Audio/Video Transport Working Group
Internet
Draft
SPIRIT DSP
Intended status:
Informational
February 25, 2009

# RTP Payload Format for IP-MR Speech Codec draft-ietf-avt-rtp-ipmr-02.txt

## Status of this Memo

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

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.

## 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 25
ms.

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)...f(n+4) denote a sequence of speech frames, and p(n-1)...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:

* Wideband, 16 kHz, speech codec

* Adaptive multi rate with six modes from about 7.7 to about 34.2 kbps

* Bit rate scalable

* Variable bit rate changing in accordance with actual speech content

* Discontinuous Transmission (DTX), silence suppression and comfort noise generation

* 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:

```
+---------+--------+--------+- - - - - - +- - - - - - +- - - - - - +
| payload | speech | speech | redundancy | redundancy | redundancy |
| header  | TOC    | data   | header     | TOC        | data       |
+---------+--------+--------+- - - - - - +- - - - - - +- - - - - - +
```

### 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|
+-+-+-+-+-+-+-+-+-+-+-+-+
```

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

* 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.

* 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.

* D (1 bit): indicates if the DTX mode is allowed or not.

* 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).

* GR (2 bits): number of frames in packet (grouping size). Actual grouping size is GR + 1,
thus maximum grouping supported is 4.

* 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|
+-+
```

* 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|
  +-+
```

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

* 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*2 = 8.

* 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 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:

```
   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
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |0|CR=1 |BR=0 |0|0|0 0|0|1|sp(0)                                |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                                                               |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                                                               |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                                                               |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                                                               |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                                                               |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |                      sp(193)|P|
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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.

### 4.2. Payload Carrying Multiple Frames with Redundancy

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 first speech frame consists of bits sp1(0) to sp1(92). After that 3 bits are added for byte
alignment. The second frame does not contain any speech information that is represented in the
payload by its TOC entry. The third frame consists of bits sp3(0) to sp3(171).

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.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|CR=0 |BR=0 |1|1|1 0|1|1 0 1|P|sp1(0)                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  sp1(92)|P|P|P|sp3(0)                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                              sp3(171)|P|P|P|P|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|CL1=2|CL2=1|1 1 1|0 1 1|red1_1(0)                   red1_1(19)|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|red1_2(0)                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   red1_2(38)|red1_3(0)                                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         red1_3(34)|red2_2(0)          red2_2(14)|red2_3(0)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             red2_3(18)|P|P|P|P|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

## 5. Media Type Registration

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

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

Type name: audio

Subtype name:  ip-mr_v2.5

Required parameters:  none

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

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) [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 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 [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.

## 7. IANA Considerations

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

## 8. Normative References

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

## 9. Author's Information

Sergey Ikonin

Russia
**109004 B.Kommunisticheskaya st. 27**
Tel: +7 495 661-2178
Fax: +7 495 912-6786
Email: ikonin@spiritdsp.com

## 10. Expiration date

This Internet-Draft will expire on August 25, 2009.

## 11. Legal Terms

shall have any effect and shall be null and void, whether published or posted by such
Contributor, or included with or in such Contribution.