SMPTE 2021 ATC Manuscript template

Standards & Specifications for Carriage of JPEG XS in RTP for IP Networks

Thomas Edwards

Amazon Web Services

Written for presentation at the

SMPTE 2021 Annual Technical Conference & Exhibition

Abstract. *JPEG XS is a low-latency, low-complexity wavelet codec that is promising for IP transport of professional media both on-premises and on the cloud. A combination of standards and specifications, including SMPTE ST 2110-22, ISO/IEC 21122-3, IETF RFC 9134, and VSF TR-08 define the transport of JPEG XS in RTP over IP. This paper describes the essential details of that transport.*

Keywords. *JPEG XS, IP, RTP, TR-08, RFC 9134, ST 2110, ISO/IEC 21122, cloud*

Description of JPEG XS Codec

JPEG XS (ISO/IEC 21122) is a low-latency, low-complexity wavelet codec optimized for visually lossless compression (as evaluated per ISO/IEC 29170-2 [1]) for both natural and synthetic images. It uses a wavelet transform to separate the image into coefficients of different spatial frequency bands. The coefficients are efficiently quantized, and the quantized values are then entropy coded.

While the basic elements of JPEG XS are similar to the JPEG 2000 codec widely used in digital cinema and WAN contribution of live events, there are several key differences. JPEG 2000 is typically implemented as a “full frame” codec, resulting in latencies of 2 or more frames. In comparison, JPEG XS can have end-to-end latency as low as 32 lines of video. JPEG 2000 is only practically implemented for live compression on specialized hardware, such as FPGAs. JPEG XS was designed such that 4K 60p video could be encoded in real time with software on an Intel Core i7 processor.

JPEG XS has very low multi-generation loss, less than 1 dB PSNR-Y’ over 10 coding/decoding cycles. JPEG XS has comparable PSNR-Y’ encoding quality to JPEG 2000 at the cost of about 20%-35% additional data rate [2]. The author has seen JPEG XS utilized for broadcast use cases by customers at 5:1 to 10:1 compression levels (110-220 Mbps for 720p59.94 format, for example).

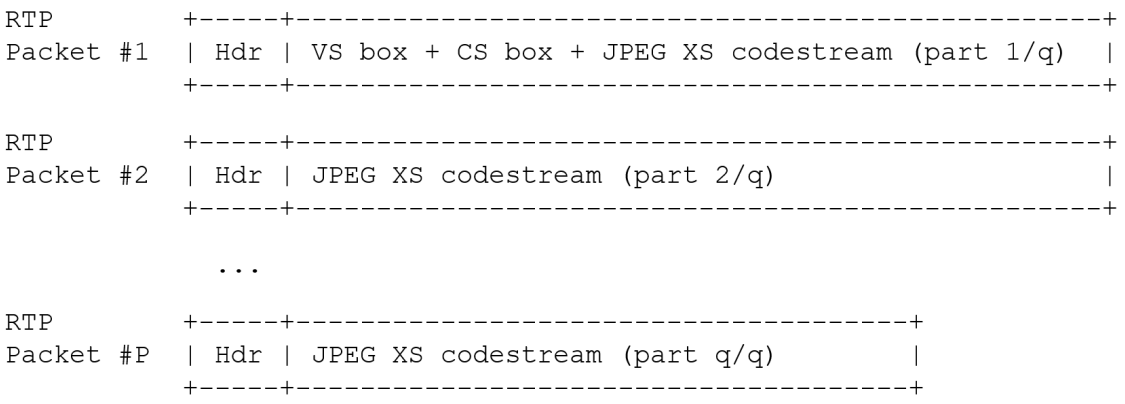
JPEG XS can be transported over IP within the SMPTE ST 2110 ecosystem as per ST 2110-22 “Professional Media Over Managed IP Networks: Constant Bit-Rate Compressed Video”. The RTP payload for JPEG XS is defined by IETF RFC 9134. Important details regarding JPEG XS transport, such as box definitions, can be found in ISO/IEC 21122-3. Furthermore, VSF TR-08 defines constraints on the transport of JPEG XS video in ST 2110-22 to improve interoperability.

