AVT	J. Lennox, Ed.
Internet-Draft	Vidyo
Intended status: Standards Track	E. Ivov
Expires: January 06, 2012	Jitsi
	E. Marocco
	Telecom Italia
	July 05, 2011

A Real-Time Transport Protocol (RTP) Header Extension for Client-to-Mixer Audio Level Indication

draft-ietf-avtext-client-to-mixer-audio-level-03

Abstract

This document defines a mechanism by which packets of Real-Time Transport Protocol (RTP) audio streams can indicate, in an RTP header extension, the audio level of the audio sample carried in the RTP packet. In large conferences, this can reduce the load on an audio mixer or other middlebox which wants to forward only a few of the loudest audio streams, without requiring it to decode and measure every stream that is received.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

In a centralized Real-Time Transport Protocol (RTP) *[RFC3550]* audio conference, an audio mixer or forwarder receives audio streams from many or all of the conference participants. It then selectively forwards some of them to other participants in the conference. In large conferences, it is possible that such a server might be receiving a large number of streams, of which only a few should be forwarded to the other conference participants.

In such a scenario, in order to pick the audio streams to forward, a centralized server needs to decode, measure audio levels, and possibly perform voice activity detection on audio data from a large number of streams. The need for such processing limits the size or number of conferences such a server can support.

As an alternative, this document defines an RTP header extension *[RFC5285]* through which senders of audio packets can indicate the audio level of the packets' payload, reducing the processing load for a server.

The header extension in this draft is different than, but complementary with, the one defined in [I-D.ietf-avtext-mixer-to-client-audio-level], which defines a mechanism by which audio mixers can indicate to clients the levels of the contributing sources that made up the mixed audio.

2. 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 *[RFC2119]* and indicate requirement levels for compliant implementations.

3. Audio Levels

The audio level header extension carries the level of the audio in the RTP payload of the packet it is associated with. This information is carried in an RTP header extension element as defined by the "General Mechanism for RTP Header Extensions" *[RFC5285]*.

The payload of the audio level header extension element can be encoded using the one or the two-byte header defined in [RFC5285]. Figure 1 and Figure 2 show sample audio level encodings with each of them.

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ID   | len=0 |V| level       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Sample audio level encoding using the one-byte header format

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |      ID       |     len=1     |V|    level    |    0 (pad)    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Sample audio level encoding using the two-byte header format

Note that, as indicated in [RFC5285] length field in the one-byte header format takes the value 0 to indicate that 1 byte follows. In the two-byte header format on the other hand it takes the value of 1.

The magnitude of the audio level itself is packed into the seven least significant bits of the single byte of the header extension, shown in Figure 1 and Figure 2. The least significant bit of the audio level magnitude is packed into the least significant bit of the byte. The most significant bit of the byte is used as a separate flag bit "V", defined below.

The audio level is expressed in -dBov, with values from 0 to 127 representing 0 to -127 dBov. dBov is the level, in decibels, relative to the overload point of the system, i.e. the maximum-amplitude signal that can be handled by the system without clipping. (Note: Representation relative to the overload point of a system is particularly useful for digital implementations, since one does not need to know the relative calibration of the analog circuitry.) For example, in the case of u-law (audio/pcmu) audio *[ITU.G711.1988]*, the 0 dBov reference would be a square wave with values +/- 8031. (This translates to 6.18 dBm0, relative to u-law's dBm0 definition in Table 6 of G.711.)

The audio level for digital silence, for example for a muted audio source, MUST be represented as 127 (-127 dBov), regardless of the dynamic range of the encoded audio format.

The audio level header extension only carries the level of the audio in the RTP payload of the packet it is associated with, with no long-term averaging or smoothing applied. That level is measured as a root mean square of all the samples in the measured range.

To simplify implementation of the encoding procedures described here, the reference implementation section in [I-D.ietf-avtext-mixer-to-client-audio-level] provides a sample Java implementation of an audio level calculator that helps obtain such values from raw linear PCM audio samples.

In addition, a flag bit (labeled V) indicates whether the encoder believes the audio packet contains voice activity (1) or does not (0). The voice activity detection algorithm is unspecified and left implementation-specific.

