Working Group PAYLOAD
Internet Draft
Intended status: Standards Track
Expires: July 2012
C. Hoene
University of Tuebingen
F. de Bont
Philips Electronics
January 4, 2012

RTP Payload Format for Bluetooth's SBC Audio Codec
draft-ietf-payload-rtp-sbc-02

Status of this Memo

Copyright Notice

Abstract

This document specifies a Real-time Transport Protocol (RTP) payload format to be used for the low complexity subband codec (SBC), which is the mandatory audio codec of the Advanced Audio Distribution Profile (A2DP) Specification written by the Bluetooth(r) Special Interest Group (SIG). The payload format is designed to be able to interoperate with existing Bluetooth A2DP devices, to provide high streaming audio quality, interactive audio transmission over the internet, and ultra-low delay coding for jam sessions on the internet. This document contains also a media type registration which specifies the use of the RTP payload format.

Table of Contents

1. Introduction

The Bluetooth(r) Special Interest Group (SIG) specifies in the Advanced Audio Distribution Profile (A2DP) [A2DPV12] a mono and stereo high quality audio subband codec (SBC). This document specifies the payload format for the encapsulation of SBC encoded audio frames into the Real-time Transport Protocol (RTP).

SBC has a low computational complexity at modest compression rates. Its bit rate can be controlled widely. Recommended operational modes range from 127 to 345 kb/s, for mono and stereo audio signals. SBC's algorithmic delay can be as low as 16 samples making it ideal for ensembles playing music over the network requiring ultra low acoustic delays.

2. 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 [RFC2119].

The following acronyms are used in this document:

A2DP - Audio Distribution Profile
AAC - Advanced Audio Coding
ATRAC - Adaptive Transform Acoustic Coding
DCCP - Datagram Congestion Control Protocol
MP3 - MPEG-1 Audio Layer 3
SBC - SubBand Codec
SIG - Special Interest Group

3. Background

The A2DP specification [A2DPV12] is intended for streaming of music content to headphones, headsets, or speakers over Bluetooth wireless channels. A2DP supports multiple audio coding including MP3, AAC, ATRAC, which are all non-mandatory. To ensure interoperability, the SBC codec has been specified, in appendix B of the A2DP specification, which shall be included into all A2DP Bluetooth devices.

SBC is a low complexity subband codec based on earlier work presented in [Bon1995] and [Rault1989]. It has a moderate

compression ratio. The SBC encoder has filter banks splitting the audio signal into 4 or 8 subbands. Then the codec decides with how many bits each subband is encoded and finally quantizes the subband signals blockwise. An SBC frame can have different block sizes. The size of a block can be 4, 8, 12 or 16. Both decoder and encoder shall support all four block sizes.

SBC can operate at four different sampling frequencies. The sampling frequency can be selected from a set of 16, 32, 44.1, and 48 kHz. It is mandatory that each SBC decoder can operate at the frequencies 44.1 and 48 kHz. Each SBC encoder shall work at least at a sampling rate of 44.1 or 48 kHz.

Four channel modes are supported, which are mono, dual channel, stereo, and joint-stereo. The decoder shall support all four of them; the encoder shall support mono and at least one additional mode.

SBC can use four or eight subbands. The decoder shall support both; the encoder shall support at least 8 subbands.

The bit allocation modes of SBC can be either based on signal to noise ratio or on loudness. The decoder shall support both modes; the encoder shall support at least the loudness mode.

The SBC encoder reduces one block to a given number of bits. The bit-pool variable defines how many bits are used per block. The A2DP profile defines the range of valid bit-pool values by providing minimum and maximum bit-pool values. The bit-pool values shall range from 2 to 250 but shall not be larger than number of subbands times 16 for the mono and dual and times 32 for the stereo and joint-stereo channel modes.

SBC encoders according to the A2DP profile may be capable of changing the bit-pool parameter dynamically during the encoding process. For example, algorithms were invented that change the number of bits depending on the current acoustic content [Pilati2008].

