RTP Payload Format for ISO/IEC 21122 (JPEG XS)
draft-ietf-payload-rtp-jpegxs-02

Abstract

This document specifies a Real-Time Transport Protocol (RTP) payload format to be used for transporting JPEG XS (ISO/IEC 21122) encoded video. JPEG XS is a low-latency, lightweight image coding system. Compared to an uncompressed video use case, it allows higher resolutions and frame rates, while offering visually lossless quality, reduced power consumption, and end-to-end latency confined to a fraction of a frame.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 11, 2020.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents
1. Introduction

This document specifies a payload format for packetization of JPEG XS encoded video signals into the Real-time Transport Protocol (RTP) [RFC3550].

The JPEG XS coding system offers compression and recompression of image sequences with very moderate computational resources while remaining robust under multiple compression and decompression cycles and mixing of content sources, e.g. embedding of subtitles, overlays or logos. Typical target compression ratios ensuring visually...
lossless quality are in the range of 2:1 to 10:1, depending on the nature of the source material. The end-to-end latency can be confined to a fraction of a frame, typically between a small number of lines down to below a single line.

2. Conventions, Definitions, and Abbreviations

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Application Data Unit (ADU)
The unit of source data provided as payload to the transport layer, and corresponding, in this RTP payload definition, to a single JPEG XS frame.

Colour specification box
A ISO colour specification box defined in ISO/IEC 21122-3 [ISO21122-3] that includes colour-related metadata required to correctly display JPEG XS frames, such as colour primaries, transfer characteristics and matrix coefficients.

JPEG XS codestream
A sequence of bytes representing a compressed image formatted according to JPEG XS Part 1 [ISO21122-1], except the End-Of-Codestream (EOC) marker which is omitted in this payload format.

JPEG XS codestream header
A sequence of bytes at the beginning of each JPEG XS codestream encoded in multiple markers and marker segments that does not carry entropy coded data, but metadata such as the frame dimension and component precision.

JPEG XS frame
The concatenation of a video support box, as defined in JPEG XS Part 3 [ISO21122-3], a colour specification box, as defined as well in JPEG XS Part 3 [ISO21122-3] and a JPEG XS codestream.

JPEG XS header segment
The concatenation of a video support box, as defined in JPEG XS Part 3 [ISO21122-3], a colour specification box, as defined as well in JPEG XS Part 3 [ISO21122-3] and a JPEG XS codestream header.

JPEG XS stream
A sequence of JPEG XS frames
A two-byte functional sequence that is part of a JPEG XS codestream starting with a 0xff byte and a subsequent byte defining its function.

Marker segment
A marker along with a 16-bit marker size and payload data following the size.

Slice
The smallest independently decodable unit of a JPEG XS codestream, bearing in mind that it decodes to wavelet coefficients which still require inverse wavelet filtering to give an image.

SOC marker
A marker that consists of the two bytes 0xff10 indicating the start of a JPEG XS codestream.

Video support box
A ISO video support box defined in ISO/IEC 21122-3 [ISO21122-3] that includes metadata required to play back a JPEG XS stream, such as its maximum bitrate, its subsampling structure, its buffer model and its frame rate.

3. Media Format Description

3.1. Image Data Structures

JPEG XS is a low-latency lightweight image coding system for coding continuous-tone grayscale or continuous-tone colour digital images.

This coding system provides an efficient representation of image signals through the mathematical tool of wavelet analysis. The wavelet filter process separates each component into multiple bands, where each band consists of multiple coefficients describing the image signal of a given component within a frequency domain specific to the wavelet filter type, i.e. the particular filter corresponding to the band.

Wavelet coefficients are grouped into precincts, where each precinct includes all coefficients over all bands that contribute to a spatial region of the image.

One or multiple precincts are furthermore combined into slices consisting of an integral number of precincts. Precincts do not cross slice boundaries, and wavelet coefficients in precincts that are part of different slices can be decoded independently from each other. Note, however, that the wavelet transformation runs across
slice boundaries. A slice always extends over the full width of the image, but may only cover parts of its height.

Each JPEG XS frame consists of a JPEG XS header segment followed by one or multiple slices completely describing a single frame.

3.2. Codestream

The overall codestream format, including the definition of all markers, is further defined in ISO/IEC 21122-1 [ISO21122-1]. It represents sample values of a single frame, bare any interpretation relative to a colour space.

