## RTP payload format for the future scalable and wideband extension of G.729 audio codec
### draft-sollaud-avt-rtp-g729-scal-wb-ext-02

Status of this Memo

Copyright Notice

Abstract

   This document specifies a real-time transport protocol (RTP) payload
   format to be used for the future scalable and wideband extension of
   the International Telecommunication Union (ITU-T) G.729 audio codec.
   A media type registration is included for this payload format.

Table of Contents

## 1.  Introduction

The International Telecommunication Union (ITU-T) is working on a scalable and wideband extension of its recommendation G.729 [7]. This future audio codec will be called G.729EV in the following text. This document specifies the payload format for packetization of G.729EV encoded audio signals into the real-time transport protocol (RTP).

The payload format itself and the handling of variable bit rate are described in Section 4.  A media type registration and the details for the use of G.729EV with SDP are given in Section 5.

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 [1].

## 2.  Background

G.729EV is mainly designed to be used as a speech codec, but it can be used for music at the highest bit rates.  The sampling frequency is 16000 Hz and the frame size is 20 ms.

This G.729-based codec produces an embedded bitstream providing an improved narrow band quality [300, 3400 Hz] at 12 kbps, and an enhanced and gracefully improving wideband quality [50, 7000 Hz] from 14 kbps to 32 kbps, by steps of 2 kbps.  At 8 kbps it generates a G.729 bitstream.

It has been mainly designed for packetized wideband voice applications (Voice over IP or ATM, Telephony over IP, private networks...) and particularly for those requiring scalable bandwidth, enhanced quality above G.729, and easy integration into existing infrastructures.

G.729EV is also designed to cope with other services like high quality audio/video conferencing, archival, messaging, etc.

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

G.729EV produces frames that are said embedded because they are composed of embedded layers.  The first layer is called the core layer and is bitstream compatible with the ITU-T G.729 with annex B coder.  Upper layers are added while bit rate increases, to improve

quality and enlarge audio bandwidth from narrowband to wideband.  As
a result, a received frame can be decoded at its original bit rate or
at any lower bit rate corresponding to lower layers which are
embedded.  Only the core layer is mandatory to decode understandable
speech, upper layers provide quality enhancement and wideband
enlargement.

Audio codecs often support voice activity detection (VAD) and comfort
noise generation (CNG).  During silence periods, the coder may
significantly decrease the transmitted bit rate by sending only
comfort noise parameters in special small frames called silence
insertion descriptors (SID).  The receiver's decoder will generate
comfort noise according to the SID information.  This operation of
sending low bit rate comfort noise parameters during silence periods
is usually called discontinuous transmission (DTX).

G.729EV will be first released without support for DTX.  Anyway, this
functionality is planned and will be defined in a separate annex
later.  Thus this specification provides DTX signalling, even if the
size of a SID frame is not yet standardized.


## 3.  RTP header usage

The format of the RTP header is specified in RFC 3550 [2].  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 sampling
frequency, that is 16 kHz.  So the timestamp unit is in samples.

The duration of one frame is 20 ms, corresponding to 320 samples per
frame.  Thus the timestamp is increased by 320 for each consecutive
frame.

The M bit should be set as specified in the applicable RTP profile,
for example, RFC 3551 [3].

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 5.2).


## 4.  Payload format

## 4.1.  Payload structure

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

```
 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  |                                               |
+-+-+-+-+-+-+-+-+                                               +
:                one ore more frames at the same bit rate       :
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
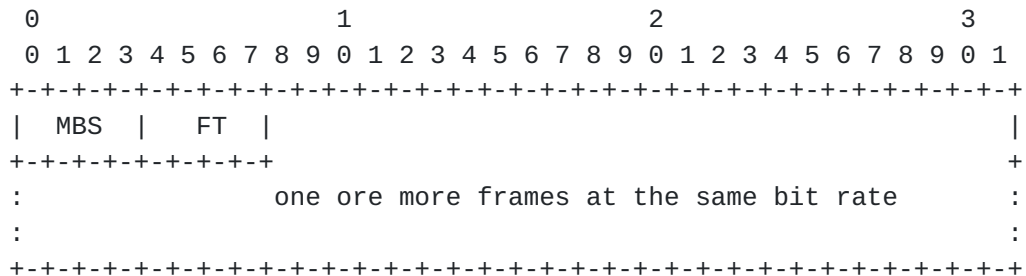
## 4.2.  Payload Header: MBS field

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 follow the received MBS.  It MUST NOT send
frames at a bit rate higher than the received MBS.  Thanks to the
embedded property of the coding scheme, note that it can send frames
at the MBS rate or any lower rate.  As long as it does not exceed the
MBS, it can change its bit rate at any time without previous notice.

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 an invalid MBS value is received, the MBS MUST be ignored.

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 field MUST be set to 15 for packets sent to 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.

## 4.3.  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 an invalid FT value is received, the whole payload MUST be ignored.

4.4.  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 the older first.

   The actual number of frame is easy to infer from the size of the
   audio data part:

      nb_frames = (size_of_audio_data) / (size_of_one_frame).

   This is compatible with DTX, with the restriction that the SID frame
   MUST be at the end of the payload (it is consistent with the payload
   format of G.729 described in section 4.5.6 of RFC 3551 [3]).  Since
   the SID frame is much smaller than any other frame, it will not
   hinder the calculation of the number of frames at the receiver side
   and can be easily detected.  Actually the presence of a SID frame
   will be inferred by the result of the above division not being an
   integer.

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


5.  Payload format parameters

   This section defines the parameters that may be used to configure
   optional features in the G.729EV RTP transmission.

   The parameters are defined here as part of the media subtype
   registration for the G.729EV codec.  A mapping of the parameters into
   the Session Description Protocol (SDP) [4] 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.  Media type registration

   This registration is done using the template defined in [8] and
   following RFC 3555 [9]

   Type name: audio

   Subtype name: G729EV

   Required parameters: none

   Optional parameters:

dtx: indicates that discontinuous transmission (DTX) is used or
   preferred.  DTX means voice activity detection and non
   transmission of silent frames.  Permissible values are 0 and 1. 0
   means no DTX. 0 is implied if this parameter is omitted.  The
   first version of G.729EV will not support DTX.

mbs: indicates an initial value of MBS, that is the current maximum
   codec bit rate supported as a receiver.  Permissible values are
   between 0 and 11 (see table in Section 4.2 of RFC XXXX).  The
   maximum MBS, that is 11, is implied if this parameter is omitted.
   Note that this parameter will be dynamically updated by the MBS
   field of the RTP packets sent, it is not an absolute value for the
   session.  The goal is to announce this value, prior to the sending
   of any packet, to avoid the remote sender to exceed the MBS at the
   beginning of the session.

ptime: the recommended length of time in milliseconds represented by
   the media in a packet.  See RFC 2327 [4].

maxptime: the maximum length of time in milliseconds which can be
   encapsulated in a packet.

Encoding considerations: This media type is framed and contains
binary data.

Security considerations: See Section 6 of RFC XXXX

Interoperability considerations: none

Published specification: RFC XXXX

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@francetelecom.com

Intended usage: COMMON

Restrictions on usage: This media type depends on RTP framing, and
hence is only defined for transfer via RTP [2].

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

## 5.2.  Mapping to SDP parameters

   The information carried in the media type specification has a
   specific mapping to fields in the Session Description Protocol (SDP)
   [4], which is commonly used to describe RTP sessions.  When SDP is
   used to specify sessions employing the G.729EV codec, the mapping is
   as follows:

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

   o  The media subtype ("G729EV") goes in SDP "a=rtpmap" as the
      encoding name.  The RTP clock rate in "a=rtpmap" MUST be 16000 for
      G.729EV.

   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.729EV encodings
   follow.

   Example 1: default parameters

      m=audio 53146 RTP/AVP 98
      a=rtpmap:98 G729EV/16000

   Example 2: recommended packet duration of 40 ms (=2 frames), DTX off,
   and initial MBS set to 26 kbps

      m=audio 51258 RTP/AVP 99
      a=rtpmap:99 G729EV/16000
      a=fmtp:99 dtx=0; mbs=8
      a=ptime:40

## 5.3.  Offer-answer model considerations

   The following considerations apply when using SDP offer-answer
   procedures to negotiate the use of G.729EV payload in RTP:

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

      Below is an example of such an offer:

