Internet Engineering Task Force
INTERNET-DRAFT
draft-ietf-avt-dtmf-01.ps

AVT WG
Schulzrinne
Columbia U.

November 18, 1998
Expires: May 15, 1999

RTP Payload for DTMF Digits

Status of this Memo

Copyright Notice

Abstract

This memo describes how to carry dual-tone multifrequency (DTMF) signaling and other tone signals in RTP packets.

1 Introduction

This memo defines a payload type for carrying dual-tone multifrequency (DTMF) digits in RTP packets. A separate payload type is desirable since low-rate voice codecs cannot be guaranteed to accurately reproduce DTMF. Defining a separate payload type also permits higher redundancy while maintaining a low bit rate.

The DTMF payload type must be suitable for both a gateway and end-to-end scenario. In the gateway scenario, a gateway connecting a packet voice network with the PSTN recreates the DTMF tones and injects them into the PSTN. Since DTMF digit recognition takes several tens of milliseconds, careful time and power (volume) alignment is needed to avoid generating spurious digits. For interactive voice response (IVR) systems directly connected to the packet voice network, time alignment and volume levels are not important, since the unit will not perform any signal analysis to detect DTMF tones from the audio stream.

DTMF digits are carried as part of the audio stream, and SHOULD use the same sequence number and time-stamp base as the regular audio channel to simplify recreation of analog audio at a gateway. The default clock frequency is 8000 Hz, but the clock frequency can be redefined when assigning the dynamic payload type.

This format achieves a higher redundancy even in the case of sustained packet loss than the method proposed for the *Voice over Frame Relay Implementation Agreement* [1].

In circumstances where exact timing alignment between the audio stream and the DTMF digits is not important and data is sent unicast, such as the IVR example mentioned earlier, it may be preferable to use a reliable control stream such as H.245.

A source MAY send coded DTMF and coded audio packets for the same time instants, using DTMF as the redundant encoding for the audio stream, or it MAY block outgoing audio while DTMF tones are active and only send DTMF digits as both the primary and redundant encodings.

2 Payload 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R|  digit  |R R| volume    |          duration             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

digit: The DTMF digits are encoded as follows:

DTMF digit	encoding (decimal)
0	0
1	1
2	2
...	...
9	9
*	10
#	11
A	12
B	13
C	14
D	15
Flash	16

volume: The power level of the digit, expressed in dBm0 after dropping the sign, with range from 0 to -63 dBm0. The range of valid DTMF is from 0 to -36 dBm0 (must accept); lower than -55 dBm0 must be rejected (TR-TSY-000181, ITU-T Q.24A). Thus, larger values denote lower volume.

Note: since the acceptable dip is 10 dB and the minimum detectable loudness variation is 3 dB, this field could be compressed by at least a bit by reducing resolution to 2 dB, if needed.

duration: Duration of this digit, in timestamp units.

For a sampling rate of 8000 Hz, this field is sufficient to express digit durations of upto approximately 8 seconds.

R: This field is reserved for future use. The sender MUST set it to zero, the receiver MUST ignore it.

An audio source SHOULD start transmitting DTMF digit packets as soon as it recognizes a DTMF digit and every 50 ms thereafter. (Precise spacing between DTMF digit packets is not necessary.)

Q.24 [2], Table A-1, indicates that all administrations surveyed use a minimum signal duration of 40 ms, with signaling velocity (tone and pause) of no less than 93 ms.

If a digit continues for more than one period, it should send a new DTMF packet with the RTP timestamp value corresponding to the beginning of the digit and the duration of the digit increased correspondingly. (The RTP sequence number is incremented by one for each packet.) If there has been no new digit in the last interval, the digit SHOULD be retransmitted three times (or until the next digit is recognized) to ensure some measure of reliability for the last digit.