When this header extension is used with RTP data sent using the RTP Payload for Redundant Audio Data *[RFC2198]*, the header's data describes the contents of the primary encoding.

Note: This audio level is defined in the same manner as is audio noise level in the RTP Payload Comfort Noise specification *[RFC3389]*. In the comfort noice specification, the overall magnitude of the noise level in comfort noise is encoded into the first byte of the payload, with spectral information about the noise in subsequent bytes. This specification's audio level parameter is defined so as to be identical to the comfort noise payload's noise-level byte.

4. Signaling (Setup) Information

The URI for declaring this header extension in an extmap attribute is "urn:ietf:params:rtp-hdrext:ssrc-audio-level". There is no additional setup information needed for this extension (i.e. no extensionattributes).

5. Considerations on Use

Mixers and forwarders generally should not base audio forwarding decisions directly on packet-by-packet audio level information, but rather should apply some analysis of the audio levels and trends. This general rule applies whether audio levels are provided by endpoints (as defined in this document), or are calculated at a server, as would be done in the absence of this information. This section discusses several issues that mixers and forwarders may wish to take into account. (Note that this section provides design guidance only, and is not normative.) First of all, audio levels should generally be measured over longer intervals than that of a single audio packet. In order to avoid false-positives for short bursts of sound (such as a cough or a dropped microphone), it is often useful to require that a participant's audio level be maintained for some period of time before considering it to be "real", i.e. some type of low-pass filter should be applied to the audio levels. Note, though, that such filtering must be balanced with the need to avoid clipping of the beginning of a speaker's speech. Additionally, different participants may have their audio input set differently. It may be useful to apply some sort of automatic gain control to the audio levels. There are a number of possible approaches to acheiving this, e.g. by measuring peak audio levels, by average audio levels during speech, or by measuring background audio levels (average audio level levels during non-speech).

6. Security Considerations

A malicious endpoint could choose to set the values in this header extension falsely, so as to falsely claim that audio or voice is or is not present. It is not clear what could be gained by falsely claiming that audio is not present, but an endpoint falsely claiming that audio is present could perform a denial-of-service attack on an audio conference, so as to send silence to suppress other conference members' audio. Thus, a device relying on audio level data from untrusted endpoints SHOULD periodically audit the level information transmitted, taking appropriate corrective action if endpoints appear to be sending incorrect data. (Note that as it is valid for an endpoint to choose to measure audio levels prior to encoding, some degree of discrepancy SHOULD be tolerated.)

In the Secure Real-Time Transport Protocol (SRTP) *[RFC3711]*, RTP header extensions are authenticated but not encrypted. When this header extension is used, audio levels are therefore visible on a packet-by-packet basis to an attacker passively observing the audio stream. As discussed in [I-D.perkins-avt-srtp-vbr-audio], such an attacker might be able to infer information about the conversation, possibly with phoneme-level resolution. In scenarios where this is a concern, additional mechanisms SHOULD be used to protect the confidentiality of the header extension. One solution is header extension encryption *[I-D.lennox-avtcore-srtp-encrypted-header-ext]*.

7. IANA Considerations

This document defines a new extension URI to the RTP Compact Header Extensions subregistry of the Real-Time Transport Protocol (RTP) Parameters registry, according to the following data:

Extension URI: urn:ietf:params:rtp-hdrext:ssrc-audio-level

Description: Audio Level

Contact: jonathan@vidyo.com

Reference: RFC XXXX

Note to RFC Editor: please replace "RFC XXXX" with the number of this RFC.

