Network Working Group
Request for Comments: 2431
Category: Standards Track

D. Tynan
Claddagh Films

October 1998

RTP Payload Format for BT.656 Video Encoding

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (1998). All Rights Reserved.

Abstract

This document specifies the RTP payload format for encapsulating ITU Recommendation BT.656-3 video streams in the Real-Time Transport Protocol (RTP). Each RTP packet contains all or a portion of one scan line as defined by ITU Recommendation BT.601-5, and includes fragmentation, decoding and positioning information.

1. Introduction

This document describes a scheme to packetize uncompressed, studio-quality video streams as defined by BT.656 for transport using RTP [1]. A BT.656 video stream is defined by ITU-R Recommendation BT.656-3 [2], as a means of interconnecting digital television equipment operating on the 525-line or 625-line standards, and complying with the 4:2:2 encoding parameters as defined in ITU-R Recommendation BT.601-5 (formerly CCIR-601) [3], Part A.

RTP is defined by the Internet Engineering Task Force (IETF) to provide end-to-end network transport functions suitable for applications transmitting real-time data over multicast or unicast network services. The complete specification of RTP for a particular application requires the RTP protocol document [1], a profile specification document [4], and a payload format specification. This document is intended to serve as the payload format specification for studio-quality video streams.

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 [5].

2. Definitions

For the purposes of this document, the following definitions apply:

Y: An 8-bit or 10-bit coded "luminance" sample. Luminance in this context refers to the BT.601-5 [3] definition which is not the same as a true CIE luminance value. The value of "luminance" refers specifically to video luma. However, in order to avoid confusion with the BT.656 and BT.601 standards, the video luma value is referenced in this document as luminance. Each value has 220 quantization levels with the black level corresponding to level 16 and the peak white level corresponding to 235.

Cb, Cr: An 8-bit or 10-bit coded color-difference sample (as per BT.601-5). Each color-difference value has 225 quantization levels in the centre part of the quantization scale with a color-difference of zero having an encoded value of 128.

True Black: BT.601-5 defines a true black level as the quad-sample sequence 0x80, 0x10, 0x80, 0x10, representing color-difference values of 128 (0x80) and a luminance value of 16 (0x10).

SAV, EAV: Video timing reference codes which appear at the start and end of a BT.656 scan line.

3. Payload Design

ITU Recommendation BT.656-3 defines a schema for the digital interconnection of television video signals in conjunction with BT.601-5 which defines the digital representation of the original analog signal. While BT.601-5 refers to images with or without color subsampling, the interconnection standard (BT.656-3) specifically requires 4:2:2 subsampling. This specification also requires 4:2:2 subsampling such that the luminance stream occupies twice the bandwidth of each of the two color-difference streams. For normal 4:3 aspect ratio images, this results in 720 luminance samples per scan line, and 360 samples of each of the two chrominance channels. The total number of samples per scan line in this case is 1440. While this payload format specification can accomodate various image sizes and frame rates, only those in accordance with BT.601-5 are currently supported.

Due to the lack of any form of video compression within the payload and sampling-rate compliance with BT.601-5, the resultant video stream can be considered "studio quality". However, such a stream can require approximately 20 megabytes per second of network bandwidth. In order to maximize packet size within a given MTU, and to optimize scan line decoding, each video scan line is encoded within one or more RTP packets.

To allow for scan line synchronization, each packet includes certain flag bits (as defined in BT.656-3) and a unique scan line number. The SAV and EAV timing reference codes are removed. Furthermore, no line blanking samples are included, so no ancillary data can be included in the line blanking period. It is the responsibility of the receiver to generate the timing reference codes, and to insert the correct number of line blanking samples.

Similarly, there is no requirement that the frame blanking samples be provided. However, it is possible to include frame blanking samples if such samples contain relevant information, such as a vertical-interlace time code (VITC), or teletext data. In the absence of frame blanking samples, the receiver MUST generate true black levels as defined above, to complete the correct number of scan lines per field. If frame blanking samples are provided, they MUST be copied without modification into the resultant BT.656-3 stream.

Scan lines MUST be sent in sequential order. Error concealment for missing scan lines or fragments of scan lines is at the discretion of the receiver.

