Network Working Group
Request for Comments: 5577
Obsoletes: 3047
Category: Standards Track
P. Luthi
Tandberg
R. Even
Gesher Erove Ltd
July 2009

RTP Payload Format for ITU-T Recommendation G.722.1

Abstract

International Telecommunication Union (ITU-T) Recommendation G.722.1 is a wide-band audio codec. This document describes the payload format for including G.722.1-generated bit streams within an RTP packet. The document also describes the syntax and semantics of the Session Description Protocol (SDP) parameters needed to support G.722.1 audio codec.

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

Table of Contents

1. Introduction

ITU-T G.722.1 [ITU.G7221] is a low-complexity coder; it compresses 50 Hz - 14 kHz audio signals into one of the following bit rates: 24 kbit/s, 32 kbit/s, or 48 kbit/s.

The coder may be used for speech, music, and other types of audio.

Some of the applications for which this coder is suitable are:

- o Real-time communications such as videoconferencing and telephony
- o Streaming audio
- o Archival and messaging

ITU-T G.722.1 [ITU.G7221] uses 20-ms frames and a sampling rate clock of 16 kHz or 32kHz. The encoding and decoding algorithm can change the bit rate at any 20-ms frame boundary, but no bit rate change notification is provided in-band with the bit stream.

For any given bit rate, the number of bits in a frame is a constant. Within this fixed frame, G.722.1 uses variable-length coding (e.g., Huffman coding) to represent most of the encoded parameters. All variable-length codes are transmitted in order from the leftmost bit (most significant bit -- MSB) to the rightmost bit (least significant bit -- LSB), see [ITU.G7221] for more details.

The ITU-T standardized bit rates for G.722.1 are 24 kbit/s or 32kbit/s at 16 Khz sample rate, and 24 kbit/s, 32 kbit/s, or 48 kbit/s at 32 Khz sample rate. However, the coding algorithm itself has the capability to run at any user-specified bit rate (not just 24, 32, and 48 kbit/s) while maintaining an audio bandwidth of 50 Hz to 14 kHz. This rate change is accomplished by a linear scaling of the codec operation, resulting in frames with size in bits equal to 1/50 of the corresponding bit rate.

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 RTP implementations.

3. RTP Usage for G.722.1

3.1. RTP G.722.1 Header Fields

The RTP header is defined in the RTP specification [RFC3550]. This section defines how fields in the RTP header are used.

Payload Type (PT): The assignment of an RTP payload type for this packet format is outside the scope of this document; it is specified by the RTP profile under which this payload format is used, or it is signaled dynamically out-of-band (e.g., using SDP).

Marker (M) bit: The M bit is set to zero.

Extension (X) bit: Defined by the RTP profile used.

Timestamp: A 32-bit word that corresponds to the sampling instant for the first frame in the RTP packet. The sampling frequency can be 16 Khz or 32 Khz. The RTP timestamp clock frequency of 32 Khz SHOULD be used unless only an RTP stream sampled at 16 Khz is going to be sent.

3.2. RTP Payload Format for G.722.1

The RTP payload for G.722.1 has the format shown in Figure 1. No additional header fields specific to this payload format are required.

```
 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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                      RTP Header                               |
 +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
 |                                                               |
 +                 one or more frames of G.722.1                 |
 |                             ....                              |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: RTP payload for G.722.1.

Because bit rate is not signaled in-band, a separate out-of-band method is REQUIRED to indicate the bit rate (see Section 5 for an example of signaling bit rate information using SDP). For the payload format specified here, the bit rate MUST remain constant for a particular payload type value. An application MAY switch bit rates and clock rates from packet to packet by defining different payload type values and switching between them.

The use of Huffman coding means that it is not possible to identify the various encoded parameters/fields contained within the bit stream without first completely decoding the entire frame. For the purposes of packetizing the bit stream in RTP, it is only necessary to consider the sequence of bits as output by the G.722.1 encoder and to present the same sequence to the decoder. The payload format described here maintains this sequence.

When operating at 24 kbit/s, 480 bits (60 octets) are produced per frame. When operating at 32 kbit/s, 640 bits (80 octets) are produced per frame. When operating at 48 kbit/s, 960 bits (120 octets) are produced per frame. Thus, all three bit rates allow for octet alignment without the need for padding bits.

Figure 2 illustrates how the G.722.1 bit stream MUST be mapped into an octet-aligned RTP payload.

```
 first bit                                                last bit
 transmitted                                           transmitted
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                                 |
 + sequence of bits (480, 640, or 960) generated by the           |
 |             G.722.1 encoder for transmission                    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


 |           |           |                      |           |
 |           |           |      ...             |           |
 |           |           |                      |           |


 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+ ... +-+-+-+-+-+-+-+-+-+-+-+-+-+
 |MSB...   LSB|MSB...   LSB|                     |MSB...   LSB|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+ ... +-+-+-+-+-+-+-+-+-+-+-+-+-+
   RTP          RTP                                 RTP
   octet 1      octet 2                             octet
                                                 60, 80, 120
```

Figure 2: The G.722.1 encoder bit stream is split into a sequence of octets (60, 80, or 120 depending on the bit rate), and each octet is in turn mapped into an RTP octet.

When operating at non-standard rates, the payload format MUST follow the guidelines illustrated in Figure 2. It is RECOMMENDED that values in the range 16000 to 48000 be used. Non-standard rates MUST have a value that is a multiple of 400 (this maintains octet alignment and does not then require (undefined) padding bits for each frame if not octet aligned). For example, a bit rate of 16.4 kbit/s will result in a frame of size 328 bits or 41 octets, which is mapped into RTP per Figure 2.

3.3. Multiple G.722.1 Frames in an RTP Packet

A sender may include more than one consecutive G.722.1 frame in a single RTP packet.

Senders have the following additional restrictions:

- o Sender SHOULD NOT include more G.722.1 frames in a single RTP packet than will fit in the MTU of the RTP transport protocol.

- o All frames contained in a single RTP packet MUST be of the same length and sampled at the same clock rate. They MUST have the same bit rate (octets per frame).

o Frames MUST NOT be split between RTP packets.

It is RECOMMENDED that the number of frames contained within an RTP packet be consistent with the application. For example, in a telephony application where delay is important, the fewer frames per packet, the lower the delay; whereas for a delay-insensitive streaming or messaging application, many frames per packet would be acceptable.

3.4. Computing the Number of G.722.1 Frames

Information describing the number of frames contained in an RTP packet is not transmitted as part of the RTP payload. The only way to determine the number of G.722.1 frames is to count the total number of octets within the RTP packet and divide the octet count by the number of expected octets per frame. This expected octet-per-frame count is derived from the bit rate and is therefore a function of the payload type.

4. IANA Considerations

This document updates the G7221 media type described in RFC 3047.

4.1. Media Type Registration

This section describes the media types and names associated with this payload format. The section registers the media types, as per RFC 4288 [RFC4288]

4.1.1. Registration of Media Type audio/G7221

Media type name: audio

Media subtype name: G7221

Required parameters:

bitrate: the data rate for the audio bit stream. This parameter is mandatory because the bit rate is not signaled within the G.722.1 bit stream. At the standard G.722.1 bit rates, the value MUST be either 24000 or 32000 at 16 Khz sample rate, and 24000, 32000, or 48000 at 32 Khz sample rate. If using the non-standard bit rates, then it is RECOMMENDED that values in the range 16000 to 48000 be used. Non-standard rates MUST have a value that is a multiple of 400 (this maintains octet alignment and does not then require (undefined) padding bits for each frame if not octet aligned).

Optional parameters:

rate: RTP timestamp clock rate, which is equal to the sampling rate. If the parameter is not specified, a clock rate of 16 Khz is assumed.

ptime: see RFC 4566. SHOULD be a multiple of 20 ms.

maxptime: see RFC 4566. SHOULD be a multiple of 20 ms.

Encoding considerations:

This media type is framed and binary, see Section 4.8 in [RFC4288].

Security considerations: See Section 6

Interoperability considerations:

Terminals SHOULD offer a media type at 16 Khz sample rate in order to interoperate with terminals that do not support the new 32 Khz sample rate.

Published specification: RFC 5577.

Applications that use this media type:

Audio and Video streaming and conferencing applications.

Additional information: none

Person and email address to contact for further information :

Roni Even: ron.even.tlv@gmail.com

Intended usage: COMMON

Restrictions on usage:

This media type depends on RTP framing, and hence is only defined for transfer via RTP [RFC3550]. Transport within other framing protocols is not defined at this time.

Author: Roni Even

Change controller:

IETF Audio/Video Transport working group delegated from the IESG.

5. SDP Parameters

The media types audio/G7221 are mapped to fields in the Session Description Protocol (SDP) [RFC4566] as follows:

- o The media name in the "m=" line of SDP MUST be audio.
- o The encoding name in the "a=rtpmap" line of SDP MUST be G7221 (the media subtype).
- o The parameter "rate" goes in "a=rtpmap" as clock rate parameter.
- o Only one bitrate SHALL be defined (using the "bitrate=" parameter in the a=fmtp line) for each payload type.

5.1. Usage with the SDP Offer/Answer Model

When offering G.722.1 over RTP using SDP in an Offer/Answer model [RFC3264], the following considerations are necessary.

The combination of the clock rate in the rtpmap and the bitrate parameter defines the configuration of each payload type. Each configuration intended to be used MUST be declared.

There are two sampling clock rates defined for G.722.1 in this document. RFC 3047 [RFC3047] supports only the 16 Khz clock rate. Therefore, a system that wants to use G.722.1 SHOULD offer a payload type with clock rate of 16000 for backward interoperability.

An example of an offer that includes a 16000 and 32000 clock rate is:

```
m=audio 49000 RTP/AVP 121 122
a=rtpmap:121 G7221/16000
a=fmtp:121 bitrate=24000
a=rtpmap:122 G7221/32000
a=fmtp:122 bitrate=48000
```

6. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [RFC3550] and any appropriate RTP profile. The main security considerations for the RTP packet carrying the RTP payload format defined within this memo are confidentiality, integrity, and source authenticity. Confidentiality is achieved by encryption of the RTP payload. Integrity of the RTP packets is achieved through a suitable cryptographic integrity-protection mechanism. Such a cryptographic system may also allow the authentication of the source of the payload. A suitable security mechanism for this RTP payload format should provide confidentiality, integrity protection, and at least source authentication capable of determining if an RTP packet is from a member of the RTP session.

Note that the appropriate mechanism to provide security to RTP and payloads following this memo may vary. It is dependent on the application, the transport, and the signaling protocol employed. Therefore, a single mechanism is not sufficient; although, if suitable, usage of the Secure Real-time Transport Protocol (SRTP) is [RFC3711] recommended. Another mechanism that may be used is IPsec [RFC4301] Transport Layer Security (TLS) [RFC5246] (RTP over TCP); other alternatives may exist.

This RTP payload format and its media decoder do not exhibit any significant non-uniformity in the receiver-side computational complexity for packet processing, and thus are unlikely to pose a denial-of-service threat due to the receipt of pathological data. Nor does the RTP payload format contain any active content.

7. Changes from RFC 3047

This specification obsoletes RFC 3047, adding the support for the Superwideband (14 Khz) audio support defined in annex C of the new revision of ITU-T G.722.1.

Other changes:

Updated the text to be in line with the current rules for RFCs and with media type registration conforming to RFC 4288.

8. Acknowledgments

The authors would like to thank Tom Taylor for his contribution to this work.

9. References

9.1. Normative References

[ITU.G7221] International Telecommunications Union, "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss", ITU-T Recommendation G.722.1, 2005.

[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.

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

9.2. Informative References

[RFC3047] Luthi, P., "RTP Payload Format for ITU-T Recommendation G.722.1", RFC 3047, January 2001.

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

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

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.

Authors' Addresses

Patrick Luthi
Tandberg
Philip Pedersens vei 22
1366 Lysaker
Norway

EMail: patrick.luthi@tandberg.no

Roni Even
Gesher Erove Ltd
14 David Hamelech
Tel Aviv 64953
Israel

EMail: ron.even.tlv@gmail.com