3.3. Video support box and colour specification box

While the information defined in the codestream is sufficient to reconstruct the sample values of one video frame, the interpretation of the samples remains undefined by the codestream itself. This interpretation is given by the video support box and the colour specification box which contain significant information to correctly play the JPEG XS stream. The layout and syntax of these boxes, together with their content, are defined in ISO/IEC 21122-3 [ISO21122-3]. The video support box provides information on the maximum bitrate, the frame rate, the subsampling image format, the timecode of the current JPEG XS frame, the profile, level and sublevel used (as defined in ISO/IEC 21122-2 [ISO21122-2]), and optionally on the buffer model and the mastering display metadata. The colour specification box indicates the colour primaries, transfer characteristics, matrix coefficients and video full range flag needed to specify the colour space of the video stream.

4. Payload Format

This section specifies the payload format for JPEG XS streams over the Real-time Transport Protocol (RTP) [RFC3550].

In order to be transported over RTP, each JPEG XS stream is transported in a distinct RTP stream, identified by a distinct SSRC.

A JPEG XS stream is divided into Application Data Units (ADUs), each ADU corresponding to a single JPEG XS frame.

An ADU is split into multiple RTP packet payloads. Figure 1 shows an example of how a JPEG XS frame fits into the payload of RTP packets ("Hdr" denotes a RTP packet header). As seen there, each packet contains either part of the JPEG XS header segment or part of a single slice. Both may extend over multiple packets. The payload of every packet shall have the same size (based e.g. on the Maximum
Transfer Unit of the network), except (possibly) the last packet of the JPEG XS header segment or a slice. The boundaries of the JPEG XS header segment and of every slice shall coincide with the boundaries of the payload of a packet, i.e. the first (resp. last) byte of the JPEG XS header segment or a slice shall be the first (resp. last) byte of the payload.

RTP
Packet #1 | Hdr | JPEG XS header segment |
          +------------------------+
RTP
Packet #2 | Hdr | Slice 0                  |
          +------------------------+
RTP
Packet #3 | Hdr | Slice 1 (part 1/3)       |
          +------------------------+
RTP
Packet #4 | Hdr | Slice 1 (part 2/3)       |
          +------------------------+
RTP
Packet #5 | Hdr | Slice 1 (part 3/3)       |
          +------------------------+

... ...

RTP
Packet #P | Hdr | Slice N-1 (part q/q)     |
          +------------------------+

Figure 1: Example of ADU defining a single JPEG XS frame

4.1. Payload Header

Figure 2 illustrates the RTP payload header used in order to transport a JPEG XS stream.
The version (V), padding (P), extension (X), CSRC count (CC), sequence number, synchronization source (SSRC) and contributing source (CSRC) fields follow their respective definitions in RFC 3550 [RFC3550].

The timestamp SHOULD be based on a 90 kHz clock reference.

As per specified in RFC 3550 [RFC3550] and RFC 4175 [RFC4175], the RTP timestamp designates the sampling instant of the first octet of the frame to which the RTP packet belongs. Packets shall not include data from multiple frames, and all packets belonging to the same frame shall have the same timestamp. Several successive RTP packets will consequently have equal timestamps if they belong to the same frame (that is until the marker bit is set to 1, marking the last packet of the frame), and the timestamp is only increased when a new frame begins.

If the sampling instant does not correspond to an integer value of the clock, the value shall be truncated to the next lowest integer, with no ambiguity.

The remaining fields are defined as follows:

Marker (M) [1 bit]:
The M bit is used to indicate the last packet of a frame. This enables a decoder to finish decoding the frame, where it otherwise may need to wait for the next packet to explicitly know that the frame is finished.
Payload Type (PT) [7 bits]:
A dynamically allocated payload type field that designates the payload as JPEG XS video.

Last (L) [1 bit]:
The L bit is set to indicate the last packet of the JPEG XS header segment or a slice. It enables the decoder to already start decoding a slice without having to wait for the full frame to finish, and thus allows low-latency decoding. As the end of the frame also ends the packet containing the last slice of the frame, the L bit is set whenever the M bit is set.

Frame counter [7 bits]:
This field identifies the frame number modulo 128 to which a packet belongs. Frame numbers increment by 1 for each frame transmitted. The frame number, in addition to the time stamp, may help the decoder to manage its input buffer and to bring packets back into their natural order.