An SBC decoder according to the A2DP profile shall support all possible bit-pool values that do not result in excess of maximum bit rate, which is 320kb/s for mono and 512kb/s for two-channel modes. The encoder is required to support at least one possible bit-pool value. The A2DP profile recommends the encoding parameters given in Table 1.

SBC encoder settings at Medium Quality				
	Mono		Joint Stereo	
Sampling frequency (kHz)	44.1	48	44.1	48
Bitpool value	19	18	35	33
Resulting frame length (bytes)	46	44	83	79
Resulting bit rate (kb/s)	127	132	229	237
SBC encoder settings at High Quality				
	Mono		Joint Stereo	
Sampling frequency (kHz)	44.1	48	44.1	48
Bitpool value	31	29	53	51
Resulting frame length (bytes)	70	66	119	115
Resulting bit rate (kb/s)	193	198	328	345
Other settings: Block length = 16, loudness, subbands = 8				

Table 1: Recommended sets of SBC parameters in the SRC device as given in [A2DPV12]

3.1. SBC Media Payload Format

The A2DP V1.2 specification describes a media payload format, which is adopted in this document one-to-one without any change. In the following, for the sake of clarity, the payload format definition is repeated in the following.

3.2. SBC Fragmentation

The payload MUST consist of one media payload format header described in Section 3.3 and SBC frames described in Section 3.4. Either an integral number of SBC frames or one fragment of an SBC frame can be transmitted:

```
 (a) When the payload contains an integral number of SBC frames
+--------+----------+----------   -+
| Header | SBC frame | SBC frame ... |
+--------+----------+----------   -+

(b) When the SBC frame is fragmented
+--------+---------------------------------------+
| Header | First fragment of SBC frame           |
+--------+---------------------------------------+

+--------+---------------------------------------+
| Header | Subsequent fragments of the SBC frame |
+--------+---------------------------------------+
```

A media payload always starts with an 8-bit header, which is placed before the SBC data.

An SBC frame can be fragmented across several media payloads. All fragmented packets, except the last one, MUST have the same total data packet size.

This payload fragmentation CAN be preferred against the fragmentation mechanisms of lower layers (e.g., IP) because the packetisation delay and thus the acoustic latency are reduced and the error robustness is increased because parts of the SBC frame can be considered for decoding.

3.3. Media Payload Format Header

The following figure shows the format of media payload header, which consists of one byte.

```
0 1 2 3   4 5 6 7
+-+-+-+---+-+-+-+-+
|F|S|L|RFA|#frames|
+-+-+-+---+-+-+-+-+
```

F bit - Set to 1 if the SBC frame is fragmented, otherwise set to 0.

S bit - Set to 1 for the starting packet of a fragmented SBC frame, otherwise set to 0.

L bit - Set to 1 for the last packet of a fragmented SBC frame, otherwise set to 0.

RFA - SHOULD be zero, reserved for future addition.

#frames (4 bits) - If the F bit is set to 0, this field indicates the number of frames contained in this packet. If the F bit is set to 1, this field indicates the number of remaining fragments, including the current fragment. Thus the last counter value MUST be one. For example, if there are three fragments then the counter has value 3, 2 and 1 for subsequent fragments.

3.4. SBC Frame Structure

An SBC frame consists of a frame header, scale factors, audio samples, and padding bits. The following diagram shows the general SBC frame format layout:

```
+--------------+--------------+--------------+---------+
| frame_header | scale_factors | audio_samples | padding |
+--------------+--------------+--------------+---------+
```

The following sections describe the audio format, which consists of bits stored in a bandwidth-efficient, compact mode.

