

Network Working Group
Request for Comments: 4298
Category: Standards Track

J.-H. Chen
W. Lee
J. Thyssen
Broadcom Corporation
December 2005

RTP Payload Format for BroadVoice Speech Codecs

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 (2005).

Abstract

This document describes the RTP payload format for the BroadVoice(R) narrowband and wideband speech codecs. The narrowband codec, called BroadVoice16, or BV16, has been selected by CableLabs as a mandatory codec in PacketCable 1.5 and has a CableLabs specification. The document also provides specifications for the use of BroadVoice with MIME and the Session Description Protocol (SDP).

Table of Contents

1. Introduction	2
2. Background	2
3. RTP Payload Format for BroadVoice16 Narrowband Codec	3
3.1. BroadVoice16 Bit Stream Definition	4
3.2. Multiple BroadVoice16 Frames in an RTP Packet	5
4. RTP Payload Format for BroadVoice32 Wideband Codec	6
4.1. BroadVoice32 Bit Stream Definition	6
4.2. Multiple BroadVoice32 Frames in an RTP Packet	8
5. IANA Considerations	8
5.1. MIME Registration of BroadVoice16 for RTP	9
5.2. MIME Registration of BroadVoice32 for RTP	9
6. Mapping to SDP Parameters	10
6.1. Offer-Answer Model Considerations	11
7. Security Considerations	11
8. Congestion Control	11
9. Acknowledgements	11
10. References	12
10.1. Normative References	12
10.2. Informative References	12

1. Introduction

This document specifies the payload format for sending BroadVoice encoded speech or audio signals using the Real-time Transport Protocol (RTP) [1]. The sender may send one or more BroadVoice codec data frames per packet, depending on the application scenario, based on network conditions, bandwidth availability, delay requirements, and packet-loss tolerance.

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

BroadVoice is a speech codec family developed for VoIP (Voice over Internet Protocol) applications, including Voice over Cable, Voice over DSL, and IP phone applications. BroadVoice achieves high speech quality with a low coding delay and relatively low codec complexity.

The BroadVoice codec family contains two codec versions. The narrowband version of BroadVoice, called BroadVoice16 [3], or BV16 for short, encodes 8 kHz-sampled narrowband speech at a bit rate of 16 kilobits/second, or 16 kbit/s. The wideband version of BroadVoice, called BroadVoice32, or BV32, encodes 16 kHz-sampled wideband speech at a bit rate of 32 kbit/s. The BV16 and BV32 use

very similar (but not identical) coding algorithms; they share most of their algorithm modules.

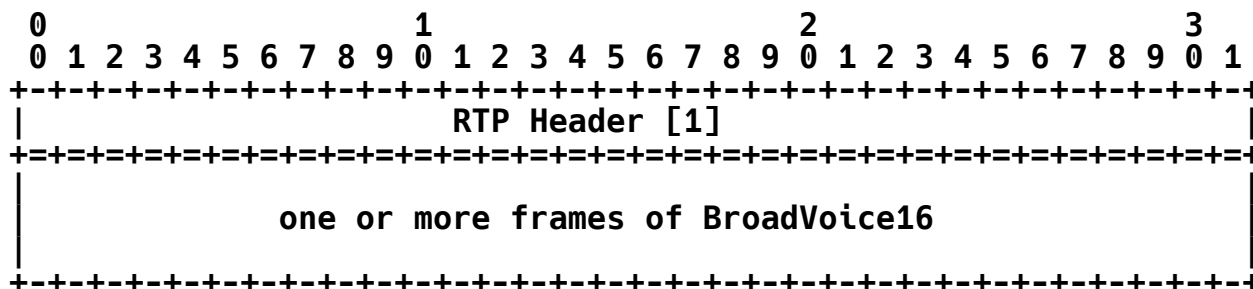
To minimize the delay in real-time two-way communications, both the BV16 and BV32 encode speech with a very small frame size of 5 ms without using any look ahead. By using a packet size as small as 5 ms if necessary, this allows VoIP systems based on BroadVoice to have a very low end-to-end system delay.

