Status of this Memo

This document is an Internet-Draft and is subject to all provisions of section 3 of RFC 3667. By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Abstract

This document describes an RTP payload format for transporting AC-3 encoded audio data. AC-3 is a high quality, multichannel audio coding system used in US HDTV, DVD, cable and satellite television and other media. The RTP payload format as presented in this document includes support for data fragmentation.
1. Introduction

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

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, itself. Section 5 discusses MIME types and SDP usage. Security considerations are covered in Section 6 and IANA considerations in Section 7. References are given in Sections 8 and 9.

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 [1].

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 [2], [7], [8]. The ".1" refers to a band-limited, optional, low-frequency enhancement channel. AC-3 was designed for signals sampled at rates of 32, 44.1, or 48 kHz. Data rates can vary between 64 kbps and 640 kbps, depending the number of channels and desired quality.

AC-3 exploits psychoacoustic 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 spectral component associated with it. 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 Sync Information (SI) field, a Bit Stream Information (BSI) field, and 6 audio blocks (AB), each representing 256 PCM samples for each channel. The entire frame represents a time duration of 1536 PCM samples across all coded channels (e.g., 32 msec @ 48kHz sample rate) [2]. 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 [2]. Each audio block also 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. Figure 2 shows the structure of an AC-3 audio block. Note that field sizes vary depending on the coded data.

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

Figure 2. AC-3 Audio Block Format

3. RTP Header Fields

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).

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.

Extension (X) bit: Defined by the RTP profile used.

Timestamp: A 32-bit word that corresponds to the sampling instant for the first AC-3 frame in the RTP packet. AC-3 encodes audio sampled at 32 kHz, 44.1 kHz, and 48 kHz. 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 it increases linearly according to the number of samples included in each frame (i.e. by 1576 for each frame).

4. RTP AC-3 Payload Format

According to [5], RTP payload formats should contain an integral number of application data units (ADUs). An ADU shall be equivalent to an AC-3 frame. To simplify the implementation of RTP receivers, each RTP packet MUST contain an integral number of complete AC-3 frames, or one 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 following payload header. Figure 3 shows the format of this header.
Frame Type (FT): 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.)

Number of frames/fragments (NF): 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 of fragments (and therefore packets) that make up the current frame. NF MUST be identical for packets containing fragments of the same frame.

Must Be Zero (MBZ): Bits marked MBZ MUST have the value zero and are reserved.

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

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 of 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 [2]. 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.
When an AC-3 frame is fragmented, it MAY be fragmented such that the first 5/8ths of the frame data is in the first fragment to provide greater resilience to packet loss. This initial portion of a frame is guaranteed to contain the data necessary to decode the first two blocks of the frame. Any other frame fragments 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 [2].

5. Types and Names

5.1 MIME Type Registration

MIME media type name: audio
MIME subtype name: vnd.dolby.dd-rtp

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

BSID: This parameter is a repetition of the Bit Stream Identification field in the AC-3 bit stream [2]. It indicates the version number of the bit stream. An AC-3 decoder is capable of decoding bit streams of a given version number and all lower version numbers. An AC-3 decoder is not capable of decoding bit streams with higher version number. In the AC-3 specification, BSID is a 5-bit unsigned integer, so the maximum allowed value is 31. BSID of 8 corresponds to AC-3 as defined in [2].

Optional parameters:
Channels: The number of channels present in the AC3 stream. This MUST be a number between 1 and 6. The LFE (".1") channel MUST be counted as one channel.

Ptime: The duration of time in milliseconds represented by the AC-3 frame(s) in the packet.

Maxptime: The maximum duration of media which can be encapsulated in each RTP packet, expressed as time in milliseconds.

Encoding considerations:
This MIME subtype is defined for RTP transport only. The AC-3 bit stream MUST be generated according to the AC-3 specification [2]. The RTP packets MUST be packetized according to the RTP payload format defined in this document.
Security considerations: See Section 6 of this Document.

Interoperability considerations: none

Published specification: This payload format specification and See [2].

Applications: Multichannel audio compression of audio and audio for video.

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.

File extension(s): .ac3
Macintosh File Type Code(s): none
Object Identifier(s) or OID(s): none

Person & email address to contact for further information: Brian Link <bdl@dolby.com>
IETF AVT working group.

Intended Usage: COMMON

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) [3], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing AC-3, the mapping is as follows:

- The MIME type ("audio") goes in SDP "m=" as the media name.
- The MIME subtype ("vnd.dolby.dd-rtp") goes in SDP "a=rtpmap" as the encoding name.
- The required parameter "rate" also goes in "a=rtpmap" as the clock rate, optionally followed by the parameter "channel".
- The required parameter "BSID" goes in SDP "a=fmtp".
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 vnd.dolby.dd-rtp/48000/6
a=fmtp:100 BSID=8
```

No special considerations are required when SDP is used in offer/answer mode. No parameter negotiations are necessary because the design of AC-3 ensures that any AC-3 stream compliant to [2] is decodable by any decoder also compliant to [2]. For example, a stereo AC-3 decoder, in addition to decoding stereo content, is required to render a downmixed, stereo version of multichannel AC-3 content.

6. Security Considerations

In order to protect copyrighted material, certain security precautions may be necessary. The payload format described in this document is subject to the security considerations defined in any applicable RTP profile (e.g. [RFC3551]) and in [4]. The security considerations discussed in [4] imply the usage of encryption to protect the confidentiality of content. Such an encryption scheme is harmless to the encoded audio data presuming the data is decrypted before being sent to the decoder. A possible encryption mechanism is described in the profile in [9].

7. IANA Considerations

Registration of a new MIME subtype for AC-3 is requested (see Section 5.)

8. Normative References


9. Informative References


Authors’ Addresses

Brian Link
Dolby Laboratories
100 Potrero Ave
San Francisco, CA 94103
Phone: +1 415 558 0200
Email: bd1@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
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.

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.