INTERNET-DRAFT   Katsushi Kobayashi
draft-kobayashi-dv-video-00.txt   Communication Research Laboratory
                                                   Akimichi Ogawa
                                                   Keio University
                                                    Stephen Casner
                                                     Cisco Systems
                                                   Carsten Bormann
                                           Universitaet Bremen TZI
                                                 February 25, 1999
                                              Expires August 1999

RTP Payload Format for DV Format Video

Status of this Memo

1. Abstract

   This document specifies the packetization scheme for encapsulating
   the digital video data streams defined by the HD Digital VCR
   Conference, commonly known as "DV", into a payload format for the
   Real-Time Transport Protocol (RTP). The RTP payload format specified
   in this document supports three quality levels of digital video
   identified as SD-VCR, HD-VCR and SDL-VCR.

2. Introduction

   The HD Digital VCR Conference has published a digital video
   specification set entitled "Specification of Consumer-Use Digital
   VCRs using 6.3mm magnetic tape" [1,2]. The specification set

consists of two subset specifications, the first of which is
"Specification of Consumer-Use Digital VCRs". That subset comprises
the whole specification for consumer-use digital video including
mechanical specifications of a cassette, helical magnetic recording
format, error correction in the magnetic tape, DCT video encoding
format, and audio encoding format. The digital video format defined
by that specification is commonly known as "DV" format.

The second subset is "Specification of Digital Interface for Consumer
Electronic Audio/Video Equipment" (abbreviated hereafter as the
Digital Interface). That subset defines the communication protocol
for carrying DV video and audio over the IEEE 1394 high performance
serial bus [3]. The IEEE 1394 bus may be used to interconnect
digital video cameras, digital VCRs, computers and other devices.

This document specifies the RTP payload format for encapsulating the
DV format data streams obtained via the Digital Interface into the
Real-time Transport Protocol (RTP), version 2 [4].

The HD Digital VCR Conference specification set supports several
video formats: SD-VCR (including 525/60, 625/50), HD-VCR (1125/60,
1250/50), SDL-VCR (525/60, 625/50), PALplus, DVB (Digital Video
Broadcast) and ATV (Advanced Television). However, the Digital
Interface specifies the IEEE1394 communication protocol for only a
subset of these video formats. The RTP payload format defined here
covers only those video formats that are included in the Digital
Interface.

Furthermore, some formats defined by the HD Digital VCR Conference,
e.g. DVB and ATV, are based on MPEG2. The payload format for
encapsulating MPEG2 into RTP has already been defined in RFC 2250.
That payload format is more suitable for transmission of MPEG2 over
the Internet than would be a packetization of MPEG2 first into the
IEEE 1394 protocol and then into RTP. Therefore, packetization of DV
formats based on MPEG2 is outside the scope of this document.

Consequently, the payload format specified in this document will
support the original six video formats of the HD Digital VCR
Conference: SD-VCR (525/60, 625/50), HD-VCR (1125/60, 1250/50) and
SDL-VCR (525/60, 625/50).

The HD Digital VCR Conference is also standardizing an audio and
video device control protocol, that is, a command set for video
equipment operation and status queries to video devices. This
document does not address these control functions.

Throughout this specification, we make extensive use of the VCR
Conference terminology. The reader should consult the Digital

Interface references for definitions of these terms.

3. DV format encoding

The DV format is designed for magnetic tape applications and is
optimized in helical magnetic recording on tape media. All video
data including audio and other system data are managed within the
picture frame unit of video.

The video encoding consists 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. The DCT macro block
consists of 6 square 8x8 DCT blocks, four of which represent Y
picture component and the remaining two 2 represent Cr and Cb.

Audio is encoded with sampled data. Its frequency is 32 kHz, 44.1
kHz or 48 kHz, its quantization is 16-bit linear or 12-bit non-
linear, and the number of channels may range from 2 to 8. Only
certain combinations of these parameters are allowed depending upon
the video format, as specified in [1].

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
fixed length (80-byte) DIF blocks. 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
data and Video data.

3.1 Transmission of DV format over IEEE 1394

The specification of the Digital Interface defines a transport
protocol for transmission of video stream data in the isochronous
stream mode of IEEE 1394 called "real time data transmission
protocol". The protocol defines the general Common Isochronous
Packet (CIP) header that does not depend on the encoding format of
the payload. Several real time transmission encodings have been
defined on CIP, including MPEG2 and MIDI in addition to DV format
[1,2].

A DIF block is the basic unit for all transmission on the IEEE 1394.
Each IEEE 1394 isochronous stream packet is composed of an integral
number of DIF blocks, assembled without regard to DIF sequence
boundaries, up to the limit of the MTU for IEEE 1394.

