AVT	J. Lennox, Ed.
Internet-Draft	Vidyo
Intended status: Standards Track	E. Ivov
Expires: September 15, 2011	Jitsi
	E. Marocco
	Telecom Italia
	March 14, 2011

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

draft-ietf-avtext-client-to-mixer-audio-level-01

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 to, 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 both the level of the audio carried in the RTP payload of the packet it is associated with, as well as an indication as to whether voice activity has been detected in the packet.

The form of the audio level extension block is as follows:

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

The length field takes the value 0 to indicate that 1 byte follows.

The audio level is defined in the same manner as is audio noise level in the RTP Comfort Noise *[RFC3389]* specification. In that specification, the overall magnitude of the noise level 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.

The magnitude of the audio level is packed into the seven least significant bits of the single byte of the header extension, shown in Figure 1. 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 reference implementation section in [I-D.ietf-avtext-mixer-to-client-audio-level] provides a sample implementation of an audio level calculator that helps obtain such values from raw 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.

The audio level for digital silence (e.g. all-0 pcmu audio), for example for a muted audio source, MAY be represented as 127 (-127 dBov), regardless of the dynamic range of the encoded audio format.

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.

4. Signaling (Setup) Information

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

The audio levels carried by the extension header defined by this document are defined as dBov, decibels below system overload.
In principle, it could be more useful to have, instead, dB SPL, decibels of sound pressure level. In traditional telephony systems, telephone handsets were calibrated such that a particular (e.g.) u-law audio level, or analog voltage, corresponded to a particular sound pressure level at the handset's mouthpiece.
However, in many environments, this information is not available. Notably, PC soundcard hardware can only determine the levels of mic- or line-in at the hardware input, and operating systems usually allow further adjustments of audio input levels without providing information about these transformations to applications. Furthermore, in many circumstances, such as speech synthesis or mixed audio, an "audio" signal may in fact never have actually existed as sound pressure at all.
Thus, while information about the correspondance between dB SPL and dBov, or encoded audio, could be useful, this document does not attempt to define it. If there are circumstances in which this information would be useful, a separate header extension would be straightforward to define. (The information carried by such a header extension could indeed be useful independently from the information in the header extension defined by this document.)

7. 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 endpoints MAY choose to measure audio levels prior to encoding, so 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 would be header extension encryption *[I-D.lennox-avt-srtp-encrypted-extension-headers]*.

8. 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:audio-level

Description: Audio Level

Contact: jonathan@vidyo.com

Reference: RFC XXXX

9. References

9.1. 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.

9.2. 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.
[ITU.P56.1993]	

	International Telecommunications Union , "Objective Measurement of Active Speech Level", ITU-T Recommendation P.56, March 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-avt-srtp-encrypted-extension-headers]	Lennox, J, "Encryption of Header Extensions in the Secure Real-Time Transport Protocol (SRTP)", Internet-Draft draft-lennox-avt-srtp-encrypted-extension-headers-02, October 2010.

Appendix A. Open issues

- In order to more accurately determine signal-to-noise ratio, it would be useful for a sender to also send its estimate of its current audio noise floor. If so, it's unclear whether this would be better as a separate header extension element, or added to this header extension element.

- It has been suggested to reference ITU P.56 *[ITU.P56.1993]* for level measurement. This needs to be investigated.

Appendix B. Changes From Earlier Versions

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

Appendix B.1. Changes From Draft -01

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

- Changed affiliation for Emil Ivov.

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