Internet Engineering Task Force
INTERNET-DRAFT
draft-ietf-avt-uncomp-video-01.txt

AVT WG
Ladan Gharai
Colin Perkins
USC/ISI

3 November 2002
Expires: May 2003

RTP Payload Format for Uncompressed Video

Status of this Memo

Abstract

This memo specifies a packetization scheme for encapsulating uncompressed video into a payload format for the Real-time Transport Protocol, RTP. It supports a range of standard- and high-definition video formats, including common television formats such as ITU BT.601, SMPTE 274M and SMPTE 296M. The format is designed to be extensible as new video formats are developed.

1. Introduction

This memo defines a scheme to packetize uncompressed, studio-quality, video streams for transport using RTP [RTP]. It supports a range of standard and high definition video formats, including ITU-R BT.601 [601], SMPTE 274M [274] and SMPTE 296M [296].

Formats for uncompressed standard definition television are defined by ITU Recommendation BT.601 [601] along with bit-serial and parallel interfaces in Recommendation BT.656 [656]. These formats allow both 625 line and 525 line operation, with 720 samples per digital active line, 4:2:2 color sub-sampling, and 8- or 10-bit digital representation.

The representation of uncompressed high definition television is specified in SMPTE standards 274M [274] and 296M [296]. SMPTE 274M defines a family of scanning systems with an image format of 1920x1080 pixels with progressive and interlaced scanning, while SMPTE 296M standard defines systems with an image size of 1280x720 pixels and only progressive scanning. In progressive scanning, scan lines are displayed in sequence from top to bottom of a full frame. In interlaced scanning, a frame is divided into its odd and even scan lines (called a field) and the two fields are displayed in succession.

SMPTE 274M and 296M define images with aspect ratios of 16:9, and define the digital representation for RGB and YCbCr components. In the case of YCbCr components, the Cb and Cr components are horizontally sub-sampled by a factor of two (4:2:2 color encoding).

Although these formats differ in their details, they are structurally very similar. This memo specifies a payload format to encapsulate these, and other similar, video formats for transport within RTP.

2. Conventions Used in this Document

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[2119].

3. Applicability Statement

This RTP payload format is designed to transport uncompressed, studio quality, video streams. Such content can be very high bandwidth and, by definition, is not congestion controlled. The intended use of this format is within a production facility or on a suitably connected private network that is specifically engineered to support this content. This format is NOT RECOMMENDED for use on public network links, unless

those links support appropriate quality of service guarantees. See also Section 10 "Security Considerations".

4. Payload Design

Each scan line of digital video is packetized into one or more (depending on the current MTU) RTP packets. A single RTP packet MAY also contain data for more than one scan line. Only the active samples are included in the RTP payload, inactive samples and the contents of horizontal and vertical blanking SHOULD NOT be transported. Scan line numbers are included in the RTP payload header, along with a field identifier for interlaced video.

> For SMPTE 296M format video, valid scan line numbers are from 26 through 745, inclusive. For progressive scan SMPTE 274M format video, valid scan lines are from scan line 42 through 1121 inclusive. For interlaced scan, valid scan line numbers for field one (F=0) are from 21 to 560 and valid scan line numbers for the second field (F=1) are from 584 to 1123. For ITU-R BT.601 format video, the blanking intervals defined in BT.656 are used: for 625 line video, lines 24 to 310 of field one (F=0) and 337 to 623 of the second field (F=1) are valid; for 525 line video, lines 21 to 263 of the first field, and 284 to 525 of the second field are valid. Other formats (e.g. [372]) may define different ranges of active lines.

It is desirable for the video to be both octet aligned when packetized, and to adhere to the principles of application level framing [ALF] by ensuring that the samples relating to a single pixel are not fragmented across two packets.

Samples may be transfered as 8, 10, 12 or 16 bit values. For 10 bit and 12 bit payloads, care must be taken to pack an appropriate number of samples per packet, such that the payload is also octet aligned. For RGB video, it is desirable that the samples corresponding to a single pixel are not fragmented across packets. Similarly, for YCrCb video, it is desirable that luminance and chrominance values are not fragmented across packets.