Packetization

IETF RFC 2736 introduces the concept of an “Application Data Unit”, or ADU, which is a unit that contains sufficient information to be processed by the receiver immediately. RFC 9134 defines the ADU for JPEG XS as a single video frame.

An ADU is spread across several packetization units. If a packetization unit is larger than the maximum size of an RTP packet payload, the packetization unit is split across multiple RTP packets. The payload of every RTP packet has the same size, except possibly the last packet of a packetization unit.

In codestream packetization mode, the packetization unit is the entire JPEG XS picture segment. A JPEG XS picture segment is the concatenation of a video support box, a color specification box, and a JPEG XS codestream (see Figure 1). A progressive frame will have a single packetization unit, while an interlaced frame will have two packetization units.



In slice packetization mode, the packetization unit is a slice. A slice is made up of an integral number of precincts. A precinct is a collection of quantized coefficients of all spatial frequency bands contributing to a given spatial region of the image. A slice always extends over the full width of the image, but may only cover parts of its height.

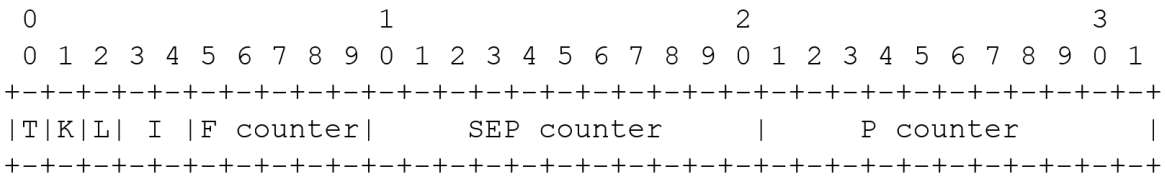
VSF TR-08 limits the packetization mode to codestream, thus this paper will not further address slice packetization mode.

RTP and Payload Header

The RTP header specified by IETF RFC 3550 is interpreted as follows:

Marker (M) [1 bit]: The marker bit is set to 1 to indicate the last packet of a video frame (for progressive video) or the last packet of a field (for interlaced video).

Timestamp [32 bits]: A 90 kHz clock rate is used for the timestamp, which designates the sampling instant of the first octet of the video frame to which the RTP packet belongs.



The JPEG XS payload header (Figure 2) consists of these fields:

Transmission mode (T) [1 bit]: The T bit is set to indicate that packets are sent sequentially by the transmitter. If T=0, nothing can be assumed about the transmission order of packets. If T=1, packets are sent sequentially by the transmitter.

packetization mode (K) [1 bit]: K=0 indicates codestream packetization mode (as required by VSF TR-08). K=1 indicates slice packetization mode.

Last (L) [1 bit]: The L bit is set to indicate the last packet of a packetization unit. In codestream packetization mode, the entire JPEG XS picture segment is a packetization unit, thus the L bit and the M bit are both set for the last packet of the frame or field, and zero in all other packets.

Interlaced information (I) [2 bit]:

00: Progressively scanned.

01: Reserved for future use.

10: The first JPEG XS picture segment of an interlaced video frame.

11: The second JPEG XS picture segment of an interlaced video frame.

F counter [5 bits]: The frame (F) counter identifies the video frame number modulo 32.

SEP counter [11 bits]: The Slice and Extended Packet (SEP) counter in codestream packetization mode increments by 1 whenever the Packet (P) counter overruns, and resets whenever the Packet counter resets.

P counter [11 bits]: The packet (P) counter identifies the packet number modulo 2048 within the current packetization unit. It is set to zero at the start of the packetization unit and incremented by 1 for every subsequent packet belonging to the same unit. For codestream packetization mode, this field practically counts the packets within a JPEG XS picture segment (as extended by the SEP counter).