4. Usage of RTP

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

4.1 RTP header usage

The meaning of RTP header fields that are specific to the DV format
is described in the following:

Payload type (PT): The payload type is dynamically assigned by means
outside the scope of this document. Details of the encoding format,
such as audio sampling rate and video scan rate, are given in the
AAUX and VAUX data embedded in the data stream. However, the same
information SHOULD be provided as part of the dynamic payload type
assignment. If multiple encoding formats are to be used within one
RTP session, then multiple dynamic payload types MUST be assigned,
one for each encoding format. The sender MUST change to the
corresponding payload type whenever the encoding format is changed.
The sender MUST NOT expect to notify the receiver of an encoding
format change with the information included in AAUX or VAUX because
the packet carrying this information might be dropped and would not
be available to the receiver until the next AAUX or VAUX packet is
received.

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	Framerate (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

The progress of video frame times MAY be monitored using the SYT
timestamp carried in the CIP header, as described in Appendix A.

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 done by differences in RTP timestamp.

4.2 DV data encapsulation into RTP payload

All of the information in the IEEE 1394 CIP header is either implicit
in the RTP payload format or supplanted by information in the RTP
header, so the CIP header is not required. For this payload format,
the CIP header MUST be removed from IEEE 1394 packet, leaving just a
sequence of DIF blocks. Integral DIF blocks are placed into the RTP
payload beginning immediately after the RTP header. DIF blocks
carried by different IEEE 1394 packets 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 there is more payload space
remaining. This requirement stems from the fact the transition from
one video frame to the next is indicated by a change in the RTP
timestamp. It also reduces the processing complexity at 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. 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.

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
generated.

When audio and video are sent in separate RTP streams, or when only
one medium is sent, then only the DIF blocks corresponding to the
selected medium are included. If VAUX DIF blocks are included, they
MUST only be sent in the video stream.

When sending a separate audio stream in the 16-bit encoding, it is
RECOMMENDED that the audio stream data be extracted from the DIF
blocks and repackaged in the L16 payload format defined in RFC 1890
[5] in order to maximize interoperability with non-DV-capable
receivers.

When sending separate video and audio streams with both in DV format,
the same timestamp SHOULD be used for both audio and video data
within the same frame in order to simplify lip synchronization at the
receiver. Lip synchronization may also be achieved using reference

timestamps passed in RTCP as described in [4].

The sender MAY send null AAUX information and omit VAUX DIF blocks if
the VAUX/AAUX information remains constant during the session.
However, the VAUX/AAUX information in the DV stream includes source
encoding parameters, such as video display aspect ratio, audio
quantization and number of audio channels, which are required to
decode the stream. Therefore, if VAUX/AAUX information is not
transmitted in the stream, the equivalent parameters essential to
playout MUST be provided by some out of band means beyond the scope
of this document.

The receiver MUST be able to process a data stream with null AAUX
information and null or omitted VAUX DIF blocks if the equivalent
parameters are provided out of band. Therefore, if the RTP receiver
is feeding the DV stream to a device that requires AAUX information
and VAUX DIF blocks, the receiver MUST be able to generate AAUX
within audio DIF blocks and VAUX DIF blocks for the device using the
parameters provided out of band.

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 by repeating the corresponding
DIF block from the previous image.

5. Security Considerations

RTP packets using the payload format defined in this specification
are subject to the security considerations discussed in the RTP
specification [4], 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
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 [6] and in multicast routing protocols to allow a
receiver to select which sources are allowed to reach it.

6. Full Copyright Statement

7. Authors' Addresses

Katsushi Kobayashi
Communication Research Laboratory
4-2-1 Nukii-kita machi, Koganei
Tokyo 184-8795
JAPAN
EMail: ikob@koganei.wide.ad.jp

Akimichi Ogawa

Keio University
5322 Endo, Fujisawa
Kanagawa 252
JAPAN
EMail: akimichi@sfc.wide.ad.jp

Stephen L. Casner
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
United States
EMail: casner@cisco.com

Carsten Bormann
Universitaet Bremen FB3 TZI
Postfach 330440
D-28334 Bremen, GERMANY
Phone: +49.421.218-7024
Fax: +49.421.218-7000
EMail: cabo@tzi.org

8. Bibliography

[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] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A
    transport protocol for real-time applications. IETF Audio/Video
    Transport Working Group, January 1996. RFC1889.

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

[6] Deering, S., "Host Extensions for IP Multicasting", STD 5,
    RFC 1112, August 1989.

Appendix A.

   In the Digital Interface specification, two types of 8-byte CIP
   headers are defined, one type including the SYT field, and the other
   without the SYT field. The SYT field is a 16-bit timestamp copied
   from lower 16 bits of CYCLE_TIME register defined in IEEE 1394. The
   CYCLE_TIME register is incremented by a 24.576 MHz clock, but the

lower 12 bits count to a maximum of 3071 before wrapping around to
zero and adding a carry to the high 4 bits. Therefore, the SYT
timestamp is not linear.

If the encoding format requires synchronization between devices, it
should adopt the CIP header with SYT. The DV format selects the CIP
header type including the SYT field, but only requires that the SYT
field contain a valid timestamp for one CIP header in every video frame
period. In the remaining CIP headers, the SYT field may contain the
special "no information" value (all ones).