For example, in YCrCb video with 4:2:0 color subsampling, each group of 4 pixels is represented by 6 values, Y1 Y2 Y3 Y4 Cr Cb. These should be packetized such that these values are not fragmented across a packet boundary. With 10 bit words this is a 60 bit value which is not octet aligned. To be both octet aligned, and appropriately framed, pixels must be framed in 2 groups of 4, thereby becoming octet aligned on a 15 octet

boundary. This length is referred to as the pixel group ("pgroup"), and it is conveyed in the SDP parameters. Tables 1 to 4 display the pgroup values for a range of color samplings and word lengths.

Color Subsampling	Pixels	10 bit words #words	octet alignment	pgroup
monochrome	4	4x10	40/8 = 5	5
4:2:0	4	6x10	2x60/8 = 15	15
4:2:2	2	4x10	40/8 = 5	5
4:4:4	1	3x10	4x30/8 = 15	15
4:4:4:4	1	4x10	40/8 = 5	5

Table 1: pgroup values for 10 bit sampling

Color Subsampling	Pixels	8 bit words #words	octet alignment	pgroup
monochrome	1	1x8	8/8 = 1	1
4:2:0	4	6x8	6x8/8 = 6	6
4:2:2	2	4x8	4x8/8 = 8	4
4:4:4	1	3x8	3x8/8 = 3	3
4:4:4:4	1	4x8	4x8/8 = 4	4

Table 2: pgroup values for 8 bit sampling

Color Subsampling	Pixels	12 bit words #words	octet alignment	pgroup
monochrome	2	2x12	2x12/8 = 3	3
4:2:0	4	6x12	72/8 = 9	9
4:2:2	2	4x12	48/8 = 6	6
4:4:4	2	6x12	2x36/8 = 9	9
4:4:4:4	1	4x12	48/8 = 6	6

Table 3: pgroup values for 12 bit sampling

Color Subsampling	Pixels	16 bit words #words	octet alignment	pgroup
monochrome	1	1x16	16/8 = 2	2
4:2:0	4	6x16	6x16/8 = 12	12
4:2:2	2	4x16	4x16/8 = 8	8
4:4:4	1	3x16	3x16/8 = 6	6
4:4:4:4	1	4x16	4x16/8 = 8	8

Table 4: pgroup values for 16 bit sampling

When packetizing digital active line content, video data MUST NOT be fragmented within a pgroup.

Video content is almost always associated with additional information such as audio tracks, time code, etc. In professional digital video applications this data is commonly embedde d in non-video portions of the data stream (horizontal and vertical blanking periods) so that precise and robust synchronization is maintained. This payload format envisions that applications requiring such synchronized ancillary data should deliver it in separate RTP sessions which operate concurrently with the video session. The normal RTP mechanisms SHOULD be used to synchronize the media.

5. RTP Packetization

The standard RTP header is followed by an 8 octet payload header for each line (or partial line) of video included. One or more lines, or partial lines, of payload data follow. For example, if two lines of video are encapsulated, the payload format will be as shown in Figure 1.

