Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 12, 2013

T. le Grand
Google
P. Jones
P. Huart
Cisco Systems
T. Shabestary
H. Alvestrand, Ed.
Google

February 8, 2013

RTP Payload Format for the iSAC Codec
draft-ietf-avt-rtp-isac-04

Abstract

iSAC is a proprietary wideband speech and audio codec developed by Global IP Solutions (now part of Google), suitable for use in Voice over IP applications. This document describes the payload format for iSAC generated bit streams within a Real-Time Protocol (RTP) packet. Also included here are the necessary details for the use of iSAC with the Session Description Protocol (SDP).

Requirements Language

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].

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on August 12, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

This document gives a general description of the iSAC wideband speech codec and specifies the iSAC payload format for usage in RTP packets. Also included here are the necessary details for the use of iSAC with the Session Description Protocol (SDP).

2. iSAC Codec Description

The iSAC codec is an adaptive wideband/superwideband speech and audio codec that operates with short delay, making it suitable for high quality real time communication. It is specially designed to deliver wideband speech quality in both low and medium bit rate applications. It also handles non-speech audio well, such as music and background noise. A freely available reference implementation exists [iSAC].

The iSAC codec compresses speech frames of 16 kHz, 16-bit sampled input speech, each frame containing 30 or 60 ms of speech. It also has a superwideband mode which allows a 32 kHz sampling rate. In super-wideband mode the input signal is split into wideband (0-8 kHz) and upper (8-16 kHz) signal. Each sub-band is encoded independently, and their associated payloads concatenated, c.f. Figure 2, to construct the overall iSAC super-wideband RTP payload. Note that the same encoder/decoder is used for the wideband part for both wideband and super-wideband modes.

The codec runs in one of two different modes called channel-adaptive mode and channel-independent mode. In both modes iSAC is aiming at a target bit rate, which is neither the average nor the maximum bit rate that will be reach by iSAC, but corresponds to the average bit rate during peaks in speech activity. The bit rate will sometimes exceed the target bit rate, but most of the time will be below. The average bit rate obtained is on average about a factor of 1.2 times lower than the target bit rate on continuous speech, and will be lower on speech with pauses.

In channel-adaptive mode the target bit rate is adapted to give a bit rate corresponding to the available bandwidth on the channel. Even at dial-up modem data rates (including IP, UDP, and RTP overhead) iSAC delivers high quality by automatically adjusting transmission rates to give the best possible listening experience over the available bandwidth.

In channel-independent mode a target bit rate has to be provided to iSAC prior to encoding; the target bit rate can be changed over the time of the call.

After encoding the speech signal the iSAC coder uses lossless coding to further reduce the size of each packet, and hence the total bit rate used.

The adaptation and the lossless coding described above both result in a variation of packet size, depending both of the nature of speech and the available bandwidth. Therefore, the iSAC codec, in wideband mode, operates at transmission rates from about 10 kbps to about 32 kbps. In super-wideband mode, the transmission rate is in the range of 10 kbps to 56 kbps. If operating in super-wideband mode, the iSAC codec automatically adjusts the effective encoded audio bandwidth for the best experience.

Bit Rate [kbps]	10 - 32	32 - 38	38 - 45	45 - 50	50 - 56
Effective Bandwidth [kHz]	0 - 8 kHz	0 - 8 operating at 32 kbps	0 - 12	0 - 12 operating at 45 kbps	0 - 16

The main characteristics can be summarized as follows:

- Wideband or superwideband, 16 kHz or 32 kHz respectively, speech and audio codec
- Variable bit rate, which depends on the input signal
- Adaptive rate with two modes: channel-adaptive or channel-independent mode
- Bit rate range from around 10 kbps to 32 kbps when operating on wideband input. For input audio sampled at 32 kHz, the bit rate range 10 kbps to 56 kbps.
- Operates on 30 or 60 ms of speech for wideband inputs, and only 30 ms for super-wideband inputs.
- In super-wideband mode, depending on the target bit rate, the effective bandwidth is adjusted for the optimal experience.

3. RTP Payload Format

The iSAC codec in wideband mode uses a sampling rate clock of 16 kHz, so the RTP timestamp MUST be in units of 1/16000 of a second. In super-wideband mode, the iSAC codec uses a sampling rate clock of 32 kHz, so the RTP timestamp MUST be in units of 1/32000 of a second.

