An RTP Payload Format for EVRC Speech

1. Introduction

This document describes how compressed EVRC speech as produced by the EVRC CODEC [1] may be formatted for use as an RTP payload type. A method is provided to interleave the output of the compressor to reduce quality degradation due to lost packets. Furthermore, the sender may choose various interleave settings based on the importance of low end-to-end delay versus greater tolerance for lost packets.

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 [3].

2. Background


The EVRC CODEC [1] compresses each 20 milliseconds of 8000 Hz, 16-bit sampled input speech into one of three different size output frames: Rate 1 (171 bits), Rate 1/2 (80 bits), or Rate 1/8 (16 bits).
The CODEC chooses the output frame rate based on analysis of the input speech and the current operating mode (either normal or one of several reduced rates). For typical speech patterns, this results in an average output of 4.2 K bits/sec for normal mode and lower for reduced rate modes.

3. RTP/EVRC Packet Format

The RTP timestamp is in 1/8000 of a second units. The RTP payload data for the EVRC CODEC the following two types.

3.1 Type 1 RTP/EVRC Packet Format

This format is for the situation that the sender and the receiver intending to uses interleaving and/or bundling to send one or more than one codec frames per packet. There are two formats of Type 1 packets.

3.1.1 Type 1 (Bundled Format)

The first format is a bundled format, where one or more codec data frames can be bundled and transmitted in one RTP packet. For this case, the RTP packet format is shown as follows.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      RTP Header [2]                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         one or more codec data frames         +-------+-+|
|                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The RTP header has the expected values as described in [2]. The extension bit is not set. The codec data frames are aligned on octet boundaries. When multiple codec data frames are present in a single RTP packet, the timestamp is, as always, that of the oldest data represented in the RTP packet.

3.1.2 Type 1 (Interleaved Format)

For the case where interleaving is in use and one or multiple codec data frames are present in a single RTP packet. The RTP packet for this format is as follows:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      RTP Header [2]                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|I|R| LLL | NNN |                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         one or more codec data frames         +-------+-+|
|                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The RTP header is the same as described in 3.1.1. The fields of the additional interleaving byte have the following meaning:

Interleave Disabled (I): 1 bit
MUST be set to zero by sender.

Reserved (RR): 2 bit
MUST be set to zero by sender, SHOULD be ignored by receiver.

Interleave (LLL): 3 bits
MUST have a value between 0 and 7 inclusive.

Interleave Index (NNN): 3 bits
MUST have a value less than or equal to the value of LLL. Values of NNN greater than the value of LLL are invalid.

3.1.3 Detection between the two Type 1 formats

The bundled and interleaved format of Type 1 packets can be distinguished at the receiver by detecting the presence of a 1 in the first bit of the RTP packet payload.

The interleaved format packets always have 0 at that bit from the RR bits above. The bundled format packets always have 1 at that bit from the first bit of the codec data frame header (see Section 4.1).

3.2 Type 2 RTP/EVRC Packet Format

The Type 2 RTP/EVRC Packet Format are designed for maximum efficiency in transmission of the EVRC codec data. Only one codec data frame is sent with each RTP packet, and there is no codec data frame header prefix the codec data. The EVRC codec rate of the data frame can be found out at the receiver from the length of the codec frame, since there is only one codec data frame in each RTP packet for this type.

3.3 Detection between the Type 1 and Type 2 packets

All receivers MUST be able to process both types of packets. The sender may choose to use one or both types of packets.

The packets of the two types can be distinguished by the payload type field in the RTP header. The association of payload type number with the packet type is done out-of-band, for example by SDP during the setup of a session.

4. CODEC data frame format

The codec data frame consists of a codec data frame header followed by the codec data. The codec data frame header will not be used when the codec data is transmitted by Type 2 RTP/EVRC packet.

4.1 Codec data frame header

The codec data frame header preceeds the codec data in the Type 1
RTP/EVRC packets. The header of the CODEC data frame indicates whether interleaving is present, if rate reduction is desired, and the rate of the codec frame. The format of the octet is indicated below:

```
  0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+
|I|D| frame type|
+-+-+-+-+-+-+-+
```

Interleaving Disabled (I): 1 bit

This bit indicates whether the interleaving byte is present. This bit MUST be set to 1 if the interleaving byte is missing (i.e., interleaving/bundling is not used), otherwise it MUST be set to 0. Note: if the first bit of the first RTP payload octet is zero this byte is the interleaving byte in the interleaved format of Type 1 (as described in 3.1.2), otherwise it is octet zero of the EVRC payload in the bundled format of Type 1.