Payload Data

As mentioned previously, a JPEG XS picture segment is the concatenation of a video support box, a color specification box, and a JPEG XS codestream. A “box” is a structured collection of data describing the image or the image decoding process. Figure 3 depicts an example of the payload data for codestream packetization mode and progressive video frame.

```
+=====[ Packetization unit (PU) #1 ]====+
|           Video support box           |  SEP counter=0
|  +--------------------------------+   |  P counter=0
|  :      Sub boxes of the VS box    :   |
|  +--------------------------------+   |
+- - - - - - - - - - - - - - - - - - -+
|        Color specification box        |
|  +--------------------------------+   |
|  :      Fields of the CS box       :   |
|  +--------------------------------+   |
+- - - - - - - - - - - - - - - - - - -+
|          JPEG XS codestream           |
:             (part 1/q)                :  M=0, K=0, L=0, I=00
+---------------------------------------+
|          JPEG XS codestream           |  SEP counter=0
|             (part 2/q)                |  P counter=1
:                                       :  M=0, K=0, L=0, I=00
+---------------------------------------+
|          JPEG XS codestream           |  SEP counter=0
|             (part 3/q)                |  P counter=2
:                                       :  M=0, K=0, L=0, I=00
+---------------------------------------+
:                                       :
+---------------------------------------+
|          JPEG XS codestream           |  SEP counter=1
|            (part 2049/q)              |  P counter=0
:                                       :  M=0, K=0, L=0, I=00
+---------------------------------------+
:                                       :
+---------------------------------------+
|          JPEG XS codestream           |  SEP counter=(q-1) div 2048
|             (part q/q)                |  P counter=(q-1) mod 2048
:                                       :  M=1, K=0, L=1, I=00
+=======================================+
```

mode, progressive video frame, from RFC 9134)

Video Support Box

Figure 4 shows the organization of a box. Boxes begin with LBox, a 4 byte length field, TBox, a 4 byte box type, XLBox, an 8 byte extended length only if LBox is 1. DBox is the data contained in the box.

The JPEG XS Video Support box defined in ISO/IEC 21122-3 is a superbox (i.e., a box that contains other boxes), and it contains information relevant for using JPEG XS codestreams. Its box type is 'jpvs' (0x6A70 7673), and the sub-box structure is shown in Figure 5.

The JPEG XS Video Information box (jpvi) fields are:

brat [32 bits]: max bitrate

frat [32 bits]: frame rate including interlace mode, numerator, denomerator

schar [16 bits]: sample characteristics, bit depth, sampling

tcod [32 bits]: timecode in HHMMSSFF format

The JPEG XS Profile and Level box (jxpl) fields are:

Ppig [16 bits]: JPEG XS profile

Plev [16 bits]: JPEG XS level

The other three sub-boxes are optional, including Buffer Model Description box (‘bmdm’), Mastering Display Metadata box (‘dmon’), and JPEG XS Video Transport Parameter box (‘jptp’).

Color Specification Box

The Color Specification box (Figure 6) defines one method by which an application can interpret the colorspace of the decompressed image data and identify associated processing for correct representation on the display. The type of the Color Specification box is 'colr' (0x636F 6C72).

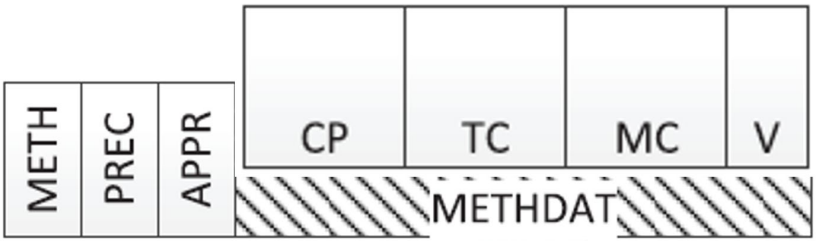


Figure 6. Fields of the Color Specification box (adapted from ISO/IEC 21122-3)

The fields of the Color Specification box are:

METH [8 bits]: Specifies the method to define the colorspace. Currently the only defined METH value is “5” which refers to the Coding Independent Code Points (CICP) as defined in Rec. ITU-T H.273, which will be found in the METHDAT field.

PREC [8 bits]: Precedence, currently undefined as shall be zero.

APPR [8 bits]: Colorspace approximation, currently shall be set to zero.