```
m=audio 55954 RTP/AVP 98 18
a=rtpmap:98 G729EV/16000
a=rtpmap:18 G729/8000
```

If the answerer supports G.729EV, it will keep the payload type 98 in its answer and the conversation will be done using G.729EV. 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 RFC 3551 [3]).

o   The "dtx" parameter concerns both sending and receiving, so both sides of a bi-directional session MUST use the same "dtx" value. If one party indicates it does not support DTX, DTX must be deactivated both ways.

o   The "mbs" parameter is not symmetric.  Values in the offer and the answer are independent and take into account local constraints. Anyway, one party MUST NOT start sending frames at a bit rate higher than the "mbs" of the other party.

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 [5].  The "maxptime" parameter MUST be handled in the same way.


6.  Security considerations

RTP packets using the payload format defined in this specification are subject to the general security considerations discussed in the RTP specification [2] and any appropriate profile (for example, RFC 3551 [3]).

As this format transports encoded speech/audio, the main security issues include confidentiality and authentication of the speech/audio itself.  The payload format itself does not have any built-in security mechanisms.  Confidentiality of the media streams is achieved by encryption, therefore external mechanisms, such as SRTP [6], MAY be used for that purpose.

This payload format and the G.729EV encoding do not exhibit any significant non-uniformity in the receiver-end computational load and thus in unlikely to pose a denial-of-service threat due to the receipt of pathological datagrams.

## 7.  IANA considerations

It is requested that one new media subtype (audio/G729EV) is registered by IANA, see Section 5.1.

## 8.  References

### 8.1.  Normative references

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

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

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

[4]   Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.

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

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

### 8.2.  Informative references

[7]   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.

[8]   Freed, N. and J. Klensin, "Media Type Specifications and Registration Procedures", draft-freed-media-type-reg-05 (work in progress), August 2005.

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

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@francetelecom.com

Intellectual Property Statement

   The IETF takes no position regarding the validity or scope of any
   Intellectual Property Rights or other rights that might be claimed to
   pertain to the implementation or use of the technology described in
   this document or the extent to which any license under such rights
   might or might not be available; nor does it represent that it has
   made any independent effort to identify any such rights.  Information
   on the procedures with respect to rights in RFC documents can be
   found in BCP 78 and BCP 79.

   Copies of IPR disclosures made to the IETF Secretariat and any
   assurances of licenses to be made available, or the result of an
   attempt made to obtain a general license or permission for the use of
   such proprietary rights by implementers or users of this
   specification can be obtained from the IETF on-line IPR repository at
   http://www.ietf.org/ipr.

   The IETF invites any interested party to bring to its attention any
   copyrights, patents or patent applications, or other proprietary
   rights that may cover technology that may be required to implement
   this standard.  Please address the information to the IETF at
   ietf-ipr@ietf.org.


Disclaimer of Validity

   This document and the information contained herein are provided on an
   "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS
   OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET
   ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED,
   INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE
   INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED
   WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.


Copyright Statement

   Copyright (C) The Internet Society (2005).  This document is subject
   to the rights, licenses and restrictions contained in BCP 78, and
   except as set forth therein, the authors retain all their rights.


Acknowledgment

   Funding for the RFC Editor function is currently provided by the
   Internet Society.