

RTP Payload Format for AC-3 Audio

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 (2005).

Abstract

This document describes an RTP payload format for transporting audio data using the AC-3 audio compression standard. AC-3 is a high quality, multichannel audio coding system that is used for United States HDTV, DVD, cable television, satellite television and other media. The RTP payload format presented in this document includes support for data fragmentation.

1. Introduction

AC-3 [ATSC] is a high-quality audio codec (audio coding format) designed to encode multiple channels of audio into a low bit-rate format. AC-3 achieves its large compression ratios via encoding a multiplicity of channels as a single entity. Dolby Digital, which is a branded version of AC-3, encodes up to 5.1 channels of audio.

AC-3 has been adopted as an audio compression scheme for many consumer and professional applications. It is a mandatory audio codec for DVD-video, Advanced Television Standards Committee (ATSC) digital terrestrial television and Digital Living Network Alliance (DLNA) home networking, as well as an optional multichannel audio format for DVD-audio.

There is a need to stream AC-3 data over IP networks. The Internet Real Time Protocol (RTP) provides a mechanism for stream

synchronization and hence serves as the best transport solution for AC-3, which is primarily used in audio-for-video applications. Applications for streaming AC-3 include streaming movies from a home media server to a display, video on demand, and multichannel Internet radio.

Section 2 gives a brief overview of the AC-3 algorithm. Section 3 specifies values for fields in the RTP header, while Section 4 specifies the AC-3 payload format. Section 5 discusses media types and SDP usage. Security considerations are covered in Section 6, congestion control in Section 7, and IANA considerations in Section 8. References are given in Sections 9 and 10.

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 [RFC2119].

2. Overview of AC-3

AC-3 can deliver up to 5.1 channels of audio at data rates approximately equal to half of one PCM channel [ATSC], [1994AC3], [1996AC3]. The ".1" refers to a band-limited, optional, low-frequency effects (LFE) channel. AC-3 was designed for signals sampled at rates of 32, 44.1, or 48 kHz. Data rates can vary between 32 kbps and 640 kbps, depending on the number of channels and the desired quality.

AC-3 exploits psycho-acoustic phenomena that cause a significant fraction of the information contained in a typical audio signal to be inaudible. Substantial data reduction occurs via the removal of inaudible information contained in an audio stream. Source coding techniques are further used to reduce the data rate.

Like most perceptual coders, AC-3 operates in the frequency domain. A 512-point TDAC transform is taken with 50% overlap, providing 256 new frequency samples. Frequency samples are then converted to exponents and mantissas. Exponents are differentially encoded. Mantissas are allocated a varying number of bits depending on the audibility of the associated spectral components. Audibility is determined via a masking curve. Bits for mantissas are allocated from a global bit pool.

2.1. AC-3 Bit Stream

AC-3 bit streams are organized into synchronization frames. Each AC-3 frame contains a Synchronization Information (SI) field, a Bit Stream Information (BSI) field, and 6 audio blocks (ABs) that each represent 256 PCM samples for all channels. The frame ends with an

optional auxiliary data field (AUX) and an error correction field (CRC). The entire frame represents the time duration of 1536 PCM samples across all coded channels [ATSC]. AC-3 encodes audio sampled at 32 kHz, 44.1 kHz, and 48 kHz. From Annex A, Part 2, of [ATSC], the time duration of an AC-3 frame varies with the sampling rate as follows:

Sampling rate		Frame duration	
48	kHz	32	ms
44.1	kHz	approx. 34.83	ms
32	kHz	48	ms

Figure 1 shows the AC-3 frame format.