METHDAT [56 bits]: Method data for CICP. METHDAT subfields are:

CP [16 bits]: COLOUR_PRIMARIES

TC [16 bits]: TRANSFER_CHARACTERISTICS

MC [16 bits]: MATRIX_COEFFICIENTS

V [8 bits]: VIDEO_FULL_RANGE_FLAG

COLOUR_PRIMARIES, TRANSFER_CHARACTERISTICS, MATRIX_COEFFICIENTS are all 16-bit fields, with values defined as per Rec. ITU-T H.273. V is a one byte (8 bit field) that contains the VIDEO_FULL_RANGE_FLAG 1 bit flag, where the value 1 indicates full range, and the flag is in the most significant bit of the V byte. The other seven least significant bits of the V byte (CICP_RESERVED) are reserved. Table 1 shows common CICP codes.

Color space	Color primaries code	Transfer characteristics code	Matrix coefficients code	Video full range flag	Notes
Rec. ITU-R BT.709-6	1	1	1	0	BT 709 SDR
Rec. ITU-R BT.2100-2	9	16	9 (Y’CbCr)	0	PQ HDR with BT 2020
Rec. ITU-R BT.2100-2	9	18	9 (Y’CbCr)	0	HLG with BT 2020

Table1. Common CICP Codes (from VSF TR-08)

JPEG XS Codestream

The JPEG XS codestream consists of three types of syntax elements: markers, marker segments and entropy coded data. Markers serve to identify the various structural parts of the codestream. Most markers start marker segments, which contain control information. Marker

segments contain a 16 bit length field immediately after the marker itself. Table 2 is an overview of the codestream format for a JPEG XS Picture Segment.

Syntax Element	Syntax Type	Code (if marker)
Start of Codestream (SOC)	marker	0xff10
Capability Marker	marker segment	0xff50
Picture Header	marker segment	0xff12
Component Table	marker segment	0xff13
Weights Table	marker segment	0xff14
Extension Marker	marker segment	0xff15
Loop over Slices {		
Slice Header	marker segment	0xff20
Loop over Precincts {		
Precinct Header	entropy coded data	
Loop over Packets {		
Packet Header	entropy coded data	
Packet Body }	entropy coded data	
Fill()		
}} End of Codesteam (EOC)	marker	0xff11

Table 2. JPEG XS Codestream Format

Additional VSF TR-08 Constraints

VSF TR-08 contains several additional constraints on the JPEG XS codestream to enhance interoperability. For example, the JPEG XS profile must be “High444.12” as specified in ISO/IEC 21122-2. There are to be 5 horizontal wavelet transforms, 2 vertical wavelet transforms, and only a uniform quantizer is to be used. TR-08 also calls for specific limitations on minimum (4 bits per pixel, or bpp) and maximum (1.5 bpp) compression ratios, and requires the number of bytes of Payload Data in a packet to be a multiple of 8 bytes. Additional constraints can be found in the TR-08 document.

Conclusions

JPEG XS represents an excellent compromise between the incredible efficiency of JPEG 2000 with the need for a lower-latency, software-friendly codec that can run on standard PC servers and virtual machines, on-premises and in the cloud. In the development of ST 2110, the drafters felt that it was best to first address uncompressed video, as SDI quality equivalence was a goal, and that time should not be wasted on debating a codec choice. ST 2110-22 opened the ST 2110 ecosystem to CBR compressed media formats, and JPEG XS appears to be a useful codec for live production in ST 2110 systems. The development of IETF RFC 9134 has taken over 3 ½ years from the first version of Internet-draft draft-lugan-rtp-jpegxs to the final approval of the 19th version of draft-ietf-payload-rtp-jpegxs. And with the interoperability

enhancing constraints of VSF TR-08, JPEG XS is now poised for wide adoption in the professional media space.

References

[1] ISO/IEC 29170-2:2015 Information technology - Advanced image coding and evaluation - Part 2: Evaluation procedure for nearly lossless coding." https://www.iso.org/standard/66094.html

[2] Edwards, T., and Smith, M., "High Throughput JPEG 2000 for Broadcast and IP-Based Applications," SMPTE Motion Imaging Journal, vol. 130, no. 4, pp. 22-35, May 2021, doi: 10.5594/JMI.2021.3066183.