Reduce Rate (D): 1 bit

Setting the ‘R’ bit indicates that this packet is requesting a reduced codec rate for the reverse direction. When the ‘R’ bit is not set the packet is requesting that the codec resume normal operation. In the case of packet loss the codec should continue to operate in the mode indicated by the last packet received. Receivers are not required to respond to the Reduce Rate signal. (See more discussion in Section 8.2).

Frame Type: 6 bits

The frame type values are described in the table below and the size of the associated packet is indicated in the table below:

<table>
<thead>
<tr>
<th>Value</th>
<th>RATE</th>
<th>TOTAL CODEC data frame size (in octets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Blank</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1/8</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1/2</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>14</td>
<td>Erasure</td>
<td>1 (SHOULD NOT be transmitted by sender)</td>
</tr>
</tbody>
</table>

Receipt of a CODEC data frame with a reserved value in octet 0 MUST be considered invalid data. All values not listed in the above table MUST be considered reserved.

4.2 The codec data

The output of the EVRC CODEC must be converted into CODEC data frames for inclusion in the RTP payload as follows:

The bits as numbered in the standard [1] from the lowest to the highest are packed into octets. The lowest numbered bit (bit 1 for Rate 1, Rate 1/2 and Rate 1/8) is placed in the most significant bit (Internet bit 0) of octet 1 of the CODEC data frame, the second lowest bit is placed in the second most significant bit of the first octet, the third lowest in the third most significant bit of the first octet, and so on. This continues until all of the bits have been placed in the CODEC data frame. The remaining unused bits of the last octet of the CODEC data frame MUST be set to zero (note that this is only applicable to rate 1 frames as the others fit completely into a
Here is a detail of how a Rate 1 frame is converted into a CODEC data frame:

Octet 0 of the data frame has value 4 (see table above) indicating the total data frame length (including octet 0) is 23 octets.

5. Bundling codec data frames in Type 1 packets

As indicated in section 3.1.1, more than one codec data frame MAY be included in a single RTP packet by a sender. Receivers may signal the maximum number of codec data frames they can handle in a single RTP packet.

Furthermore, senders have the following additional restrictions:

- MUST never bundle more more codec data frames in a single RTP packet than signaled by maxbundle in Section 9.
- SHOULD not bundle more codec data frames in a single RTP packet than will fit in the MTU of the RTP transport protocol. For the purpose of computing the maximum bundling value, all CODEC data frames should be assumed to have the Rate 1 size.

Since no count is transmitted as part of the RTP payload and the codec data frames have differing lengths, the only way to determine how many codec data frames are present in the RTP packet is to examine octet 0 of each codec data frame in sequence until the end of the RTP packet is reached.

6. Interleaving codec data frames in Type 1 packets

All receivers MUST support interleaving. Senders MAY support
interleaving.

Given a time-ordered sequence of output frames from the EVRC CODEC numbered 0..n, a bundling value B, and an interleave value L where n = B * (L+1) - 1, the output frames are placed into RTP packets as follows (the values of the fields LLL and NNN are indicated for each RTP packet):

First RTP Packet in Interleave group:
LLL=L, NNN=0
  Frame 0, Frame L+1, Frame 2(L+1), Frame 3(L+1), ... for a total of B frames

Second RTP Packet in Interleave group:
LLL=L, NNN=1
  Frame 1, Frame 1+L+1, Frame 1+2(L+1), Frame 1+3(L+1), ... for a total of B frames

This continues to the last RTP packet in the interleave group:

L+1 RTP Packet in Interleave group:
LLL=L, NNN=L
  Frame L, Frame L+L+1, Frame L+2(L+1), Frame L+3(L+1), ... for a total of B frames

Senders MUST transmit in timestamp-increasing order. Furthermore, within each interleave group, the RTP packets making up the interleave group MUST be transmitted in value-increasing order of the NNN field. While this does not guarantee reduced end-to-end delay on the receiving end, when packets are delivered in order by the underlying transport, delay will be reduced to the minimum possible.