Slice counter [12 bits]:
This field identifies the slice modulo 4096 to which the packet contributes. If the data belongs to the JPEG XS header segment, this field shall have its maximal value, namely 4095 = 0xFFF. Otherwise, it is the slice index modulo 4096. Slice indices count from 0 at the top of the frame to their maximum number.

Packet counter [12 bits]:
This field identifies the packet number modulo 4096 within the JPEG XS header segment or a slice. The packet counter is set to 0 at the start of the JPEG XS header segment and incremented by 1 for every subsequent packet (if any) of this JPEG XS header segment. The packet counter is then reset to 0 at the start of every slice, and incremented by 1 for every packet that contributes to the same slice.

4.2. Payload Data

The payload data of a JPEG XS RTP stream consists of a concatenation of multiple JPEG XS frames.

Each JPEG XS frame is the concatenation of a JPEG XS header segment followed by one or several slices completely defining a single frame. Figure 3 depicts this layout.
### JPEG XS Payload Data

---

**Video support box**

- Slice counter = 0x0fff

**Colour specification box**

- Specification method (METH = 5)
- Additional fields of the cs-box

**JPEG XS codestream header**

- SOC marker
- Additional Marker segments

---

**Slices**

- Slice counter = 0
- Slice counter = 1
- Slice counter = N-1

---

**Figure 3: JPEG XS Payload Data**
4.3. Traffic Shaping and Delivery Timing

The traffic shaping and delivery timing shall be in accordance with the Network Compatibility Model compliance definitions specified in SMPTE ST 2110-21 [SMPTE-ST2110-21] for either Narrow Linear Senders (Type NL) or Wide Senders (Type W). The session description shall include a format-specific parameter of either TP=2110TPNL or TP=2110TPW to indicate compliance with Type NL or Type W respectively.

NOTE: The Virtual Receiver Buffer Model compliance definitions of ST 2110-21 do not apply.

5. Congestion Control Considerations

Congestion control for RTP SHALL be used in accordance with RFC 3550 [RFC3550], and with any applicable RTP profile: e.g., RFC 3551 [RFC3551]. An additional requirement if best-effort service is being used is users of this payload format MUST monitor packet loss to ensure that the packet loss rate is within acceptable parameters. Circuit Breakers [RFC8083] is an update to RTP [RFC3550] that defines criteria for when one is required to stop sending RTP Packet Streams and applications implementing this standard MUST comply with it. RFC 8085 [RFC8085] provides additional information on the best practices for applying congestion control to UDP streams.

6. Payload Format Parameters

6.1. Media Type Definition

Type name: video
Subtype name: jxsv

Required parameters:

rate: The RTP timestamp clock rate. Applications using this payload format SHOULD use a value of 90000.

Optional parameters:

profile: The JPEG XS profile in use, as defined in ISO/IEC 21122-2 (JPEG XS Part 2) [ISO21122-2].

level: The JPEG XS level in use, as defined in ISO/IEC 21122-2 (JPEG XS Part 2) [ISO21122-2].
sublevel: The JPEG XS sublevel in use, as defined in ISO/IEC 21122-2 (JPEG XS Part 2) [ISO21122-2].

sampling: Signals the colour difference signal sub-sampling structure.

Signals utilizing the non-constant luminance Y’C’B C’R signal format of Recommendation ITU-R BT.601-7, Recommendation ITU-R BT.709-6, Recommendation ITU-R BT.2020-2, or Recommendation ITU-R BT.2100 shall use the appropriate one of the following values for the Media Type Parameter "sampling":

- YCbCr-4:4:4 (4:4:4 sampling)
- YCbCr-4:2:2 (4:2:2 sampling)
- YCbCr-4:2:0 (4:2:0 sampling)

Signals utilizing the Constant Luminance Y’C’BC C’RC signal format of Recommendation ITU-R BT.2020-2 shall use the appropriate one of the following values for the Media Type Parameter "sampling":

- CLYCbCr-4:4:4 (4:4:4 sampling)
- CLYCbCr-4:2:2 (4:2:2 sampling)
- CLYCbCr-4:2:0 (4:2:0 sampling)

Signals utilizing the constant intensity I CT CP signal format of Recommendation ITU-R BT.2100 shall use the appropriate one of the following values for the Media Type Parameter "sampling":

- ICTCp-4:4:4 (4:4:4 sampling)
- ICTCp-4:2:2 (4:2:2 sampling)
- ICTCp-4:2:0 (4:2:0 sampling)

Signals utilizing the 4:4:4 R’ G’ B’ or RGB signal format (such as that of Recommendation ITU-R BT.601, Recommendation ITU-R BT.709, Recommendation ITU-R BT.2020, Recommendation ITU-R BT.2100, SMPTE ST 2065-1 or ST 2065-3) shall use the following value for the Media Type Parameter sampling.

