

Network Working Group
Request for Comments: 3189
Category: Standards Track

K. Kobayashi
Communication Research Laboratory
A. Ogawa
Keio University
S. Casner
Packet Design
C. Bormann
Universitaet Bremen TZI
January 2002

RTP Payload Format for DV (IEC 61834) Video

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 (2002). All Rights Reserved.

Abstract

This document specifies the packetization scheme for encapsulating the compressed digital video data streams commonly known as "DV" into a payload format for the Real-Time Transport Protocol (RTP).

1. Introduction

This document specifies payload formats for encapsulating both consumer- and professional-use DV format data streams into the Real-time Transport Protocol (RTP), version 2 [6]. DV compression audio and video formats were designed for helical-scan magnetic tape media. The DV standards for consumer-market devices, the IEC 61883 and 61834 series, cover many aspects of consumer-use digital video, including mechanical specifications of a cassette, magnetic recording format, error correction on the magnetic tape, DCT video encoding format, and audio encoding format [1]. The digital interface part of IEC 61883 defines an interface on an IEEE 1394 network [2,3]. This specification set supports several video formats: SD-VCR (Standard Definition), HD-VCR (High Definition), SDL-VCR (Standard Definition - Long), PALPlus, DVB (Digital Video Broadcast) and ATV (Advanced Television). North American formats are indicated with a number of lines and "/60", while European formats use "/50". DV standards

extended for professional use were published by SMPTE as 306M and 314M, for different sampling systems, higher color resolution, and faster bit rates [4,5].

There are two kinds of DV, one for consumer use and the other for professional. The original "DV" specification designed for consumer-use digital VCRs is approved as the IEC 61834 standard set. The specifications for professional DV are published as SMPTE 306M and 314M. Both encoding formats are based on consumer DV and used in SMPTE D-7 and D-9 video systems. The RTP payload format specified in this document supports IEC 61834 consumer DV and professional SMPTE 306M and 314M (DV-Based) formats.

IEC 61834 also includes magnetic tape recording for digital TV broadcasting systems (such as DVB and ATV) that use MPEG2 encoding. The payload format for encapsulating MPEG2 into RTP has already been defined in RFC 2250 [7] and others.

Consequently, the payload specified in this document will support six video formats of the IEC standard: SD-VCR (525/60, 625/50), HD-VCR (1125/60, 1250/50) and SDL-VCR (525/60, 625/50), and six of the SMPTE standards: 306M (525/60, 625/50), 314M 25Mbps (525/60, 625/50) and 314M 50Mbps (525/60, 625/50). In the future it can be extended into other high-definition formats.

Throughout this specification, we make extensive use of the terminology of IEC and SMPTE standards. The reader should consult the original references for definitions of these terms.

1.1 Terminology

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 [8].

2. DV format encoding

The DV format only uses the DCT compression technique within each frame, contrasted with the interframe compression of the MPEG video standards [9,10]. All video data, including audio and other system data, are managed within the picture frame unit of video.

The DV video encoding is composed of a three-level hierarchical structure. A picture frame is divided into rectangle- or clipped-rectangle-shaped DCT super blocks. DCT super blocks are divided into 27 rectangle- or square-shaped DCT macro blocks.

Audio data is encoded with PCM format. The sampling frequency is 32 kHz, 44.1 kHz or 48 kHz and the quantization is 12-bit non-linear, 16-bit linear or 20-bit linear. The number of channels may be up to 8. Only certain combinations of these parameters are allowed depending upon the video format; the restrictions are specified in each document.

A frame of data in the DV format stream is divided into several "DIF sequences". A DIF sequence is composed of an integral number of 80-byte DIF blocks. A DIF block is the primitive unit for all treatment of DV streams. Each DIF block contains a 3-byte ID header that specifies the type of the DIF block and its position in the DIF sequence. Five types of DIF blocks are defined: DIF sequence header, Subcode, Video Auxiliary information (VAUX), Audio, and Video. Audio DIF blocks are composed of 5 bytes of Audio Auxiliary data (AAUX) and 72 bytes of audio data.

Each RTP packet starts with the RTP header as defined in RFC 1889 [6]. No additional payload-format-specific header is required for this payload format.

2.1 RTP header usage

The RTP header fields that have a meaning specific to the DV format are described as follows:

Payload type (PT): The payload type is dynamically assigned by means outside the scope of this document. If multiple DV encoding formats are to be used within one RTP session, then multiple dynamic payload types **MUST** be assigned, one for each DV encoding format. The sender **MUST** change to the corresponding payload type whenever the encoding format is changed.