Additionally, senders have the following restrictions:

- Once beginning a session with a given maximum interleaving value set by maxinterleave in Section 9, MUST NOT increase the interleaving value exceeding the maximum interleaving the value that is signaled.

- MAY change the interleaving value only between interleave groups.

6.1 Finding Interleave Group Boundaries

Given an RTP packet with sequence number S, interleave value (field LLL) L, and interleave index value (field NNN) N, the interleave group consists of RTP packets with sequence numbers from S-N to S-N+L inclusive. In other words, the Interleave group always consists of L+1 RTP packets with sequential sequence numbers. The bundling value for all RTP packets in an interleave group MUST be the same.

The receiver determines the expected bundling value for all RTP packets in an interleave group by the number of CODEC data frames bundled in the first RTP packet of the interleave group received. Note that this may not be the first RTP packet of the interleave group sent if packets are delivered out of order by the underlying transport.

On receipt of an RTP packet in an interleave group with other than the expected bundling value, the receiver MAY discard CODEC data frames off the end of the RTP packet or add erasure CODEC data frames to the end of the packet in order to manufacture a substitute packet.
with the expected bundling value. The receiver MAY instead choose to
discard the whole interleave group and play silence.

6.2 Reconstructing Interleaved Speech

Given an RTP sequence number ordered set of RTP packets in an
interleave group numbered 0..L, where L is the interleave value and B
is the bundling value, and CODEC data frames within each RTP packet
that are numbered in order from first to last with the numbers 1..B,
the original, time-ordered sequence of output frames from the CODEC
may be reconstructed as follows:

First L+1 frames:
- Frame 0 from packet 0 of interleave group
- Frame 0 from packet 1 of interleave group
- And so on up to...
- Frame 0 from packet L of interleave group

Second L+1 frames:
- Frame 1 from packet 0 of interleave group
- Frame 1 from packet 1 of interleave group
- And so on up to...
- Frame 1 from packet L of interleave group

And so on up to...

Bth L+1 frames:
- Frame B from packet 0 of interleave group
- Frame B from packet 1 of interleave group
- And so on up to...
- Frame B from packet L of interleave group

6.3 Receiving Invalid Interleaving Values

On receipt of an RTP packet with an invalid value of the LLL or NNN
field, the RTP packet MUST be treated as lost by the receiver for the
purpose of generating erasure frames as described in Section 7.

6.4 Additional Receiver Responsibility

Assume that the receiver has begun playing frames from an interleave
group. The time has come to play frame x from packet n of the
interleave group. Further assume that packet n of the interleave
group has not been received. As described in section 7, an erasure
frame will be sent to the EVRC CODEC.

Now, assume that packet n of the interleave group arrives before
frame x+1 of that packet is needed. Receivers SHOULD use frame x+1
of the newly received packet n rather than substituting an erasure
frame. In other words, just because packet n wasn’t available the
first time it was needed to reconstruct the interleaved speech, the
receiver SHOULD NOT assume it’s not available when it’s subsequently
needed for interleaved speech reconstruction.

7. Handling lost RTP packets

The EVRC CODEC supports the notion of erasure frames. These are
frames that for whatever reason are not available. When
reconstructing interleaved speech or playing back non-interleaved
speech, erasure frames MUST be fed to the EVRC CODEC for all of the
missing packets.

Receivers MUST use the timestamp clock to determine how many CODEC data frames are missing. Each CODEC data frame advances the timestamp clock EXACTLY 160 counts.

Since the bundling/interleaving value may vary, the timestamp clock is the only reliable way to calculate exactly how many CODEC data frames are missing when a packet is dropped.

Specifically when reconstructing interleaved speech, a missing RTP packet in the interleave group should be treated as containing B erasure CODEC data frames where B is the bundling value for that interleave group.

8. Implementation Issues

8.1 Interleaving Length

The EVRC CODEC interpolates the missing speech content when given an erasure frame. However, the best quality is perceived by the listener when erasure frames are not consecutive. This makes interleaving desirable as it increases speech quality when dropped packets are more likely.

