RTP Payload for DTMF Digits, Telephony Tones and Telephony Signals

Status of this Memo

This document is an Internet-Draft and is subject to all provisions of section 3 of RFC 3667. By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she become aware will be disclosed, in accordance with RFC 3668.

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/ietf/1id-abstracts.txt
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Abstract

This memo describes how to carry dual-tone multifrequency (DTMF) signaling, other tone signals and telephony events in RTP packets. This memo captures and expands upon the basic framework defined in RFC 2833, but retains only the most basic event codepoints. Other codepoints are documented separately.
Table of Contents

1. Introduction..................................................4
  1.1 Terminology................................................4
  1.2 Overview..................................................4
  1.3 Potential Applications...................................5
  1.4 Events, States, Tone Patterns, and Voice Encoded Tones...6

2. RTP Payload Format for Named Telephone Events.................7
  2.1 Introduction...............................................7
  2.2 Use of RTP Header Fields.................................8
    2.2.1 Timestamp..........................................8
    2.2.2 Marker Bit..........................................8
  2.3 Payload Format...........................................8
    2.3.1 Event Field.........................................8
    2.3.2 E ("End") Bit......................................8
    2.3.3 R Bit................................................8
    2.3.4 Volume Field.......................................9
    2.3.5 Duration Field.....................................9
  2.4 Optional MIME Parameters................................9
    2.4.1 Relationship to SDP................................10
  2.5 Procedures..............................................10
    2.5.1 Sending Procedures.................................10
      2.5.1.1 Negotiation of Payloads........................10
      2.5.1.2 Transmission of Event Packets................11
      2.5.1.3 Long Duration Events..........................12
      2.5.1.4 Retransmission of Final Packet................12
      2.5.1.5 Packing Multiple Events Into One Packet........12
    2.5.2 Receiving Procedures................................13
      2.5.2.1 Indication of Receiver Capabilities using SDP...13
      2.5.2.2 Playout of Tone Events playout................13
      2.5.2.3 Long Duration Events............................15
      2.5.2.4 Multiple Events In a Packet.....................15
      2.5.2.5 Soft States......................................16
  2.6 Reliability.............................................16
    2.6.1 Intra-Event Updates................................16
    2.6.2 Multi-Event Redundancy..............................16

3. Specification of Codepoints For Telephone Events.............17
  3.1 DTMF Events.............................................18
  3.2 Data Modem and Fax Events...............................19
    3.2.1 V.21 Events........................................20
    3.2.2 V.8 Events..........................................22
    3.2.3 V.25 Events........................................23
    3.2.4 T.30 Events........................................25

4. RTP Payload Format for Telephony Tones........................28
4.1 Introduction........................................................................28
4.2 Examples of Common Telephone Tone Signals................29
4.3 Use of RTP Header Fields.................................................30
  4.3.1 Timestamp........................................................................30
  4.3.2 Marker Bit........................................................................30
  4.3.3 Payload Format.................................................................30
  4.3.4 Optional MIME Parameters..............................................32
4.4 Procedures...........................................................................32
  4.4.1 Sending Procedures.......................................................32
  4.4.2 Receiving Procedures....................................................33
5. Application Considerations..................................................34
  5.1 Combining Tones and Named Events.................................34
  5.2 Simultaneous Generation of Audio and Events...............34
  5.3 Strategies For Handling FAX and Modem Signals...........35
  5.4 Examples...........................................................................36
    5.4.1 Use of RFC 2198 Redundancy With Named Events........36
    5.4.2 Combined Tone and Telephone-event Payloads............38
6. MIME Registration..............................................................40
  6.1 audio/telephone-event.....................................................40
  6.2 audio/tone.........................................................................41
7. Security Considerations......................................................42
8. IANA Considerations............................................................42
9. Acknowledgements..............................................................44
10. Authors ...............................................................44
12. References.............................................................................45
  12.1 Normative References....................................................45
  12.2 Informative References..................................................46
1. Introduction

1.1 Terminology

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in RFC 2119 [N-1] and indicate requirement levels for compliant implementations.

Normative references appear as [N-n], while informative references appear as [I-n]. All references are at the end of this memo.

This document uses the following abbreviations:

DTMF  Dual Tone Multifrequency
IVR   Integrated Voice Response unit
PSTN  Public Switched (circuit) Telephone Network

1.2 Overview

This memo defines two RTP [N-4] payload formats, one for carrying dual-tone multifrequency (DTMF) digits and other line and trunk signals as events (section 2), and a second one to describe general multi-frequency tones in terms only of their frequency and cadence (section 4). Separate RTP payload formats for telephony tone signals are desirable since low-rate voice codecs cannot be guaranteed to reproduce these tone signals accurately enough for automatic recognition. In addition, tone properties such as the phase reversals in the ANS51 tone will not survive speech coding. Defining separate payload formats also permits higher redundancy while maintaining a low bit rate. Finally, some telephony events such as "on-hook" occur out-of-band and cannot be transmitted as tones.

The remainder of this section provides the motivation for defining the payload types described in this document. Section 2 defines the payload format and associated procedures for use of named events. Section 3 describes the events for which codepoints are defined in this document. Section 4 describes the payload format and associated procedures for tone representations. Section 5 deals with achievement of reliable delivery through redundancy and the use of combined payloads. Section 6 provides the MIME media type registrations for the two payload formats, and also defines the IANA requirements for registration of codepoints for named telephone events. Section 7 deals with security considerations.
1.3 Potential Applications

The payload formats described here may be useful in a number of different scenarios.

