Network Working Group
Request for Comments: 3190
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 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio

Status of this Memo

Copyright Notice

Abstract

This document specifies a packetization scheme for encapsulating 12-bit nonlinear, 20-bit linear, and 24-bit linear audio data streams using the Real-time Transport Protocol (RTP). This document also specifies the format of a Session Description Protocol (SDP) parameter to indicate when audio data is preemphasized before sampling. The parameter may be used with other audio payload formats, in particular L16 (16-bit linear).

1. Introduction

This document describes the sampling of audio data in 12-bit nonlinear, 20-bit linear, and 24-bit linear encodings, and specifies the encapsulation of the audio data into the Real-time Transport Protocol (RTP), version 2 [1,2]. DAT (digital audio tape) and DV (digital video) devices [3,4] use these audio encodings in addition to 16-bit linear encoding. The packetization scheme for 16-bit linear audio (L16) is already specified [2,5]. This document specifies the packetization scheme for the other encodings following that for L16; in particular, when used with the RTP profile [2], these payload formats follow the encoding-independent rules for sample ordering and channel interleaving specified in [2] plus extensions specified here. This document also specifies out-of-band negotiation methods for the extended channel interleaving rules and for use when an analog preemphasis technique is applied to the audio data.

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 [6]

2. The need for RTP encapsulation of 12-, 20- and 24-bit audio

Many high-quality digital audio and visual systems, such as DAT and DV, adopt sample-based audio encodings. Different audio formats are used in various situations. To transport the audio data using RTP, an encapsulation needs to be defined for each specific format. Only 16-bit linear audio encapsulation (L16) has thus far been defined. Other encoding formats have already appeared, such as the 12-bit nonlinear, 20-bit linear and 24-bit linear encodings used in the DAT and DV video world. This specification defines the RTP payload encapsulation format in order to use the new encodings in the RTP environment.

3. 12-bit nonlinear audio encapsulation

IEC 61119 [3] specifies the 12-bit nonlinear audio format in DAT and DV, called LP (Long Play) audio. It would be easy to convert 12-bit nonlinear audio into 16-bit linear form at the RTP sender and transmit it using the L16 audio format already defined. However, this would consume 33% more network bandwidth than necessary. This payload format is specified as a more efficient alternative.

The 12-bit nonlinear encoding is the same as for 16-bit linear audio except for the packing of each sampled data element. Each sample of 12-bit nonlinear audio is derived from a single sample of 16-bit linear audio by a nonlinear compression. Table 1 shows the details of the conversion from 16 to 12 bits. The result is a 12-bit signed value ranging from -2048 to 2047 and it is represented in two’s complement notation. The 12-bit samples are packed contiguously into payload octets starting with the most significant bit. When the payload contains an odd number of samples, the four LSBs of the last octet are unused. Parameters other than quantization, e.g., sampling frequency and audio channel assignment, are the same as in the L16 payload format. In particular, samples are packed into the packet in time sequence beginning with the oldest sample.

32,767 (7FFFh)	Y = INT(X/64) + (600h)	2,047 (7FFh)
16,384 (4000h)		1,792 (700h)
16,383 (3FFFh)	Y = INT(X/32) + (500h)	1,791 (6FFh)
8,192 (2000h)		1,536 (600h)
8,191 (1FFFh)	Y = INT(X/16) + (400h)	1,535 (5FFh)
4,096 (1000h)		1,280 (500h)
4,095 (0FFFh)	Y = INT(X/8) + (300h)	1,279 (4FFh)
2,048 (0800h)		1,024 (400h)
2,047 (07FFh)	Y = INT(X/4) + (200h)	1,023 (3FFh)
1,024 (0400h)		768 (300h)
1,023 (03FFh)	Y = INT(X/2) + (100h)	767 (2FFh)
512 (0200h)		512 (200h)
511 (01FFh)	Y = X	511 (1FFh)
0 (0000h)		0 (000h)
-1 (FFFFh)	Y = X	-1 (FFFh)
-512 (FE00h)		-512 (E00h)
-513 (FFFFh)	Y = INT((X + 1)/2) - (101h)	-513 (DFFh)
-1,024 (FE00h)		-768 (D00h)
-1,025 (FBFFh)	Y = INT((X + 1)/4) - (201h)	-769 (CFFh)
-2,048 (F800h)		-1,024 (C00h)
-2,049 (F7FFh)	Y = INT((X + 1)/8) - (301h)	-1,025 (BFFh)
-4,096 (F000h)		-1,280 (B00h)
-4,097 (EFFFh)	Y = INT((X + 1)/16) - (401h)	-1,281 (AFFh)
-8,192 (E000h)		-1,536 (A00h)
-8,193 (DFFFh)	Y = INT((X + 1)/32) - (501h)	-1,537 (9FFh)
-16,384 (C000h)		-1,792 (900h)
-16,385 (BFFFh)	Y = INT((X + 1)/64) - (601h)	-1,793 (8FFh)
-32,768 (8000h)		-2,048 (800h)