3.5. Frame Header

The frame header consists of fields defined in [A2DPV12], which are SYNCWORD, SAMPLING_FREQUENCY, BLOCKS, CHANNEL_MODE, ALLOCATION_METHOD, SUBBANDS, BITPOOL, CRC_CHECK, optionally JOIN bit fields and a RFA. The layout of the first four bytes of the frame header is given in the following table.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| SYNCWORD      |SF.|BL.|CM.|A|S|BITPOOL        |CRC_CHECK      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
Legend: SF.=SAMPLING FREQUENCY, BL.=BLOCKS, CM.=CHANNEL_MODE,
A.=ALLOCATION_METHOD, S.=SUBBANDS
```

SYNCWORD (8 bits): The first field is the 8 bit synchronization word, which is always set to 156.

SAMPLING_FREQUENCY (2 bits): The sampling frequency field indicates with which sampling frequency the SBC frame has been encoded. The table below specifies the corresponding sampling frequencies for the bit patterns. The sampling frequency MUST NOT be changed without changing the payload type, too.

SAMPLING_FREQUENCY bit 0 1	sampling frequency (Hz)
0 0	16000
0 1	32000
1 0	44100
1 1	48000

BLOCKS (2 bits): It indicates the block size with which the stream has been encoded. The block size is selected conforming to the table below. The block size MUST NOT be changed without changing the payload type, too.

BLOCKS bit 0 1	Number of blocks
0 0	4
0 1	8
1 0	12
1 1	16

CHANNEL_MODE (2 bits): These two bits indicate with which channel mode the frame has been encoded. The number of channels depends on this information. The channel mode MUST NOT be changed without changing the payload type, too.

CHANNEL_MODE bit 0 1	channel mode	number of channels
0 0	MONO	1
0 1	DUAL_CHANNEL	2
1 0	STEREO	2
1 1	JOINT_STEREO	2

ALLOCATION_METHOD (1 bit): This bit indicates how the bit pool is allocated to different subbands. Either it is based on the loudness of the sub band signal or on the signal to noise ratio. The allocation method MUST NOT be changed without changing the payload type, too.

```
+-------------------+------------+
| ALLOCATION_METHOD | allocation |
|       bit 0       | method     |
+-------------------+------------+
|           0       |  LOUDNESS  |
|           1       |     SNR    |
+-------------------+------------+
```

SUBBANDS (1 bit): This bit indicates the number of subbands with which the frame has been encoded. The number of subband MUST NOT be changed without changing the payload type, too.