8. References

8.1. Normative References

[RFC5285]	Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, July 2008.
[RFC2198]	Perkins, C., Kouvelas, I., Hodson, O., Hardman, V., Handley, M., Bolot, J., Vega-Garcia, A. and S. Fosse-Parisis, "RTP Payload for Redundant Audio Data", RFC 2198, September 1997.
[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
[RFC3550]	Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

8.2. Informative References

[RFC3389]	Zopf, R., "Real-time Transport Protocol (RTP) Payload for Comfort Noise (CN)", RFC 3389, September 2002.
[ITU.G711.1988]	International Telecommunications Union , "Pulse Code Modulation (PCM) of Voice Frequencies", ITU-T Recommendation G.711, November 1988.
[I-D.ietf-avtext-mixer-to-client-audio-level]	Ivov, E, Marocco, E and J Lennox, "A Real-Time Transport Protocol (RTP) Header Extension for Mixer-to- Client Audio Level Indication", Internet-Draft draft-ietf-avtext-mixer-to-client-audio-level-06, November 2011.
[RFC3711]	Baugher, M., McGrew, D., Naslund, M., Carrara, E. and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.
[I-D.perkins-avt-srtp-vbr-audio]	Perkins, C and J Valin, "Guidelines for the use of Variable Bit Rate Audio with Secure RTP", Internet-Draft draft-perkins-avt-srtp-vbr-audio-05, December 2010.
[I-D.lennox-avtcore-srtp-encrypted-header-ext]	Lennox, J, "Encryption of Header Extensions in the Secure Real-Time Transport Protocol (SRTP)", Internet-Draft draft-lennox-avtcore-srtp-encrypted-header-ext-00, March 2011.

Appendix A. Changes From Earlier Versions

Note to the RFC-Editor: please remove this section prior to publication as an RFC.

Appendix A.1. Changes From Draft -02

- Changed encoding related text so that it would cover both the one-byte and the two-byte header formats.
- Clarified use of root mean square for dBov calculation
- Other minor editorial changes.

Appendix A.2. Changes From Draft -01

- Changed the URI for declaring this header extension from "urn:ietf:params:rtp-hdrext:audio-level" to "urn:ietf:params:rtp-hdrext:ssrc-audio-level" for consistency with [I-D.ietf-avtext-mixer-to-client-audio-level].

* Removed the "Limitations" section; it was discussing a potential extension that consensus indicated was out of scope of this document.

* Closed the P.56 open issue. It was agreed on IETF 80 that P.56 is mostly about speech levels and the levels transported by the extension defined here should also be able to serve as an indication for noise.

* Closed the open issue about transmitting noise floor information. Noise floor is (loosely) inferrable by observing the per-packet level information over a period of time, so the additional complexity seemed unnecessary.

* Editorial changes for consistency with [I-D.ietf-avtext-mixer-to-client-audio-level].

* Moved several descriptions of normative items that previously had only been described in informative sections of the text.

* Other editorial clarifications.

Appendix A.3. Changes From Draft -00

* Added references to the sample level calculator in [I-D.ietf-avtext-mixer-to-client-audio-level].

* Changed affiliation for Emil Ivov.

Appendix A.4. Changes From Individual Submission Draft -01

* This version is primarily a document refresh.

* Emil Ivov and Enrico Marocco have been added as co-authors.

* Additional open issues listed.

Appendix A.5. Changes From Individual Submission Draft -00

* The draft name has been changed to clarify that this document defines Client-To-Mixer Audio Levels, to more clearly distinguish it from [I-D.ietf-avtext-mixer-to-client-audio-level].

* The header extension format has been changed from a two-byte to a one-byte payload, eliminating the 7 reserved bits and the one must-be-zero bit.

* The sections Considerations on Use *[use]* and Limitations have been added.

* It has been noted that senders MAY indicate -127 dBov for digital silence, and that level measurement MAY be done prior to encoding audio.

* A reference to [I-D.lennox-avtcore-srtp-encrypted-header-ext] has been added to the security considerations.

* The term "header extension" is now used consistentenly throughout the document (as opposed to "extension header").

Authors' Addresses

Jonathan Lennox editor Lennox Vidyo, Inc. 433 Hackensack Avenue Seventh Floor Hackensack, NJ 07601 US EMail: jonathan@vidyo.com

Emil Ivov Ivov Jitsi Strasbourg, 67000 France EMail: emcho@jitsi.org

Enrico Marocco Marocco Telecom Itialia Via G. Reiss Romoli, 274 Turin, 10148 Italy EMail: enrico.marocco@telecomitalia.it