Table 1. Conversion from 16-bit linear values (X) to 12-bit nonlinear values (Y) [3]

When conveying encoding information in an SDP [7] session description, the 12-bit nonlinear audio payload format specified here is given the encoding name "DAT12". Thus, the media format representation might be:

```
m=audio 49230 RTP/AVP 97 98
a=rtpmap:97 DAT12/32000/2
a=rtpmap:98 L16/48000/2
```

4. 20- and 24-bit linear audio encapsulation

The 20- and 24-bit linear audio encodings are simply an extension of the L16 linear audio encoding [2]. The 20- or 24-bit uncompressed audio data samples are represented as signed values in two’s complement notation. The samples are packed contiguously into payload octets starting with the most significant bit. For the 20-bit encoding, when the payload contains an odd number of samples, the four LSBs of the last octet are unused. Samples are packed into the packet in time sequence beginning with the oldest sample.

When conveying encoding information in an SDP session description, the 20- and 24-bit linear audio payload formats specified here are given the encoding names "L20" and "L24", respectively. An example SDP audio media description would be:

```
m=audio 49230 RTP/AVP 99 100
a=rtpmap:99 L20/48000/2
a=rtpmap:100 L24/48000
```

5. Preemphasized audio data

In order to improve the higher frequency characteristics of audio signals, analog preemphasis is often applied to the signal before quantization. If analog preemphasis was applied before the payload data was sampled, the type of the preemphasis SHOULD be conveyed with out-of-band signaling. An "emphasis" parameter is defined for this purpose and may be conveyed either as a MIME optional parameter or using the SDP format-specific attribute (a=fmtp line) as below:

```
a=fmtp:<payload type> emphasis=<emphasis type>
```

Only one <emphasis type> value is defined for the parameter at this point:

```
50-15           <50/15 microsecond CD-type emphasis>
```

The emphasis attribute MUST NOT be included in the SDP record if preemphasis was not applied. This rule allows the emphasis attribute to be used with other audio formats, in particular L16 [2], while retaining backward compatibility with existing implementations so long as preemphasis is not applied. If an existing application that does not implement preemphasis accepts a session description with an emphasis attribute but ignores that attribute, the only penalty is that the sound will be too "bright" when receiving or "dull" when sending.

A sample SDP record showing preemphasis applied only to payload type 99 might be as follows:

```
m=audio 49230 RTP/AVP 99 100
a=rtpmap:99 L20/48000/2
a=fmtp:99 emphasis=50-15
a=rtpmap:100 L24/48000
```

6. Translation of DV audio error code

The DV video specification IEC 61834-4 [4] defines the negative full-scale audio sample value to be an audio error code indicating that no valid audio sample is available for that sample period. Such an error might occur due to a failure while reading audio data from magnetic tape. The audio error code values for each of the DV audio encodings are (in hexadecimal):

```
12-bit nonlinear:  800h
16-bit linear:     8000h
20-bit linear:     80000h
```

For the payload formats defined in this document, as well as for the L16 payload format defined in [2], no such error code is defined. That is, all possible sample values are valid. When an RTP sender accepts audio samples from a DV video system and encapsulates those samples according to one of these payload formats, the RTP sender SHOULD perform some error concealment algorithm which may depend upon whether a single sample error or multiple sample errors have occurred. The error concealment algorithm is not specified here and is left to the implementation. The RTP sender MAY treat the error code as if it were a valid audio sample, but this is likely to cause undesirable audio output.

