

Real-time Transport Protocol (RTP) Payload for Comfort Noise (CN)

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

Copyright (C) The Internet Society (2002). All Rights Reserved.

Abstract

This document describes a Real-time Transport Protocol (RTP) payload format for transporting comfort noise (CN). The CN payload type is primarily for use with audio codecs that do not support comfort noise as part of the codec itself such as ITU-T Recommendations G.711, G.726, G.727, G.728, and G.722.

1. Conventions Used in This Document

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 [7].

2. Introduction

This document describes a RTP [1] payload format for transporting comfort noise. The payload format is based on Appendix II of ITU-T Recommendation G.711 [8] which defines a comfort noise payload format (or bit-stream) for ITU-T G.711 [2] use in packet-based multimedia communication systems. The payload format is generic and may also be used with other audio codecs without built-in Discontinuous Transmission (DTX) capability such as ITU-T Recommendations G.726 [3], G.727 [4], G.728 [5], and G.722 [6]. The payload format provides a minimum interoperability specification for communication of comfort noise parameters. The comfort noise analysis and synthesis as well as the Voice Activity Detection (VAD) and DTX algorithms are unspecified and left implementation-specific.

However, an example solution for G.711 has been tested and is described in the Appendix [8]. It uses the VAD and DTX of G.729 Annex B [9] and a comfort noise generation algorithm (CNG) which is provided in the Appendix for information.

The comfort noise payload, which is also known as a Silence Insertion Descriptor (SID) frame, consists of a single octet description of the noise level and MAY contain spectral information in subsequent octets. An earlier version of the CN payload format consisting only of the noise level byte was defined in draft revisions of the RFC 1890. The extended payload format defined in this document should be backward compatible with implementations of the earlier version assuming that only the first byte is interpreted and any additional spectral information bytes are ignored.

3. CN Payload Definition

The comfort noise payload consists of a description of the noise level and spectral information in the form of reflection coefficients for an all-pole model of the noise. The inclusion of spectral information is OPTIONAL and the model order (number of coefficients) is left unspecified. The encoder may choose an appropriate model order based on such considerations as quality, complexity, expected environmental noise, and signal bandwidth. The model order is not explicitly transmitted since the number of coefficients can be derived from the length of the payload at the receiver. The decoder may reduce the model order by setting higher order reflection coefficients to zero if desired to reduce complexity or for other reasons.

3.1 Noise Level

The magnitude of the noise level is packed into the least significant bits of the noise-level byte with the most significant bit unused and always set to 0 as shown below in Figure 1. The least significant bit of the noise level magnitude is packed into the least significant bit of the byte.

The noise level is expressed in -dBov, with values from 0 to 127 representing 0 to -127 dBov. dBov is the level relative to the overload of the system. (Note: Representation relative to the overload point of a system is particularly useful for digital implementations, since one does not need to know the relative calibration of the analog circuitry.) For example, in the case of a u-law system, the reference would be a square wave with values +/- 8031, and this square wave represents 0dBov. This translates into 6.18dBm0.

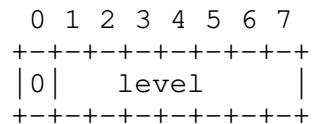


Figure 1: Noise Level Packing

3.2 Spectral Information

The spectral information is transmitted using reflection coefficients [8]. Each reflection coefficient can have values between -1 and 1 and is quantized uniformly using 8 bits. The quantized value is represented by the 8 bit index N , where $N=0\dots,254$, and index $N=255$ is reserved for future use. Each index N is packed into a separate byte with the MSB first. The quantized value of each reflection coefficient, k_i , can be obtained from its corresponding index using:

$$k_i(N_i) = \frac{258 \cdot (N_i - 127)}{32768} \quad \text{for } N_i = 0\dots 254; -1 < k_i < 1$$

3.3 Payload Packing

The first byte of the payload MUST contain the noise level as shown in Figure 1. Quantized reflection coefficients are packed in subsequent bytes in ascending order as in Figure 2 where M is the model order. The total length of the payload is $M+1$ bytes. Note that a 0th order model (i.e., no spectral envelope information) reduces to transmitting only the energy level.

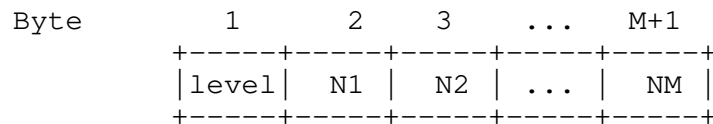


Figure 2: CN Payload Packing Format

4. Usage of RTP

The RTP header for the comfort noise packet SHOULD be constructed as if the comfort noise were an independent codec. Thus, the RTP timestamp designates the beginning of the comfort noise period. When this payload format is used under the RTP profile specified in RFC 1890 [10], a static payload type of 13 is assigned for RTP timestamp clock rate of 8,000 Hz; if other rates are needed, they MUST be defined through dynamic payload types. The RTP packet SHOULD NOT have the marker bit set.

Each RTP packet containing comfort noise MUST contain exactly one CN payload per channel. This is required since the CN payload has a variable length. If multiple audio channels are used, each channel MUST use the same spectral model order 'M'.

5. Guidelines for Use

An audio codec with DTX capabilities generally includes VAD, DTX, and CNG algorithms. The job of the VAD is to discriminate between active and inactive voice segments in the input signal. During inactive voice segments, the role of the CNG is to sufficiently describe the ambient noise while minimizing the transmission rate. A CN payload (or SID frame) containing a description of the noise is sent to the receiver to drive the CNG. The DTX algorithm determines when a CN payload is transmitted. During active voice segments, packets of the voice codec are transmitted and indicated in the RTP header by the static or dynamic payload type for that codec. At the beginning of an inactive voice segment (silence period), a CN packet is transmitted in the same RTP stream and indicated by the CN payload type. The CN packet update rate is left implementation specific. For example, the CN packet may be sent periodically or only when there is a significant change in the background noise characteristics. The CNG algorithm at the receiver uses the information in the CN payload to update its noise generation model and then produce an appropriate amount of comfort noise.

The CN payload format provides a minimum interoperability specification for communication of comfort noise parameters. The comfort noise analysis and synthesis as well as the VAD and DTX algorithms are unspecified and left implementation-specific. However, an example solution for G.711 has been tested and is described in Appendix II of ITU-T Recommendation G.711 [8]. It uses the VAD and DTX of G.729 Annex B [9] and a comfort noise generation algorithm (CNG), which is provided in the Appendix for information. Additional guidelines for use such as the factors affecting system performance in the design of the VAD/DTX/CNG algorithms are described in the Appendix.

5.1 Usage of SDP

When using the Session Description Protocol (SDP) [11] to specify RTP payload information, the use of comfort noise is indicated by the inclusion of a payload type for CN on the media description line. When using CN with the RTP/AVP profile [10] and a codec whose RTP timestamp clock rate is 8000 Hz, such as G.711 (PCMU, static payload type 0), the static payload type 13 for CN can be used:

```
m=audio 49230 RTP/AVP 0 13
```

When using CN with a codec that has a different RTP timestamp clock rate, a dynamic payload type mapping (rtpmap attribute) is required.

This example shows CN used with the G.722.1 codec (see RFC 3047 [12]):

```
m=audio 49230 RTP/AVP 101 102
a=rtpmap:101 G7221/16000
a=fmtp:121 bitrate=24000
a=rtpmap:102 CN/16000
```

Omission of a payload type for CN on the media description line implies that this comfort noise payload format will not be used, but it does not imply that silence will not be suppressed. RTP allows discontinuous transmission (silence suppression) on any audio payload format. The receiver can detect silence suppression on the first packet received after the silence by observing that the RTP timestamp is not contiguous with the end of the interval covered by the previous packet even though the RTP sequence number has incremented only by one. The RTP marker bit is also normally set on such a packet.

6. IANA Considerations

This section defines a new RTP payload name and associated MIME type, CN (audio/CN). The payload format specified in this document is also assigned payload type 13 in the RTP Payload Types table of the RTP Parameters registry maintained by the Internet Assigned Numbers Authority (IANA).

6.1 Registration of MIME media type audio/CN

MIME media type name: audio

MIME subtype name: CN

Required parameters: None

Optional parameters:

rate: specifies the RTP timestamp clock rate, which is usually (but not always) equal to the sampling rate. This parameter should have the same value as the codec used in conjunction with comfort noise. The default value is 8000.

Encoding considerations:

This type is only defined for transfer via RTP [RFC 1889].

Security considerations: see Section 7 "Security Considerations".

Interoperability considerations: none

Published specification:

This document and Appendix II of ITU-T Recommendation G.711

Applications which use this media type:

Audio and video streaming and conferencing tools.

Additional information: none

Person & email address to contact for further information:

Robert Zopf
zopf@lucent.com

Intended usage: COMMON

Author/Change controller:

Author: Robert Zopf
Change controller: IETF AVT Working Group

7. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [1]. This implies that confidentiality of the media streams is achieved by encryption. Because the payload format is arranged end-to-end, encryption MAY be performed after encapsulation so there is no conflict between the two operations.

As this format transports background noise, there are no significant security, confidentiality, or authentication concerns.

8. References

- [1] Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", RFC 1889, January 1996.
- [2] ITU Recommendation G.711 (11/88) - Pulse code modulation (PCM) of voice frequencies.

- [3] ITU Recommendation G.726 (12/90) - 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).
- [4] ITU Recommendation G.727 (12/90) - 5-, 4-, 3- and 2-bits sample embedded adaptive differential pulse code modulation (ADPCM).
- [5] ITU Recommendation G.728 (09/92) - Coding of speech at 16 kbits/s using low-delay code excited linear prediction.
- [6] ITU Recommendation G.722 (11/88) - 7 kHz audio-coding within 64 kbit/s.
- [7] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [8] Appendix II to Recommendation G.711 (02/2000) - A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems.
- [9] Annex B (08/97) to Recommendation G.729 - C source code and test vectors for implementation verification of the algorithm of the G.729 silence compression scheme.
- [10] Schulzrinne, H., "RTP Profile for Audio and Video Conferences with Minimal Control", RFC 1890, January 1996.
- [11] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.
- [12] Luthi, P., "RTP Payload Format for ITU-T Recommendation G.722.1", RFC 3047, January 2001.

9. Author's Address

Robert Zopf
Lucent Technologies
INS Access VoIP Networks
2G-234A
101 Crawfords Corner Rd
Holmdel, NJ 07733-3030 US

Phone: 1-732-949-1667
Fax: 1-732-949-7016
EMail: zopf@lucent.com

10. Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS 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.

Acknowledgement

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