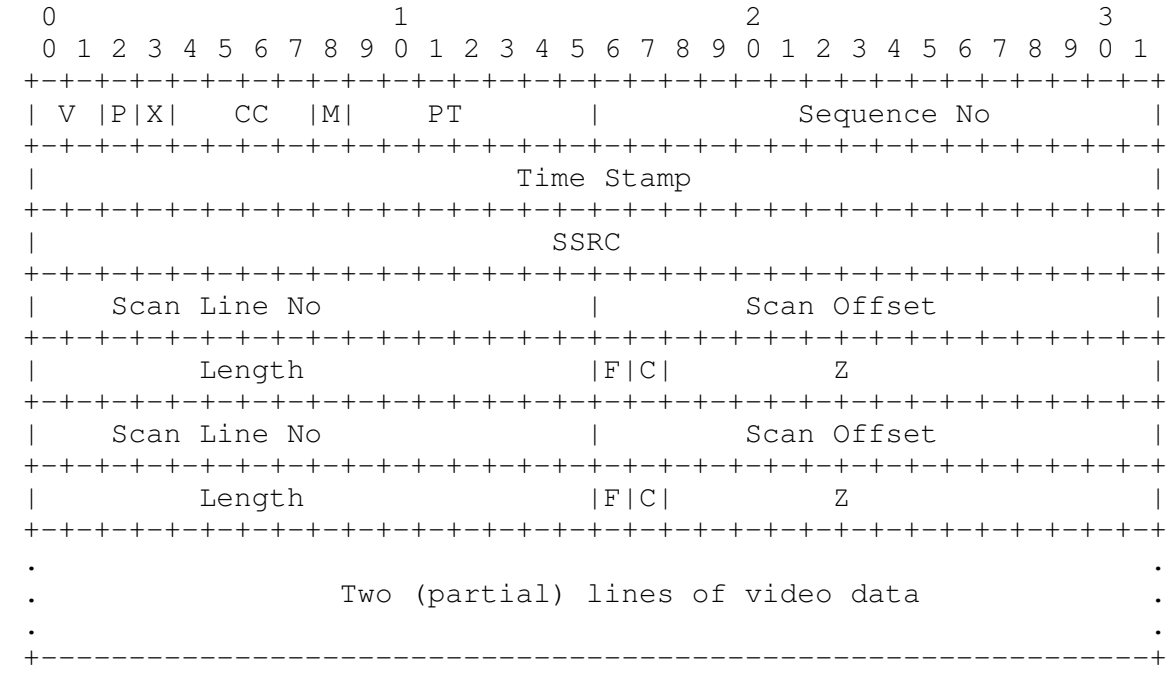

Figure 1: RTP Payload Format showing two (partial) lines of video

5.1. The RTP Header

The fields of the fixed RTP header have their usual meaning, with the following additional notes:

Payload Type (PT): 7 bits

A dynamically allocated payload type field which designates the payload as uncompressed video.

Timestamp: 32 bits

A 90 kHz timestamp MUST be used to denote the sampling instant of the video frame to which the RTP packet belongs. Packets MUST NOT include data from multiple frames, and all packets belonging to the same frame MUST have the same timestamp.

TBD: Consider whether the two fields of interlaced video MAY have distinct timestamps. In some ways this is more "natural" for true interlaced video and distinguishes it from "progressive segmented frame" (PsF) mode in which the two fields really do refer to the same time instant.

Marker bit (M): 1 bit

The Marker bit denotes the end of a video frame, and MUST be set to 1 for the last packet of the video frame. It MUST be set to 0 for other packets.

5.2. Payload Header

Scan Line No : 16 bits

Scan line number of encapsulated data in network byte order. Successive RTP packets MAY contains parts of the same scan line (with an incremented RTP sequence number, but the same timestamp), if it is necessary to fragment a line.

Scan Offset : 16 bits

Scan offset of the first sample in the payload data. If YCrCb format data is being transported, this is the offset of the co-sited luminance sample and if RGB format data is being transported it is the offset of the red sample. The value is in network byte order, and the offset has a value of zero if the first sample in the payload corresponds to the start of the line.

Length: 16 bits

Number of octets of data included. This MUST be a multiple of the pgroup value.

Field Identification (F): 1 bit

Identifies which field the scan line belongs to, for interlaced data. F=0 identifies the the first field and F=1 the second field. For progressive scan data (e.g. SMPTE 296M format video), F MUST always be set to zero.

Continuation (more lines) bit (C): 1 bit

Determines if an additional payload header follows the current header in the RTP packet. Set to 1 if an additional header follows, implying that the RTP packet is carrying data for more than one scan line. Set to 0 otherwise.