- RGB RGB or R’ G’ B’ samples

Signals utilizing the 4:4:4 X’ Y’ Z’ signal format (such as defined in SMPTE ST 428-1) shall use the following value for the Media Type Parameter sampling.

- XYZ X’ Y’ Z’ samples
Key signals as defined in SMPTE RP 157 shall use the value key for the Media Type Parameter sampling. The Key signal is represented as a single component.

**KEY** samples of the key signal

- **depth**: Determines the number of bits per sample. This is an integer with typical values including 8, 10, 12, and 16.
- **width**: Determines the number of pixels per line. This is an integer between 1 and 32767.
- **height**: Determines the number of lines per frame. This is an integer between 1 and 32767.
- **exactframerate**: Signals the frame rate in frames per second. Integer frame rates shall be signaled as a single decimal number (e.g. "25") whilst non-integer frame rates shall be signaled as a ratio of two integer decimal numbers separated by a "forward-slash" character (e.g. "30000/1001"), utilizing the numerically smallest numerator value possible.

**colorimetry**: Specifies the system colorimetry used by the image samples. Valid values and their specification are:

- **BT601-5**: ITU Recommendation BT.601-5
- **BT709-2**: ITU Recommendation BT.709-2
- **SMPTE240M**: SMPTE standard 240M
- **BT601**: as specified in Recommendation ITU-R BT.601-7
- **BT709**: as specified in Recommendation ITU-R BT.709-6
- **BT2020**: as specified in Recommendation ITU-R BT.2020-2
- **BT2100**: as specified in Recommendation ITU-R BT.2100 Table 2 titled "System colorimetry"
- **ST2065-1**: as specified in SMPTE ST 2065-1 Academy Color Encoding Specification (ACES)
- **ST2065-3**: as specified for Academy Density Exchange Encoding (ADX) in SMPTE ST 2065-3
- **XYZ**: as specified in ISO 11664-1 section titled "1931 Observer"

Signals utilizing the Recommendation ITU-R BT.2100 colorimetry should also signal the representational range using the optional parameter RANGE defined below.

**interlace**: If this OPTIONAL parameter name is present, it indicates that the video is interlaced. If this parameter name is not present, the progressive video format shall be assumed.
TCS: Transfer Characteristic System. This parameter specifies the transfer characteristic system of the image samples. Valid values and their specification are:

- **SDR** (Standard Dynamic Range) Video streams of standard dynamic range, that utilize the OETF of Recommendation ITU-R BT.709 or Recommendation ITU-R BT.2020. Such streams shall be assumed to target the EOTF specified in ITU-R BT.1886.
- **PQ** Video streams of high dynamic range video that utilize the Perceptual Quantization system of Recommendation ITU-R BT.2100
- **HLG** Video streams of high dynamic range video that utilize the Hybrid Log-Gamma system of Recommendation ITU-R BT.2100

RANGE: This parameter should be used to signal the encoding range of the sample values within the stream. When paired with ITU Rec BT.2100 colorimetry, this parameter has two allowed values NARROW and FULL, corresponding to the ranges specified in table 9 of ITU Rec BT.2100. In any other context, this parameter has three allowed values: NARROW, FULLPROTECT, and FULL, which correspond to the ranges specified in SMPTE RP 2077. In the absence of this parameter, NARROW shall be the assumed value in either case.

Encoding considerations:
This media type is framed and binary; see Section 4.8 in RFC 6838 [RFC6838].

Security considerations:
Please see the Security Considerations section in RFC XXXX

### 6.2. Mapping to SDP

#### 6.2.1. General

A Session Description Protocol (SDP) object shall be created for each RTP stream and it shall be in accordance with the provisions of SMPTE ST 2110-10 [SMPTE-ST2110-10].

The information carried in the media type specification has a specific mapping to fields in the Session Description Protocol (SDP), which is commonly used to describe RTP sessions.
6.2.2. Media type and subtype

The media type ("video") goes in SDP "m=" as the media name.