The RTP payload for iSAC has the format shown in Figure 1. No additional header fields specific to this payload format are required. For RTP based transportation of iSAC encoded audio, the standard RTP header [RFC3550] is followed by one payload data block.

The assignment of an RTP payload type for the format defined in this memo is outside the scope of this document. The RTP profiles in use currently mandate binding the payload type dynamically for this payload format.

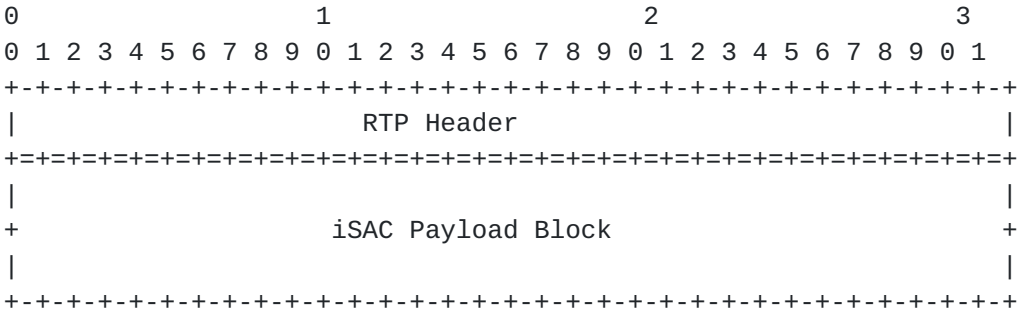

Figure 1: RTP packet format for iSAC

3.1. Payload Header

The payload header holds information for the receiver about the available bandwidth, in the form of a Bandwidth Estimation Index (BEI), and the length of the speech data in the current payload (frame length, FL). The header has the format defined in Figure 3. Note that the size of the header can vary due to the lossless encoding described in section 2 and in section 3.1. Also note that the BEI is always estimated and transmitted, even if iSAC runs in channel-independent mode.

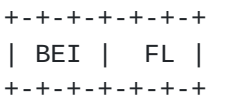

Figure 3: Payload Header

- BEI: Bandwidth Estimation Index. The bandwidth estimate that the sender estimates for a stream originated at the receiver. It is quantized into one out of 24 values. Valid values are 0 to 23; consult source code for details.

- FL: The length of the speech data (Frame Length) present in the payload, given in number of speech samples. Valid frame lengths are 480 (30 ms) and 960 (60 ms) samples.

The BEI and FL are encoded together with the data using a lossless compressed encoding, which results in a variable number of bits used

to represent the fields.

3.2. iSAC Wideband Payload Format

The iSAC payload block consists of a payload header and one or two encoded 30 ms speech frames. The iSAC payload is generated in the following manner:

- Parameters representing one or two 30 ms frames of speech data are determined by the encoder. The parameters are quantized to generate encoded data corresponding to the one or two speech frames. The length of the encoded data is variable and depends on the signal characteristics and the target bit rate.

- The payload header is generated (described in Section 3.1) and added before the encoded parameter data for the speech frame(s).

- Lossless coding is applied to the complete iSAC payload block, including payload header, to generate a compressed payload. The length depends on the length of the data generated to represent the speech and the effectiveness of the lossless coding.

No part of the payload header or the encoded speech data can be retrieved without partly or fully decoding the packet.

The following figure shows an iSAC payload block containing 60 ms of encoded speech data.

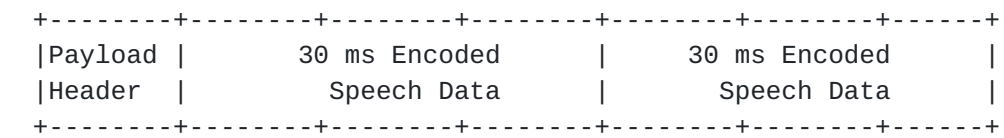

Figure 2: Payload format for iSAC

3.2.1. Encoded Speech Data

The iSAC encoded speech data consist of parameters representing one or two frames of 30 ms speech. The length of the speech data is signaled in the header (in number of samples), and the length may change at any time during a session. In channel-adaptive mode the length is changed to best utilize the available bandwidth, and extra padding is added to some packets as a bandwidth probe.

