Network Working Group   A. Sollaud
Request for Comments: 4749   France Telecom
Category: Standards Track   October 2006

RTP Payload Format for the G.729.1 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

Abstract

This document specifies a Real-time Transport Protocol (RTP) payload format to be used for the International Telecommunication Union (ITU-T) G.729.1 audio codec. A media type registration is included for this payload format.

Table of Contents

1. Introduction

The International Telecommunication Union (ITU-T) recommendation G.729.1 [1] is a scalable and wideband extension of the recommendation G.729 [9] audio codec. This document specifies the payload format for packetization of G.729.1 encoded audio signals into the Real-time Transport Protocol (RTP).

The payload format itself is described in Section 5. A media type registration and the details for the use of G.729.1 with SDP are given in Section 6.

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 [2].

2. Background

G.729.1 is an 8-32 kbps scalable wideband (50-7000 Hz) speech and audio coding algorithm interoperable with G.729, G.729 Annex A, and G.729 Annex B. It provides a standardized solution for packetized voice applications that allows a smooth transition from narrowband to wideband telephony.

The most important services addressed are IP telephony and videoconferencing, either for enterprise corporate networks or for mass market (like Public Switched Telephone Network (PSTN) emulation over DSL or wireless access). Target devices can be IP phones or other VoIP handsets, home gateways, media gateways, IP Private Branch Exchange (IPBX), trunking equipment, voice messaging servers, etc.

For all those applications, the scalability feature allows tuning the bit rate versus quality trade-off, possibly in a dynamic way during a session, taking into account service requirements and network transport constraints.

The G.729.1 coder produces an embedded bitstream structured in 12 layers corresponding to 12 available bit rates between 8 and 32 kbps. The first layer, at 8 kbps, is called the core layer and is bitstream compatible with the ITU-T G.729/G.729A coder. At 12 kbps, a second layer improves the narrowband quality. Upper layers provide wideband audio (50-7000 Hz) between 14 and 32 kbps, with a 2 kbps granularity allowing graceful quality improvements. Only the core layer is mandatory to decode understandable speech; upper layers provide quality enhancement and wideband enlargement.

The codec operates on 20-ms frames, and the default sampling rate is 16 kHz. Input and output at 8 kHz are also supported, at all bit rates.

3. Embedded Bit Rates Considerations

The embedded property of G.729.1 streams provides a mechanism to adjust the bandwidth demand. At any time, a sender can change its sending bit rate without external signalling, and the receiver will be able to properly decode the frames. It may help to control congestion, since the bandwidth can be adjusted by selecting another bit rate.

The ability to adjust the bandwidth may also help when having a fixed bandwidth link dedicated to voice calls, for example in a residential or trunking gateway. In that case, the system can change the bit rates depending on the number of simultaneous calls. This will only impact the sending bandwidth. In order to adjust the receiving bandwidth as well, we introduce an in-band signalling to request the other party to change its own sending bit rate. This in-band request is called MBS, for Maximum Bit rate Supported. It is described in Section 5.2. Note that it is only useful for two-way unicast G.729.1 traffic, because when A sends an in-band MBS to B in order to request that B modify its sending bit rate, it concerns the stream from B to A. If there is no G.729.1 stream in the reverse direction, the MBS will have no effect.

4. RTP Header Usage

The format of the RTP header is specified in RFC 3550 [3]. This payload format uses the fields of the header in a manner consistent with that specification.

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

G.729.1 has also the capability to operate with 8 kHz sampled input/output signals at all bit rates. 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 20 ms, corresponding to 320 samples at 16 kHz. Thus the timestamp is increased by 320 for each consecutive frame.

The M bit MUST be set to zero in all packets.

The assignment of an RTP payload type for this packet format is outside the scope of the 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 6.2).

5. Payload Format

5.1. Payload Structure

The complete payload consists of a payload header of 1 octet, followed by zero or more consecutive audio frames at the same bit rate.

The payload header consists of two fields: MBS (see Section 5.2) and FT (see Section 5.3).

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  MBS  |   FT  |                                               |
+-+-+-+-+-+-+-+-+                                               +
:                zero or more frames at the same bit rate       :
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5.2. Payload Header: MBS Field

MBS (4 bits): maximum bit rate supported. Indicates a maximum bit rate to the encoder at the site of the receiver of this payload. The value of the MBS field is set according to the following table:

MBS	max bit rate
0	8 kbps
1	12 kbps
2	14 kbps
3	16 kbps
4	18 kbps
5	20 kbps
6	22 kbps
7	24 kbps
8	26 kbps
9	28 kbps
10	30 kbps
11	32 kbps
12-14	(reserved)
15	NO_MBS

The MBS is used to tell the other party the maximum bit rate one can receive. The encoder MUST NOT exceed the sending rate indicated by the received MBS. Note that, due to the embedded property of the coding scheme, the encoder can send frames at the MBS rate or any lower rate. As long as it does not exceed the MBS, the encoder can change its bit rate at any time without previous notice.