BroadVoice also has relatively low codec complexity when compared with ITU-T standard speech codecs based on CELP (Coded Excited Linear Prediction), such as G.728, G.729, G.723.1, and G.722.2. Full-duplex implementations of the BV16 and BV32 take around 12 and 17 MIPS, respectively, on general-purpose 16-bit fixed-point digital signal processors (DSPs). The total memory footprints of the BV16 and BV32, including program size, data tables, and data RAM, are around 12 kwords each, or 24 kbytes.

The PacketCable(TM) project of Cable Television Laboratories, Inc. (CableLabs(R)) has chosen the BV16 codec for use in VoIP telephone services provided by cable operators. More specifically, the BV16 codec was selected as one of the mandatory audio codecs in the PacketCable(TM) 1.5 Audio/Video Codecs Specification [8] and has been implemented by multiple vendors. The wideband version (BV32) has been developed by Broadcom but has not yet appeared in a public specification; since it is technically very similar to BV16, its payload format is also defined in this document.

3. RTP Payload Format for BroadVoice16 Narrowband Codec

The BroadVoice16 uses 5 ms frames and a sampling frequency of 8 kHz, so the RTP timestamp MUST be in units of 1/8000 of a second. The RTP timestamp indicates the sampling instant of the oldest audio sample represented by the frame(s) present in the payload. The RTP payload for the BroadVoice16 has the format shown in the figure below. No additional header specific to this payload format is required.



If BroadVoice16 is used for applications with silence compression, the first BroadVoice16 packet after a silence period during which packets have not been transmitted contiguously **SHOULD** have the marker bit in the RTP data header set to one. The marker bit in all other packets is zero. Applications without silence suppression **MUST** set the marker bit to zero.

The assignment of an RTP payload type for this new packet format is outside the scope of this document, and will not be specified here. It is expected that the RTP profile for a particular class of applications will assign a payload type for this encoding, or if that is not done, then a payload type in the dynamic range shall be chosen.

3.1. BroadVoice16 Bit Stream Definition

The BroadVoice16 encoder operates on speech frames of 5 ms corresponding to 40 samples at a sampling rate of 8000 samples per second. For every 5 ms frame, the encoder encodes the 40 consecutive audio samples into 80 bits, or 10 octets. Thus, the 80-bit bit stream produced by the BroadVoice16 for each 5 ms frame is octet-aligned, and no padding bits are required. The bit allocation for the encoded parameters of the BroadVoice16 codec is listed in the following table.

Encoded Parameter	Codeword	Number of bits per frame
Line Spectrum Pairs	L0,L1	7+7=14
Pitch Lag	PL	7
Pitch Gain	PG	5
Log-Gain	LG	4
Excitation Vectors	V0,...,V9	5*10=50
Total:		80 bits

The mapping of the encoded parameters in an 80-bit BroadVoice16 data frame is defined in the following figure. This figure shows the bit packing in "network byte order", also known as big-endian order. The bits of each 32-bit word are numbered 0 to 31, with the most significant bit on the left and numbered 0. The octets (bytes) of each word are transmitted with the most significant octet first. The bits of the data field for each encoded parameter are numbered in the same order, with the most significant bit on the left.

