Internet Draft
draft-chen-rtp-bv-00.txt
June 18, 2003
Expires: December 18, 2003

Juin-Hwey Chen
Cheng-Chieh Lee
Winnie Lee
Jes Thyssen
Broadcom Corporation

RTP Payload Format for BroadVoice Speech Codecs

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
http://www.ietf.org/ietf/1id-abstracts.txt
The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html.

Abstract

This document describes the RTP payload format for the BroadVoice(TM) narrowband and wideband speech codecs developed by Broadcom Corporation. The document also provides specifications for the use of BroadVoice with MIME and 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.....3
   3.2 Multiple BroadVoice16 Frames in An RTP Packets.....4
4. RTP Payload Format for BroadVoice32 Wideband Codec.....5
   4.1 BroadVoice32 Bit Stream Definition.....5
   4.2 Multiple BroadVoice32 Frames in An RTP Packet.....7
5. Storage Format.....7
6. IANA Considerations.....8
   6.1 MIME registration of BroadVoice16.....8

6.2 MIME registration of BroadVoice32.....9

7. Mapping To SDP Parameters.....10
8. Security Considerations.....11
9. References.....11
10. Authors' Addresses.....11

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 [3] is a speech codec family developed by Broadcom for VoIP 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, 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. This allows VoIP systems based on BroadVoice to have a very low end-to-end system delay, by using a packet size as small as 5 ms if necessary.

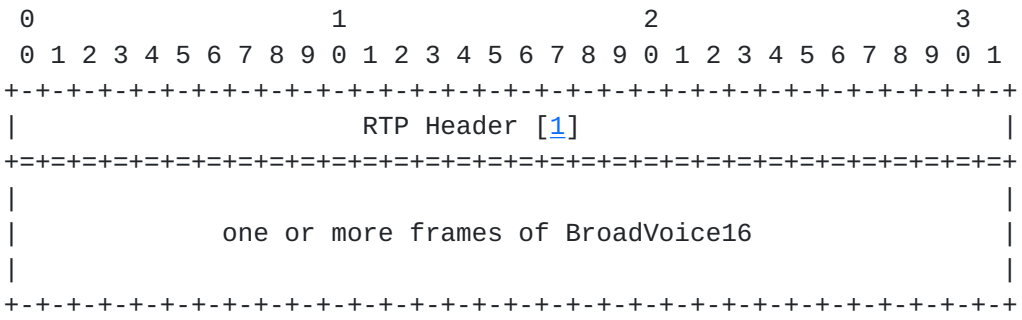
BroadVoice also has relatively low codec complexity when compared with other ITU-T standard speech codecs based on CELP (Coded Excited Linear Prediction), such as G.728, G.729, G.723.1, G.722.2, etc. Full-duplex implementations of the BV16 and BV32 take around 12 and 17 MIPS, respectively, on general-purpose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

24 kbytes.

Cable Television Laboratories (CableLabs(R)) intends to adopt BroadVoice16 as a PacketCable(TM) audio codec standard for VoIP over Cable applications.

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 payload for the BroadVoice16 has the format shown in the figure below. No additional header specific to this payload format is required.



When more than one codec data frame is present in a single RTP packet, the timestamp is, as always, that of the oldest data frame represented in the RTP packet.

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 most significant octet first. The bits of data field for each encoded parameter are numbered in the same order, with the most significant bit on the left.