Both 8-bit and 10-bit quantization types as defined by BT.601-5 are supported. 10-bit samples are considered to have two extra bits of fixed-point precision such that a binary value of 10111110.11 represents a sample value of 190.75. Using 8-bit quantization, this would give a sample value of 190. An application receiving 8-bit samples for a 10-bit device MUST consider the sample as reflecting the most-significant 8 bits. The two least-significant bits SHOULD be set to zero. Similarly, an application sending 8-bit samples from a 10-bit device MUST drop the two least-significant bits. For a 10-bit quantization payload, each pair of samples MUST be encoded into a 40-bit word (five octets) prior to transmission, as specified in Section 6.

To allow for scan lines with octet lengths larger than the path maximum transmission unit (MTU), a scan offset field is included in the packet header. Applications SHOULD attempt path MTU discovery [6] and fragment scan lines into multiple packets no larger than the MTU.

Fragmentation MUST occur on a sample-pair boundary, such that the chrominance and luminance values are not split across packets. For 8-bit quantization this gives a four-octet alignment, and a five-octet alignment for 10-bit quantization. As a result, the scan offset refers not to the byte offset within the payload, but the sample-pair offset.

4. Usage of RTP

Due to the unreliable nature of the RTP protocol, and the lack of an orderly delivery mechanism, each packet contains enough information to form a single scan line without reference to prior scan lines or prior frames. In addition to the RTP header, a fixed length payload header is included in each packet. This header is four octets in length.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           RTP Header                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Payload Header                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Payload Data                         |
|                                .                              |
|                                .                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

4.1. RTP Header usage

Each RTP packet starts with a fixed RTP header. The following fields of the RTP fixed header are used for BT.656-3 encapsulation:

Marker bit (M): The Marker bit of the RTP header is set to 1 for the last packet of a frame (or the last fragment of the last scan line if fragmented), and set to 0 on all other packets.

Payload Type (PT): The Payload Type indicates the use of the payload format defined in this document. A profile MAY assign a payload type value for this format either statically or dynamically as described in RFC 1890 [4].

Timestamp: The RTP Timestamp encodes the sampling instant of the video frame currently being rendered. All scan line packets within the same frame will have the same timestamp. The timestamp SHOULD refer to the ’Ov’ field synchronization point of the first field. For the payload format defined by this document, the RTP timestamp is based on a 90kHz clock.

5. Payload Header

The payload header is a fixed four-octet header broken down as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|V| Type  |P| Z |     Scan Line (SL)    |  Scan Offset (SO)   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

F: 1 bit
When 0, indicates the first field of a frame (line 4 through 265 inclusive for Type=0 or 2, and line 1 through 312 inclusive for Type=1 or 3). Any other scan line is considered a component of the second field, and the F bit will be set to 1. This bit is copied directly from the BT.656-compliant stream by the transmitter, and inserted into the stream by the receiver.

V: 1 bit
When 1, indicates that the scan line is part of the vertical interval. Should always be 0 unless frame blanking data is sent. In which case, the V bit SHOULD be set to 1 for scan lines which do not form an integral part of the image. This bit is copied directly from the BT.656-compliant stream by the transmitter, and inserted into the stream by the receiver. For receivers which do not receive scan lines during the vertical interval, BT.656 vertical interval data MUST be generated by repeating the quad-sample sequence 0x80, 0x10, 0x80, 0x10, representing a true black level.

Type: 4 bits
This field indicates the type of frame encoding within the payload. It MUST remain unchanged for all scan lines within the same frame. Currently only four types of encoding are defined. These are as follows;

0 - NTSC (13.5MHz sample rate; 720 samples per line; 60 fields per second; 525 lines per frame)

1 - PAL (13.5MHz sample rate; 720 samples per line; 50 fields per second; 625 lines per frame)

2 - High Definition NTSC (18MHz sample rate; 1144 samples per line; 60 fields per second; 525 lines per frame)

3 - High Definition PAL (18MHz sample rate; 1152 samples per line; 50 fields per second; 625 lines per frame)

Further encodings can only be defined through the Internet Assigned Numbers Authority (IANA). For more information refer to Section 8, "IANA Considerations".

P: 1 bit
Indicates the required sample quantization size. When 0, the payload is comprised of 8-bit samples. Otherwise, it carries 10-bit samples. This bit MUST remain unchanged for all scan lines within the same frame.

Z: 2 bits
Reserved for future use. Must be set to zero by the transmitter and ignored by the receiver.

Scan Line (SL): 12 bits
Indicates the scan line encapsulated in the payload. Valid values range from 1 through 625 inclusive. If no frame blanking data is being transmitted, only scan lines 23 through 310 inclusive, and lines 336 through 623 inclusive SHOULD be sent in the case of Type=1 or 3. For 525/60 encoding (Type=0 or 2), scan lines 10 through 263 inclusive and lines 273 through 525 SHOULD be transmitted.