On the other hand, interleaving can greatly increase the end-to-end delay. Where an interactive session is desired, the non-interleaved RTP payload type is recommended.

When end-to-end delay is not a concern, an interleaving value (field LLL) of 4 or 5 is recommended subject to MTU limitations.

The parameters maxbundle and maxinterleaving at the initial setup of the session guarantees that the receiver can allocate a well-known amount of buffer space at the beginning of the session that will be sufficient for all future reception in that session. Less buffer space may be required at some point in the future if the sender decreases the bundling value or interleaving value, but never more buffer space. This prevents the possibility of the receiver needing to allocate more buffer space (with the possible result that none is available).

8.2 Signaling of Reduce rate

The reduce rate signal requests reducing of the codec rate on the reverse direction. It is not required that all implementations to be able to react to the Reduce rate signal. If an implementation will react to the Reduce rate signal, it MUST be able to process/react to the D bit in Type 1 packets.

In additional, the Reduce rate signal may also be sent through non-RTP means, which is out of the scope of this specification.

9. The EVRC MIME Type Registration

The MIME-name for the EVRC codec is allocated from the IETF tree since EVRC is expected to be a widely used codec for voice-over-IP applications.
Media Type Name: audio

Media Subtype Name: EVRC

Required Parameters:

ptype: It is the type of the RTP/EVRC packets. The valid values are 1 or 2.

Optional parameters for RTP mode:

ptime: Defined as usual for RTP audio.

maxbundle: Maximum number of EVRC speech frames that can be bundled in one RTP packet for the type 1 packets (bundled). The bundling values used in the entire session should not exceed this maximum value. If not signalled, the default maxbundle value is 10.

maxinterleave: Maximum number for interleaving value. The interleaving values used in the entire session should not exceed this maximum value. If not signalled, the maxinterleave value is 5.

Optional parameters for storage mode: none

Encoding considerations for RTP mode: see Section 5 and Section 6 of this document.

Encoding considerations for storage mode: The EVRC speech frames are packed into consecutive compound EVRC payloads, see Section 5 and Section 6. The compound EVRC payloads must be stored in sequential order. Furthermore, missing frames and non-received frames during non-speech period must be encapsulated into a compound EVRC payload as blank frames or erasures. Each receiving entity that accepts this MIME type must be able to decode all EVRC coding modes.

Security considerations: see Section 11 "Security Considerations".

Public specification: this document.

Additional information for storage mode:

Magic number: none
File extensions: evc, EVC
Macintosh file type code: none
Object identifier or OID: none

Intended usage: COMMON. It is expected that many VoIP applications (as well as mobile applications) will use this type.

10. Mapping to SDP Parameters

Please note that this chapter applies to the RTP mode only.

Parameters are mapped to SDP [5] as usual.

Example usage in SDP:

\[ m = \text{audio 49120 RTP/AVP 97} \]
\[ a = \text{rtpmap:97 EVRC} \]
\[ a = \text{fmt:97 ptype = 1 maxbundle = 4} \]
11. Security Considerations

RTP packets using the payload format defined in this specification
are subject to the security considerations discussed in the RTP
specification [2], and any appropriate profile (for example [4]).
This implies that confidentiality of the media streams is achieved by
encryption. Because the data compression used with this payload
format is applied end-to-end, encryption may be performed after
compression so there is no conflict between the two operations.

A potential denial-of-service threat exists for data encodings using
compression techniques that have non-uniform receiver-end
computational load. The attacker can inject pathological datagrams
into the stream which are complex to decode and cause the receiver to
be overloaded. However, this encoding does not exhibit any
significant non-uniformity.

As with any IP-based protocol, in some circumstances, a receiver may
be overloaded simply by the receipt of too many packets, either
desired or undesired. Network-layer authentication may be used to
discard packets from undesired sources, but the processing cost of
the authentication itself may be too high. In a multicast
environment, pruning of specific sources may be implemented in
future versions of IGMP [6] and in multicast routing protocols to
allow a receiver to select which sources are allowed to reach it.

12. Acknowledgements

The editor thanks the following authors for contributions to this
document: J. D. Villasenor, D.S. Park, J.H. Park, K. Miller,
S. C. Greer, D. Leon, N. Leung, K. J. McKay, M. Lioy, T. Hiller,
P. J. McCann, M. D. Turner, A. Rajkumar, Dan Gal, Magnus Westerlund,
Lars-Erik Jonsson, Greg Sherwood, and Thomas Zeng.