> DTMF digits are sent incrementally to avoid having the receiver wait for the completion of the digit. Since some tones are two seconds long, this would incur a substantial delay. The transmitter does not know if digit length is important and thus needs to transmit immediately and incrementally. If the receiver application does not care about digit length, the incremental transmission mechanism avoids delay. Some applications, such as gateways into the GSTN, care about both delays and digit duration.

3 Reliability

To achieve reliability even when the network loses packets, the audio redundancy mechanism described in RFC 2198 [3] is used. The effective data rate is r times 64 bits (32 bits for the redundancy header and 32 bits for the DTMF payload) every 50 ms or r times 1280 bits/second, where r is the number of redundant DTMF digits carried in each packet. The value of r is an implementation trade-off, with a value of 5 suggested.

> The timestamp offset in this redundancy scheme has 14 bits, so that it allows a single packet to "cover" 2.048 seconds of DTMF digits at a sampling rate of 8000 Hz. Including the starting time of previous digits allows precise reconstruction of the tone sequence at a gateway. The scheme is resilient to consecutive packet losses spanning this interval of 2.048 seconds or r digits, whichever is less. Note that for previous digits, only an average loudness can be represented.

An encoder MAY treat the DTMF payload as a highly-compressed version of the current audio frame. In that mode, each RTP packet during a DTMF tone would contain the current audio codec rendition (say, G.723.1 or G.729) of this digit as well as the representation described in Section 2, plus any previous digits as before.

> This approach allows dumb gateways that do not understand this format to function. Other reasons?

3.1 Example

A typical RTP packet, where the user is just dialing the last digit of the DTMF sequence "911". The first digit was 200 ms long and started at time 0, the second digit lasted 250 ms and started at time 800 ms, the third digit was pressed at time 1.4 s and the packet shown was sent at 1.45 s. The frame duration is 50 ms. To make the parts recognizable, the figure below ignores byte alignment. Timestamp and sequence number are assumed to have been zero at the beginning of the first digit. In this example, the dynamic payload types 96 and 97 have been assigned for the redundancy mechanism and the DTMF payload, respectively.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
| 2 |0|0|   0   |0|     96      |              28               |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            timestamp                          |
|                              12000                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
|                           0x5234a8                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   block PT  |     timestamp offset      |   block length    |
|1|     97      |            12000          |         4         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   block PT  |     timestamp offset      |   block length    |
|1|     97      |             5600          |         4         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|   Block PT  |
|0|     97      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R|  digit  |R R| volume    |          duration             |
|0 0 0|    9    |0 0|     7     |             1600              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R|  digit  |R R| volume    |          duration             |
|0 0 0|    1    |0 0|    10     |             2000              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R|  digit  |R R| volume    |          duration             |
|0 0 0|    1    |0 0|    20     |              400              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

4 Compact Reliability Scheme

[This section is more speculative.] A more compact representation could be achieved by measuring DTMF tones in a different sampling rate from that of the surrounding audio codec, e.g., as multiples of 1, 10, 40 or 50 ms. Each RTP payload type should have a fixed sampling rate, so choosing a value that depends on frame interval of the surrounding codec is not recommended. For a sampling interval of 50 ms, the following payload would "cover" 8 seconds of duration and offset:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    offset     |R R R|  digit  |R R| volume    |   duration    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5 Changes Since Version -00

- Uniform interval of 50 ms, since audio frame interval may change based on codec.

6 Acknowledgements

The suggestions of the VoIP working group and Fred Burg are gratefully acknowledged.

References

[1] R. Kocen and T. Hatala, “Voice over frame relay implementation agreement,” Implementation Agreement FRF.11, Frame Relay Forum, Foster City, California, Jan. 1997.

[2] International Telecommunication Union, “Multifrequency push-button signal reception,” Recommendation Q.24, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, 1988.

[3] C. Perkins, I. Kouvelas, V. Hardman, M. Handley, and J. Bolot, “RTP payload for redundant audio data,” RFC 2198, Internet Engineering Task Force, Sept. 1997.