If a receiver is generating a BT.656-3 data stream directly from this packet, the F and V bits MUST be copied from the header rather than being generated implicitly from the scan line number. In the event of a conflict, the F and V bits have precedence.

Scan Offset (SO): 11 bits
Indicates the offset within the scan line for application-level fragmentation. After doing PMTU discovery, if the path MTU is less than the required size for one complete scan line, the data SHOULD be fragmented such that a given RTP packet does not exceed the allowable MTU. The offset for the first packet of a scan line MUST be set to zero. The scan offset refers to the sample-pair offset within the scan such that for a scan line width of 720, the maximum scan offset is 359.

6. Payload Format

In keeping with the 4:2:2 color subsampling of BT.656 and BT.601, each pair of color-difference samples will be intermixed with two luminance samples. As per BT.656, the format for transmission SHALL be Cb, Y, Cr, Y. The following is a representation of a 720 sample packet with 8-bit quantization:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Cb0      |      Y0       |      Cr0      |      Y1       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Cb1      |      Y2       |      Cr1      |      Y3       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                                .
                                .
                                .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Cb359     |     Y718      |     Cr359     |     Y719      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

1144 and 1152 sample packets SHOULD increase the packet size accordingly while maintaining the sample order.

For 10-bit quantization, each group of four samples MUST be encoded into a 40-bit word (five octets) prior to transmission. The sample order is identical to that for 8-bit quantization. The following is a representation of a 720 sample packet with 10-bit quantization:

```
            0         1         2         3
            0 2 4 6 8 0 2 4 6 8 0 2 4 6 8 0 2 4 6 8
           +---------+---------+---------+---------+
           |   Cb0   |   Y0    |   Cr0   |   Y1    |
           +---------+---------+---------+---------+
           |   Cb1   |   Y2    |   Cr1   |   Y3    |
           +---------+---------+---------+---------+
                               .
                               .
                               .
           +---------+---------+---------+---------+
           |  Cb359  |  Y718   |  Cr359  |  Y719   |
           +---------+---------+---------+---------+
             (Note that the word width is 40 bits)
           +-------+-------+-------+-------+-------+
   Octets: |   0   |   1   |   2   |   3   |   4   |
           +-------+-------+-------+-------+-------+
```

The octets shown in these diagrams are transmitted in network byte order, that is, left-to-right as shown.

7. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [1]. This implies that confidentiality of the media streams is achieved by encryption. Because the payload format is arranged end-to-end, encryption MAY be performed after encapsulation so there is no conflict between the two operations.

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.

8. IANA Considerations

The four encoding types defined by this document relate to specific schema defined by ITU-R Recommendation BT.656-3. Future revisions of the recommendation may create further encoding types which need to be supported over RTP. The "Type" field is four bits wide allowing for a total of up to sixteen possible encodings, with twelve currently reserved for future use. Due to the small number of possible encodings and given that it is very unlikely that future revisions of BT.656 will introduce any new schema, requests to extend the Type field MUST be vetted by the Internet Assigned Numbers Authority. Furthermore, implementors SHOULD check the IANA repository for new definitions of the Type field in order to comply with this document.

Applications for a new Type value MUST be submitted to the IANA and include the requestors name and contact information, the reason for requesting a new Type and references to appropriate standards, such as an updated version of ITU-R Recommendation BT.656. Furthermore, in the unlikely event that the new Type will lessen the security of a compliant implementation, such security risk MUST be detailed in the application. The application will be reviewed by a Designated Expert and if appropriate, a new Type will be assigned. This type will be listed in the IANA repository for future implementations.

9. References

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

[2] 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), ITU-R Recommendation BT.656-3, 1995.

[3] Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios, ITU-R Recommendation BT.601-5, 1995.

[4] Schulzrinne, H., "RTP Profile for Audio and Video Conference with Minimal Control", RFC 1890, January 1996.

[5] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[6] Mogul, J., and S. Deering, "Path MTU Discovery", RFC 1191, November 1990.

10. Author’s Address

Dermot Tynan
Claddagh Films Limited
3 White Oaks
Clybaun Road
Galway
Ireland

EMail: dtynan@claddagh.ie
Phone: +353 91 529944

11. Full Copyright Statement

Copyright (C) The Internet Society (1998). 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.