Timestamp: 32-bit 90 kHz timestamp representing the time at which the first data in the frame was sampled. All RTP packets within the same video frame **MUST** have the same timestamp. The timestamp **SHOULD** increment by a multiple of the nominal interval for one frame time, as given in the following table:

| Mode | Frame rate (Hz) | Increase of one frame in 90kHz timestamp |
|---------|-----------------|---|
| 525-60 | 29.97 | 3003 |
| 625-50 | 25 | 3600 |
| 1125-60 | 30 | 3000 |
| 1250-50 | 25 | 3600 |

When the DV stream is obtained from an IEEE 1394 interface, the progress of video frame times MAY be monitored using the SYT timestamp carried in the CIP header as specified in IEC 61883 [2].

Marker bit (M): The marker bit of the RTP fixed header is set to one on the last packet of a video frame, and otherwise, must be zero. The M bit allows the receiver to know that it has received the last packet of a frame so it can display the image without waiting for the first packet of the next frame to arrive to detect the frame change. However, detection of a frame change MUST NOT rely on the marker bit since the last packet of the frame might be lost. Detection of a frame change MUST be based on a difference in the RTP timestamp.

2.2 DV data encapsulation into RTP payload

Integral DIF blocks are placed into the RTP payload beginning immediately after the RTP header. Any number of DIF blocks may be packed into one RTP packet, except that all DIF blocks in one RTP packet must be from the same video frame. DIF blocks from the next video frame MUST NOT be packed into the same RTP packet even if more payload space remains. This requirement stems from the fact that the transition from one video frame to the next is indicated by a change in the RTP timestamp. It also reduces the processing complexity on the receiver. Since the RTP payload contains an integral number of DIF blocks, the length of the RTP payload will be a multiple of 80 bytes.

Audio and video data may be transmitted as one bundled RTP stream or in separate RTP streams (unbundled). The choice MUST be indicated as part of the assignment of the dynamic payload type and MUST remain unchanged for the duration of the RTP session to avoid complicated procedures of sequence number synchronization. The RTP sender MAY omit DIF-sequence header and subcode DIF blocks from a stream since the information is either known out-of-band or may not be required for RTP transport. When sending DIF-sequence header and subcode DIF blocks, both types of blocks MUST be included in the video stream.

DV streams include "source" and "source control" packs that carry information indispensable for proper decoding, such as aspect ratio, picture position, quantization of audio sampling, number of audio channels, audio channel assignment, and language of the audio. However, describing all of these attributes with a signaling protocol would require large descriptions to enumerate all the combinations. Therefore, no Session Description Protocol (SDP) [13] parameters for these attributes are defined in this document. Instead, the RTP sender MUST transmit at least those VAUX DIF blocks and/or audio DIF blocks with AAUX information bytes that include "source" and "source control" packs containing the indispensable information for decoding.

In the case of one bundled stream, DIF blocks for both audio and video are packed into RTP packets in the same order as they were encoded.

In the case of an unbundled stream, only the header, subcode, video and VAUX DIF blocks are sent within the video stream. Audio is sent in a different stream if desired, using a different RTP payload type. It is also possible to send audio duplicated in a separate stream, in addition to bundling it in with the video stream.

When using unbundled mode, it is RECOMMENDED that the audio stream data be extracted from the DIF blocks and repackaged into the corresponding RTP payload format for the audio encoding (DAT12, L16, L20) [11,12] in order to maximize interoperability with non-DV-capable receivers while maintaining the original source quality.

In the case of unbundled transmission where both audio and video are sent in the DV format, the same timestamp SHOULD be used for both audio and video data within the same frame to simplify the lip synchronization effort on the receiver. Lip synchronization may also be achieved using reference timestamps passed in RTCP as described in RFC 1889 [6].

The sender MAY reduce the video frame rate by discarding the video data and VAUX DIF blocks for some of the video frames. The RTP timestamp must still be incremented to account for the discarded frames. The sender MAY alternatively reduce bandwidth by discarding video data DIF blocks for portions of the image which are unchanged from the previous image. To enable this bandwidth reduction, receivers SHOULD implement an error concealment strategy to accommodate lost or missing DIF blocks, e.g., repeating the corresponding DIF block from the previous image.

3. SDP Signaling for RTP/DV

When using SDP (Session Description Protocol) [13] for negotiation of the RTP payload information, the format described in this document SHOULD be used. SDP descriptions will be slightly different for a bundled stream and an unbundled stream.

