Network Working Group
Request for Comments: 5391
Category: Standards Track

A. Sollaud
France Telecom

November 2008

RTP Payload Format for ITU-T Recommendation G.711.1

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This document specifies a Real-time Transport Protocol (RTP) payload format to be used for the ITU Telecommunication Standardization Sector (ITU-T) G.711.1 audio codec. Two media type registrations are also included.

Table of Contents

1. Introduction

The ITU Telecommunication Standardization Sector (ITU-T) Recommendation G.711.1 [ITU-G.711.1] is an embedded wideband extension of the Recommendation G.711 [ITU-G.711] audio codec. This document specifies a payload format for packetization of G.711.1 encoded audio signals into the Real-time Transport Protocol (RTP).

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 [RFC2119].

2. Background

G.711.1 is a G.711 embedded wideband speech and audio coding algorithm operating at 64, 80, and 96 kbps. At 64 kbps, G.711.1 is fully interoperable with G.711. Hence, an efficient deployment in existing G.711-based Voice over IP (VoIP) infrastructures is foreseen.

The codec operates on 5-ms frames, and the default sampling rate is 16 kHz. Input and output at 8 kHz are also supported for narrowband modes.

The encoder produces an embedded bitstream structured in three layers corresponding to three available bit rates: 64, 80, and 96 kbps. The bitstream can be truncated at the decoder side or by any component of the communication system to adjust, "on the fly", the bit rate to the desired value.

The following table gives more details on these layers.

Layer	Description	Bit rate
L0	G.711 compatible	64 kbps
L1	narrowband enhancement	16 kbps
L2	wideband enhancement	16 kbps

Table 1: Layers description

The combinations of these three layers results in the definition of four modes, as per the following table.

Mode	L0	L1	L2	Audio band	Bit rate
R1	x			narrowband	64 kbps
R2a	x	x		narrowband	80 kbps
R2b	x		x	wideband	80 kbps
R3	x	x	x	wideband	96 kbps

Table 2: Modes description

3. RTP Header Usage

The format of the RTP header is specified in [RFC3550]. The payload format defined in this document uses the fields of the header in a manner consistent with that specification.

marker (M):

G.711.1 does not define anything specific regarding Discontinuous Transmission (DTX), a.k.a. silence suppression. Codec-independent mechanisms may be used, like the generic comfort-noise payload format defined in [RFC3389].

For applications that send either no packets or occasional comfort-noise packets during silence, the first packet of a talkspurt -- that is, the first packet after a silence period during which packets have not been transmitted contiguously --

SHOULD be distinguished by setting the marker bit in the RTP data header to one. The marker bit in all other packets is zero. The beginning of a talkspurt MAY be used to adjust the playout delay to reflect changing network delays. Applications without silence suppression MUST set the marker bit to zero.

payload type (PT):
The assignment of an RTP payload type for this packet format is outside the scope of this document, and will not be specified here. It is expected that the RTP profile under which this payload format is being used will assign a payload type for this codec or specify that the payload type is to be bound dynamically (see Section 5.3).

timestamp:
The RTP timestamp clock frequency is the same as the default sampling frequency: 16 kHz.

G.711.1 has also the capability to operate with 8-kHz sampled input/output signals. It does not affect the bitstream, and the decoder does not require a priori knowledge about the sampling rate of the original signal at the input of the encoder. Therefore, depending on the implementation and the audio acoustic capabilities of the devices, the input of the encoder and/or the output of the decoder can be configured at 8 kHz; however, a 16-kHz RTP clock rate MUST always be used.

The duration of one frame is 5 ms, corresponding to 80 samples at 16 kHz. Thus, the timestamp is increased by 80 for each consecutive frame.

4. Payload Format

The complete payload consists of a payload header of 1 octet, followed by one or more consecutive G.711.1 audio frames of the same mode.

The mode may change between packets, but not within a packet.

4.1. Payload Header

The payload header is illustrated below.