Conversely, an RTP receiver that accepts audio packets in one of these payload formats and delivers the audio samples to a DV video system SHOULD translate the audio samples that would be interpreted as error codes into the next smaller negative audio value. Such audio samples may be present because the audio packets may have come from a source other than a DV video system. The DV video specification [4] gives the following translations for the defined audio encodings:

```
12-bit nonlinear:  800h                ->  801h
16-bit linear:     8000h               ->  8001h
20-bit linear:     80000h - 8000Fh     ->  80010h
```

For the 20-bit linear encoding, note that multiple audio sample values are translated in order to allow a 16-bit system to play 20-bit audio data by ignoring the least significant four bits. Note also that no translation is specified for 24-bit linear audio because that encoding is not included in the DV video specification.

7. Channel interleaving and non-AIFF-C audio channel convention

When multiple channels of audio, such as in a stereo program, are multiplexed into a single RTP stream, the audio samples from each channel are interleaved according to the rules specified in [2] to be consistent with the L16 payload format. That is, samples from different channels taken at the same sampling instant are packed into consecutive octets. For example, for a two-channel encoding, the sample sequence is (left channel, first sample), (right channel, first sample), (left channel, second sample), (right channel, second sample). Samples for all channels belonging to a single sampling instant MUST be contained in the same packet.

This sample order differs from the packing of samples into blocks in a native DV audio stream. Therefore, applications transmitting DV audio using the payload formats defined in this document MUST reshuffle the samples into the order specified here. This requirement is intended to enable interworking between DV systems and other digital audio systems. Applications choosing to send bundled DV audio/video streams using the native DV block structure may use the payload format defined in [8] instead.

Most of the existing RTP audio payload formats follow the AIFF-C convention for channel ordering as specified in [2] when sending more than two audio channels within a single RTP stream. However, this convention does not cover some applications. For example, some DV audio formats define a "woofer" channel, but AIFF-C does not include this frequency-dependent channel. Thus, it is necessary to specify the audio channel allocation information explicitly when the contents of the audio stream are beyond the scope of AIFF-C.

For DV audio streams of 4 or more channels, the channel order MUST be specified out-of-band. This applies both to the payload formats defined in this document and to the L16 payload format. A "channel-order" parameter is defined here for this purpose and may be conveyed either as a MIME optional parameter or with the SDP a=fmtp attribute using the following syntax:

```
a=fmtp:<payload type> channel-order=<convention>.<order>
```

The first component of the value, <convention>, specifies the type of channel assignment convention used. This allows for conventions other than AIFF-C and DV to be defined in the future. This document defines only one convention for use in the channel-order parameter:

DV

The second component of the value, <order>, indicates the arrangement of channels within the stream. The DV video specification [4] defines the types of audio channels that may be present and in what order. The symbols used to denote the channel types are reproduced in the Appendix at the end of this document. For the DV convention, the following values, which were formed from the concatenation of the individual channel symbols in the allowed channel orders, are defined for the <order> component:

```
4 channels: LRLsRs, LRCS, LRCWo
5 channels: LRLsRsC
6 channels: LRLsRsCS, LmixRmixTWoQ1Q2
8 channels: LRCWoLsRsLmixRmix, LRCWoLs1Rs1Ls2Rs2, LRCWoLsRsLcRc
```

The <convention> and <order> symbols are case-insensitive, but are shown here in mixed case to make the individual channel symbols more apparent. These concatenated symbols were deliberately constructed without separators to make clear the fact that the channels MUST NOT be assembled in other, arbitrary orders.

For interworking with DV video systems, some of the audio encodings are defined only for a subset of the channel combinations listed above. The DV video specification lists the channel combinations that are allowed for each encoding.

The channel-order parameter MUST be consistent with the number of channels specified in the MIME optional parameter "channels" or in the a=rtpmap attribute of SDP. For RTP audio streams of 1, 2 or 3 channels, the AIFF-C channel order is used and is implicit in the number of audio channels specified. To allow backward compatibility, the channel-order parameter MUST NOT be included in this case.

Note that for the DV convention with 5 channels only one channel order is allowed, but for consistency the channel-order parameter MUST be included nonetheless.