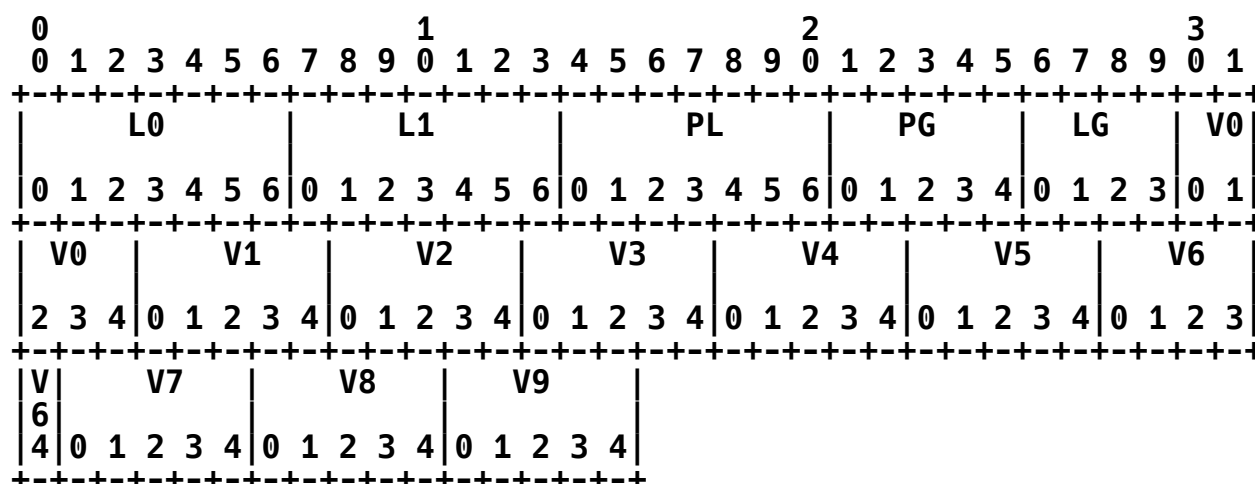


Figure 1: BroadVoice16 bit packing

3.2. Multiple BroadVoice16 Frames in an RTP Packet

More than one BroadVoice16 frame MAY be included in a single RTP packet by a sender. Senders have the following additional restrictions:

- o SHOULD NOT include more BroadVoice16 frames in a single RTP packet than will fit in the MTU of the RTP.
- o MUST NOT split a BroadVoice16 frame between RTP packets.
- o BroadVoice16 frames in an RTP packet MUST be consecutive.

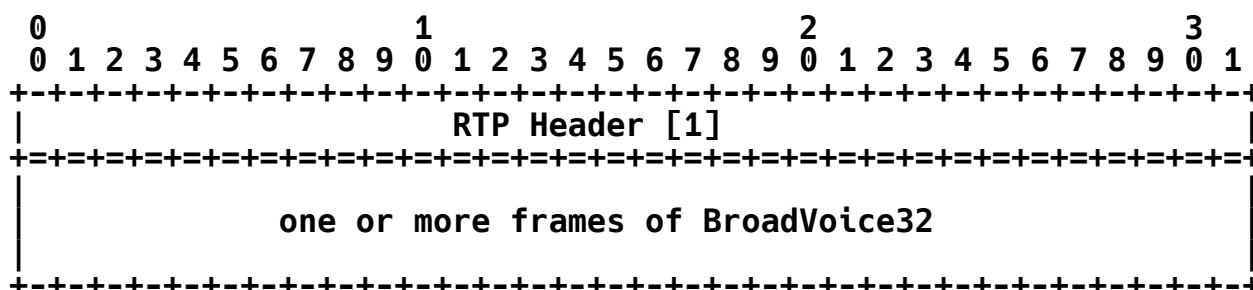
Since multiple BroadVoice16 frames in an RTP packet MUST be consecutive, and since BroadVoice16 has a fixed frame size of 5 ms, recovering the timestamps of all frames within a packet is easy. The oldest frame within an RTP packet has the same timestamp as the RTP packet, as mentioned above. To obtain the timestamp of the frame that is N frames later than the oldest frame in the packet, one simply adds 5*N ms worth of time units to the timestamp of the RTP packet.

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.

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 BroadVoice16 frames is to count the total number of octets within the RTP payload, and divide the octet count by 10.

4. RTP Payload Format for BroadVoice32 Wideband Codec

The BroadVoice32 uses 5 ms frames and a sampling frequency of 16 kHz, so the RTP timestamp MUST be in units of 1/16000 of a second. The RTP timestamp indicates the sampling instant of the oldest audio sample represented by the frame(s) present in the payload. The RTP payload for the BroadVoice32 has the format shown in the figure below. No additional header specific to this payload format is required.