The media subtype ("jxsv") goes in SDP "a=rtpmap" as the encoding name, followed by a slash ("/"), and the required parameter "rate" corresponding to the RTP timestamp clock rate (which for the payload format defined in this document MUST be 90000). The optional parameters go in the SDP "a=fmtp" attribute by copying them directly from the MIME media type string as a semicolon-separated list of parameter=value pairs.

A sample SDP mapping for JPEG XS video is as follows:

m=video 30000 RTP/AVP 112
a=rtpmap:112 jxsv/90000
  a=fmtp:112 sampling=YCbCr-4:2:2; width=1920; height=1080;
  depth=10; colorimetry=BT709; TCS=SDR;
  RANGE=FULL; TP=2110TPNL

In this example, a JPEG XS RTP stream is being sent to UDP destination port 30000, with an RTP dynamic payload type of 112 and a media clock rate of 90000 Hz. Note that the "a=fmtp:" line has been wrapped to fit this page, and will be a single long line in the SDP file.

6.2.3. Traffic shaping

The SDP object shall include the TP parameter (either 2110TPNL or 2110TPW as specified in Section 4.3) and may include the CMAX parameter as specified in SMPTE ST 2110-21 [SMPTE-ST2110-21].

6.2.4. Offer/Answer Considerations

The following considerations apply when using SDP offer/answer procedures [RFC3264] to negotiate the use of the JPEG XS payload in RTP:

- The "encode" parameter can be used for sendrecv, sendonly, and recvonly streams. Each encode type MUST use a separate payload type number.
- Any unknown parameter in an offer MUST be ignored by the receiver and MUST NOT be included in the answer.
7. IANA Considerations

This memo requests that IANA registers video/jxsv as specified in Section 6.1. The media type is also requested to be added to the IANA registry for "RTP Payload Format MIME types" [1].

8. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [RFC3550] and in any applicable RTP profile such as RTP/AVP [RFC3551], RTP/AVPF [RFC4585], RTP/SAVP [RFC3711], or RTP/SAVPF [RFC5124]. This implies that confidentiality of the media streams is achieved by encryption.

However, as "Securing the RTP Framework: Why RTP Does Not Mandate a Single Media Security Solution" [RFC7202] discusses, it is not an RTP payload format’s responsibility to discuss or mandate what solutions are used to meet the basic security goals like confidentiality, integrity, and source authenticity for RTP in general. This responsibility lies on anyone using RTP in an application. They can find guidance on available security mechanisms and important considerations in "Options for Securing RTP Sessions" [RFC7201]. Applications SHOULD use one or more appropriate strong security mechanisms.

This payload format and the JPEG XS encoding do not exhibit any substantial non-uniformity, either in output or in complexity to perform the decoding operation and thus are unlikely to pose a denial-of-service threat due to the receipt of pathological datagrams.

It is important to note that HD or UHDTV JPEG XS-encoded video can have significant bandwidth requirements (typically more than 1 Gbps for ultra high-definition video, especially if using high framerate). This is sufficient to cause potential for denial-of-service if transmitted onto most currently available Internet paths.

Accordingly, if best-effort service is being used, users of this payload format MUST monitor packet loss to ensure that the packet loss rate is within acceptable parameters. Packet loss is considered acceptable if a TCP flow across the same network path, and experiencing the same network conditions, would achieve an average throughput, measured on a reasonable timescale, that is not less than the RTP flow is achieving. This condition can be satisfied by implementing congestion control mechanisms to adapt the transmission rate (or the number of layers subscribed for a layered multicast...
session), or by arranging for a receiver to leave the session if the loss rate is unacceptably high.

This payload format may also be used in networks that provide quality-of-service guarantees. If enhanced service is being used, receivers SHOULD monitor packet loss to ensure that the service that was requested is actually being delivered. If it is not, then they SHOULD assume that they are receiving best-effort service and behave accordingly.

9. RFC Editor Considerations

Note to RFC Editor: This section may be removed after carrying out all the instructions of this section.

RFC XXXXX is to be replaced by the RFC number this specification receives when published.

10. References

10.1. Normative References

[ISO21122-1]

[ISO21122-2]

[ISO21122-3]


10.2. Informative References


10.3. URIs


Authors’ Addresses

Sebastien Lugan
intoPIX S.A.
Rue Emile Francqui, 9
1435 Mont-Saint-Guibert
Belgium

Phone: +32 10 23 84 70
Email: D313B41E@dynmail.crt1.net
URI: http://www.intopix.com