An example of an SDP session description using the channel-order parameter is:

```
v=0
o=ikob 2890844526 2890842807 IN IP4 126.16.64.4
s=POI (Audio only)
i=A Seminar on making 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 112 113
a=rtpmap:112 L16/48000/2
a=rtpmap:113 DAT12/32000/4
a=fmtp:113 emphasis=50-15; channel-order=DV.LRCWO
```

This session description shows the audio medium being transmitted in two formats, L16 and DAT12, using payload type numbers 112 and 113, respectively. For the L16 format, the audio data contains 2-channel stereo following the implicit, default AIFF-C convention for left channel first and right channel second. For the DAT12 format, the audio data contains 4 channels following the DV audio convention for the channels left, right, center, and woofer.

This example also shows how multiple MIME optional parameters ("emphasis" and "channel-order") for these payload formats are mapped to a single a=fmtp attribute as a semicolon separated list of parameter=value pairs.

The channel-order parameter described here provides a generic out-of-band mechanism to define alternatives to the AIFF-C channel order. However, if multi-channel audio data could be sent following the AIFF-C channel convention after simple processing, such as a data shuffling on the sender side, the alternative channel order SHOULD NOT be defined and instead the AIFF-C order SHOULD be employed to maximize interoperability.

Moreover, multiple channels of audio data should only be multiplexed within a single RTP stream when all of the audio channels are intended to be reproduced together, such as the left and right channels in a stereo program. Independent audio channels, such as multiple language translations, SHOULD be sent in separate RTP sessions. This reduces bandwidth requirements by allowing receivers to tune in to only those channels which are desired.

8. MIME registration

This document defines some new RTP payload format names which are also registered as MIME subtypes: DAT12, L20 and L24. The registration forms for these MIME subtypes are provided in the next sections.

8.1 MIME registration form for audio/DAT12

MIME media type name: audio

MIME subtype name: DAT12

Required parameters:

rate: number of samples per second -- RECOMMENDED values for rate are 8000, 11025, 16000, 22050, 24000, 32000, 44100 and 48000 samples per second. Other values are permissible.

Optional parameters:

channels: how many audio streams are interleaved -- defaults to 1; stereo would be 2, etc. Interleaving takes place between individual 12-bit samples.

emphasis: analog preemphasis applied to the data before quantization. The only emphasis value defined here is emphasis=50-15 to indicate 50/15 microsecond preemphasis. This parameter MUST be omitted if no analog preemphasis was applied.

channel-order: specifies the sample interleaving order for multiple-channel audio streams (see RFC 3190 Section 7). Permissible values are DV.LRLsRs, DV.LRCS, DV.LRCWo, DV.LRLsRsC, DV.LRLsRsCS, DV.LmixRmixTWoQ1Q2, DV.LRCWoLsRsLmixRmix, DV.LRCWoLs1Rs1Ls2Rs2, DV.LRCWoLsRsLcRc. For interoperation with DV video systems, only a subset of these channel combinations is specified for use with 12-bit nonlinear encoding in the DV video specification [4]; that subset is all of the above except DV.LmixRmixTWoQ1Q2. This parameter MUST be omitted when the AIFF-C channel order convention is in use.

Encoding considerations:

DAT12 audio can be transmitted with RTP as specified in RFC 3190.

Security considerations: See Section 9.

Interoperability considerations: NONE

Published specification:
IEC 61119 Standard [4] and RFC 3190.

Applications which use this media type:
Audio communication.

Additional information:
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

8.2 MIME registration form for audio/L20

MIME media type name: audio

MIME subtype name: L20

Required parameters:
rate: number of samples per second -- RECOMMENDED values for rate are 8000, 11025, 16000, 22050, 24000, 32000, 44100 and 48000 samples per second. Other values are permissible.

Optional parameters:
channels: how many audio streams are interleaved -- defaults to 1; stereo would be 2, etc. Interleaving takes place between individual 20-bit samples.

emphasis: analog preemphasis applied to the data before quantization. The only emphasis value defined here is emphasis=50-15 to indicate 50/15 microsecond preemphasis. This parameter MUST be omitted if no analog preemphasis was applied.

channel-order: specifies the sample interleaving order for multiple-channel audio streams (see RFC 3190 Section 7). Permissible values are DV.LRLsRs, DV.LRCS, DV.LRCWo, DV.LRLsRsC, DV.LRLsRsCS, DV.LmixRmixTWoQ1Q2, DV.LRCWoLsRsLmixRmix, DV.LRCWoLs1Rs1Ls2Rs2, DV.LRCWoLsRsLcRc. For interoperation with DV video systems, none of these channel combinations is specified for use with 20-bit linear encoding in the DV video specification [4]; only mono and stereo are allowed. This parameter MUST be omitted when the AIFF-C channel order convention is in use.