```
+----------+-----------+
| SUBBANDS | number of |
|   bit 0  | subbands  |
+----------+-----------+
|       0  |      4    |
|       1  |      8    |
+----------+-----------+
```

BITPOOL (8 bits): This unsigned integer indicates the size of the bit allocation pool that has been used for encoding the current block. The value of the bit-pool field MUST NOT exceed 16 times the number of subbands for the MONO and DUAL_CHANNEL channel modes and 32 times the number of subbands for the STEREO and JOINT_STEREO channel modes. The bitpool value MAY change from SBC frame to the next. In addition, the bitpool value MUST be restricted such that it does not result in excess of maximum bit rate, which is 320kb/s for mono and 512kb/s for two-channel modes.

The remaining part of the header consists of CRC_CHECK, optionally JOIN bit fields and a RFA.

3.6. Remaining Frame Part

The remaining part of the frame includes scale factors and audio sample data, which are processed by the codec as described in [A2DPV12].

4. Usage Scenarios

As compared to many other encoding schemes, the SBC codec is general enough to support multiple, quite diverse usage scenarios. Thus, it

might be required to change the behavior of the encoding and transmission to achieve a good performance for a given usage scenario. Thus, three main scenarios are listed and their quality requirements and impact on encoding and transmission are described.

4.1. Scenario 1: Interconnection of A2DP Devices

This scenario is intended for interconnecting Bluetooth A2DP devices. RTP frames generated by an A2DP device can be transmitted directly via this RTP profile. Vice versa, an A2DP device should be able to receive the RTP profile by default. Thus, the payload format describe in this RFC MUST be fully interoperable with any A2DP device.

The transmission between two A2DP devices has a constant frame rate with a sender-controlled bit rate. It is not anticipated that the transmission is adapted to congestion and bandwidth variation.

4.2. Scenario 2: High Quality Interactive Audio Transmissions

In the second scenario a telephone call is considered having a very good audio quality at modest acoustic one-way latencies ranging from 50 and 150 ms [ITUG107], so that music can be listened over the telephone while two persons talk together interactively.

In addition, the reliability of the audio transmission should be high, even in cases of low and varying bandwidth.

This second scenario assumes that the SBC transmission is used on top of a transport protocol that implements a congestion control algorithm. Using the SBC encoding, the sampling, bit, and frame rates should be controlled to cope with congestion. For example, if the available transmission bandwidth is too low to allow SBC to transmit audio at a high quality, the application can lower the sampling, bit, or frame rate of the stream at the cost of higher algorithmic delay or a degraded audio quality. In this case, changing the sampling or frame rate may cause a short acoustic artifact because SBC's internal filters must be reset.

The A2DP media format does not allow a dynamic change of the encoding parameters beside the bit-pool value. The encoding parameters can only be altered with the "Change Parameters" procedure, which is defined in [GAVDPV12]. Such a change will cause a hearable interruption and thus shall be avoided.

If an application using RTP wants to switch between different sets of encoding parameters, then these set of parameter CAN be either negotiate beforehand (as described in Section 7.2.) or an renegotiation similar to the "Change Parameters" procedure CAN take place. An application MUST NOT change the sampling frequency, block length, encoding mode or the number of subbands within one RTP session having the same RTP payload identifier.

4.3. Scenario 3: Ensembles performing over a Network

In some usage scenarios, users want to act simultaneously and not just interactively. For example, if persons sing in a chorus, if musicians jam, or if e-sportsmen play computer games in a team together, they need to acoustically communicate.

In these scenarios, the latency requirements are much harder than for interactive usages. For example, if two musicians are placed more than 10 meters apart, they can hardly keep synchronized. Empirical studies [Gurevich2004] have shown that if ensembles playing over networks, the optimal acoustic latency is around 11.5 ms with targeted range from 10 to 25 ms.

To fulfill such requirements, it might be necessary to further reduce the algorithmic coding delay by varying the block length parameter. The default value of the block length parameter is chosen such that the coding efficiency is maximized. For example, at 44.1 kHz and using 8 subbands and a block length of 16, the algorithmic delay is 4.72 ms (208 samples). The value of the block length parameter can be decreased, at the expense of a higher bit rate or lower quality, to lower the latency to fulfill the very stringent latency requirements of this scenario.