When a DV stream is sent to port 31394 using RTP payload type identifier 111, the m=??line will be like:

```
m=video 31394 RTP/AVP 111
```

The a=rtpmap attribute will be like:

```
a=rtpmap:111 DV/90000
```

"DV" is the encoding name for the DV video payload format defined in this document. The "90000" specifies the RTP timestamp clock rate, which for the payload format defined in this document is a 90kHz clock.

In SDP, format-specific parameters are defined as a=fmtp, as below:

```
a=fmtp:<format> <format-specific parameters>
```

In the DV video payload format, the a=fmtp line will be used to show the encoding type within the DV video and will be used as below:

```
a=fmtp:<payload type> encode=<DV-video encoding>
```

The required parameter <DV-video encoding> specifies which type of DV format is used. The DV format name will be one of the following:

```
SD-VCR/525-60
SD-VCR/625-50
HD-VCR/1125-60
HD-VCR/1250-50
SDL-VCR/525-60
SDL-VCR/625-50
306M/525-60
306M/625-50
314M-25/525-60
314M-25/625-50
314M-50/525-60
314M-50/625-50
```

In order to show whether the audio data is bundled into the DV stream or not, a format specific parameter is defined as below:

```
a=fmtp:<payload type> audio=<audio bundled>
```

The optional parameter <audio bundled> will be one of the following:

```
bundled
none      (default)
```

If the fmtp audio parameter is not present, then audio data MUST NOT be bundled into the DV video stream.

3.1 SDP description for unbundled streams

When using unbundled mode, the RTP streams for video and audio will be sent separately to different ports or different multicast groups. When this is done, SDP carries several m=? lines, one for each media type of the session (see RFC 2327 [13]).

An example SDP description using these attributes is:

```
v=0
o=ikob 2890844526 2890842807 IN IP4 126.16.64.4
s=POI Seminar
i=A Seminar on how to make Presentations on the Internet
u=http://www.koganei.wide.ad.jp/~ikob/POI/index.html
e=ikob@koganei.wide.ad.jp (Katsushi Kobayashi)
c=IN IP4 224.2.17.12/127
t=2873397496 2873404696
m=audio 49170 RTP/AVP 112
a=rtpmap:112 L16/32000/2
m=video 50000 RTP/AVP 113
a=rtpmap:113 DV/90000
a=fmtp:113 encode=SD-VCR/525-60
a=fmtp:113 audio=none
```

This describes a session where audio and video streams are sent separately. The session is sent to a multicast group 224.2.17.12. The audio is sent using L16 format, and the video is sent using SD-VCR 525/60 format which corresponds to NTSC format in consumer DV.

3.2 SDP description for bundled streams

When sending a bundled stream, all the DIF blocks including system data will be sent through a single RTP stream. An example SDP description for a bundled DV stream is:

```
v=0
o=ikob 2890844526 2890842807 IN IP4 126.16.64.4
s=POI Seminar
i=A Seminar on how to make Presentations on the Internet
u=http://www.koganei.wide.ad.jp/~ikob/POI/index.html
e=ikob@koganei.wide.ad.jp (Katsushi Kobayashi)
c=IN IP4 224.2.17.12/127
t=2873397496 2873404696
m=video 49170 RTP/AVP 112 113
a=rtpmap:112 DV/90000
a=fmtp: 112 encode=SD-VCR/525-60
a=fmtp: 112 audio=bundled
a=fmtp: 113 encode=306M/525-60
a=fmtp: 113 audio=bundled
```

This SDP record describes a session where audio and video streams are sent bundled. The session is sent to a multicast group 224.2.17.12. The video is sent using both 525/60 consumer DV and SMPTE standard 306M formats, when the payload type is 112 and 113, respectively.

4. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [6], and any appropriate RTP profile. This implies that confidentiality of the media streams is achieved by encryption. Because the data compression used with this payload format is applied to end-to-end, encryption may be performed after compression so there is no conflict between the two operations.

A potential denial-of-service threat exists for data encodings using compression techniques that have non-uniform receiver-end computational load. The attacker can inject pathological datagrams into the stream which are complex to decode and cause the receiver to be overloaded. However, this encoding does not exhibit any significant non-uniformity.

As with any IP-based protocol, in some circumstances a receiver may be overloaded simply by the receipt of too many packets, either desired or undesired. Network-layer authentication may be used to discard packets from undesired sources, but the processing cost of the authentication itself may be too high. In a multicast environment, pruning of specific sources may be implemented in future versions of IGMP [14] and in multicast routing protocols to allow a receiver to select which sources are allowed to reach it.

