORE WG Draft?

Support for MRST, MRMT 1/2

- SRST: All layers are conveyed in a single RTP session single transport (identified by 5 tuple). SFU removes layers by acting as RTP mixer with SRST input, multiple SRST outputs tailored towards receiver capabilities
- MRST: Layers are conveyed as individual RTP streams (identified by SSRC) in a single transport (5 tuple). RFC 7798 (HEVC payload format) never fully specified the SDP signaling for MRST support—unusable in IETF protocol context (outside rtcweb). (Jury is out whether it would work inside rtcweb, where we have SSRC multiplexing)
- MRMT: Layers are conveyed as individual RTP streams in multiple transports (5 tuples), typically envisioned as IP multicast groups. Enables original "Receiver-driven layered multicast" concept. Requires codepoints with associated IP (receiver buffer size signaling).

Support for MRST, MRMT 2/2

- MRST has never been implemented in RFC 7798 context, nor in any other scalable video coding context we are aware of
- Need to define SDP signaling as mentioned in RFC 7798. However, that's easier now compared to RFC 7798-times due to SSRC multiplexing work in rtcweb
- **Q1: Should we include MRST now? Companion document? -bis?**
- **Q.2: General preference of lean-and-mean vs. comprehensive and profiled (as in RFC 7798)?**
- **Q.3: MRMT: purely academic exercise?**
- **Q.4: Does anyone here still care about receiver-driven layered multicast?**

Need for PACI

- History: When developing RFC 7798 we found a need to provide shortcut info about temporal layer structures easily accessible for FSUs. We put this into an optional extension of the RTP payload header.
- While we were at it, we invented an RFC7798-generic extension mechanism for RFC 7798 payload headers
- One use of this "PACI" extension was temporal layer structure signaling. No other PACI use was ever identified.
- Today, we have framemarking, serving a similar purpose as PACI, but cleaner and better in a security context
- Generic PACI, in 20/20 hindsight, was over-design
- **Q. 5: Is there a value in a generic payload header extension mechanism? Even if we know of no use for it today?**

Remove Interleaving/DOND signaling

- Raised by Yago Sanchez (Fraunhofer HHI) in private communication
- Earlier RTP payload formats for ITU-T/MPEG video specs included interleaving support
  - Key target: Error resilience, random packet losses, leave parts of a picture intact and use error concealment to obtain useful quality levels
- Better video codec designs -> less redundancy -> less likelihood of successful error concealment. Decision by JVET: forego mechanisms for in-picture error resilience.
- That has also been a general trend in implementations of H.264/H.265: forego in-picture error resilience, rely on FEC, temporal scalability, …
- Removal of DOND may also remove disclosed IP
- **Q. 6: Do we care about in-picture error resilience support in this spec?**
- **Q.7: Remove DOND + associated SDP signaling?**