Note that the MBS is a codec bit rate; the actual network bit rate is higher and depends on the overhead of the underlying protocols.

The MBS received is valid until the next MBS is received, i.e., a newly received MBS value overrides the previous one.

If a payload with a reserved MBS value is received, the MBS MUST be ignored.

The MBS field MUST be set to 15 for packets sent to a multicast group and MUST be ignored on packets received from a multicast group.

The MBS field MUST be set to 15 in all packets when the actual MBS value is sent through non-RTP means. This is out of the scope of this specification.

See Sections 3 and 7 for more details on the use of MBS for congestion control.

5.3. Payload Header: FT Field

FT (4 bits): Frame type of the frame(s) in this packet, as per the following table:

FT	encoding rate	frame size
0	8 kbps	20 octets
1	12 kbps	30 octets
2	14 kbps	35 octets
3	16 kbps	40 octets
4	18 kbps	45 octets
5	20 kbps	50 octets
6	22 kbps	55 octets
7	24 kbps	60 octets
8	26 kbps	65 octets
9	28 kbps	70 octets
10	30 kbps	75 octets
11	32 kbps	80 octets
12-14	(reserved)	
15	NO_DATA	0

The FT value 15 (NO_DATA) indicates that there is no audio data in the payload. This MAY be used to update the MBS value when there is no audio frame to transmit. The payload will then be reduced to the payload header.

If a payload with a reserved FT value is received, the whole payload MUST be ignored.

5.4. Audio Data

Audio data of a payload contains one or more consecutive audio frames at the same bit rate. The audio frames are packed in order of time, that is, oldest first.

The size of one frame is given by the FT field, as per the table in Section 5.3, 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.

Note that if FT=15, there will be no audio frame in the payload.

6. Payload Format Parameters

This section defines the parameters that may be used to configure optional features in the G.729.1 RTP transmission.

The parameters are defined here as part of the media subtype registration for the G.729.1 codec. A mapping of the parameters into the Session Description Protocol (SDP) [5] 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.

6.1. Media Type Registration

This registration is done using the template defined in RFC 4288 [6] and following RFC 3555 [7].

Type name: audio

Subtype name: G7291

Required parameters: none

Optional parameters:

maxbitrate: the absolute maximum codec bit rate for the session, in bits per second. Permissible values are 8000, 12000, 14000, 16000, 18000, 20000, 22000, 24000, 26000, 28000, 30000, and 32000. 32000 is implied if this parameter is omitted. The maxbitrate restricts the range of bit rates which can be used. The bit rates indicated by FT and MBS fields in the RTP packets MUST NOT exceed maxbitrate.

mbs: the current maximum codec bit rate supported as a receiver, in bits per second. Permissible values are in the same set as for the maxbitrate parameter, with the constraint that mbs MUST be lower or equal to maxbitrate. If the mbs parameter is omitted, it is set to the maxbitrate value. So if both mbs and maxbitrate are omitted, they are both set to 32000. The mbs parameter corresponds to a MBS value in the RTP packets as per table in Section 5.2 of RFC 4749. Note that this parameter may be dynamically updated by the MBS field of the RTP packets sent; it is not an absolute value for the session.

ptime: the recommended length of time (in milliseconds) represented by the media in a packet. See Section 6 of RFC 4566 [5].

maxptime: the maximum length of time (in milliseconds) that can be encapsulated in a packet. See Section 6 of RFC 4566 [5]

Encoding considerations: This media type is framed and contains binary data; see Section 4.8 of RFC 4288 [6].

Security considerations: See Section 8 of RFC 4749

Interoperability considerations: none

Published specification: RFC 4749

Applications which 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 [3].

Author: Aurelien Sollaud

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

6.2. Mapping to SDP Parameters

The information carried in the media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [5], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing the G.729.1 codec, the mapping is as follows:

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

o The media subtype ("G7291") goes in SDP "a=rtpmap" as the encoding name. The RTP clock rate in "a=rtpmap" MUST be 16000 for G.729.1.

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

- o Any remaining parameters go in the SDP "a=fmtp" attribute by copying them directly from the media type string as a semicolon separated list of parameter=value pairs.

Some example SDP session descriptions utilizing G.729.1 encodings follow.