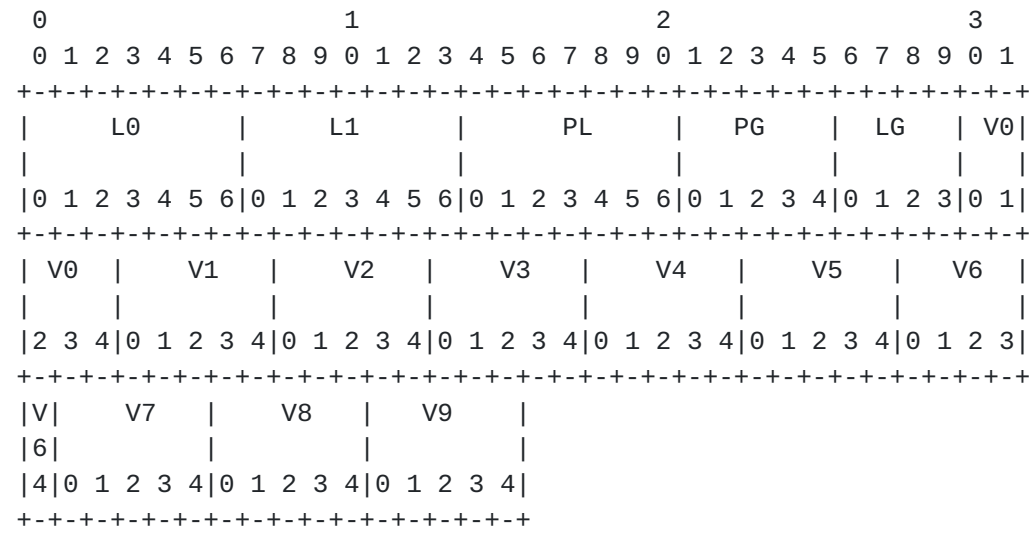


Figure 1: BroadVoice16 bit packing

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:

- SHOULD NOT include more BroadVoice16 frames in a single RTP packet than will fit in the MTU of the RTP transport protocol.

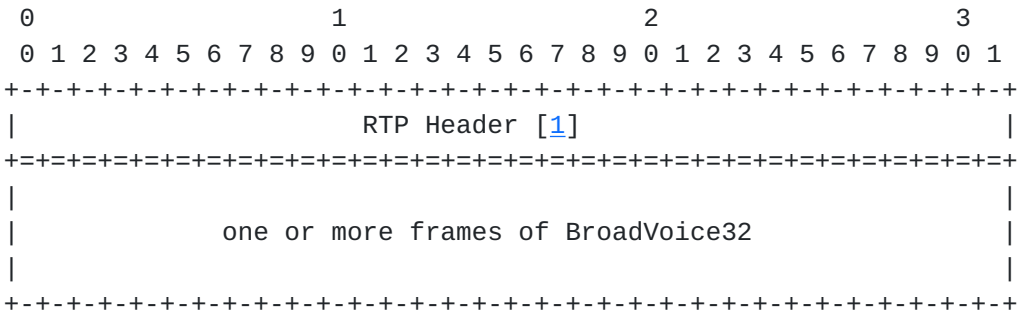
- MUST NOT split a BroadVoice16 frame between RTP packets.

It is RECOMMENDED that the number of frames contained within an RTP packet is 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 packet, 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 payload for the BroadVoice32 has the format shown in the figure below. No additional header specific to this payload format is required.



When more than one codec data frame is present in a single RTP packet, the timestamp is, as always, that of the oldest data frame represented in the RTP packet.

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 the encoded parameters of the BroadVoice32 codec is listed in the following table.

Encoded Parameter	Codeword	Number of bits per frame
Line Spectrum Pairs	L0,L1,L2	7+5+5=17
Pitch Lag	PL	8
Pitch Gain	PG	5
Log-Gains (1st & 2nd subframes)	LG0,LG1	5+5=10
Excitation Vectors (1st subframe)	VA0,...,VA9	6*10=60
Excitation Vectors (2nd subframe)	VB0,...,VB9	6*10=60
Total:		160 bits

The mapping of the encoded parameters in a 160-bit BroadVoice32 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 most significant octet first. The bits of data field for each encoded parameter are numbered in the same order, with the most significant bit on the left.