Still, given the speed of light as the fundamental limit of speed of information exchange, distributed ensembles can perform only regionally if latency budget of 25 ms must keep. Typically, an optical fiber has a refractive index of 1.46 and thus in an optical fiber bits travel about 5136 km one-way in 25 ms.

5. Header Usage

The format of the RTP header is specified in [RFC3550]. The payload format defined in this document uses the fields of the header in a manner fully consistent with that specification.

marker (M): In accordance with [A2DPV12] the marker bit MUST be set to zero.

payload type (PT): 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 6.2).

timestamp (TS): The RTP timestamp clock frequency MUST be the same as the sampling frequency, which has been negotiated for the current RTP session (see Section 6.2). If a media payload consists of multiple SBC frames, the TS of the media packet header represents the TS of the first SBC frame. The TS of the following SBC frames MUST be calculated using the sampling rate and the number of samples per frame per channel. A change in sampling frequency MUST NOT occur within one media packet.
A SBC frame may be fragmented into multiple media packets to reduce the packetisation delay. Then, all packets that make up a fragmented SBC frame MUST use the same TS.

6. Payload Format

The format of the payload MUST follow exactly the description given in appendix B of [A2DPV12] and which is repeated in Section 3. If appendix B of [A2DPV12] and the description in Section 3 differ, the former standard is normative.

If the payload format parameters have been negotiated and a restricted set of encoding and decoding modes have been selected, than any SBC frame that describes a coding mode that has not been chosen MUST be ignored.

7. Payload Format Parameters

This section defines the parameters that MAY be used to configure optional features in the SBC payload format over RTP transmission.

The parameters are defined here as part of the media subtype registrations for the SBC codec. A mapping of the parameters into the Session Description Protocol (SDP) [RFC4566] 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.

7.1. Media Type Registration for SBC

[Note to RFC Editor: Please replace all occurrences of RFC XXXX by the RFC number assigned to this document]

This registration is done using the template defined in [RFC4288] and following [RFC4855].

Media type name: audio

Subtype name: SBC

Required parameters:

Rate: The RTP timestamp clock rate. See Section 5 for usage details.

Optional parameters:

Channels: Specifies the number of audio channels: 2 for stereo (refer to RFC 4566 [RFC4566]) and 1 for mono, accordingly the SBC channel mode. If one channel is used, this parameter can be omitted.

Capabilities: The capabilities of the encoder and decoder are described by a parameter string that MUST start with an octet written as two hexadecimal digits. This octet is called VERSION and MUST be identical to the SYNCWORD that will be used in the SBC frames. It is used to distinguish different negotiation procedures.
The interpretation of the following characters depends on the value of the VERSION octet. Refer to Section 7.1.1. and Section 7.1.2. to find a description.

Encoding considerations: This media type is framed and contains binary data; see Section 4.8 of RFC 4288.

Security considerations: See Section 9 of RFC XXXX

Interoperability considerations: none

Published specification: RFC XXXX

Applications which use this media type: Audio and video conferencing tools, distributed orchestras

Additional information: none

Person & email address to contact for further information:
See Authors' Addresses at the end of RFC XXXX

Intended usage: COMMON

Restrictions on usage: none

Author: See Authors' Addresses at the end of RFC XXXX

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

7.1.1. Capabilities: A2DP Modes

The capabilities of the encoder and decoder MUST start with the hexadecimal value of 9C, followed by a comma and four comma-separated hexadecimal octets. These four octets called Octet 1, 2, 3, and 4 share a similar meaning as those defined in Section 4.3.2 of [A2DPV12]. However, because sampling frequency and number of channels are already given in the SDP parameter "a=rtpmap", bit 0 up to and including bit 3 of Octet 1 MUST BE ignored if received. The meaning of the bits and the octets are described in the following enumeration. The bit numbering follows the network bit order having the highest bit first.