The iSAC payload is padded to whole octets, and has a variable length depending on the input source signal, number of 30 ms speech frames, and target bit rate.

The number of octets used to describe one frame of 30 ms speech

typically varies from around 50 to around 120 octets. For the case of 60 ms speech (two 30 ms speech frames), the number of octets varies from around 100 to around 240 octets. The absolute maximum allowed payload length is 400 octets. The sender can choose to limit the packet size further when transmitting. The minimum useful limit for the payload length is 100 octets.

The sensitivity to bit errors is equal for all bits in the payload.

3.3. iSAC Superwideband Payload Format

In super-wideband mode, payloads associated with each sub-band (wideband 0-8 kHz and upper-band 8-16 kHz) are constructed independently and concatenated as depicted in Figure 2. Note that in super-wideband mode only one 30 ms frame is encoded in each payload.

The receiver will know from negotiation whether wideband or super-wideband is sent; it can also verify this for each packet by verifying the CRC checksum.

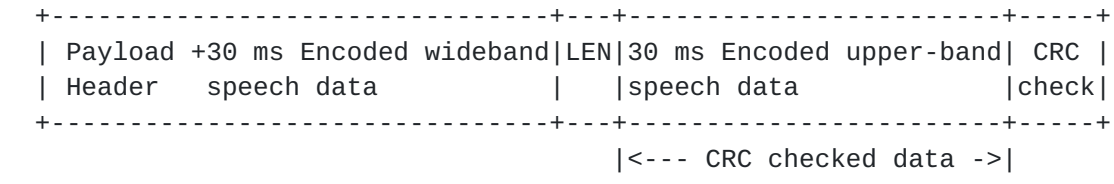

Figure 4: Super-Wideband payload format

Payloads of wideband and upper-band are encoded independently, allowing the encoder to simply concatenate two payloads to construct one iSAC super-wideband payload. The RTP payload of the iSAC super-wideband codec starts with the payload of the wideband part, which is padded to whole octets, followed by one byte (LEN in Figure 4) representing the length of the remaining sequence, payload of the upper-band plus 4 bytes for CRC sequence.

If LEN_UB denotes the length of the upper-band payload, then LEN = 1 + LEN_UB + 4. If this value would exceed 255 at encoding, the upper-band payload is omitted.

The CRC check is added to distinguish between upper-band payload and random bit-stream padding that can be added for probing available network bandwidth.

At the receive side, a super-wideband payload is first given to the wideband decoder. The wideband decoder decodes as many parameters as required to uniquely reproduce the encoded wideband audio. The next byte in the payload should hold the value of LEN. This provides a

sanity check that the decoding process has not failed. Thereafter, the receiver runs a CRC check over the upper-band payload and compares the results with the last 4 bytes in the packet.

If the computed CRC and the last four bytes of the payload don't match, the remaining bits are assumed to be added for probing the network. Hence, the upper-band signal is replaced by zeros and combined with the wideband signal to generate the super-wideband signal.

If the two CRCs match, then the upper-band payload is given to the upper-band decoder. Thereby, the output of the upper-band decoder is combined with the wide-band decoded audio to generate the super-wideband signal.

It might be that for a given packet, the wideband decoder uses all the given payload. This can be the case when a super-wideband encoder is operating at low rates and has adjusted the effective bandwidth to wideband. In this case, the decoder inserts zeros as the reconstructed upper-band and combines both bands to reproduce the super-wideband signal.

3.3.1. Encoded Upper-band Speech Data

The iSAC encoded upper-band speech data consists of parameters representing one frame of 30 ms speech. Depending on the target rate the upper-band encoder might choose to only encode the sub-band of 8 kHz to 12 kHz.

3.4. Padding

Padding, which consists of randomly generated bits, may be added at the end of the payload in both wideband and superwideband modes. It can be used by the sender for bandwidth probing, and is always ignored by the receiver.

In wideband mode, padding simply follows the payload, preceded by a length field.

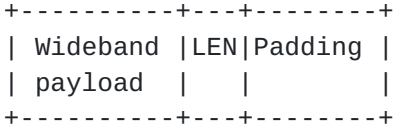

Figure 5: Wideband payload format with padding.

LEN is the length of the padding in bytes + 1: LEN = LEN_PAD + 1

In superwideband mode, the format of a packet with padding looks like the following.