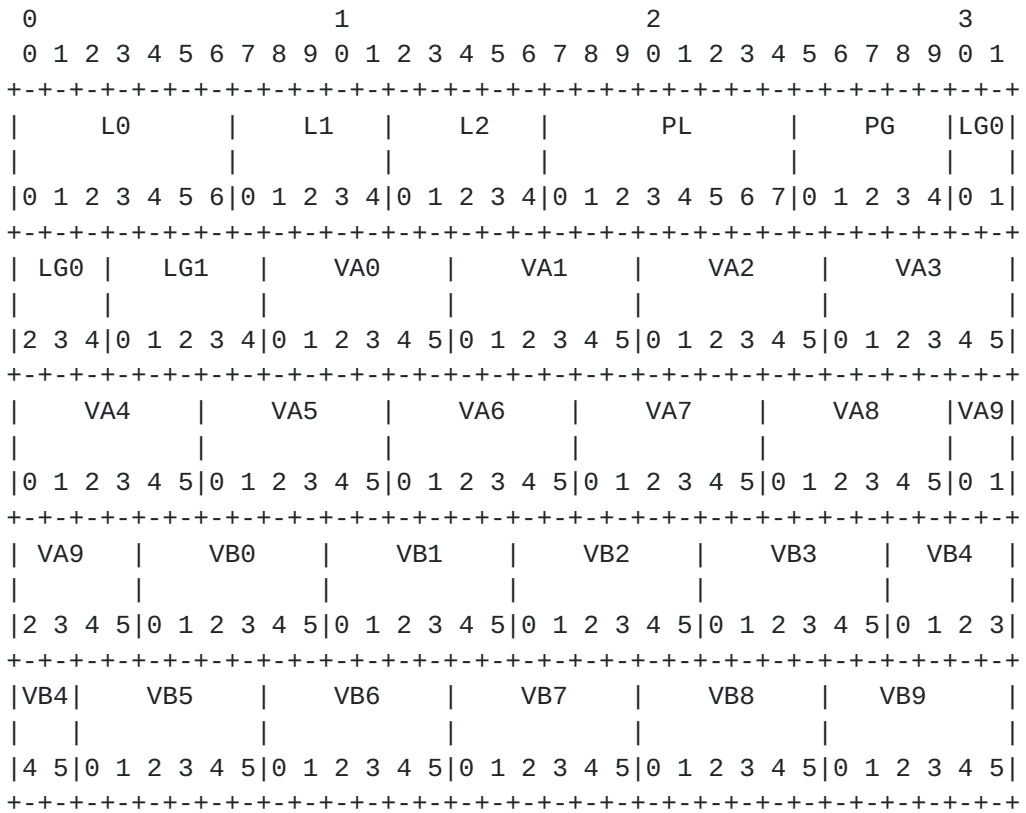


Figure 2: BroadVoice32 bit packing

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:

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

It is RECOMMENDED that the number of frames contained within an RTP packet is 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.

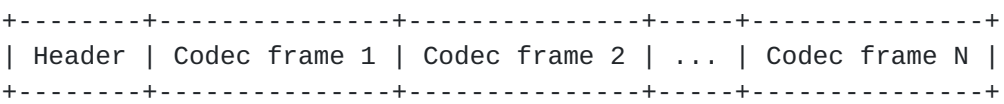
Information describing the number of frames contained in an RTP packet is not transmitted as part of the RTP payload. The only way to determine the number of BroadVoice32 frames is to count the total number of octets within the RTP packet, and divide the octet count by 20.

5. Storage Format

The storage format is used for storing speech frames, e.g., as a file or e-mail attachment.

The file begins with a header that includes only a magic number to identify the codec that is used. The magic number for the BroadVoice16 narrowband codec MUST correspond to the ASCII character string "#!BV16\n", or "0x23 0x21 0x42 0x56 0x31 0x36 0x0A" in hexadecimal format. The magic number for the BroadVoice32 wideband codec MUST correspond to the ASCII character string "#!BV32\n", or "0x23 0x21 0x42 0x56 0x33 0x32 0x0A". A file contains the encoded bit stream of either BroadVoice16 or BroadVoice32, but not both.

After the header that contains the magic number identifying the codec used, the encoded codec data frames are stored in a sequential order, as shown below.



6. IANA Considerations

Two new MIME sub-types as described in this section are to be registered.

