Internet Engineering Task Force
INTERNET-DRAFT
draft-gharai-avt-uncomp-video-00.txt

AVT WG
Ladan Gharai
Colin Perkins
USC/ISI

22 June 2002
Expires: December 2002

RTP Payload Format for Uncompressed Video

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Abstract

This memo specifies a packetization scheme for encapsulating uncompressed HDTV as defined by SMPTE 274M and SMPTE 296M into a payload format for the Real-Time Transport Protocol (RTP). SMPTE 274M and SMPTE 296M define the analog and digital representation of HDTV with image formats of 1920x1080 and 1280x720, respectively. The payload has been designed such that it may scale to future higher resolutions, suhc as Digital Cinema.

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 [1] and SMPTE 296M [2].

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 colour sub-sampling, and 8- or 10-bit digital representation.

The representation of uncompressed high definition television is specified in SMPTE standards 274M [1] and 296M [2]. 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[10].

3. 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 may define different ranges of active lines.

Sample values for pixels may be transfered as 8 bit or 10 bit values. For 10 bit payloads, care must be taken such that the payload is also octet aligned.

However, for video content it is desirable for the video to be both octet aligned when packetized and also adhere to the principles of application level framing [11]. For YCrCb video, the ALF principle translates into not fragmenting related luminance and chrominance values across packets. For example, with 4:2:0 color subsampling each group of 4 pixels is represented by 6 values, Y1 Y2 Y3 Y4 Cr Cb, and video content 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 and 2 display the pgroup value for 4:2:2 and 4:4:4 color samplings, for 10 bit and 8 bit words.

Color Subsampling	Pixels	10 bit words #words	octet alignment	pgroup
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

Table 1: pgroup values for 10 bit sampling

Color Subsampling	Pixels	8 bit words #words	octet alignment	pgroup
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

Table 2: pgroup values for 8 bit sampling

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

4. RTP Packetization

The standard RTP header is followed by a 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.

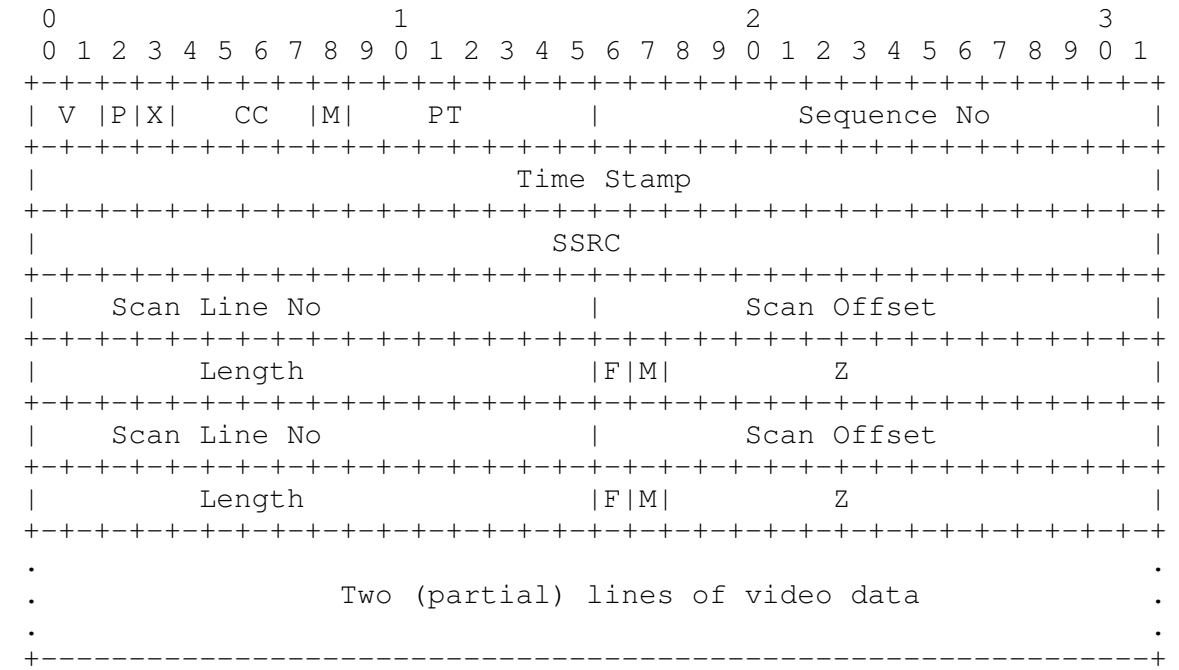

Figure 1: RTP Payload Format showing two (partial) lines of video

4.1. 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.

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.

4.2. 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

Sample number of the co-sited luminance sample (if YUV format data is being transported), or the red sample (if RGB format data is transported) where the scan line is fragmented, in network byte order.

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 data (SMPTE 296M) F MUST always be set to zero.

Follow On (more lines) bit (M): 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.

4.3. 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.

If the video is in YUV format, the packing of samples into the payload depends on the colour 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. 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 4:4:4 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)

5. Required Parameters

(tbd)

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

6. 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.

7. 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>

8. Mapping to SDP Parameters

Parameters are mapped to SDP [12] 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.

9. 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 to be noted that uncompressed video can have immense bandwidth requirements (270 Mbps for standard definition video, and approximately 1 Gbps for high definition video). This is sufficient to cause potential for denial-of-service if transmitted onto most currently available Internet paths. In the absence from the standards track of a suitable congestion control mechanism for flows of this sort, use of the payload SHOULD be narrowly limited to suitably connected network endpoints, or to networks where QoS guarantees are available, and great care taken with the scope of multicast transmissions. This potential threat is common to all high bit rate applications without congestion control.

10. Relation to RFC 2431

(tbd) [13]

11. Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works.

However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR

FITNESS FOR A PARTICULAR PURPOSE."

12. Authors' Addresses

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

USC Information Sciences Institute
3811 N. Fairfax Drive
Arlington, VA 22203-1695
USA

Bibliography

[1] Society of Motion Picture and Television Engineers, Bit-Serial Digital Interface for High-Definition Television Systems, SMPTE 292M-1998.

[2] Society of Motion Picture and Television Engineers, Digital Representation and Bit-Parallel Interface - 1125/60 High-Definition Production System, SMPTE 260M-1999.

[3] Society of Motion Picture and Television Engineers, 1920x1080 50Hz, Scanning and Interface, SMPTE 295M-1997.

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

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

[6] ISO/IEC International Standard 13818-2; "Generic coding of moving pictures and associated audio information: Video", 1996.

[7] ATSC Digital Television Standard Document A/53, September 1995, http://www.atsc.org

[8] ISO/IEC International Standard 13818-1; "Generic coding of moving pictures and associated audio information: Systems",1996.

[9] Hoffman, Fernando, Goyal, Civanlar, "RTP Payload Format for MPEG1/MPEG2 Video", RFC 2250, IETF, January 1998.

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

[11] 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.

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

[13] 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.

[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.