5. IANA Considerations

This document defines a new RTP payload name and associated MIME type, DV. The registration forms for the MIME types for both video and audio are shown in the next sections.

5.1 DV video MIME registration form

MIME media type name: video

MIME subtype name: DV

Required parameters:

 encode: type of DV format. Permissible values for encode are
 SD-VCR/525-60, SD-VCR/625-50, HD-VCR/1125-60 HD-VCR/1250-50,
 SDL-VCR/525-60, SDL-VCR/625-50, 306M/525-60, 306M/625-50,
 314M-25/525-60, 314M-25/625-50, 314M-50/525-60, and
 314M-50/625-50.

Optional parameters:

 audio: whether the DV stream includes audio data or not.
 Permissible values for audio are bundled and none. Defaults to
 none.

Encoding considerations:

 DV video can be transmitted with RTP as specified in RFC 3189.
 Other transport methods are not specified.

Security considerations:

 See Section 4 of RFC 3189.

Interoperability considerations: NONE

Published specification: IEC 61834 Standard

 SMPTE 306M

 SMPTE 314M

 RFC 3189

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:

Katsushi Kobayashi
e-mail: ikob@koganei.wide.ad.jp

Intended usage: COMMON

Author/Change controller:

Katsushi Kobayashi
e-mail: ikob@koganei.wide.ad.jp

5.2 DV audio MIME registration form

MIME media type name: audio

MIME subtype name: DV

Required parameters:

 encode: type of DV format. Permissible values for encode are
 SD-VCR/525-60, SD-VCR/625-50, HD-VCR/1125-60 HD-VCR/1250-50,
 SDL-VCR/525-60, SDL-VCR/625-50, 306M/525-60, 306M/625-50,
 314M-25/525-60, 314M-25/625-50, 314M-50/525-60, and
 314M-50/625-50.

Optional parameters: NONE

Encoding considerations:

 DV audio can be transmitted with RTP as specified in RFC 3189.
 Other transport methods are not specified.

Security considerations:

 See Section 4 of RFC 3189.

Interoperability considerations: NONE

Published specification: IEC 61834 Standard

 SMPTE 306M
 SMPTE 314M
 RFC 3189

Applications which use this media type:

 Audio 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:

Katsushi Kobayashi

e-mail: ikob@koganei.wide.ad.jp

Intended usage: COMMON

Author/Change controller:

Katsushi Kobayashi

e-mail: ikob@koganei.wide.ad.jp

6. References

- [1] IEC 61834, Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems).
- [2] IEC 61883, Consumer audio/video equipment - Digital interface.
- [3] IEEE Std 1394-1995, Standard for a High Performance Serial Bus
- [4] SMPTE 306M, 6.35-mm type D-7 component format - video compression at 25Mb/s -525/60 and 625/50.
- [5] SMPTE 314M, Data structure for DV-based audio and compressed video 25 and 50Mb/s.
- [6] Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson, "RTP: A transport protocol for real-time applications", RFC 1889, January 1996.
- [7] Hoffman, D., Fernando, G., Goyal, V. and M. Civanlar, "RTP Payload Format for MPEG1/MPEG2 Video", RFC 2250, January 1998.
- [8] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [9] ISO/IEC 11172, Coding of moving pictures and associated audio for digital storage media up to about 1,5 Mbits/s.
- [10] ISO/IEC 13818, Generic coding of moving pictures and associated audio information.
- [11] Schulzrinne, H., "RTP Profile for Audio and Video Conferences with Minimal Control", RFC 1890, January 1996.
- [12] Kobayashi, K., Ogawa, A., Casner S. and C. Bormann, "RTP Payload Format for 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio", RFC 3190, January 2002.

- [13] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.
- [14] Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC 1112, August 1989.

7. Authors' Addresses

Katsushi Kobayashi
Communication Research Laboratory
4-2-1 Nukii-kitamachi,
Koganei Tokyo 184-8795 JAPAN

EEmail: ikob@koganei.wide.ad.jp

Akimichi Ogawa
Keio University
5322 Endo,
Fujisawa Kanagawa 252 JAPAN

EEmail: akimichi@sfc.wide.ad.jp

Stephen L. Casner
Packet Design
2465 Latham Street
Mountain View, CA 94040 United States

EEmail: casner@acm.org

Carsten Bormann
Universitaet Bremen TZI
Postfach 330440
D-28334 Bremen, Germany

Phone: +49 421 218 7024
Fax: +49 421 218 7000
EEmail: cabo@tzi.orgEEmail: cabo@tzi.org

8. 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.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

