

Audio/Video Transport Working Group
Internet Draft
Intended status: Informational

SPIRIT DSP
April 14, 2009

RTP Payload Format for IP-MR Speech Codec [draft-ietf-avt-rtp-ipmr-03.txt](#)

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Copyright (c) 2009 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 in effect on the date of publication of this document (<http://trustee.ietf.org/license-info>). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

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/1id-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on October 14, 2009.

Abstract

This document specifies the payload format for packetization of SPIRIT IP-MR encoded speech signals into the Real-time Transport Protocol (RTP). The payload format supports transmission of multiple frames per payload and introduced redundancy for robustness against packet loss.

Table of Contents

1.	Introduction.....	3
2.	P-MR Codec Description.....	3
3.	Payload Format.....	4
3.1.	Payload Format Structure.....	4
3.2.	Payload Header.....	4
3.3.	Speech Table of Contents.....	5
3.4.	Speech Data.....	6
3.5.	Redundancy Header.....	6
3.6.	Redundancy Table of Contents.....	7
3.7.	Redundancy Data.....	7
4.	Payload Examples.....	7
4.1.	Payload Carrying a Single Frame.....	7
4.2.	Payload Carrying Multiple Frames with Redundancy.....	8
5.	Media Type Registration.....	9
5.1.	Registration of media subtype audio/ip-mr_v2.5.....	9
5.2.	Mapping Media Type Parameters into SDP.....	10
6.	Security Considerations.....	11
7.	IANA Considerations.....	11
8.	Normative References.....	11
9.	Author's Information	11
10.	Disclaimer.....	12
11.	Legal Terms.....	12

1. Introduction

This document specifies the payload format for packetization of SPIRIT IP-MR encoded speech signals into the Real-time Transport Protocol (RTP). The payload format supports transmission of multiple frames per payload and introduced redundancy for robustness against packet loss.

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](#) [[RFC 2119](#)].

2. IP-MR Codec Description

The IP-MR codec is scalable adaptive multi-rate wideband speech codec designed by SPIRIT for use in IP based networks. These codec is suitable for real time communications such as telephony and videoconferencing.

The codec operates on 20 ms frames at 16 kHz sampling rate and has an algorithmic delay of 25ms.

The IP-MR supports six wide band speech coding modes with respective bit rates ranging from about 7.7 to about 34.2 kbps. The coding mode can be changed at any 20 ms frame boundary making possible to dynamically adjust the speech encoding rate during a session to adapt to the varying transmission conditions.

The coded frame consists of multiple coding layers-base (or core) layer and several enhancement layers which are coded independently. These enhancement layers can be omitted and remaining base layer can be meaningfully decoded without artifacts. This making bit stream scalable and allows reduce bit rate during transmission without re-encoding.

This memo specifies an optional form of redundancy coding within RTP for protection against packet loss. It is based on commonly known scheme when previously transmitted frames are aggregated together with new ones. Each frame is retransmitted once in the following RTP payload packet. $f(n-2) \dots f(n+4)$ denote a sequence of speech frames, and $p(n-1) \dots p(n+4)$ a sequence of payload packets:

```

--+-+-----+-----+-----+-----+-----+-----+-----+-----+
  | f(n-2) | f(n-1) | f(n)  | f(n+1) | f(n+2) | f(n+3) | f(n+4) |
--+-+-----+-----+-----+-----+-----+-----+-----+-----+

    <---- p(n-1) ---->
      <----- p(n) ----->
        <---- p(n+1) ---->
          <---- p(n+2) ---->
            <---- p(n+3) ---->
              <---- p(n+4) ---->

```


But because of scalable nature of IP-MR codec there is no need to duplicate a whole previous frame - only core layer may be retransmitted. This reduces redundancy overhead while keeping efficiency. Moreover, the speech bits encoded in core layer are divided on six classes (from A to F) of perceptual sensitivity to errors. Using these classes as introduced redundancy make possible to adjust trade-off between overhead and robustness against packet loss.

The mechanism described does not really require signaling at the session setup. The sender is responsible for selecting an appropriate amount of redundancy based on feedback about the channel conditions.

The main codec characteristics can be summarized as follows:

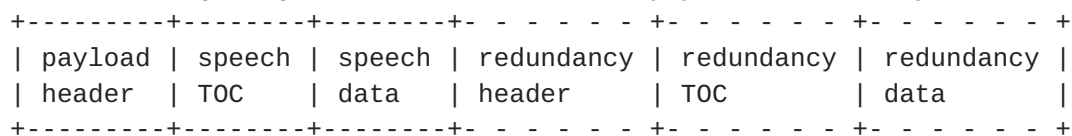
- o Wideband, 16 kHz, speech codec
- o Adaptive multi rate with six modes from about 7.7 to about 34.2 kbps
- o Bit rate scalable
- o Variable bit rate changing in accordance with actual speech content
- o Discontinuous Transmission (DTX), silence suppression and comfort noise generation
- o In-band redundancy scheme for protection against packet loss

3. Payload Format

3.1. Payload Format Structure

The IP-MR payload format consists of a payload header with general information about packet, a speech table of contents (TOC), and speech data. An optional redundancy section follows after speech data. The redundancy section consists of redundancy header, redundancy TOC and redundancy data payload.