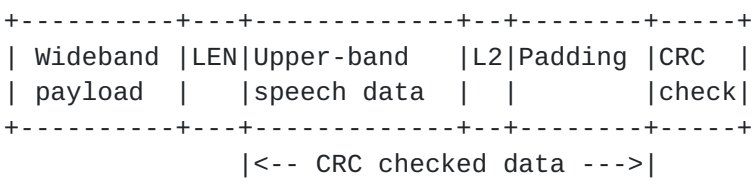

Figure 6: Super-Wideband payload format

LEN is 1 + LEN_UB + 1 + LEN_PAD + 4, where LEN_UB is the length of the upper-band speech data in bytes, and LEN_PAD is the length of the padding in bytes.

L2 is LEN_PAD + 1.

The CRC check runs over the upper-band speech data, L2 and the padding.

3.5. Multiple iSAC frames in an RTP packet

More than one iSAC payload block MUST NOT be included in an RTP packet by a sender.

Further, iSAC payload blocks MUST NOT be split between RTP packets.

4. Congestion Control

When ISAC is used in an environment where congestion control is useful, there are two properties of importance:

- The ISAC format has the ability to pad packets. This allows a sender to probe a channel with more bits per second than is strictly needed for the transmission of current data, so that it can check for the possibility of sending bigger packets without incurring increased packet loss.

- The iSAC encoder (in channel-adaptive mode) can continuously tune its encoding parameters so as to adapt the encoding to the available bandwidth, without introducing switching artifacts into the audio stream.

- In the case where two parties have one audio channel in each direction, they can use the BEI field of the A->B audio flow as a feedback channel for the B->A audio flow.

Coupled with a feedback channel (which may be of any type), the sender can send some packets of larger size than necessary; the recipient can then figure out if this increased size led to increased packet loss or delay, and can send back information about this to the sender.

The sender can then change its encoding parameters to produce smaller or larger packets; when in wideband mode, it can also switch between 30-ms and 60-ms mode.

In the particular case of one audio channel in each direction, both using iSAC, iSAC defines the BEI field as a feedback channel. The available bandwidth is continuously estimated at the receiving iSAC; the receiver will signal the sender in-band in the iSAC bit stream, using the BEI field, what its estimate is. If the sending iSAC is running in channel-adaptive mode, it will adjust its bitrate accordingly.

This specification does not specify any particular feedback mechanism for any other use case.

Note: This mechanism is only capable of reducing iSAC traffic to the lowest available setting for iSAC. If there is congestion that makes even less bandwidth available, other mechanisms, such as dropping the call, will have to be used to escape from the congestion situation.

5. IANA Considerations

This RTP payload format is identified using the media type audio/isac, which is registered in accordance with [RFC4855] and uses the template of [RFC6838].

Type name: audio

Subtype name: isac

Required parameters: None

Optional parameters:

* ibitrate: The parameter indicates the upper bound in bits per second of the initial target bit rate (counting only payload bits) the device would like to receive. A sender SHOULD set its initial target bitrate to a value less than or equal to this parameter. An acceptable value for ibitrate is in the range of

20000 to 32000 (bits per second). In the absence of the parameter, the sender can choose any value up to the maximum bitrate possible.

* maxbitrate: The parameter indicates the maximum bit rate the endpoint expects to receive. The recipient of this parameter SHOULD NOT transmit at a higher bit rate. The default maximum value is 53400 bits per second, which is the maximum bitrate possible for iSAC.

Encoding considerations:
This media format is framed and binary.

Security considerations: See Section 7

Interoperability considerations: None

Published specification: RFC XXXX

Applications which use this media type:
This media type is suitable for use in numerous applications needing to transport encoded voice or other audio. Some examples include Voice over IP, Streaming Media, Voice Messaging, and Conferencing.

Fragment identifier considerations The meaning of fragment identifiers is not defined by this specification.

Additional information: None

Person to contact for further information:
Tina le Grand [tlegrand@google.com]

Intended usage: COMMON

Other Information/General Comment:
iSAC is a speech and audio codec owned by Google. The codec operates on 30 or 60 ms speech frames at a sampling rate clock of 16 kHz or 32 kHz.

Restrictions on usage:
This media type depends on RTP framing, and hence is only defined for transfer via RTP [RFC3550]. Transport within other framing protocols is not defined at this time.