13. References

Option 3 for Wideband Spread Spectrum Digital Systems", January
1997.

[2] Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson,
1889, January 1996.

[3] Bradner, S., "Key words for use in RFCs to Indicate Requirement

[4] Schulzrinne, H., "RTP Profile for Audio and Video Conferences
with Minimal Control", RFC 1890, January 1996.

RFC 2327, April 1998.

1112, August 1989.

14. Authors’ Address
Adam H. Li
Image Communication Lab
Electrical Engineering Department
University of California
Los Angeles, CA 90095
USA
Phone: +1 310 825 5178
EMail: adamli@icsl.ucla.edu

John D. Villasenor
Image Communication Lab
Electrical Engineering Department
University of California
Los Angeles, CA 90095
USA
Phone: +1 310 825 0228
EMail: villa@icsl.ucla.edu

Dong-Seek Park
Samsung Electronics
Suwon, Kyungki 442-742
Korea
Phone: +82 31 200 3674
Email: dspark@samsung.com

Jeong-Hoon Park
Samsung Electronics
Suwon, Kyungki 442-742
Korea
Phone: +82 31 200 3747
Email: dspark@samsung.com

Keith Miller
Nokia
6000 Connection Drive
Irving, Texas 75039
USA
Phone: +1 972 894 4296
Email: keith.miller@nokia.com

S. Craig Greer
Nokia
6000 Connection Drive
Irving, Texas 75039
USA
Phone: +1 972 894 4867
Email: craig.greer@nokia.com

David Leon
Nokia
6000 Connection Drive
Irving, Texas 75039
USA
Phone: +1 972 374 1860
Email: david.leon@nokia.com

Marcello Lioy
QUALCOMM, Incorporated
5775 Morehouse Drive
San Diego, CA 92121
USA
Nikolai Leung  
QUALCOMM, Incorporated  
7710 Takoma Ave.  
Takoma Park, MD 20912  
USA  
Phone: +1 703 346 8351  
Email: nleung@qualcomm.com

Kyle J. McKay  
QUALCOMM, Incorporated  
5775 Morehouse Drive  
San Diego, CA 92121-1714  
USA  
Phone: +1 858 587 1121  
EMail: kylem@qualcomm.com

Tom Hiller  
Lucent Technologies  
Room 2F-218  
263 Shuman Drive  
Naperville, IL 60137  
USA  
Phone: +1 630 979 7673  
Email: tom.hiller@lucent.com

Peter J. McCann  
Lucent Technologies  
Room 2Z-305  
263 Shuman Drive  
Naperville, IL 60137  
USA  
Phone: +1 630 713 9359  
Email: mccap@lucent.com

Michael D. Turner  
Lucent Technologies  
Room 2A-203  
67 Whippany Rd  
Whippany, NJ 07981  
USA  
Phone: +1 973 386 3579  
Email: mdturner@lucent.com

Ajay Rajkumar  
Lucent Technologies  
Room 1A-235  
67 Whippany Rd  
Whippany, NJ 07981  
USA  
Phone: +1 973 386 5249  
Email: ajayrajkumar@lucent.com

Dan Gal  
Lucent Technologies  
67 Whippany Rd  
Whippany, NJ 07981  
USA  
Phone: +1 973 428 7734
Email: dgal@lucent.com

Magnus Westerlund
Ericsson Research
Ericsson Radio Systems AB
Torshamnsgatan 23
SE-164 80 Stockholm
Sweden
Phone: +46 8 4048287
Email: magnus.westerlund@ericsson.com

Lars-Erik Jonsson
Ericsson Erisoft AB
Box 920
SE-971 28 Luleå
Sweden
Phone: +46 920 20 21 07
Email: lars-erik.jonsson@ericsson.com

Greg Sherwood
PacketVideo Corporation
4820 Eastgate Mall
San Diego, CA 92121
USA
Email: sherwood@packetvideo.com

Thomas Zeng
PacketVideo Corporation
4820 Eastgate Mall
San Diego, CA 92121
USA
Email: zeng@packetvideo.com