Encoding considerations:
L20 audio can be transmitted with RTP as specified in RFC 3190.

Security considerations: See Section 9.

Interoperability considerations: NONE

Published specification:
RFC 3190.

Applications which use this media type:
Audio communication.

Additional information:
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

8.3 MIME registration form for audio/L24

MIME media type name: audio

MIME subtype name: L24

Required parameters:
rate: number of samples per second -- RECOMMENDED values for rate are 8000, 11025, 16000, 22050, 24000, 32000, 44100 and 48000 samples per second. Other values are permissible.

Optional parameters:
channels: how many audio streams are interleaved -- defaults to 1; stereo would be 2, etc. Interleaving takes place between individual 24-bit samples.

emphasis: analog preemphasis applied to the data before quantization. The only emphasis value defined here is emphasis=50-15 to indicate 50/15 microsecond preemphasis. This parameter MUST be omitted if no analog preemphasis was applied.

channel-order: specifies the sample interleaving order for multiple-channel audio streams (see Section 7). Permissible values are DV.LRLsRs, DV.LRCS, DV.LRCWo, DV.LRLsRsC, DV.LRLsRsCS, DV.LmixRmixTWoQ1Q2, DV.LRCWoLsRsLmixRmix, DV.LRCWoLs1Rs1Ls2Rs2, DV.LRCWoLsRsLcRc. This parameter MUST be omitted when the AIFF-C channel order convention is in use.

Encoding considerations:
L24 audio can be transmitted with RTP as specified in RFC 3190.

Security considerations: See Section 9.

Interoperability considerations: NONE

Published specification:
RFC 3190.

Applications which use this media type:
Audio communication.

Additional information:
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

9. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [1], and any appropriate RTP profile [2]. This implies that confidentiality of the media streams is achieved by encryption. Because the data compression used along 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 [9] and in multicast routing protocols to allow a receiver to select which sources are allowed to reach it.

10. IANA Considerations

This document defines two new MIME subtype-specific optional parameters "emphasis" and "channel-order", which are specified with the sets of permissible values for those parameters in Sections 5 and 7, respectively. Section 8 includes registrations for three new MIME subtypes that use those optional parameters. These registrations define the subset of the optional parameter values allowed for each subtype. It is expected that the number of additional values that will need to be defined for these optional parameters is small. Therefore, anyone wishing to define new values is required to produce a revision of this document to be vetted through the normal Internet Standards process.

11. References

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

[2] H. Schulzrinne, "RTP profile for audio and video conferences with minimal control", RFC 1890, January 1996.

[3] IEC 61119, Digital audio tape cassette system (DAT), November 1992.

[4] 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), August 1998.

[5] Salsman, J., "The Audio/L16 MIME content type", RFC 2586, May 1999.

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

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

[8] Kobayashi, K., Ogawa, A., Casner, S. and C. Bormann, "RTP Payload Format for DV (IEC 61834) Video", RFC 3189, January 2002.

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

Appendix

The DV audio channel symbols are specified in Table 2. These symbols are taken from the notation used in the DV video specification IEC 61834-4 [4], chapter 8.1. For the exact meaning of each symbol, the original DV video specification should be consulted.

```
L: Left channel of stereo
R: Right channel of stereo
M: Monaural signal
C: Center channel of 3,4,6 or 8 channel audio
S: Surround channel of 4 channel audio
Ls, Ls1, Ls2: Left surround channel
Rs, Rs1, Rs2: Right surround channel
Lc: Left center channel of 8 channel audio
Rc: Right center channel of 8 channel audio
Wo: Woofer channel
Lmix: L + 0.7071C + 0.7071LS
Rmix: R + 0.7071C + 0.7071RS
T: 0.7071C
Q1: 0.7071LS + 0.7071RS
Q2: 0.7071LS - 0.7071RS
```

Table 2. Channel symbols for audio channels in DV video [4]

Authors' Addresses

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

Phone: +81 42 327 6174
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
Packet Design
2465 Latham Street
Mountain View, CA 94040
United States

Phone: +1 650-943-1843
EMail: casner@acm.org

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

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

Full Copyright Statement

Acknowledgement

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