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|0 0 0 0 0|  MI |
+-+-+-+-+-+-+-+-+
```

The five most significant bits are reserved for further extension and MUST be set to zero and MUST be ignored by receivers.

The Mode Index (MI) field (3 bits) gives the mode of the following frame(s) as per the table:

Mode Index	G.711.1 mode	Frame size
1	R1	40 octets
2	R2a	50 octets
3	R2b	50 octets
4	R3	60 octets

Table 3: Modes in payload header

All other values of MI are reserved for future use and MUST NOT be used.

Payloads received with an undefined MI value MUST be discarded.

If a restricted mode-set has been set up by the signaling (see Section 5), payloads received with an MI value not in this set MUST be discarded.

4.2. Audio Data

After this payload header, the consecutive audio frames are packed in order of time, that is, oldest first. All frames MUST be of the same mode, indicated by the MI field of the payload header.

Within a frame, layers are always packed in the same order: L0 then L1 for mode R2a, L0 then L2 for mode R2b, L0 then L1 then L2 for mode R3. This is illustrated below.

```
     +-------------------------------+
R1   |              L0               |
     +-------------------------------+

     +-------------------------------+--------+
R2a  |              L0               |   L1   |
     +-------------------------------+--------+

     +-------------------------------+--------+
R2b  |              L0               |   L2   |
     +-------------------------------+--------+

     +-------------------------------+--------+--------+
R3   |              L0               |   L1   |   L2   |
     +-------------------------------+--------+--------+
```

The size of one frame is given by the mode, as per Table 3, and the actual number of frames is easy to infer from the size of the audio data part:

nb_frames = (size_of_audio_data) / (size_of_one_frame).

Only full frames must be considered. So if there is a remainder to the division above, the corresponding remaining bytes in the received payload MUST be ignored.

5. Payload Format Parameters

This section defines the parameters that may be used to configure optional features in the G.711.1 RTP transmission.

Both A-law and mu-law G.711 are supported in the core layer L0, but there is no interoperability between A-law and mu-law. So two media types with the same parameters will be defined: audio/PCMA-WB for A-law core, and audio/PCMU-WB for mu-law core. This is consistent with the audio/PCMA and audio/PCMU media types separation for G.711 audio.

The parameters are defined here as part of the media subtype registrations for the G.711.1 codec. A mapping of the parameters into the Session Description Protocol (SDP) [RFC4566] is also provided for those applications that use SDP. In control protocols that do not use MIME or SDP, the media type parameters must be mapped to the appropriate format used with that control protocol.

5.1. PCMA-WB Media Type Registration

This registration is done using the template defined in [RFC4288] and following [RFC4855].

Type name: audio

Subtype name: PCMA-WB

Required parameters: none

Optional parameters:

mode-set: restricts the active codec mode set to a subset of all modes. Possible values are a comma-separated list of modes from the set: 1, 2, 3, 4 (see Mode Index in Table 3 of RFC 5391). The modes are listed in order of preference; first is preferred. If mode-set is specified, frames encoded with modes outside of the subset MUST NOT be sent in any RTP payload. If not present, all codec modes are allowed.

ptime: the recommended length of time (in milliseconds) represented by the media in a packet. It should be an integer multiple of 5 ms (the frame size). See Section 6 of RFC 4566.

maxptime: the maximum length of time (in milliseconds) that can be encapsulated in a packet. It should be an integer multiple of 5 ms (the frame size). See Section 6 of RFC 4566.

Encoding considerations: This media type is framed and contains binary data. See Section 4.8 of RFC 4288.

Security considerations: See Section 8 of RFC 5391.

Interoperability considerations: none

Published specification: RFC 5391

Applications that use this media type: Audio and video conferencing tools.

Additional information: none

Person & email address to contact for further information: Aurelien Sollaud, aurelien.sollaud@orange-ftgroup.com

Intended usage: COMMON

Restrictions on usage: This media type depends on RTP framing, and hence is only defined for transfer via RTP.

Author: Aurelien Sollaud

Change controller: IETF Audio/Video Transport working group delegated from the IESG

5.2. PCMU-WB Media Type Registration

This registration is done using the template defined in [RFC4288] and following [RFC4855].

Type name: audio

Subtype name: PCMU-WB

Required parameters: none

Optional parameters:

mode-set: restricts the active codec mode-set to a subset of all modes. Possible values are a comma-separated list of modes from the set: 1, 2, 3, 4 (see Mode Index in Table 3 of RFC 5391). The modes are listed in order of preference; first is preferred. If mode-set is specified, frames encoded with modes outside of the subset MUST NOT be sent in any RTP payload. If not present, all codec modes are allowed.

ptime: the recommended length of time (in milliseconds) represented by the media in a packet. It should be an integer multiple of 5 ms (the frame size). See Section 6 of RFC 4566.

maxptime: the maximum length of time (in milliseconds) that can be encapsulated in a packet. It should be an integer multiple of 5 ms (the frame size). See Section 6 of RFC 4566.

Encoding considerations: This media type is framed and contains binary data. See Section 4.8 of RFC 4288.

Security considerations: See Section 8 of RFC 5391.

Interoperability considerations: none

Published specification: RFC 5391

Applications that use this media type: Audio and video conferencing tools.

Additional information: none

Person & email address to contact for further information: Aurelien Sollaud, aurelien.sollaud@orange-ftgroup.com

Intended usage: COMMON

Restrictions on usage: This media type depends on RTP framing, and hence is only defined for transfer via RTP.

Author: Aurelien Sollaud

Change controller: IETF Audio/Video Transport working group delegated from the IESG

5.3. Mapping to SDP Parameters

The information carried in the media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [RFC4566], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing the G.711.1 codec, the mapping is as follows:

- o The media type ("audio") goes in SDP "m=" as the media name.

- o The media subtype ("PCMA-WB" or "PCMU-WB") goes in SDP "a=rtpmap" as the encoding name. The RTP clock rate in "a=rtpmap" MUST be 16000 for G.711.1.

- o The parameter "mode-set" goes in the SDP "a=fmtp" attribute by copying it as a "mode-set=<value>" string.

- o The parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

5.3.1. Offer-Answer Model Considerations

The following considerations apply when using SDP offer-answer procedures [RFC3264] to negotiate the use of G.711.1 payload in RTP:

- o Since G.711.1 is an extension of G.711, the offerer SHOULD announce G.711 support in its "m=audio" line, with G.711.1 preferred. This will allow interoperability with both G.711.1 and G.711-only capable parties. This is done by offering the PCMA media subtype in addition to PCMA-WB, and/or PCMU in addition to PCMU-WB.

Below is an example part of such an offer, for A-law:

```
m=audio 54874 RTP/AVP 96 8
a=rtpmap:96 PCMA-WB/16000
a=rtpmap:8 PCMA/8000
```

As a reminder, the payload format for G.711 is defined in Section 4.5.14 of [RFC3551].

- o The "mode-set" parameter is bi-directional; i.e., the restricted mode-set applies to media both to be received and sent by the declaring entity. If a mode-set was supplied in the offer, the answerer MUST return either the same mode-set or a subset of this mode-set. The answerer MAY change the preference order. If no mode-set was supplied in the offer, the anwerer MAY return a mode-set to restrict the possible modes. In any case, the mode-set in the answer then applies for both offerer and answerer. The offerer MUST NOT send frames of a mode that has been removed by the answerer.

  For multicast sessions, if "mode-set" is supplied in the offer, the answerer SHALL only participate in the session if it supports the offered mode-set.

- o The parameters "ptime" and "maxptime" will in most cases not affect interoperability. The SDP offer-answer handling of the "ptime" parameter is described in [RFC3264]. The "maxptime" parameter MUST be handled in the same way.

- o Any unknown parameter in an offer MUST be ignored by the receiver and MUST NOT be included in the answer.

Below are some example parts of SDP offer-answer exchanges.

- o Example 1

  Offer: G.711.1 all modes, with G.711 fallback, prefers mu-law

  ```
  m=audio 54874 RTP/AVP 96 97 0 8
  a=rtpmap:96 PCMU-WB/16000
  a=rtpmap:97 PCMA-WB/16000
  a=rtpmap:0 PCMU/8000
  a=rtpmap:8 PCMA/8000
  ```

  Answer: all modes accepted, both mu- and A-law.

  ```
  m=audio 59452 RTP/AVP 96 97
  a=rtpmap:96 PCMU-WB/16000
  a=rtpmap:97 PCMA-WB/16000
  ```

o Example 2

Offer: G.711.1 all modes, with G.711 fallback, prefers A-law

```
m=audio 54874 RTP/AVP 96 97 8 0
a=rtpmap:96 PCMA-WB/16000
a=rtpmap:97 PCMU-WB/16000
```

Answer: wants only A-law mode R3

```
m=audio 59452 RTP/AVP 96
a=rtpmap:96 PCMA-WB/16000
a=fmtp:96 mode-set=4
```

o Example 3

Offer: G.711.1 A-law with two modes, R2b and R3, with R3 preferred

```
m=audio 54874 RTP/AVP 96
a=rtpmap:96 PCMA-WB/16000
a=fmtp:96 mode-set=4,3
```

Answer: accepted

```
m=audio 59452 RTP/AVP 96
a=rtpmap:96 PCMA-WB/16000
a=fmtp:96 mode-set=4,3
```

If the answerer had wanted to restrict to one mode, it would have answered with only one value in the mode-set, for example mode-set=3 for mode R2b.

5.3.2. Declarative SDP Considerations

For declarative use of SDP, nothing specific is defined for this payload format. The configuration given by the SDP MUST be used when sending and/or receiving media in the session.

6. G.711 Interoperability

The L0 layer of G.711.1 is fully interoperable with G.711, and is embedded in all modes of G.711.1. This provides an easy G.711.1 - G.711 transcoding process.

A gateway or any other network device receiving a G.711.1 packet can easily extract a G.711-compatible payload, without the need to decode and re-encode the audio signal. It simply has to take the audio data of the payload, and strip the upper layers (L1 and/or L2), if any.

If a G.711.1 packet contains several frames, the concatenation of the L0 layers of each frame will form a G.711-compatible payload.

7. Congestion Control

Congestion control for RTP SHALL be used in accordance with [RFC3550] and any appropriate profile (for example, [RFC3551]).

The embedded nature of G.711.1 audio data can be helpful for congestion control, since a coding mode with a lower bit rate can be selected when needed. This property is usable only when multiple modes have been negotiated (either no "mode-set" parameter in the SDP or a "mode-set" with at least two modes).

The number of frames encapsulated in each RTP payload influences the overall bandwidth of the RTP stream, due to the header overhead. Packing more frames in each RTP payload can reduce the number of packets sent and hence the header overhead, at the expense of increased delay and reduced error robustness.

8. Security Considerations

RTP packets using the payload format defined in this specification are subject to the general security considerations discussed in the RTP specification [RFC3550] and any appropriate profile (for example, [RFC3551]).

As this format transports encoded speech/audio, the main security issues include confidentiality, integrity protection, and authentication of the speech/audio itself. The payload format itself does not have any built-in security mechanisms. Any suitable external mechanisms, such as the Secure Real-time Transport Protocol (SRTP) [RFC3711], MAY be used.

This payload format and the G.711.1 encoding do not exhibit any significant non-uniformity in the receiver-end computational load, and thus they are unlikely to pose a denial-of-service threat due to the receipt of pathological datagrams. In addition, they do not contain any type of active content such as scripts.

9. IANA Considerations

Two new media subtypes (audio/PCMA-WB and audio/PCMU-WB) have been registered by IANA. See Sections 5.1 and 5.2.

10. References

10.1. Normative References

[ITU-G.711.1] International Telecommunications Union, "Wideband embedded extension for G.711 pulse code modulation", ITU-T Recommendation G.711.1, March 2008.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, June 2002.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.

[RFC4288] Freed, N. and J. Klensin, "Media Type Specifications and Registration Procedures", BCP 13, RFC 4288, December 2005.

[RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, July 2006.

[RFC4855] Casner, S., "Media Type Registration of RTP Payload Formats", RFC 4855, February 2007.

10.2. Informative References

[ITU-G.711] International Telecommunications Union, "Pulse code modulation (PCM) of voice frequencies", ITU-T Recommendation G.711, November 1988.

[RFC3389] Zopf, R., "Real-time Transport Protocol (RTP) Payload for Comfort Noise (CN)", RFC 3389, September 2002.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

Author's Address

Aurelien Sollaud
France Telecom
2 avenue Pierre Marzin
Lannion Cedex 22307
France

Phone: +33 2 96 05 15 06
EMail: aurelien.sollaud@orange-ftgroup.com