If BroadVoice32 is used for applications with silence compression, the first BroadVoice32 packet after a silence period during which packets have not been transmitted contiguously SHOULD have the marker bit in the RTP data header set to one. The marker bit in all other packets is zero. Applications without silence suppression MUST set the marker bit to zero.

The assignment of an RTP payload type for this new packet format is outside the scope of this document, and will not be specified here. It is expected that the RTP profile for a particular class of applications will assign a payload type for this encoding, or if that is not done, then a payload type in the dynamic range shall be chosen.

4.1. BroadVoice32 Bit Stream Definition

The BroadVoice32 encoder operates on speech frames of 5 ms corresponding to 80 samples at a sampling rate of 16000 samples per second. For every 5 ms frame, the encoder encodes the 80 consecutive audio samples into 160 bits, or 20 octets. Thus, the 160-bit bit stream produced by the BroadVoice32 for each 5 ms frame is octet-aligned, and no padding bits are required. The bit allocation for

4.2. Multiple BroadVoice32 Frames in an RTP Packet

More than one BroadVoice32 frame MAY be included in a single RTP packet by a sender. Senders have the following additional restrictions:

- o SHOULD NOT include more BroadVoice32 frames in a single RTP packet than will fit in the MTU of the RTP.
- o MUST NOT split a BroadVoice32 frame between RTP packets.
- o BroadVoice32 frames in an RTP packet MUST be consecutive.

Since multiple BroadVoice32 frames in an RTP packet MUST be consecutive, and since BroadVoice32 has a fixed frame size of 5 ms, recovering the timestamps of all frames within a packet is easy. The oldest frame within an RTP packet has the same timestamp as the RTP packet, as mentioned above. To obtain the timestamp of the frame that is N frames later than the oldest frame in the packet, one simply adds $5 \times N$ ms worth of time units to the timestamp of the RTP packet.

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.

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 BroadVoice32 frames is to count the total number of octets within the RTP payload, and divide the octet count by 20.

5. IANA Considerations

Two new MIME sub-types, as described in this section, have been registered.

The MIME names for the BV16 and BV32 codecs have been allocated from the IETF tree since these two codecs are expected to be widely used for Voice-over-IP applications, especially in Voice over Cable applications.

5.1. MIME Registration of BroadVoice16 for RTP

MIME media type name: audio

MIME media subtype name: BV16

Required parameter: none

Optional parameters:

ptime: Defined as usual for RTP audio (see RFC 2327 [4]).

maxptime: See RFC 3267 [7] for its definition. The maxptime SHOULD be a multiple of the duration of a single codec data frame (5 ms).

Encoding considerations:

This type is defined for transferring BV16-encoded data via RTP using the payload format specified in Section 3 of RFC 4298. Audio data is binary data and must be encoded for non-binary transport; the Base64 encoding is suitable for Email.

Security considerations:

See Section 7 "Security Considerations" of RFC 4298.

Public specification:

The BroadVoice16 codec has been specified in [3].

Intended usage:

COMMON. It is expected that many VoIP applications, especially Voice over Cable applications, will use this type.

Person & email address to contact for further information:

Juin-Hwey (Raymond) Chen
rchen@broadcom.com

Author/Change controller:

Author: Juin-Hwey (Raymond) Chen, rchen@broadcom.com
Change Controller: IETF Audio/Video Transport Working Group
delegated from the IESG

5.2. MIME Registration of BroadVoice32 for RTP

MIME media type name: audio

MIME media subtype name: BV32

Required parameter: none

Optional parameters:

ptime: Defined as usual for RTP audio (see RFC 2327 [4]).

maxptime: See RFC 3267 [7] for its definition. The maxptime **SHOULD** be a multiple of the duration of a single codec data frame (5 ms).