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|SI|BSI| AB0 | AB1 | AB2 | AB3 | AB4 | AB5 |AUX|CRC|
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---

```

Figure 1. AC-3 Frame Format

The Synchronization Information field contains information needed to acquire and maintain codec synchronization. The Bit Stream Information field contains parameters that describe the coded audio service [ATSC]. Each audio block contains fields that indicate the use of various coding tools: block switching, dither, coupling, and exponent strategy. They also contain metadata, optionally used to enhance the playback, such as dynamic range control. Finally, the exponents and bit allocation data needed to decode the mantissas into audio data, and the mantissas themselves, are included. The placement of these fields in an AC-3 audio block is shown in Figure 2. The fields shown in this figure are described in detail in [ATSC]. Note that field sizes vary depending on the coded data.

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---
| Block |Dither |Dynamic   |Coupling |Coupling   |Exponent |
| Switch|Flags  |Range Ctrl|Strategy|Coordinates|Strategy |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---
|      Exponents      | Bit Allocation |      Mantissas      |
|                      | Parameters    |                      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---

```

Figure 2. AC-3 Audio Block Format

3. RTP AC-3 Header Fields

- o Payload Type (PT): The assignment of an RTP payload type for this packet format is outside the scope of this document. It is specified by the RTP profile under which this payload format is used, or signaled dynamically out-of-band (e.g., using SDP).
- o Marker (M) bit: The M bit is set to one to indicate that the RTP packet payload contains at least one complete AC-3 frame or contains the final fragment of an AC-3 frame.
- o Extension (X) bit: Defined by the RTP profile used.
- o Timestamp: A 32-bit word that corresponds to the sampling instant for the first AC-3 frame in the RTP packet. Packets containing fragments of the same frame MUST have the same time stamp. The timestamp of the first RTP packet sent SHOULD be selected at random. Thereafter, the timestamp increases linearly with the number of samples included in each frame (i.e., by 1536 for each frame).

4. RTP AC-3 Payload Format

This payload format is defined for AC-3, as defined in the main part and Annex D of [ATSC]. It is not defined for E-AC-3, as defined in Annex E of [ATSC], and it MUST NOT be used to carry E-AC-3.

According to [RFC2736], RTP payload formats should contain an integral number of application data units (ADUs). The AC-3 frame corresponds to an ADU, in the context of this payload format. Each RTP payload MUST start with the two-byte payload header followed by an integral number of complete AC-3 frames or by a single fragment of an AC-3 frame.

If an AC-3 frame exceeds the MTU for a network, it SHOULD be fragmented for transmission within an RTP packet. Section 4.2 provides guidelines for creating frame fragments.

4.1. Payload-Specific Header

There is a two-octet Payload Header at the beginning of each payload.

4.1.1. Payload Header

Each AC-3 RTP payload MUST begin with the payload header shown in Figure 3.

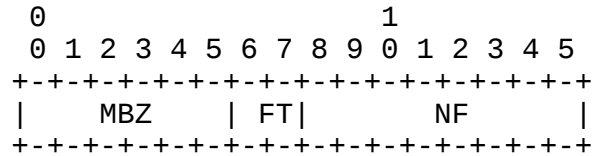


Figure 3. AC-3 RTP Payload Header

- o MBZ (Must Be Zero): Bits marked MBZ SHALL be set to the value zero and SHALL be ignored by receivers. The bits are reserved for future extensions.
- o FT (Frame Type): This two-bit field indicates the type of frame(s) present in the payload. It takes the following values:
 - 0 - One or more complete frames.
 - 1 - Initial fragment of frame which includes the first 5/8ths of the frame. (See Section 4.2.)
 - 2 - Initial fragment of frame, which does not include the first 5/8ths of the frame.
 - 3 - Fragment of frame other than initial fragment. (Note that M bit in RTP header is set for final fragment).
- o NF (Number of frames/fragments): An 8-bit field whose meaning depends on the Frame Type (FT) in this payload. For complete frames (FT of 0), it is used to indicate the number of AC-3 frames in the RTP payload. For frame fragments (FT of 1, 2, or 3), it is used to indicate the number fragments (and therefore packets) that make up the current frame. NF MUST be identical for packets containing fragments of the same frame.

Figure 4 shows the full AC-3 RTP payload format.