Reserved (Z): 14 bits

These bits SHOULD be set to zero by the sender and MUST be ignored by receivers.

5.3. Payload Data

Depending on the video format, each RTP packet can include either a single complete scan line, a single fragment of a scan line, or one (or more) complete scan lines plus a fragment of a scan line. Every scan line or scan line fragment MUST begin at an octet boundary in the payload data.

If the video is in YUV format, the packing of samples into the payload depends on the color sub-sampling used. For RGB format video, there is a single packing scheme.

For RGB format video, samples are packed in order Red-Green-Blue. All samples are the same bit size, which may be 8, 10, 12, or 16 bits. If 8 bit samples are used, the pgroup is 3 octets. If 10 bit samples are used, samples from adjacent pixels are packed with no padding, and the pgroup is 15 octets (4 pixels). Refer to Tables 1 thru 4.

For RGBA format video, samples are packed in order Red-Green-Blue-Alpha. All samples are the same bit size, which may be 8, 10, 12, or 16 bits. Refer to Tables 1 thru 4.

For YUV 4:4:4 format video, samples are packed in order Cb-Y-Cr. Each sample is either an 8 bit or a 10 bit value. If 8 bit samples are used, the pgroup is 3 octets. If 10 bit samples are used, samples from adjacent pixels are packed with no padding, and the pgroup is 15 octets (4 pixels).

For YUV 4:2:2 format video, the Cb and Cr components are horizontally sub-sampled by a factor of two (each Cb and Cr samples corresponds to two Y components). Samples are packed in order Cb0-Y0-Cr0-Y1. If 8 bit samples are used, the pgroup is 4 octets. If 10 bit samples are used, the pgroup is 5 octets.

(tbd: YUV 4:2:0 format video)

It is possible that the scan line length is not evenly divisible by the number of pixels in a pgroup, so the final pixel data of a scan line does not align to either an octet or pgroup boundary. Nonetheless the payload MUST contain a whole number of pgroups; the sender MUST fill the remaining bits of the final pgroup with zero and the receiver MUST ignore the fill data. (In effect, the trailing edge of the image is black-filled to a pgroup boundary.)

6. Required Parameters

(tbd)

Parameters are: color mode (RGB/YUV), color sub-sampling (4:4:4, 4:2:2, 4:2:0), lines per frame, pixels per line, bits per sample and scan mode (progressive or interlaced). Propose to map these to SDP a=fmtp: values.

Optional parameters are: colorimetry (primaries, whitepoint, reference medium), transfer function (log, gamma, toe treatment, black offset), image orientation, capture temporal mode (field integration, frame integration, spot scan, pushbroom scan). [286], [22028]

7. RTCP Considerations

RFC1889 recommends transmission of RTCP packets every 5 seconds or at a reduced minimum in seconds of 360 divided by the session bandwidth in kilobits/seconds. At the 1.485 Gbps (uncompressed HDTV rate) the reduced minimum interval computes to 0.2ms or 4028 packets per second.

It should be noted that the sender's octet count in SR packets wraps around in 23 seconds, and that the cumulative number of packets lost wraps around in 93 seconds. This means these two fields cannot accurately represent octet count and number of packets lost since the beginning of transmission, as defined in RFC 1889. Therefore for network monitoring purposes other means of keeping track of these variables

SHOULD be used.

8. IANA Considerations

This memo defines a new RTP payload format and associated MIME type.
The MIME registration form is enclosed below:

MIME media type name: video

MIME subtype name: raw

Required parameters: rate

Optional parameters: (tbd)

Encoding considerations: Uncompressed video can be transmitted with RTP as specified in RFC XXXX

Security considerations: See section 9 of RFC XXXX

Interoperability considerations: NONE

Published specification: RFC XXXX

Applications which use this media type: Video communication.

Additional information: None

Magic number(s): None

File extension(s): None

Macintosh File Type Code(s): None

Person & email address to contact for further information:
Ladan Gharai <ladan@isi.edu>
IETF AVT working group.

