Network Working Group
Internet-Draft
Expire in six months
Category: Informational

Ross Finlayson
LIVE.COM
1999.01.24

A More Loss-Tolerant RTP Payload Format for MP3 Audio

<draft-finlayson-rtp-mp3-00.txt>

1. Status of this Memo

2. Abstract

While the RTP payload format defined in RFC 2250 is generally applicable to all forms of MPEG audio or video, it is less suitable for MPEG 1 or 2, layer III audio (commonly known as "MP3"). The reason for this is that an MP3 frame is not a true "Application Data Unit" - it contains a back-pointer to data in earlier frames, and so cannot be decoded independently of these earlier frames. Because RFC 2250 defines that packet boundaries coincide with frame boundaries, it handles packet loss inefficiently when carrying MP3 data. The loss of an MP3 frame will render some data in previous (or future) frames useless, even if they are received without loss.

In this document we define a new RTP payload format for MP3 audio. The new format is essentially the same as that defined in RFC 2250, except for a data-preserving rearrangement of the original MPEG frames, so that packet boundaries now coincide with true MP3 "Application Data Units". This new format is therefore more data efficient in the face of packet loss.

3. The Structure of MP3 Frames

In this section we give a brief overview of the structure of a MP3 frame. (For more detailed description, see the official MPEG 1 audio [2] and MPEG 2 audio [3] specifications.)

Each MPEG audio frame begins with a 4-byte header. Information defined by this header includes:
- Whether the audio is MPEG 1 or MPEG 2.
- Whether the audio is layer I, II, or III.
      (The remainder of this document assumes layer III, i.e., "MP3")
- Whether the audio is mono or stereo.
- Whether or not there is a 2-byte CRC field following the header.
- (indirectly) The size of the frame.

The following structures appear after the header:
- (optionally) A 2-byte CRC field
- A "side info" structure. This has the following length:
    - 32 bytes for MPEG 1 stereo
    - 17 bytes for MPEG 1 mono, or for MPEG 2 stereo
    - 9 bytes for MPEG 2 mono
- Encoded audio data (filling out the rest of the frame)

For the purpose of this document, the "side info" structure is the most important, because it defines the location and size of the "Application Data Unit" (ADU) that an MP3 decoder will process. In particular, the "side info" structure defines:
- "main_data_begin": This is a back-pointer (in bytes) to the start of the ADU. The back-pointer is counted from the beginning of the frame, and counts only encoded audio data (i.e., ignoring any header, CRC, or "side info" fields).
- Several "part2_3_length" fields. These fields - each of which counts bits - are added together to form the length (in bits) of the ADU. Like the back-pointer, this length counts only encoded audio data - not header, CRC, or "side info" fields.

An MP3 decoder processes each ADU independently. The ADUs will generally vary in length, but their average length will, of course, be that of the of the MP3 frames (minus the length of the header, CRC, and "side info" fields).

4. A New Payload Format

As noted in [4], a payload format should be designed so that packet boundaries coincide with "codec frame boundaries" - i.e., with ADUs. The new payload format for MP3 is exactly the same as that defined by RFC 2250 for MPEG audio [1], EXCEPT:
1/ Instead of containing MP3 frames, each packet contains "ADU frames", where an "ADU frame" is defined as:
    - The 4-byte MPEG header
    - The optional 2-byte CRC field
    - The "side info" structure
    - The complete sequence of encoded data for the ADU (padded at the

end by zero-bits to fill out a byte boundary)
2/ The (static) payload type 14 that was defined for MPEG audio [5] MUST NOT be used. Instead, a different, dynamic payload type MUST be used.

Apart from using a different payload type, the use of the RTP header (and the MPEG audio-specific header for fragmentation) is exactly the same as defined in RFC 2250.

Note that no information is lost by converting a sequence of MP3 frames to a corresponding sequence of "ADU frames", so a receiving RTP implementation can either feed the ADU frames directly to an appropriately modified MP3 decoder, or convert them back into a sequence of MP3 frames.

5. SDP payload format description

Pending any future standardization of this payload format, SDP "rtpmap" attributes [6] use the name "X-MP3" to denote this format.

6. Security Considerations

The security considerations for this payload format are identical to those noted for RFC 2250 [1].

7. References

[1] Hoffman, D., Fernando, G., Goyal, V., and Civanlar, M. "RTP Payload Format for MPEG1/MPEG2 Video", RFC 2250, January 1998.
[2] ISO/IEC International Standard 11172-3; "Coding of moving pictures and associated audio for digital storage media up to about 1,5 Mbits/s - Part 3: Audio", 1993.
[3] ISO/IEC International Standard 13818-3; "Generic coding of moving pictures and associated audio information - Part 3: Audio", 1998.
[4] Handley, M. "Guidelines for Writers of RTP Payload Format Specifications" Work-in-Progress, Internet-Draft "draft-ietf-avt-rtp-format-guidelines-01.txt", November 1998.
[5] Schulzrinne, H. "RTP Profile for Audio and Video Conferences with Minimal Control" RFC 1890, January 1996.
[6] Handley, M., Jacobson, V., "SDP: Session Description Protocol", RFC 2327, April, 1998.

8. Author's Address

Ross Finlayson,
Live Networks, Inc. (LIVE.COM)
email: finlayson@live.com
WWW: http://www.live.com/