- o Octet 1: Bit 0 (aka 2^7): If one, then the sampling frequency 16000 Hz is supported (ignored during SDP negotiations but SHOULD be set if the clock rate is 16000 and MUST be cleared otherwise).

- o Octet 1: Bit 1: If one, then the sampling frequency 32000 Hz is supported (ignored during SDP negotiations but SHOULD be set if the clock rate is 32000 and MUST be cleared otherwise).

- o Octet 1: Bit 2: If one, then the sampling frequency 44100 Hz is supported (ignored during SDP negotiations but SHOULD be set if the clock rate is 44100 and MUST be cleared otherwise).

- o Octet 1: Bit 3: If one, then the sampling frequency 48000 Hz is supported (ignored during SDP negotiations but SHOULD be set if the clock rate is 48000 and MUST be cleared otherwise).

- o Octet 1: Bit 4: If one, then the channel mode MONO is supported (ignored during SDP negotiations but SHOULD be set if the number of channels is one and MUST be cleared otherwise).

- o Octet 1: Bit 5: If one, then the channel mode DUAL_CHANNEL is supported (*).

- o Octet 1: Bit 6: If one, then the channel mode STEREO is supported (*).

- o Octet 1: Bit 7 (aka 2^0): If one, then the channel mode JOINT_STEREO is supported (*).

- o Octet 2: Bit 0: If one, the block length can be 4.

- o Octet 2: Bit 1: If one, the block length can be 8.

- o Octet 2: Bit 2: If one, the block length can be 12.

- o Octet 2: Bit 3: If one, the block length can be 16.

- o Octet 2: Bit 4: If one, the number of subband can be 4.

- o Octet 2: Bit 5: If one, the number of subband can be 8.

- o Octet 2: Bit 6: If one, the allocation mode based on signal to noise ratio is supported.

- o Octet 2: Bit 7: If one, the allocation mode based on loudness is supported.

- o Octet 3: Unsigned integer: The minimal bit-pool value that the device supports. MUST be larger or equal than 2 and less or equal than the maximal bit-pool value.

- o Octet 4: Unsigned integer: The maximal bit-pool value that the device supports MUST be equal or lower than 250.

(*) At least one of the bits 5, 6 or 7 of Octet 1 MUST be set if the number of channels is set to two in the SDP parameter "a=rtpmap".

7.1.2. Capabilities: Other Modes

If the value of the VERSION octet is not equal to a known SYNCWORD value, then the capabilities MUST be ignored.

7.2. 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. When SDP is used to specify sessions employing the SBC codec, the mapping is as follows:

- o The media type ("audio") goes in SDP "m=" as the media name.
- o The media subtype ("SBC") goes in SDP "a=rtpmap" as the encoding name.
- o The required parameter "rate" goes in SDP "a=rtpmap" as the RTP <clock rate>.
- o The optional parameter "channels", if present, goes in SDP as the "a=rtpmap" RTP <encoding parameters>.
- o The optional parameter "capabilities", if present, goes in the SDP "a=fmtp" by the capabilities description as described in Section 7.1.

7.2.1. Offer-Answer Model Considerations

The Bluetooth standard document [AVDTPV12] describes how an A2DP source and an A2DP sink negotiate their capabilities. Prior to the establishment of the audio stream, one A2DP device can query the service capabilities of the other device using the "Get Capabilities Procedure". In any case, the coding mode is set using the "Set Configuration" procedure. Only after a successful configuration, the stream connection can be established.

In addition to the Bluetooth negotiation procedure, the SDP negotiation MUST NOT agree on one single configuration but CAN agree that multiple configuration modes, which are identified by different payload type values, are supported.

The following considerations apply when using SDP offer-answer procedures [RFC3264] to negotiate the use of SBC payload in RTP:

- o The "capabilities" parameter is bi-directional, i.e., the restricted mode set applies to media both to be received and sent by the declaring entity. If the capabilities were supplied in the offer, the answerer MUST return either the same mode-set or a subset of this mode-set. If no capabilities were supplied in the offer, the answerer MAY return capabilities to restrict the possible modes. In any case, the capabilities in the answer then apply for both offerer and answerer. The offerer MUST NOT send frames of a mode that has been removed by the answerer. The

negotiation is finished if the offerer and the answerer have agreed upon explicit capabilities for each payload type number. The number of blocks and subbands and the kind of allocation method and channel mode MUST have been negotiated unambiguously.

- Any unknown parameter in an offer MUST be ignored by the receiver and MUST NOT be included in the answer.

Below are some example parts of SDP offer-answer exchanges.

- Example 1

  Offer: SBC all A2DP modes