Intended usage: COMMON

Author/Change controller:
Ladan Gharai <ladan@isi.edu>

9. Mapping to SDP Parameters

Parameters are mapped to SDP [SDP] as follows:

```
m=video 30000 RTP/AVP 111
a=rtpmap:111 raw/90000
a=fmtp:111 (tbd)
```

In this example, a dynamic payload type 111 is used for uncompressed video. The RTP sampling clock is 90kHz.

10. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification, and any appropriate RTP profile. This implies that confidentiality of the media streams is achieved by encryption.

This payload type does not exhibit any significant non-uniformity in the receiver side computational complexity for packet processing to cause a potential denial-of-service threat.

It is important to be note that uncompressed video can have immense bandwidth requirements (270 Mbps for standard definition video, and approximately 1 Gbps for high definition video), and is not congestion controlled. This is sufficient to cause potential for denial-of-service if transmitted onto most currently available Internet paths. Use of the payload format defined here MUST be narrowly limited to suitably connected private networks, or to networks where quality of service guarantees are available. This potential threat is common to all high rate applications without congestion control.

11. Relation to RFC 2431

In comparison with RFC 2431 this memo specifies support for a wider variety of uncompressed video, in terms of frame size, color subsampling and sample sizes. While [BT656] can transport up to 4096 scan lines and 2048 pixels per line, our payload type can support up to 64k scan lines and pixels per line. Also, RFC 2431 only address 4:2:2 YUV data, while this memo covers YUV and RGB and most common color subsampling schemes. Given the variety of video types that we cover, this memo also assumes out-of-band signaling for sample size and data types (RFC 2431 uses in band signaling).

12. Full Copyright Statement

13. Acknowledgements

The authors are grateful to Philippe Gentric and Chuck Harrison for their feedback.

14. Authors' Addresses

Ladan Gharai <ladan@isi.edu>
Colin Perkins <csp@csperkins.org>

USC Information Sciences Institute
3811 N. Fairfax Drive, #200
Arlington, VA 22203
USA

Bibliography

[274] Society of Motion Picture and Television Engineers, 1920x1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates, SMPTE 274M-1998.

[268] Society of Motion Picture and Television Engineers, File Format for Digital Moving Picture Exchange (DPX), SMPTE 268M-1994. (Currently under revision.)

[296] Society of Motion Picture and Television Engineers, 1280x720 Scanning, Analog and Digital Representation and Analog Interfaces, SMPTE 296M-1998.

[372] Society of Motion Picture and Television Engineers, Dual Link 292M Interface for 1920 x 1080 Picture Raster, SMPTE 372M-2002.

[2119] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119.

[ALF] Clark, D. D., and Tennenhouse, D. L., "Architectural Considerations for a New Generation of Protocols", In Proceedings of SIGCOMM '90 (Philadelphia, PA, Sept. 1990), ACM.

[SDP] M. Handley and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.

[BT656] D. Tynan, "RTP Payload Format for BT.656 Video Encoding", Internet Engineering Task Force, RFC 2431, October 1998.

[RTP] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", Internet Engineering Task Force, RFC 1889, January 1996.

[292RTP] L. Gharai et al., "RTP Payload Format for SMPTE 292M Video", Internet Draft, draft-ietf-avt-smpte292-video-07.txt, Work in progress.

[601] International Telecommunication Union, "Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios", Recommendation BT.601, October 1995.

[656] International Telecommunication Union, "Interfaces for Digital Component Video Signals in 525-line and 625-line Television

Systems Operating at the 4:2:2 Level of Recommendation ITU-R BT.601 (Part A)", Recommendation BT.656, April 1998.

[22028] ISO TC42 (Photography), Photography and graphic technology - Extended colour encodings for digital image storage, manipulation and interchange - Part 1: Architecture and requirements, ISO/CD 22028-1, Work in Progress.