The following diagram shows the standard payload format layout:



3.2. Payload Header

The payload header has the following format:

0

1

0 1 2 3 4 5 6 7 8 9 0 1
+--+--+--+--+--+--+--+--+--+--+
|T| CR | BR |D|A|GR |R|
+--+--+--+--+--+--+--+--+--+--+

- o T (1 bit): Reserved compatibility with future extensions. SHOULD be set to 0.

- o CR (3 bits): coding rate of frame(s) in this packet, as per the following table:

CR	avg. bitrate
0	7.7 kbps
1	9.8 kbps
2	14.3 kbps
3	20.8 kbps
4	27.9 kbps
5	34.2 kbps
6	(reserved)
7	NO_DATA

The CR value 7 (NO_DATA) indicates that there is no speech data (and speech TOC accordingly) in the payload. This MAY be used to transmit redundancy data only. The value 6 is reserved. If receiving this value the packet SHOULD be discarded.

- o BR (3 bits): base rate for core layer of frame(s) in this packet. Values in the range 0-5 indicate bitrates for core layer, same as for packet SHOULD be discarded. The base rate is the lowest rate for scalability, so speech payload can be scaled down not lower than BR value. If a received packet has $BR > CR$ then during decoding it will be assumed that $BR = CR$.
- o D (1 bit): indicates if the DTX mode is allowed or not.
- o A (1 bit): byte-aligned payload. If A=1 then all speech frames MUST be byte-aligned. This mode speeds up speech data access. The A=0 value specifies bandwidth-efficient mode with no byte alignment(including end of header).
- o GR (2 bits): number of frames in packet (grouping size). Actual grouping size is $GR + 1$, thus maximum grouping supported is 4.
- o R (1 bit): redundancy presence bit. If R=1 then the packet contains redundancy information for lost packets recovery. In this case after speech data the redundancy section is present.

3.3. Speech Table of Contents

The speech TOC contains entries for each frame in packet (grouping size in total). Each entry contains a single field:

0
+-+
|E|
+-+

SPIRIT DSP

Expires October 14, 2009

[Page 5]

- o E (1 bit): frame existence indicator. If set to 0, this indicates the corresponding frame is absent and the receiver should set special LOST_FRAME flag for decoder. This can be followed by the lost frame itself or by empty frames generated by the encoder during silence intervals in DTX mode.

Note that if CR flag from payload header is 7 (NO_DATA) then speech TOC is empty.

3.4. Speech Data

Speech data of a payload contains one or more speech frames or comfort noise frames, as specified in the speech TOC of the payload.

Each speech frame represents 20 ms of speech encoded with the rate indicated in the CR and base rate indicated in BR field of the payload header. The length of the speech frame is variable due to the nature of the codec and can be calculated after decoding.

3.5. Redundancy Header

If a packet contains redundancy (R field of payload header is 1) the speech data is followed by redundancy header:

```

0 1 2 3 4 5
+--+--+--+--+
| CL1 | CL2 |
+--+--+--+--+

```

Redundancy header consists of two fields. Each field contains class specifier for amount of redundancy partly taken from the preceding packet (CL1) and pre-preceding packet (CL2), e.g. distant from the current packet by 1 and 2 packets accordingly. The values are listed in the table below:

CL	amount redundancy
0	NONE
1	CLASS A
2	CLASS B
3	CLASS C
4	CLASS D
5	CLASS E
6	CLASS F
7	(reserved)

Each specifier takes 3 bits, thus the total redundancy header size is 6 bits.

3.6. Redundancy Table of Contents

```

+--+--+--+--+--+--+--+--+--+--+
| Pkt1 Entries| Pkt2 Entries|
+--+--+--+--+--+--+--+--+--+--+

```

The redundancy TOC contains entries for redundancy frames from preceding and pre-preceding packets. Each entry takes 1 bit like speech TOC entry (3.3):

```

0
+--+
|E|
+--+

```

- o E (1 bit): frame existence indicator. If set to 0, this indicates the corresponding frame is absent.
- o For each preceding and pre-preceding packet the number of entries is equal to the grouping size of the current packet. E.g. maximum number of entries is $4 \times 2 = 8$.
- o If class specifier in the redundancy header is CL=0 (NO_DATA) then there is no entries for corresponding packet redundancy.