```
 m=audio 54874 RTP/AVP 96
 a=rtpmap:96 SBC/48000/2
 a=fmtp:96 capabilities=9C,17,FF,02,FA
m=audio 54874 RTP/AVP 97
 a=rtpmap:97 SBC/48000
 a=fmtp:97 capabilities=9C,18,FF,02,FA
 m=audio 54874 RTP/AVP 98
 a=rtpmap:98 SBC/44100/2
 a=fmtp:98 capabilities=9C,27,FF,02,FA
m=audio 54874 RTP/AVP 99
 a=rtpmap:99 SBC/44100
 a=fmtp:99 capabilities=9C,28,FF,02,FA
 m=audio 54874 RTP/AVP 100
 a=rtpmap:100 SBC/32000/2
 a=fmtp:101 capabilities=9C,47,FF,02,FA
m=audio 54874 RTP/AVP 102
 a=rtpmap:102 SBC/32000
 a=fmtp:102 capabilities=9C,48,FF,02,FA
 m=audio 54874 RTP/AVP 103
 a=rtpmap:103 SBC/16000/2
 a=fmtp:103 capabilities=9C,87,FF,02,FA
m=audio 54874 RTP/AVP 104
 a=rtpmap:104 SBC/48000
 a=fmtp:104 capabilities=9C,88,FF,02,FA
```

  Answer: 48 kHz, JOINT_STEREO, 16 blocks, 8 subbands, LOUDNESS