Author Tina Le Grand and the listed authors of RFC XXXX

Change controller: The IETF Payload working group delegated from the IESG.

Provisional registration? No

Note to the RFC Editor / IANA: Please replace "RFC XXXX" above with the number of this RFC when published, and remove this note.

6. 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 iSAC codec, the mapping is as follows:

- The media type ("audio") goes in SDP "m=" as the media name.
- The media subtype (payload format name) goes in SDP "a=rtpmap" as the encoding name.
- The clock rate is 16000 for wideband, and 32000 for superwideband.
- Any remaining parameters go in the SDP "a=fmtp" attribute by copying them directly from the media type string as a semicolon separated list of parameter=value pairs.

The optional parameter ibitrate MUST NOT be higher than the parameter maxbitrate.

The iSAC parameters in an SDP offer are completely independent from those in the SDP answer. For both ibitrate and maxbitrate it is legal for the answer to contain a value that is different than what is provided in an offer. The parameter may be present in the answer, even if absent in the offer.

When conveying information by SDP, the encoding name SHALL be "isac" (the same as the media subtype).

6.1. Example Initial Target Bit Rate

The offer indicates that it wishes to receive a wideband bitstream with an initial target rate of 20000 bits per second. The remote party should change its initial target rate to the requested value or less.

```
m=audio 10000 RTP/AVP 98
a=rtpmap: 98 isac/16000
a=fmtp:98 ibitrate=20000
```

6.2. Example Max Bit Rate

The offer indicates that it wishes to receive a superwideband bitstream with an initial target rate of 20000 bits per second, and a maximum bit rate of 45000 bits per second. The remote party should change its initial target rate to 20000 bits per second or less, and should not transmit at a higher rate than 45000.

```
m=audio 10000 RTP/AVP 98
a=rtpmap: 98 isac/32000
a=fmtp:98 ibitrate=20000;maxbitrate=45000
```

6.3. Example with both WB and SWB offered

This offer indicates willingness to receive both wideband and superwideband iSAC encodings, with default values for ibitrate and bitrate. Superwideband is preferred.

```
m=audio 10000 RTP/AVP 98 99
a=rtpmap: 98 isac/32000
a=rtpmap: 99 isac/16000
```

7. Security Considerations

RTP packets using the payload format defined in this specification are subject to the general security considerations discussed in RFC 3550 section 8.1.

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

Since iSAC is a variable rate codec, the attack using the length of encoded packets described in [RFC6562] is of interest. When using RTP for transport, the padding approach described in that document is usable; when such padding is not available or not feasible, the iSAC padding mechanism can be used to the same effect.

8. Acknowledgments

Special thanks to Roni Even for his thorough review of the document, and to Colin Perkins for additional review.

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

The present version is prepared using xml2rfc and xxe-xml2rfc.

9. References

9.1. Normative References

[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", RFC 3550, July 2003.

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

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

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

[RFC6838] Freed, N., Klensin, J., and T. Hansen, "Media Type Specifications and Registration Procedures", BCP 13, RFC 6838, January 2013.

9.2. Informative References

[RFC6562] Perkins, C. and JM. Valin, "Guidelines for the Use of Variable Bit Rate Audio with Secure RTP", RFC 6562, March 2012.

[iSAC] GIPS / Google, "iSAC reference implementation".

Available at http://code.google.com/p/webrtc/source - directory src/modules/audio_coding/codecs/isac

Authors' Addresses

Tina le Grand
Google
Kungsbron 2
Stockholm, 11122
Sweden

Paul E. Jones
Cisco Systems
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
USA

Phone: +1 919 476 2048
Fax:
Email: paulej@packetizer.com
URI:

Pascal Huart
Cisco Systems
400, Avenue Roumanille, Batiment T3
Biot - Sophia Antipolis, 06410
France

Phone: +33 4 9723 2643
Fax:
Email: phuart@cisco.com
URI:

Turaj Zakizadeh Shabestary
Google
1950 Charleston Road
Mountain View, CA 94043
USA

Phone:
Fax:
Email: turajs@google.com
URI:

Harald Alvestrand (editor)
Google
Kungsbron 2
Stockholm, 11122
Sweden

Phone:
Fax:
Email: hta@google.com
URI: