

Audio/Video Transport Working Group
Internet
Draft
SPIRIT DSP
Intended status:
Informational
February 10, 2009

RTP Payload Format for SPIRIT IP-MR Speech Codec Software [draft-ietf-avt-rtp-ipmr-01.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 (<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.

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 August 10, 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, introduced redundancy for robustness against packet loss, and payload format extension for future versions compatibility.

Table of Contents

1. Introduction	2
2. IP-MR RTP Payload Formats	2
2.1. Standard Payload Format	3
2.1.1. Payload Format Structure	3
2.1.2. Payload Header	3
2.1.3. Speech Table of Contents	4
2.1.4. Speech Data	4
2.1.5. Redundancy Header	5
2.1.6. Redundancy Table of Contents	5
2.1.7. Redundancy Data	6
2.2. Payload Examples	6
2.2.1. Standard Payload Carrying a Single Frame	6
2.2.2. Standard Payload Carrying Multiple Frames with Redundancy	7
2.2.3. Extended Payload Carrying a Single Frame	8
3. Media Type Registration	8
3.1. Registration of MIME media type audio/ip-mr_v2.5	8
3.2. Mapping Media Type Parameters into SDP	9
4. Security Considerations	10
6. Normative References	10
Author's Addresses	10
Expiration date	10
Legal Terms	10

[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, introduced redundancy for robustness against packet loss, and payload format extension for future versions compatibility.

[2. IP-MR RTP Payload Formats](#)

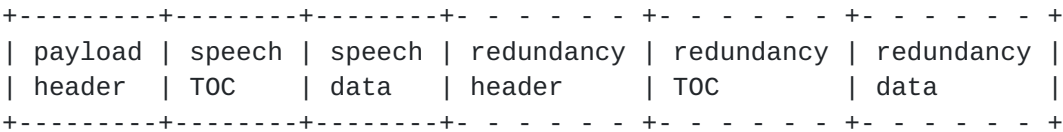
The payload has two formats: standard optimized for current use-cases and extended for future versions compatibility. The payload format is defined by first bit of header. Both of these formats will be described bellow.

2.1. Standard Payload Format

2.1.1. Payload Format Structure

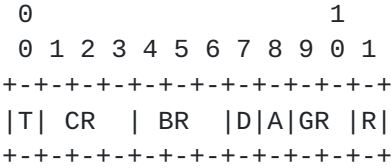
The standard payload consists of a payload header with general information about packet, a speech table of contents (TOC), and speech data. An optional redundancy section follows after speech data. The redundancy section consists of redundancy header, redundancy TOC and redundancy data payload.

The following diagram shows the standard payload format layout:



2.1.2. Payload Header

The payload header has the following format:



- 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
+-----+-----+	

Table 1 Coding rates of IP-MR codec

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 CR. Values 6 and 7 are reserved. If one of these values is received the 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. If greater grouping size is required the extended payload format (sec. 2.2) MAY be used.

- 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 TOC redundancy flags and TOC sections are present. If R=0 then speech TOC is the last section of payload header.