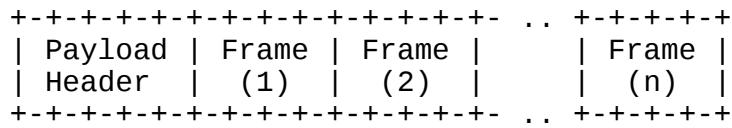


Figure 4. Full AC-3 RTP payload

When receiving AC-3 payloads with $FT = 0$ and more than a single frame ($NF > 1$), a receiver needs to use the "frmsizecod" field in the Synchronization Information (syncinfo) block in each AC-3 frame to determine the frame's length. That way a receiver can determine the boundary of the next frame. Note that the frame length may vary from frame to frame.

4.2. Fragmentation of AC-3 Frames

The size of an AC-3 frame depends on the sample rate of the audio and the data rate used by the encoder (which are indicated in the "Synchronization Information" header in the AC-3 frame). The size of a frame, for a given sample rate and data rate, is specified in Table 5.18 ("Frame Size Code Table") of [ATSC]. This table shows that AC-3 frames range in size from a minimum of 128 bytes to a maximum of 3840 bytes. If the size of an AC-3 frame exceeds the MTU size, the frame SHOULD be fragmented at the RTP level. The fragmentation MAY be performed at any byte boundary in the frame. RTP packets containing fragments of the same AC-3 frame SHALL be sent in consecutive order, from first to last fragment. This enables a receiver to assemble the fragments in correct order.

When an AC-3 frame is fragmented, it MAY be fragmented such that at least the first 5/8ths of the frame data is in the first fragment. This provides greater resilience to packet loss. This initial portion of any frame is guaranteed to contain the data necessary to decode the first two blocks of the frame. Any frame fragments other than those containing the first 5/8ths of frame data are only decodable once the complete frame is received. The 5/8ths point of the frame is defined in Table 7.34 ("5/8_frame Size Table") of [ATSC].

5. Types and Names

5.1. Media Type Registration

This registration uses the template defined in [DRAFT-FREED] and follows RFC 3555 [RFC3555].

o Type name: audio

o Subtype name: ac3

o Required parameters:

rate: The RTP timestamp clock rate that is equal to the audio sampling rate. Permitted rates are 32000, 44100, and 48000.

o Optional parameters:

channels: From a sender, the maximum number of channels present in the AC3 stream. From a receiver, the maximum number of output channels the receiver will deliver. This MUST be a number between 1 and 6. The LFE (".1") channel MUST be counted as one channel. Note that the channel order used in AC-3 differs from the channel order scheme in [RFC3551]. The AC-3 channel order scheme can be found in Table 5.8 of [ATSC].

ptime: See RFC 2327 [RFC2327].

maxptime: See RFC 3267 [RFC3267].

o Encoding considerations:

This media type is framed (see section 4.8 in [DRAFT-FREED]) and contains binary data.

o Security considerations:

See Section 6 of this document.

o Interoperability considerations:

None

o Published specification:

This payload format specification and see [ATSC].

- o Applications that use this media type:

- Multichannel audio compression of audio and audio for video.

- o Additional Information:

- Magic number(s):

- The first two octets of an AC-3 frame are always the synchronization word, which has the hex value 0x0B77.

- o Person & email address to contact for further information:

- Brian Link <bd1@dolby.com>
IETF AVT working group.

- o Intended Usage:

- COMMON

- o Restrictions on usage:

- This media type depends on RTP framing, and hence is only defined for transfer via RTP [RFC3550]. Transport within other framing protocols is not defined at this time.

Author/Change controller:

IETF Audio/Video Transport Working Group delegated from the IESG.

5.2. SDP Usage

The information carried in the MIME media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [RFC2327], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing AC-3, the mapping is as follows:

- o The Media type ("audio") goes in SDP "m=" as the media name.
- o The Media subtype ("ac3") goes in SDP "a=rtpmap" as the encoding name.
- o The required parameter "rate" also goes in "a=rtpmap" as the clock rate, optionally followed by the parameter "channel".
- o The optional parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

An example of the SDP data for AC-3:

```
m=audio 49111 RTP/AVP 100
a=rtpmap:100 ac3/48000/6
```

Certain considerations are needed when SDP is used to perform offer/answer exchanges [RFC3264].