```
 m=audio 59452 RTP/AVP 96
 a=rtpmap:96 SBC/48000/2
 a=fmtp:96 capabilities=9C,11,15,02,FA
```

- Example 2

  Offer: The A2DP SBC 48 kHz modes with mono or joint stereo, 8 subbands, loudness allocation method. In addition an unknown mode

called AD is offered.

```
m=audio 54874 RTP/AVP 96
a=rtpmap:96 SBC/48000/2
a=fmtp:96 capabilities=9C,11,F5,02,FA
m=audio 54874 RTP/AVP 97
a=rtpmap:97 SBC/48000/1
a=fmtp:97 capabilities=9C, 18,F5,02,FA
m=audio 54874 RTP/AVP 98
a=rtpmap:98 SBC/16000/1
a=fmtp:98 capabilities=AD
```

Answer: both A2DP modes are accepted but the unknown mode AD is ignored.

```
m=audio 59452 RTP/AVP 96
a=rtpmap:96 SBC/48000/2
a=fmtp:96 capabilities=9C,11,F5,02,FA
m=audio 59452 RTP/AVP 9
a=rtpmap:97 SBC/48000/1
a=fmtp:97 capabilities=9C,18,F5,02,FA
```

7.2.2. Declarative SDP Considerations

For declarative use of SDP nothing specific is defined for this payload format. The configuration given by the SDP MUST be used when sending and/or receiving media in the session.

8. Congestion Control

One Bluetooth links, bandwidth can be reserved and thus the A2DP specification does not consider any kind of congestion control. However, congestion control is an important issue for any usage in non-dedicated networks such as the Internet. Thus, congestion control for RTP MUST be used in accordance with [RFC3550] and any appropriate profile (for example, [RFC3551]). An additional requirement if best-effort service is being used is: users of this payload format MUST monitor packet loss to ensure that the packet loss rate is within acceptable parameters.

Reducing the session bandwidth is possible by one or more of the following means, which all will have negative impact to the users' experience as he can notice a higher latency or a degraded audio quality. The selection of the following means depends on current usage scenario, the congestion control protocol, and the perceptual assessment of the audio transmission and is not subject of this specification.

1. If the bandwidth and frame rate shall be reduced, the sampling rate can be lowered [Boutremans2004,Hoene2005].

2. If the gross bandwidth and the frame rate shall be reduced, more blocks can be put into one SBC frame and more SBC frames can be placed in one RTP payload.

3. If the bandwidth shall be reduced, then the bit-pool value can be reduced, so that the frames get smaller or the mono mode can be selected.

4. If the bandwidth is very low, instead of an ongoing transmission, a push-to-talk like service with temporary transmission interruptions and a high delay can be applied.

5. If the packet loss rate is very high, the session shall be terminated because the quality of the audio transmission is too bad to be useful [Widmer2002].

Because the SBC encoding can be tuned with many parameters, it is especially useful for rate adaptive transport protocols such as DCCP [RFC4340] or TCP [RFC4571]. The report [Hoene2009] describes, which SBC coding mode gives the best speech and audio quality under known bandwidth and time constrains.

9. Packet Loss Concealment

In order to cope with packet losses, the SBC decoder SHOULD be extended by a packet loss concealment algorithm. The packet loss concealment algorithm SHOULD provide a good audio quality in case of losses. Otherwise, the congestion control algorithm can not trade off well the quality impairment due to packet losses versus the quality impairment caused by different encoding modes. It is RECOMMENDED that at a least the reserve order replicated pitch periods (RORPP) algorithm as defined in [Hoene2009] or any better is used.

If this requirement is not meet, then the congestion control cannot predict the impact of packet loss on the audio quality and thus will not be able to control the encoding parameters optimally.

10. Security Considerations

RTP packets using the payload format defined in this specification are subject to the general security considerations discussed in the

RTP specification [RFC3550] and any appropriate profile (for example, [RFC3551]).

As this format transports encoded speech/audio, the main security issues include confidentiality, integrity protection, and authentication of the speech/audio itself. The payload format itself does not have any built-in security mechanisms. Any suitable external mechanisms, such as SRTP [RFC3711], MAY be used.

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

11. IANA Considerations

It is requested that one new media subtype (audio/SBC) and one optional parameter for this media subtype ("capabilities") are registered by IANA, see Section 5.1 and Section 5.2.

12. References

12.1. Normative References

[A2DPV12] Bluetooth SIG, "Advanced Audio Distribution Profile", Audio Video WG, adopted specification, revision V1.2, April 16th, 2007.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3264] Rosenberg, J. and Schulzrinne, H., "An Offer/Answer Modelwith Session Description Protocol (SDP)", RFC 3264, June 2002.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC3551] Schulzrinne, H. and Casner, S., "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.

[RFC4288] Freed, N. and Klensin, J., "Media Type Specifications and Registration Procedures", BCP 13, RFC 4288, December 2005.

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

[RFC4855] Casner, S., "Media Type Registration of RTP Payload Formats", RFC 4855, February 2007.

12.2. Informative References

[AVDTPV12] Bluetooth SIG, "Audio/Video Distribution Transport Protocol Specification", Audio Video WG, adopted specification, revision V12, April 16th, 2007.

[Bon1995] de Bont, F., Groenewegen, M., and Oomen, W., "A High Quality Audio-Coding System at 128 kb/s", 98th AES Convention, February 25 - 28, 1995.

[Boutremans2004] Boutremans, C., Le Boudec J.-Y., and Widmer, J., "End-to-end congestion control for tcp-friendly flows with variable packet size", ACM Computer Communication Review, Vol. 31, No. 2, pp. 137-151, 2004.

[Pilati2008] Pilati, L., Zadissa, M., "Enhancements to the SBC CODEC for Voice Communication in Mobile Devices", AES Convention 124, No. 7347, May 2008.

[Hoene2009] Hoene, C., Hyder, M.. "Considering bluetooth's subband codec (SBC) for wideband speech and audio on the internet". Technical Report WSI-2009-3, Universitaet Tuebingen - WSI, 72076 Tuebingen, Germany, October 2009.

[GAVDPV12] Bluetooth SIG, "Generic Audio/Video Distribution Profile", Audio Video WG, adopted specification, revision V12, April 16th, 2007.

[Gurevich2004] Gurevich, M., Chafe, C., Leslie, G., and Tyan, S., "Simulation of Networked Ensemble Performance with Varying Time Delays: Characterization of Ensemble Accuracy", Proceedings of the 2004 International Computer Music Conference, Miami, USA, 2004.

[Hoene2005] Hoene, C., and Karl, H., and Wolisz, A., "A perceptual quality model intended for adaptive VoIP applications", International Journal of Communication Systems, Wiley, August 2005.

[ITUG107] ITU-T G.107, "The E-model, a computational model for use in transmission planning", ITU-T Recommendation G.107, May 2000.

[Rault1989] Rault, J., Dehery, Y., Roudaut, J., Bruekers, A., and Veldhuis, R., "Digital transmission system using subband coding of a digital signal", Publication number: EP0400755 (B1).

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[RFC4340] Kohler, E., Handley, M., and Floyd, S., "Datagram Congestion Control Protocol (DCCP)", RFC 4340, March 2006.

[RFC4571] Lazzaro, J., "Framing Real-time Transport Protocol (RTP) and RTP Control Protocol (RTCP) Packets over Connection-Oriented Transport", RFC4571, July 2006.

[Widmer2002] Widmer, J., Mauve, M., and Damm, J., "Probabilistic congestion control for non-adaptable flows", In 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV), Miami, FL, USA, May 2002.

13. Acknowledgments

Funding for this draft has been provided by the University of Tuebingen within the "Projektfoerderung fuer Nachwuchswissenschaftler".

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Christian Hoene
University of Tuebingen
Wilhelm-Schickard-Institute
Sand 13
72076 Tuebingen
DE

Phone: +49 7071 29 70532
Email: hoene@uni-tuebingen.de

Frans de Bont
Philips Electronics
High Tech Campus 36
5656 AE Eindhoven
NL

Phone: +31 40 2740234
Email: frans.de.bont@philips.com