2.1.3. 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|
+--+
```

- o E (1 bit): frame existence indicator. If set to 0, this indicates the corresponding frame is absent and the receiver should set the `teRxFrType` to `LOST_FRAME`. 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 field from coding flags is 7 (`NO_DATA`) then speech TOC is empty.

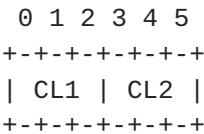
2.1.4. Speech Data

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 or by using `GetFrameInfo` function detailed in [1].

2.1.5. Redundancy Header

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

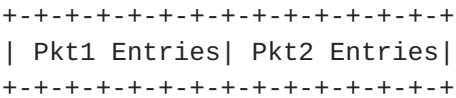


Redundancy header consists of two fields. Each field contains class specifier for redundancy partly taken from the preceding packet (CL1) and pre-preceding packet (CL2), e.g. distant from the current packet by 1 and 2 packets accordingly. The values are listed in the table below:

+-----+-----+	
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. In case of redundancy usage followed by preceding (or pre- preceding) packet loss the receiver sets the special flag for decoder with CL class specifier.

2.1.6. Redundancy Table of Contents



The redundancy TOC contains entries for redundancy frames from preceding and pre-preceding packets. Each entry takes 1 bit like speech TOC entry (2.1.3):



o E (1 bit): frame existence indicator. If set to 0, this indicates the corresponding frame is absent.

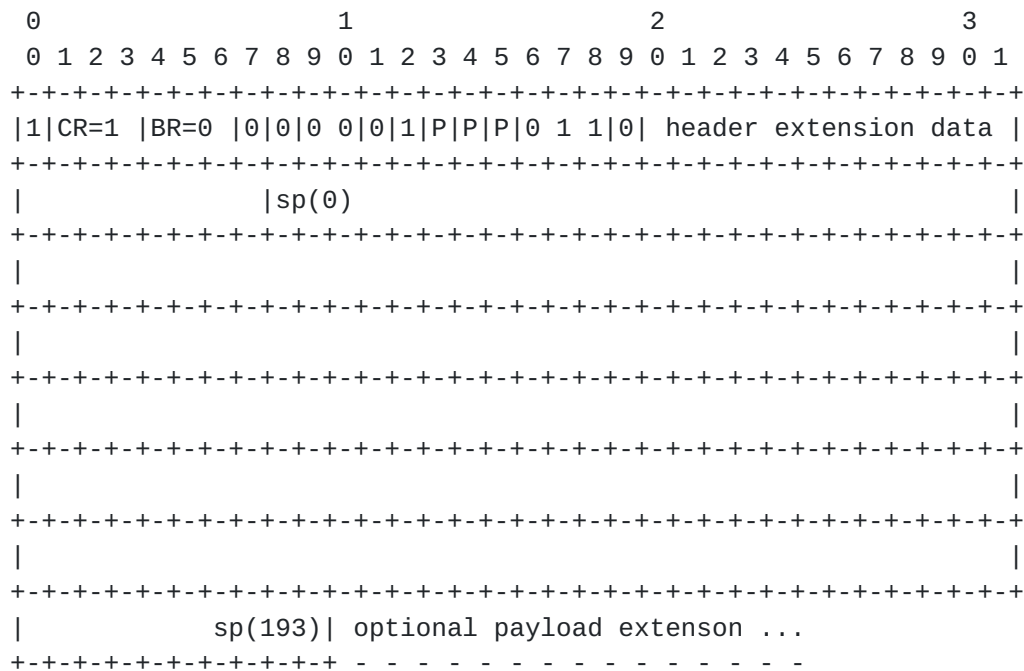
The following diagram shows a payload that contains three frames, one of them with no speech data. The coding rate is 7.7 kbps (CR=0), the base rate is 7.7 kbps (BR=0), and the DTX mode is on. The speech frames are byte aligned (A=1), so 1 zero bit is added at the end of the header. Besides the speech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

[illegible]

| red2_3(18)|P|P|P|P|
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

The following diagram shows an extended IP-MR payload carrying a single speech frame without redundancy:



The standard header is the same as in example 2.3.1 except for the first bit that is set to 1 to reflect extended payload type. The standard header is padded with zeros to achieve byte alignment. After that the size of header extension follows (HESZ=3). Then the header extension data is placed. In 3 bytes (HESZ) from header extension beginning, the standard speech payload starts. After that, the optional payload extension MAY be added.

This section describes the media types and names associated with this payload format.

Type name: audio

Subtype name: ip-mr_v2.5

Required parameters: none

Optional parameters:

o 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](#)

Applications that use this media type:

Audio and video streaming and conferencing tools.

Additional information: none

Intended usage: COMMON

Restrictions on usage:

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

3.2. 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) [[RFC4566](#)], 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:

- * 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 RTP clock rate in "a=rtpmap" MUST 16000.
- * The parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.

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.

4. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [[RFC3550](#)], 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.

6. Normative References

- [1] SPIRIT IP-MR v2.5 User Guide, website <http://spiritdsp.com>
- [2] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [3] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, [RFC 3550](#), July 2003.
- [4] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", [RFC 4566](#), July 2006.

Author's Addresses

SPIRIT DSP
Russia
[109004](#) B.Kommunisticheskaya st. 27
Tel: +7 495 661-2178
Fax: +7 495 912-6786
Email: Elena Berlizova berlizova@spiritdsp.com

Expiration date

This Internet-Draft will expire on August 09, 2009.

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.

