RTP Payload Format for AC-3 Audio
<draft-gharai-ac3-01.txt>

Status of this Memo

Abstract

This document specifies a packetization scheme for encapsulating AC-3 audio streams into a payload format for the Real-Time Transport Protocol (RTP).

1. Introduction

AC-3, also known as Dolby Digital or Dolby AC-3, is a flexible audio data compression technology. It has been in use in feature films since 1992 **and has also been selected as the audio format of HDTV. The AC-3** digital compression algorithm can encode 1 to 5.1 audio channels in PCM representation into a single serial bit stream. Encoding multiple channels as a single entity is more efficient than individually encoding each channel, resulting in an overall lower bit rate.

The syntax for AC-3 is fully described in [1] by the Advanced Television Standards Committee (ATSC). The audio compression system used by HDTV is a restricted subset of this specification, where the restrictions are specified in Annex B of the Digital Television Standard [2].

2. AC-3 Digital Audio

An AC-3 audio stream is constructed as a sequence of synchronization frames also called the sync frame. Each frame is completely self contained and is made up of:

- o a synchronization information (SI) header, which includes:
  - - a sync word, used for acquiring and maintaining synchronization
  - - an indication of the sampling rate, 48kHz, 44.1kHz or 32kHz
  - - and the size of the sync frame
- o a bit stream information (BSI) header which includes the sync frames' timestamp,
- o 6 audio blocks (AB), each block represents 256 new audio samples,
- o an auxiliary data field (Aux),
- o and finally, an error check field CRC.

```
   -------------------------------------------------------------
   | SI | BSI |  AB0 | AB1 | AB2 | AB3 | AB4 | AB5 | Aux | CRC |
   -------------------------------------------------------------
     Figure 1. An AC-3 synchronization frame (not to scale).
```

All sync frames within a sequence are the same size. Frame sizes range from 128bytes to 3840bytes. Table 5.13 in [1] lists all possible frame sizes per bit rate and sampling frequency. At 48kHz each sync frame represents 32ms of audio data (each audio block is 5.33ms).

Each sync frame is a complete independent data unit, it does not require any other data to be decoded. A complete sync frame MUST be presented to the decoder for decompression. An incomplete sync frame will not pass the decoder's error detection test causing the decoder to mute. At 48kHz this can cause a maximum of 64ms of muted audio (if decoder is unable to synchronize with the immediate next sync word).

3. RTP Packetization

When feasible, a RTP packet will contain an integral number of sync frames. However, depending on the path-MTU, a sync frame may require multiple RTP packets, in which case the sync frame will be fragmented across multiple RTP packets. Multiple RTP packets transferring a fragmented sync frame must have the same timestamp, which reflects the

sampling instance of the sync frame. Fragmented sync frames are reassembled via the RTP timestamp and sequence number.

An RTP packet should not carry fragments of different sync frames, or a fragment of one sync frame and an other complete sync frame. Once received fragmented sync frames MUST be reassembled before being presented to the decoder.

The fields of the RTP fixed header are used as follows:

Marker bit (M): The Marker bit of the RTP header is set to 1 for the last packet of a sync frame and set to 0 on all other packets.

Payload Type (PT): The Payload Type indicates the use of the payload format defined in this document. A profile may assign a payload type value for this format either statically or dynamically as described in RFC 1890 [4].

Timestamp: A 32bit 48kHz, 44.1kHz or 32kHz (corresponding with the sampling frequency of the audio) timestamp which encodes the sampling instant of the first sync frame in the RTP packet. All packets transferring a fragmented sync frame MUST have the same timestamp.

4. SDP Payload Format Description

With a dynamic payload type (say 96) and using the encoding name AC-3, the rtpmap for an AC-3 audio stream sampled at 48kHz is as follows:

```
a=rtpmap:96 AC-3/48000
```

5. Data Resiliency

With a transfer rate of 32kbps (the lowest transfer rate suggested in table 5.13 of the ATSC standard) the size of sync frames for audio sampled at 32kHz, 44.1kHz and 48kHz are 192bytes, 138bytes and 128bytes respectively.

Given the "all or nothing" nature of AC-3 sync frame, fragmented sync frames are highly susceptible to network loss, i.e. the loss of one RTP packet carrying part of a sync frame renders the other packets useless.

Augmenting the RTP stream with AC-3 sync frames compressed at 32kbps increases the resiliency of the data stream, particularly for large fragmented sync frames. The two audio streams can be interleaved into an RTP stream. The application will first attempt de-packetize and (if necessary) reassemble the higher quality AC-3 sync frames. However for

missing or incomplete sync frames the lower quality sync frames shall be presented to the decoder.

6. 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.

7. IANA Considerations

None.

8. To Do

The AC-3 stream is likely to be well-served by a repair vector similar to that proposed for AAC audio.

9. Full Copyright Statement

10. Authors' Address

Ladan Gharai ladan@isi.edu

11. Bibliography

[1] ATSC Digital Audio Compression Standard (AC-3) Document A/52, Sep. 1995, http://www.atsc.org.

[2] ATSC Digital Television Standard Document A/53, September 1995, http://www.atsc.org

[3] Schulzrinne, Casner, Frederick, Jacobson, "RTP: A transport protocal for real time Applications", RFC 1889, IETF, January 1996.

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