3.7. Redundancy Data

Redundancy data of a payload contains redundancy information for one or more speech frames or comfort noise frames that may be lost during transition, as specified in the redundancy TOC of the payload. Actually redundancy is the most important part of preceding frames representing **20 ms of speech. This data MAY be used for partial reconstruction of** lost frames. The amount of available redundancy is specified by CL flag in redundancy header section (3.5). This flag SHOULD be passed to decoder. The length of redundancy frame is variable and can be 281 calculated after decoding.

4. Payload Examples

A few examples to highlight the payload format follow.

4.1. Payload Carrying a Single Frame

The following diagram shows a standard IP-MR payload carrying a single

speech frame without redundancy:

SPIRIT DSP

Expires October 14, 2009

[Page 7]

In the payload the speech frame is not damaged at the IP origin (E=1), the coding rate is 9.7 kbps (CR=1), the base rate is 7.8 kbps (BR=0), and the DTX mode is off. There is no byte alignment (A=0) and no redundancy (R=0). The encoded speech bits - s(0) to s(193) - are placed immediately after TOC. Finally, one zero bit is added at the end as padding to make the payload byte aligned.

The following diagram shows a payload that contains three frames, one of them with no speech data. The coding rate is 7.7 kbps (CR=0), the base rate is 7.7 kbps (BR=0), and the DTX mode is on. The speech frames are byte aligned (A=1), so 1 zero bit is added at the end of the header. Besides the speech frames the payload contains six redundancy frames (three per each delayed packet).

The redundancy header follows after speech data. The one-packet- delayed redundancy contains class A+B bits (CL1=2), and two-packet- delayed redundancy contains class A bits (CL2=1). The one-packet- delayed redundancy contains three frames with 20, 39 and 35 bits respectively. The first frame of two-packet-delayed redundancy is absent, it is represented in its TOC entry, and two other frames have sizes 15 and 19 bits.

Note that all speech frames are padded with zero bits for byte alignment.

5. Media Type Registration

5.1. Registration of media subtype audio/ip-mr_v2.5

Optional parameters:

* ptime: Gives the length of time in milliseconds represented by the media in a packet. Allowed values are: 20, 40, 60 and 80.

Encoding considerations: This media type is framed binary data (see RFC 4288, [Section 4.8](#)).

Security considerations: See [RFC 3550](#) [[RFC 3550](#)]

Interoperability considerations: none

Published specification: RFC XXXX

Applications that use this media type: Real-time audio applications like voice over IP and teleconference, and multi-media streaming.

Additional information: none

Person & email address to contact for further information:

Elena Berlizova

berlizova@spiritdsp.com

Intended usage: COMMON

Restrictions on usage: This media type depends on RTP framing, and hence is only defined for transfer via RTP [[RFC 3550](#)].

Author:

Sergey Ikonin <ikonin@spiritdsp.com>

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

[5.2. Mapping Media Type Parameters into SDP](#)

The information carried in the media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [[RFC 4566](#)], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing the IP-MR codec, the mapping is as follows:

- o The media type ("audio") goes in SDP "m=" as the media name.
- o The media subtype (payload format name) goes in SDP "a=rtpmap" as the encoding name. The RTP clock rate in "a=rtpmap" MUST 16000.
- o The parameter "ptime" goes in the SDP "a=ptime" attributes.

Any remaining parameters go in the SDP "a=fmtp" attribute by copying them directly from the media type parameter string as a semicolon-separated list of parameter=value pairs.

Note that the payload format (encoding) names are commonly shown in

upper case. Media subtypes are commonly shown in lower case. These names are case-insensitive in both places.

6. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [[RFC 3550](#)], and any appropriate RTP profile. This implies that confidentiality of the media streams is achieved by encryption. Encryption may be performed after compression so there is no conflict between the two operations.

This payload format does not exhibit any significant non-uniformity in the receiver side computational complexity for packet processing, and thus is unlikely to pose a denial-of-service threat due to the receipt of pathological data.

7. IANA Considerations

One media type has been defined and needs registration in the media types registry.

8. Normative References

- [RFC 2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC 3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, [RFC 3550](#), July 2003.
- [RFC 4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", [RFC 4566](#), July 2006.

9. Author(s) Information

Sergey Ikonin email: ikonin@spiritdsp.com

Russia 109004
Building 27, A. Solgenizyn street
Tel: +7 495 661-2178
Fax: +7 495 912-6786

10. Disclaimer

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, **2008**. **The person(s) controlling the copyright in some of this material** may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

11. Legal Terms

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST 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 THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

The IETF Trust 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 any IETF 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.

Copies of Intellectual Property 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 any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that

published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of [RFC 5378](#). No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under [RFC 5378](#), shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