Encoding considerations:

This type is defined for transferring BV32-encoded data via RTP using the payload format specified in Section 4 of RFC 4298. Audio data is binary data and must be encoded for non-binary transport; the Base64 encoding is suitable for Email.

Security considerations:

See Section 7 "Security Considerations" of RFC 4298.

Intended usage:

COMMON. It is expected that many VoIP applications, especially Voice over Cable applications, will use this type.

Person & email address to contact for further information:

Juin-Hwey (Raymond) Chen
rchen@broadcom.com

Author/Change controller:

Author: Juin-Hwey (Raymond) Chen, rchen@broadcom.com
Change Controller: IETF Audio/Video Transport Working Group
delegated from the IESG

6. Mapping to SDP Parameters

The information carried in the MIME 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 BroadVoice16 or BroadVoice32 codec, the mapping is as follows:

- The MIME type ("audio") goes in SDP "m=" as the media name.
- The MIME subtype (payload format name) goes in SDP "a=rtpmap" as the encoding name. The RTP clock rate in "a=rtpmap" **MUST** be 8000 for BV16 and 16000 for BV32.
- The parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

An example of the media representation in SDP for describing BV16 might be:

```
m=audio 49120 RTP/AVP 97
a=rtpmap:97 BV16/8000
```

An example of the media representation in SDP for describing BV32 might be:

```
m=audio 49122 RTP/AVP 99
a=rtpmap:99 BV32/16000
```

6.1. Offer-Answer Model Considerations

No special considerations are needed for using the SDP Offer/Answer model [5] with the BV16 and BV32 RTP payload formats.

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] and any appropriate profile (for example, [6]). This implies that confidentiality of the media streams is achieved by encryption.

A potential denial-of-service threat exists for data encoding using compression techniques that have non-uniform receiver-end computational load. The attacker can inject pathological datagrams into the stream that are complex to decode and cause the receiver to become overloaded. However, the encodings covered in this document do not exhibit any significant non-uniformity.

8. Congestion Control

The general congestion control considerations for transporting RTP data apply to BV16 and BV32 audio over RTP as well (see RTP [1]) and any applicable RTP profile like AVP [6]. BV16 and BV32 do not have any built-in mechanism for reducing the bandwidth. Packing more frames in each RTP payload can reduce the number of packets sent, and hence the overhead from IP/UDP/RTP headers, at the expense of increased delay and reduced error robustness against packet losses.

9. Acknowledgements

The authors would like to thank Magnus Westerlund, Colin Perkins, Allison Mankin, and Jean-Francois Mule for their review of this document.

10. References

10.1. Normative References

- [1] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.
- [2] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [3] Cable Television Laboratories, Inc., BroadVoice(TM)16 Speech Codec Specification, Revision 1.2, October 30, 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] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.
- [7] Sjöberg, J., Westerlund, M., Lankaniemi, A., and Q. Xie, "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", RFC 3267, June 2002.

10.2. Informative References

- [8] Cable Television Laboratories, Inc., PacketCable(TM) 1.5 Audio/Video Codecs Specification, PKT-SP-CODEC1.5-I01-050128, January 28, 2005.
<http://www.cablelabs.com/specifications/archives/>

Authors' Addresses

Juin-Hwey (Raymond) Chen
Broadcom Corporation
Room A3020
16215 Alton Parkway
Irvine, CA 92618
USA

Phone: +1 949 926 6288
EMail: rchen@broadcom.com

Winnie Lee
Broadcom Corporation
Room A2012E
200-13711 International Place
Richmond, British Columbia V6V 2Z8
Canada

Phone: +1 604 233 8605
EMail: wlee@broadcom.com

Jes Thyssen
Broadcom Corporation
Room A3018
16215 Alton Parkway
Irvine, CA 92618
USA

Phone: +1 949 926 5768
EMail: jthyssen@broadcom.com

Full 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.

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.

Intellectual Property

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.

Acknowledgement

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