Example 1: default parameters

```
m=audio 53146 RTP/AVP 98
a=rtpmap:98 G7291/16000
```

Example 2: recommended packet duration of 40 ms (=2 frames), maximum bit rate is 12 kbps, and initial MBS set to 8 kbps. It could be a loaded PSTN gateway which can operate at 12 kbps but asks to initially reduce the bit rate to 8 kbps.

```
m=audio 51258 RTP/AVP 99
a=rtpmap:99 G7291/16000
a=fmtp:99 maxbitrate=12000; mbs=8000
a=ptime:40
```

6.2.1. Offer-Answer Model Considerations

The following considerations apply when using SDP offer-answer procedures [8] to negotiate the use of G.729.1 payload in RTP:

- o Since G.729.1 is an extension of G.729, the offerer SHOULD announce G.729 support in its "m=audio" line, with G.729.1 preferred. This will allow interoperability with both G.729.1 and G.729-only capable parties.

  Below is an example of such an offer:

  ```
  m=audio 55954 RTP/AVP 98 18
  a=rtpmap:98 G7291/16000
  a=rtpmap:18 G729/8000
  ```

  If the answerer supports G.729.1, it will keep the payload type 98 in its answer, and the conversation will be done using G.729.1. Else, if the answerer supports only G.729, it will leave only the payload type 18 in its answer, and the conversation will be done using G.729 (the payload format for G.729 is defined in Section 4.5.6 of RFC 3551 [4]).

Note that when used at 8 kbps in G.729-compatible mode, the G.729.1 decoder supports G.729 Annex B. Therefore, Annex B can be advertised (by default, annexb=yes for G729 media type; see Section 4.1.9 of RFC 3555 [7]).

- o The "maxbitrate" parameter is bi-directional. If the offerer sets a maxbitrate value, the answerer MUST reply with a smaller or equal value. The actual maximum bit rate for the session will be the minimum.

- o If the received value for "maxbitrate" is between 8000 and 32000 but not in the permissible values set, it SHOULD be read as the closest lower valid value. If the received value is lower than 8000 or greater than 32000, the session MUST be rejected.

- o The "mbs" parameter is not symmetric. Values in the offer and the answer are independent and take into account local constraints. One party MUST NOT start sending frames at a bit rate higher than the "mbs" of the other party. The parameter allows announcing this value, prior to the sending of any packet, to prevent the remote sender from exceeding the MBS at the beginning of the session.

- o If the received value for "mbs" is greater or equal to 8000 but not in the permissible values set, it SHOULD be read as the closest lower valid value. If the received value is lower than 8000, the session MUST be rejected.

- 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 RFC 3264 [8]. 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.

Some special rules apply for mono-directional traffic:

- o For sendonly streams, the "mbs" parameter is useless and SHOULD NOT be used.

- o For recvonly streams, the "mbs" parameter is the only way to communicate the MBS to the sender, since there is no RTP stream towards it. So to request a bit rate change, the receiver will need to use an out-of-band mechanism, like a SIP RE-INVITE.

Some special rules apply for multicast:

- o The "mbs" parameter MUST NOT be used.

- o The "maxbitrate" parameter becomes declarative and MUST NOT be negotiated. This parameter is fixed, and a participant MUST use the configuration that is provided for the session.

6.2.2. Declarative SDP Considerations

For declarative use of SDP such as in SAP [10] and RTSP [11], the following considerations apply:

- o The "mbs" parameter MUST NOT be used.

- o The "maxbitrate" parameter is declarative and provides the parameter that SHALL be used when receiving and/or sending the configured stream.

7. Congestion Control

Congestion control for RTP SHALL be used in accordance with RFC 3550 [3] and any appropriate profile (for example, RFC 3551 [4]). The embedded and variable bit rates capability of G.729.1 provides a mechanism that may help to control congestion; see Section 3 for more details.

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 [3] and any appropriate profile (for example, RFC 3551 [4]).

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 SRTP [12], MAY be used.

This payload format and the G.729.1 encoding do not exhibit any significant non-uniformity in the receiver-end computational load and thus are unlikely to pose a denial-of-service threat due to the receipt of pathological datagrams.

9. IANA Considerations

IANA has registered audio/G7291 as a media subtype; see Section 6.1.

10. References

10.1. Normative References

[1] International Telecommunications Union, "G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729", ITU-T Recommendation G.729.1, May 2006.

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

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

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

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

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

[7] Casner, S. and P. Hoschka, "MIME Type Registration of RTP Payload Formats", RFC 3555, July 2003.

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

10.2. Informative References

[9] International Telecommunications Union, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)", ITU-T Recommendation G.729, March 1996.

[10] Handley, M., Perkins, C., and E. Whelan, "Session Announcement Protocol", RFC 2974, October 2000.

[11] Schulzrinne, H., Rao, A., and R. Lanphier, "Real Time Streaming Protocol (RTSP)", RFC 2326, April 1998.

[12] 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

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).