- o The "rate" is a symmetric parameter, and the answer MUST use the same value or remove the payload type.
- o The "channels" parameter is declarative and indicates, for recvonly or sendrecv, the desired channel configuration to receive, and for sendonly, the intended channel configuration to transmit. All receivers are capable of receiving any of the defined channel configurations, and the parameter exchange might be used to help optimize the transmission to the number of channels the receiver requests. If the "channels" parameter is omitted, a default maximum value of 6 is implied.
- o The "ptime" and "maxptime" parameters are negotiated as defined for "ptime" in RFC 3264 [RFC3264].

6. Security Considerations

The payload format described in this document is subject to the security considerations defined in the RTP specification [RFC3550] and in any applicable RTP profile (e.g., [RFC3551]). To protect the user's privacy and any copyrighted material, confidentiality protection would have to be applied. To also protect against modification by intermediate entities and ensure the authenticity of the stream, integrity protection and authentication would be required. Confidentiality, integrity protection, and authentication have to be provided by a mechanism external to this payload format, e.g., SRTP [RFC3711].

The AC-3 format is designed so that the validity of data frames can be determined by decoders. A decoder that encounters a malformed frame discards the malformed data and conceals the errors in the audio output until a valid frame is detected and decoded. This is expected to prevent crashes and other abnormal decoder behavior in response to errors or attacks.

7. Congestion Control

The general congestion control considerations for transporting RTP data apply to AC-3 audio over RTP as well. See the RTP specification [RFC3550] and any applicable RTP profile (e.g., [RFC3551]).

AC-3 encoders may use a range of bit rates to encode audio data, so it is possible to adapt network bandwidth by adjusting the encoder bit rate in real time or by having multiple copies of content encoded at different bit rates. Additionally, packing more frames in each RTP payload can reduce the number of packets sent and hence the overhead from IP/UDP/RTP headers, at the expense of increased delay and reduced robustness against packet losses.

8. IANA Considerations

A new media subtype has been assigned for AC-3; see Section 5.1.

9. Normative References

- [RFC2119] Bradner, S., "Key Words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.
- [ATSC] U.S. Advanced Television Systems Committee (ATSC), "ATSC Standard: Digital Audio Compression (AC-3), Revision B," Doc A/52B, June 2005.
- [RFC2327] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.
- [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.
- [RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, June 2002.
- [RFC3267] Sjöberg, J., Westerlund, M., Lakaniemi, A., and Q. Xie, "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", RFC 3267, June 2002.
- [RFC3555] Casner, S. and P. Hoschka, "MIME Type Registration of RTP Payload Formats", RFC 3555, July 2003.

10. Informative References

- [RFC2736] Handley, M. and C. Perkins, "Guidelines for Writers of RTP Payload Format Specifications", BCP 36, RFC 2736, December 1999.
- [RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.
- [1994AC3] Todd, C., et al., "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," Preprint 3796, Presented at the 96th Convention of the Audio Engineering Society, May 1994.
- [1996AC3] Fielder, L., et al., "AC-2 and AC-3: Low-Complexity Transform-Based Audio Coding," Collected Papers on Digital Audio Bit-Rate Reduction, pp. 54-72, Audio Engineering Society, September 1996.
- [RFC3711] Baugher, M., et al., "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.
- [DRAFT-FREED] Freed, N. and Klensin, J., "Media Type Specifications and Registration Procedures", Work in Progress, April 2005.

Authors' Addresses

Brian Link
Dolby Laboratories
100 Potrero Ave
San Francisco, CA 94103

Phone: +1 415 558 0200
EMail: bdl@dolby.com

Todd Hager
Dolby Laboratories
100 Potrero Ave
San Francisco, CA 94103

Phone: +1 415 558 0136
EMail: thh@dolby.com

Jason Flaks
Microsoft Corporation
1 Microsoft Way
Redmond, WA 98052

Phone: +1 425 722 2543
EMail: jasonfl@microsoft.com

Full Copyright Statement

Copyright (C) The Internet Society (2005).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM 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.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgement

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

