                RTP Payload Format for PureVoice(tm) Audio

Status of this Memo

   This document specifies an Internet standards track protocol for the
   Internet community, and requests discussion and suggestions for
   improvements. Please refer to the current edition of the "Internet
   Official Protocol Standards" (STD 1) for the standardization state
   and status of this protocol. Distribution of this memo is unlimited.

ABSTRACT

   This document describes the RTP payload format for PureVoice(tm)
   Audio. The packet format supports variable interleaving to reduce
   the effect of packet loss on audio quality.

1 Introduction

   This document describes how compressed PureVoice audio as produced by
   the Qualcomm PureVoice 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
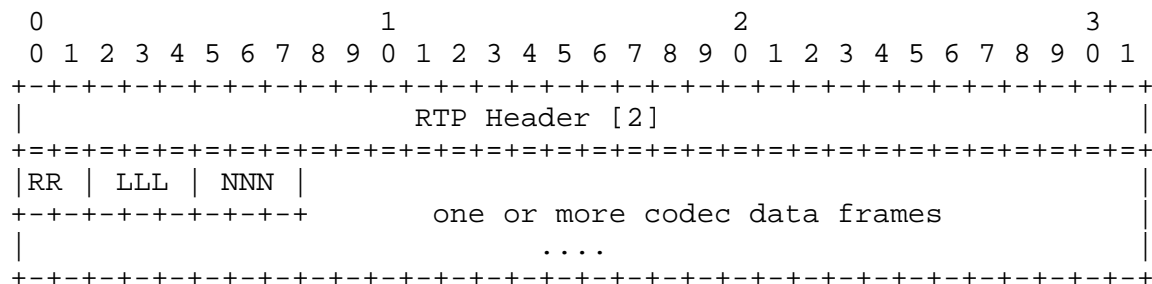   document are to be interpreted as described in RFC 2119 [3].

2 Background

   The Electronic Industries Association (EIA) & Telecommunications
   Industry Association (TIA) standard IS-733 [1] defines an audio
   compression algorithm for use in CDMA applications. In addition to
   being the standard CODEC for all wireless CDMA terminals, the
   Qualcomm PureVoice CODEC (a.k.a. Qcelp) is used in several Internet
   applications most notably JFax(tm), Apple(r) QuickTime(tm), and
   Eudora(r).

   The Qcelp CODEC [1] compresses each 20 milliseconds of 8000 Hz, 16-
   bit sampled input speech into one of four different size output
   frames: Rate 1 (266 bits), Rate 1/2 (124 bits), Rate 1/4 (54 bits)
   or Rate 1/8 (20 bits). The CODEC chooses the output frame rate based
   on analysis of the input speech and the current operating mode
   (either normal or reduced rate). For typical speech patterns, this
   results in an average output of 6.8 k bits/sec for normal mode and
   4.7 k bits/sec for reduced rate mode.