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

6.1 MIME registration of BroadVoice16

MIME media type name: audio

MIME media subtype name: BV16

Required Parameter: none

Optional parameters:
The following parameters apply to RTP transfer only.

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

maxptime: The maximum amount of media which can be encapsulated in each packet, expressed as time in milliseconds. The time SHALL be calculated as the sum of the time the media present in the packet represents. The time SHOULD be a multiple of the duration of a single codec data frame (5 ms). If not signaled, the default maxptime value SHALL be 200 milliseconds.

Encoding considerations:
This type is defined for transfer of BV16-encoded data via RTP using the payload format specified in Sections 3 of RFC xxxx. It is also defined for other transfer methods using the storage format specified in Section 5 of RFC xxxx. Audio data is binary data, and must be encoded for non-binary transport; the Base64 encoding is suitable for Email.

Security considerations:
See Section 8 "Security Considerations" of RFC xxxx.

Public specification:
The BroadVoice16 codec will be specified in a CableLabs PacketCable standard document.

Additional information:
The following information applies to storage format only.

Magic number: ASCII character string "#!BV16\n" (or "0x23 0x21 0x42 0x56 0x31 0x36 0x0A" in hexadecimal)

File extensions: bvn, BVN (stands for "BroadVoice, Narrowband")

Macintosh file type code: none

Object identifier or OID: none

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

6.2 MIME registration of BroadVoice32

MIME media type name: audio

MIME media subtype name: BV32

Required Parameter: none

Optional parameters:
The following parameters apply to RTP transfer only.

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

maxptime: The maximum amount of media which can be encapsulated in each packet, expressed as time in milliseconds. The time SHALL be calculated as the sum of the time the media present in the packet represents. The time MUST be a multiple of the duration of a single codec data frame (5 ms). If not signaled, the default maxptime value SHALL be 200 milliseconds.

Encoding considerations:
This type is defined for transfer of BV32-encoded data via RTP using the payload format specified in Sections 4 of RFC xxxx. It is also defined for other transfer methods using the storage format specified in Section 5 of RFC xxxx. Audio data is binary data, and must be encoded for non-binary transport; the Base64

encoding is suitable for Email.

Security considerations:
See Section 8 "Security Considerations" of RFC xxxx.

Additional information:
The following information applies to storage format only.

Magic number: ASCII character string "#!BV32\n" (or "0x23 0x21 0x42 0x56 0x33 0x32 0x0A" in hexadecimal)

File extensions: bvw, BVW (stands for "BroadVoice, Wideband")

Macintosh file type code: none

Object identifier or OID: none

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

7. Mapping To SDP Parameters

Parameters are mapped to SDP [4] in a standard way. When conveying information by SDP, the encoding name SHALL be "BV16" for the BroadVoice16 narrowband codec and "BV32" for the BroadVoice32 wideband codec (the same as the MIME media subtype names).

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
```

8. 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, [5]). This implies that confidentiality of the media streams is achieved by encryption. Because the data compression used with this payload format is applied end-to-end, encryption may be performed after compression so there is no conflict between the two operations.

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 which 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.

9. References

[1] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", IETF RFC 1889, January 1996.

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

[3] PacketCable(TM) Audio/Video Codecs Specification, Cable Television Laboratories, Inc.

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

[5] H. Schulzrinne, "RTP Profile for Audio and Video Conferences with Minimal Control" IETF RFC 1890, January 1996.

10. Authors' Addresses

Juin-Hwey (Raymond) Chen
Broadcom Corporation
Room A3032
16215 Alton Parkway
Irvine, CA 92618
USA
Phone: +1 949-585-6288
Email: rchen@broadcom.com

Cheng-Chieh Lee
Broadcom Corporation
Room A3086
16215 Alton Parkway
Irvine, CA 92618
USA
Phone: +1 949-585-6467
Email: cclee@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 A3053
16215 Alton Parkway
Irvine, CA 92618
USA
Phone: +1 949-585-5768
Email: jthyssen@broadcom.com