Notes on H.264 mapping

H.264 splits the compressed video data into two broad categories:

1. Picture and slice data. Relates directly to a single compressed picture or piece of a picture (slice).
2. Meta Data. Data that might be useful to multiple pictures or slices, such as:
    - picture resolution
    - compression parameters
    - presentation and decoding times
    - etc.

The metadata is captured in a couple of primary data units:

- Sequence Parameter Set (SPS): Carries information necessary to re-construct a sequence of pictures/slices.
- Video Usability Information (VUI): Carries information necessary to reconstruct video from the reconstructed picture/slices.
- Picture Parameter Set (PPS): Carries information necessary to reconstruct an individual picture/slice.
- Supplemental Enhancement Information (SEI). Carries supplemental information to enhance the video. Example: Closed captioning.

In general when transmitting or packaging the H.264 compressed video, the picture data is (obviously) included. However, the Meta data (PPS, SPS and SEI) may or may not be. The H.264 standard has been written to allow the metadata to be sent separately from the picture data, as long as it arrives at the decoder sufficiently in advance. Some of it (like PPS data) could even be preconfigured and remain static throughout the encoding process, and therefore never be sent at all.

The primary part of the H.264 standard does not include metadata in the stream syntax. It only provides the syntax for the picture/slide data stream. Appendix B of the H.264 standard does provide a mechanism for carrying a complete network stream with all of the information necessary for a decoder to recreate the full video, including SPS, PPS, VUI, SEI and other data units.

Appendix B is the syntax used by MPEG2-TS to carry H.264. The Blu-Ray discs, DVB and the ATSC use this mapping for carrying H.264 in MPEG2-TS. The SCTE uses a restricted form of appendix B for transmission on US cable systems. It restricts the time between I- or IDR-frames (contained in an SCTE Random Access Package (SRAP)) to be less than 1.0 seconds, in order to provide reasonable channel change times.

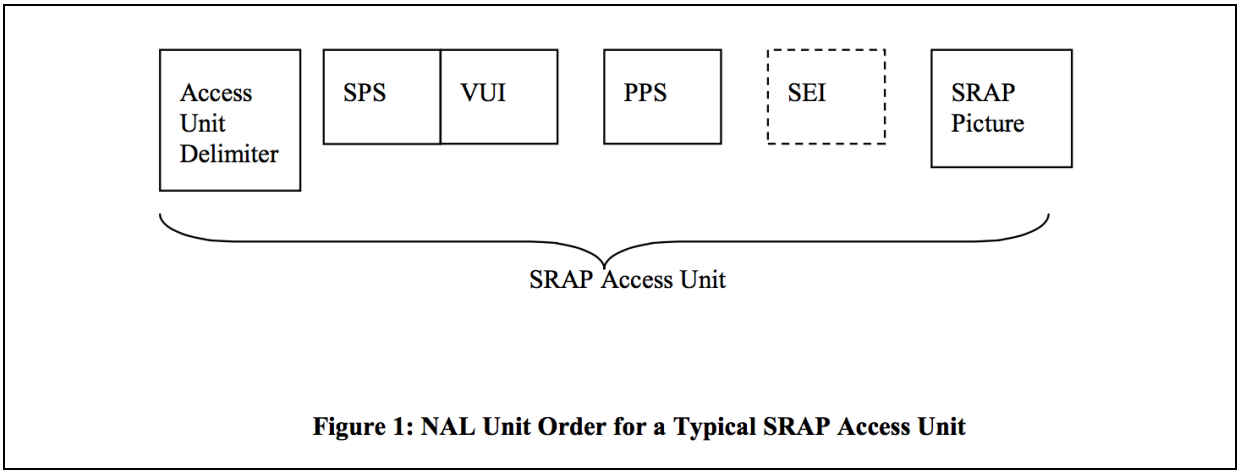

Figure 1: NAL Unit Order for a Typical SRAP Access Unit

The SCTE SRAP

What does this have to do with 1722? RTP does not encapsulate an appendix 2 byte stream. It builds its own encapsulation of picture data and meta data.

3. Scope

This payload specification can only be used to carry the "naked" H.264 NAL unit stream over RTP and not the bitstream format discussed in Annex B of H.264. Likely, the first applications of this specification will be in the conversational multimedia field, video telephony or video conferencing, but the payload format also covers other applications, such as Internet streaming and TV over IP.

RTP Scope

RTP carries the picture data and the metadata using several different payload structures working in conjunction with key items in the RTP header. In particular the M and timestamp fields of the RTP header are important when using an AVB network for transmission.

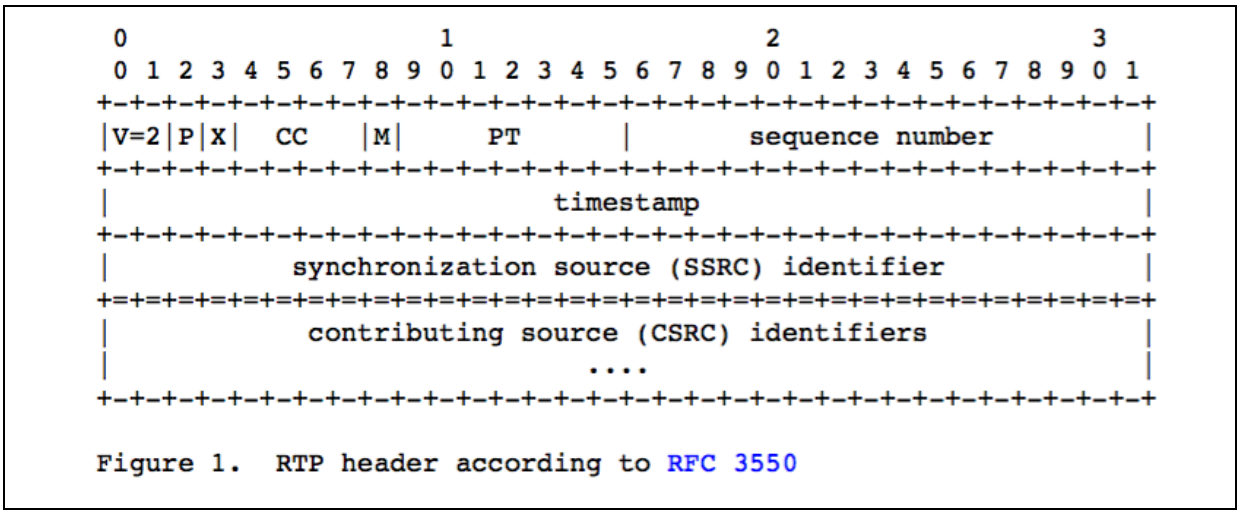

Figure 1. RTP header according to RFC 3550

RTP Header

- M: Indicates the final packet of an H.264 access unit.
- timestamp: Indicates the *sampling timestamp* of the content packet. It's value is very dependent on the type of packet being carried in the RTP payload. There are a lot of rules. As Christian has already pointed out, sometimes this timestamp arrives out of order.
- Sequence number: This is used by the RTP decoder to determine the order it needs to process the packets. It does not need to be transmitted in an AVB network, since packets will always arrive in order. Note that the listener needs to recreate a valid series of sequence IDs when delivering RTP packets to its RTP decoder.

Conclusion: It's important that we keep the RTP timestamp *intact* in the AVTP mapping for H.264 video.