Network Working Group
Request for Comments: 5459
Updates: 4749
Category: Standards Track
A. Sollaud
France Telecom
January 2009

G.729.1 RTP Payload Format Update: Discontinuous Transmission (DTX) Support

Status of This Memo

Copyright Notice

Abstract

This document updates the Real-time Transport Protocol (RTP) payload format to be used for the International Telecommunication Union (ITU-T) Recommendation G.729.1 audio codec. It adds Discontinuous Transmission (DTX) support to the RFC 4749 specification, in a backward-compatible way. An updated media type registration is included for this payload format.

Table of Contents

1. Introduction

The International Telecommunication Union (ITU-T) Recommendation G.729.1 [ITU-G.729.1] is a scalable and wideband extension of the Recommendation G.729 [ITU-G.729] audio codec. [RFC4749] specifies the payload format for packetization of G.729.1-encoded audio signals into the Real-time Transport Protocol (RTP) [RFC3550].

Annex C of G.729.1 [ITU-G.729.1-C] adds Discontinuous Transmission (DTX) support to G.729.1. This document updates the RTP payload format to allow usage of this Annex.

Only changes or additions to [RFC4749] will be described in the following sections.

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.729.1 supports Discontinuous Transmission (DTX), a.k.a. "silence suppression". It means that the coder includes a Voice Activity Detection (VAD) algorithm to determine if an audio frame contains silence or actual audio. During silence periods, the coder may significantly decrease the transmitted bit rate by sending a small frame called a Silence Insertion Descriptor (SID) and then stop transmission. The receiver's decoder will generate comfort noise (CNG) according to the parameters contained in the SID. This DTX/CNG scheme is specified in [ITU-G.729.1-C].

The G.729.1 SID has an embedded structure. The core SID is the same as the legacy G.729 SID [ITU-G.729-B]. The first enhancement layer adds some parameters for narrowband comfort noise, while a second enhancement layer adds wideband information. The G.729.1 SID can be 2, 3, or 6 octets long.

3. RTP Header Usage

The fields of the RTP header must be used as described in [RFC4749], except for the Marker (M) bit.

If DTX is used, the first packet of a talkspurt -- that is, the first packet after a silence period during which packets have not been transmitted contiguously -- MUST be distinguished by setting the M bit in the RTP data header to 1. The M bit in all other packets MUST be set to 0. The beginning of a talkspurt MAY be used to adjust the playout delay to reflect changing network delays.

If DTX is not used, the M bit MUST be set to 0 in all packets.

4. Payload Format

The payload format is the same as in [RFC4749], with the option to add a SID at the end.

So the complete payload consists of a payload header of 1 octet (MBS (maximum bit rate supported) and FT (frame type) fields), followed by zero or more consecutive audio frames at the same bit rate, followed by zero or one SID.

> Note that this is consistent with the payload format of G.729 described in section 4.5.6 of [RFC3551].

To be able to transport a SID alone -- that is, without actual audio frames -- we assign the FT value 14 to the SID. When using FT=14, only a single SID frame SHALL be included in the payload. The actual SID size (2, 3, or 6 octets) is inferred from the payload size: it is the size of what is left after the payload header.

When a SID is appended to actual audio frames, the FT value remains the one describing the encoding rate of the audio frames. Since the SID is much smaller than any other frame, it can be easily detected at the receiver side, and it will not hinder the calculation of the number of frames. The actual SID size is inferred from the payload size: it is the size of what is left after the audio frames.

Section 5.4 of [RFC4749] mandates to ignore the remaining bytes after complete frames. This document overrides that statement: the receiver of the payload must consider these remaining bytes as a SID frame. If the size of this remainder is not a valid SID frame size (2, 3, or 6 octets), the receiver MUST ignore these bytes.

The full FT table is given for convenience:

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
12-13	(reserved)	-
14	SID	2, 3, or 6 octets
15	NO_DATA	0

DTX has no impact on the MBS definition and use.

5. Payload Format Parameters

Parameters defined in [RFC4749] are not modified. We add a new optional parameter to configure DTX.

5.1. Media Type Registration Update

We add a new optional parameter to the audio/G7291 media subtype:

dtx: indicates that Discontinuous Transmission (DTX) is used or preferred. Permissible values are 0 and 1. 0 means no DTX. 1 means DTX support, as described in Annex C of ITU-T Recommendation G.729.1. 0 is implied if this parameter is omitted.

When DTX is turned off, the RTP payload MUST NOT contain a SID, and the FT value 14 MUST NOT be used.

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) [RFC4566], which is commonly used to describe RTP sessions.

The mapping described in [RFC4749] remains unchanged.

The "dtx" parameter goes in the SDP "a=fmtp" attribute.

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

Example 1: default parameters (DTX off)

```
m=audio 57586 RTP/AVP 96
a=rtpmap:96 G7291/16000
```

Example 2: recommended packet duration of 40 ms (=2 frames), maximum bit rate is 20 kbps, DTX supported and preferred.

```
m=audio 49987 RTP/AVP 97
a=rtpmap:97 G7291/16000
a=fmtp:97 maxbitrate=20000; dtx=1
a=ptime:40
```

5.2.1. DTX Offer-Answer Model Considerations

The offer-answer model considerations of [RFC4749] fully apply. In this section, we only define the management of the "dtx" parameter.

The "dtx" parameter concerns both sending and receiving, so both sides of a bi-directional session MUST have the same DTX setting. If one party indicates that it does not support DTX, DTX must be deactivated both ways. In other words, DTX is actually activated if, and only if, "dtx=1" in the offer and in the answer.

A special rule applies for multicast: the "dtx" 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.

An RTP receiver compliant with only RFC 4749 and not this specification will ignore the "dtx" parameter and will not include it in its answer, so DTX will not be activated in this case. As a remark, if that happened, this kind of receiver would not be confused by an RTP stream with DTX because RFC 4749 requires that the bytes that are now used for SID frames be ignored. But during the silence period, the receiver would then react using packet loss concealment instead of comfort noise generation, leading to bad audio quality. This justifies the use of the "dtx" parameter, even if the payload format is backward-compatible at the binary level.

5.2.2. DTX Declarative SDP Considerations

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

6. Congestion Control

The congestion control considerations of [RFC4749] apply.

The use of DTX can help congestion control by reducing the number of transmitted RTP packets and the average bandwidth of audio streams.

7. Security Considerations

The security considerations of [RFC4749] apply.

By observing the RTP flow with DTX, an attacker could see when and for how long people are speaking. This is a general fact for DTX, and G.729.1 DTX introduces no new specific issue.

8. IANA Considerations

IANA has assigned a new optional parameter for media subtype (audio/G7291); see Section 5.1.

9. References

9.1. Normative References

[ITU-G.729.1] 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.

[ITU-G.729.1-C] International Telecommunications Union, "G.729.1 DTX/CNG scheme", ITU-T Recommendation G.729.1 Annex C, May 2008.

[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.

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

[RFC4749] Sollaud, A., "RTP Payload Format for the G.729.1 Audio Codec", RFC 4749, October 2006.

9.2. Informative References

[ITU-G.729] 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.

[ITU-G.729-B] International Telecommunications Union, "A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70", ITU-T Recommendation G.729 Annex B, October 1996.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, 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@orange-ftgroup.com