3 RTP/Qcelp Packet Format

   The RTP timestamp is in 1/8000 of a second units. The RTP payload
   data for the Qcelp CODEC has the following format:

```
    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]                           |
   +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
   |RR | LLL | NNN |                                               |
   +-+-+-+-+-+-+-+-+        one or more codec data frames          |
   |                           ....                                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   The RTP header has the expected values as described in [2]. The
   extension bit is not set and this payload type never sets the marker
   bit. The codec data frames are aligned on octet boundaries. When
   interleaving is in use and/or 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. The other fields have the
   following meaning:

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

   Interleave (LLL): 3 bits
     MUST have a value between 0 and 5 inclusive. The remaining two
     values (6 and 7) MUST not be used by senders. If this field is
     non-zero, interleaving is enabled. All receivers MUST support
     interleaving. Senders MAY support interleaving. Senders that do
     not support interleaving MUST set field LLL and NNN to zero.

   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 Receiving Invalid 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 4.

3.2 CODEC data frame format

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

   a. Octet 0 of the CODEC data frame indicates the rate and total size
      of the CODEC data frame as indicated in this table:

OCTET 0	RATE	TOTAL CODEC data frame size (in octets)
0	Blank	1
1	1/8	4
2	1/4	8
3	1/2	17
4	1	35
5	reserved	8 (SHOULD be treated as a reserved value)
14	Erasure	1 (SHOULD NOT be transmitted by sender)
other	n/a	reserved

      Receipt of a CODEC data frame with a reserved value in octet 0
      MUST be considered invalid data as described in 3.1.

   b. The bits as numbered in the standard [1] from highest to lowest
      are packed into octets. The highest numbered bit (265 for Rate 1,
      123 for Rate 1/2, 53 for Rate 1/4 and 19 for Rate 1/8) is placed
      in the most significant bit (Internet bit 0) of octet 1 of the
      CODEC data frame. The second highest numbered bit (264 for Rate
      1, etc.) is placed in the second most significant bit (Internet
      bit 1) of octet 1 of the data frame. This continues so that bit
      258 from the standard Rate 1 frame is placed in the least
      significant bit of octet 1. Bit 257 from the standard is placed
      in the most significant bit of octet 2 and so on until bit 0 from
      the standard Rate 1 frame is placed in Internet bit 1 of octet 34
      of the CODEC data frame. The remaining unused bits of the last
      octet of the CODEC data frame MUST be set to zero.

Here is a detail of how a Rate 1/8 frame is converted into a CODEC data frame:

CODEC data frame

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               |1|1|1|1|1|1|1|1|1|1| | | | | | | | | | | | | | |
| 1 (Rate 1/8)  |9|8|7|6|5|4|3|2|1|0|9|8|7|6|5|4|3|2|1|0|Z|Z|Z|Z|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Octet 0 of the data frame has value 1 (see table above) indicating the total data frame length (including octet 0) is 4 octets. Bits 19 through 0 from the standard Rate 1/8 frame are placed as indicated with bits marked with "Z" being set to zero. The Rate 1, 1/4 and 1/2 standard frames are converted similarly.

3.3 Bundling CODEC data frames

   As indicated in section 3, more than one CODEC data frame MAY be included in a single RTP packet by a sender. Receivers MUST handle bundles of up to 10 CODEC data frames in a single RTP packet.

   Furthermore, senders have the following additional restrictions:

   o MUST 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.

   o MUST never bundle more than 10 CODEC data frames in a single RTP packet.

   o Once beginning transmission with a given SSRC and given bundling value, MUST NOT increase the bundling value. If the bundling value needs to be increased, a new SSRC number MUST be used.

   o MAY decrease the bundling value only between interleave groups (see section 3.4). If the bundling value is decreased, it MUST NOT be increased (even to the original value), although it may be decreased again at a later time.

3.3.1 Determining the number of bundled CODEC data frames

   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.

3.4 Interleaving CODEC data frames

   Interleaving is meaningful only when more than one CODEC data frame
   is bundled into a single RTP packet.

   All receivers MUST support interleaving. Senders MAY support
   interleaving.

   Given a time-ordered sequence of output frames from the Qcelp 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:

o Once beginning transmission with a given SSRC and given interleave
   value, MUST NOT increase the interleave value. If the interleave
   value needs to be increased, a new SSRC number MUST be used.

o MAY decrease the interleave value only between interleave groups.
   If the interleave value is decreased, it MUST NOT be increased
   (even to the original value), although it may be decreased again
   at a later time.

3.5 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.

3.6 Reconstructing Interleaved Audio

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

3.6.1 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 4, an erasure
    frame will be sent to the Qcelp 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 audio, the
    receiver SHOULD NOT assume it's not available when it's subsequently
    needed for interleaved audio reconstruction.

4 Handling lost RTP packets

    The Qcelp CODEC supports the notion of erasure frames. These are
    frames that for whatever reason are not available. When
    reconstructing interleaved audio or playing back non-interleaved
    audio, erasure frames MUST be fed to the Qcelp 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 value may vary (it can only decrease), 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 audio, 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.

5 Discussion

The Qcelp CODEC interpolates the missing audio 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 audio 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, an interleave (field
LLL) value of 0 or 1 and a bundling factor of 4 or less is
recommended.

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

The restrictions on senders set forth in sections 3.3 and 3.4
guarantee that after receipt of the first payload packet from the
sender, the receiver can allocate a well-known amount of buffer space
that will be sufficient for all future reception from the same SSRC
value. Less buffer space may be required at some point in the future
if the sender decreases the bundling value or interleave, 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) should the bundling value or interleave value be
increased by the sender. Also, were the interleave or bundling value
to increase, the receiver could be forced to pause playback while it
receives the additional packets necessary for playback at an
increased bundling value or increased interleave.

6 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 [5] and in multicast routing protocols to allow a receiver to select which sources are allowed to reach it.

7 References

    [1]   TIA/EIA/IS-733. TR45: High Rate Speech Service Option for
          Wideband Spread Spectrum Communications Systems. Available from
          Global Engineering +1 800 854 7179 or +1 303 792 2181. May also
          be ordered online at http://www.eia.org/eng/.

    [2]   Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson,
          "RTP: A Transport Protocol for Real-Time Applications", RFC
          1889, January 1996.

    [3]   Bradner, S., "Key words for use in RFCs to Indicate Requirement
          Levels", BCP 14, RFC 2119, March 1997.

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

    [5]   Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC
          1112, August 1989.

8 Author's Address

    Kyle J. McKay
    QUALCOMM Incorporated
    5775 Morehouse Drive
    San Diego, CA 92121-1714
    USA

    Phone: +1 858 587 1121
    EMail: kylem@qualcomm.com

9 Full Copyright Statement

Acknowledgement