On the sending side, there are two basic possibilities: either the sending side is an end system which originates the signals itself, or it is a gateway with the task of propagating incoming telephone signals into the Internet.

On the receiving side there are more possibilities. The first is that the receiver must propagate tone signalling accurately into the PSTN for machine consumption. One example of this is a gateway passing DTMF tones to an IVR. In this scenario, frequencies, amplitudes, tone durations, and the durations of pauses between tones are all significant, and individual tone signals must be delivered reliably and in order.

In the second scenario, the receiver must play out tones for human consumption. Typically, rather than a series of tone signals each with its own meaning, the content will consist of a single sequence of tones and possibly silence, played out continuously or repeated cyclically for some period of time. Often the end of the tone playout will be triggered by an event fed back in the other direction, using either in- or out-of-band means. Examples of this are dial tone or busy tone.

The relationship between locality and the tones to be played out is a complicating factor in this scenario. In the phone network, tones are generated at different places, depending on the switching technology and the nature of the tone. This determines, for example, whether a person making a call to a foreign country hears her local tones she is familiar with or the tones as used in the country called.

For analog lines, dial tone is always generated by the local switch. ISDN terminals may generate dial tone locally and then send a Q.931 SETUP message containing the dialed digits. If the terminal just sends a SETUP message without any Called Party digits, then the switch does digit collection, provided by the terminal as KEYPAD messages, and provides dial tone over the B-channel. The terminal can either use the audio signal on the B-channel or can use the Q.931 messages to trigger locally generated dial tone.

Ringing tone (also called ringback tone) is generated by the local switch at the callee, with a one-way voice path opened up as soon as the callee’s phone rings. (This reduces the chance of clipping the called party’s response just after answer. It also permits pre-answer announcements or in-band call-progress indications to reach the
caller before or in lieu of a ringing tone.) Congestion tone and special information tones can be generated by any of the switches along the way, and may be generated by the caller’s switch based on ISUP messages received. Busy tone is generated by the caller’s switch, triggered by the appropriate ISUP message, for analog instruments, or the ISDN terminal.

In the third scenario, an end system is directly connected to the Internet and does not need to generate tone signals again, so that time alignment and power levels are not relevant. These systems rely on PSTN gateways or Internet end systems to generate DTMF events and do not perform their own audio waveform analysis. An example of such a system is an Internet interactive voice-response (IVR) system.

In circumstances where exact timing alignment between the audio stream and the DTMF digits or other events is not important and data is sent unicast, such as the IVR example mentioned earlier, it may be preferable to use a reliable control protocol rather than RTP packets. In those circumstances, this payload format would not be used.

Note that in a number of these cases it is possible that the gateway or end system will be both a sender and receiver of telephone signals. Sometimes the same class of signals will be sent as received -- in the case of "RTP trunking" or voiceband data, for instance. In other cases, such as that of an end system serving analogue lines, the signals sent will be in a different class from those received.

1.4 Events, States, Tone Patterns, and Voice Encoded Tones

This document provides the means for in-band transport over the Internet of two broad classes of signalling information: in-band tones or tone sequences, and signals sent out-of-band in the PSTN. Three methods, two of which are defined by this document, are available for carrying tone signals; only one of the three can be used to carry out-of-band PSTN signals. Depending on the application, it may be desirable to carry the signalling information in more than one form at once. Section 5 discusses when and how this should be done.

1) The gateway or end system can upspeed to a higher-bandwidth codec such as G.711 [I-3] when tone signals are to be conveyed. Alternatively, for FAX or modem signals respectively, a specialized transport such as T.38 [I-8], RFC 2793 [I-1], or V.150.1 modem relay [I-17] may be used.

2) The sending gateway can simply measure the frequency components of the voice band signals and transmit this information to the RTP...
receiver using the tone representation defined in this document (section 4). In this mode, the gateway makes no attempt to discern the meaning of the tones, but simply distinguishes tones from speech signals. An end system may use the same approach using configured rather than measured frequencies.

All tone signals in use in the PSTN and meant for human consumption are sequences of simple combinations of sine waves, either added or modulated. (There is at least one tone, however, the ANSam tone [N-11] used for indicating data transmission over voice lines, that makes use of periodic phase reversals.)

3) As a third option, a gateway can recognize the tones and translate them into a name, such as ringing or busy tone or DTMF digit '0' (section 2). The receiver then produces a tone signal or other indication appropriate to the signal. Generally, since the recognition of signals at the sender often depends on their on/off pattern or the sequence of several tones, this recognition can take several seconds. On the other hand, the gateway may have access to the actual signaling information that generates the tones and thus can generate the RTP packet immediately, without the detour through acoustic signals.

The use of named events is the only feasible method for transmitting out-of-band PSTN signals as content within RTP sessions.

2. RTP Payload Format for Named Telephone Events

2.1 Introduction

The RTP payload format for named telephone events is designated as "telephone-event", the MIME type as "audio/telephone-event". In accordance with current practice, this payload format does not have a static payload type number, but uses a RTP payload type number established dynamically and out-of-band. The default clock frequency is 8000 Hz, but the clock frequency can be redefined when assigning the dynamic payload type.

Named telephone events are carried as part of the audio stream, and MUST use the same sequence number and time-stamp base as the regular audio channel to simplify the generation of audio waveforms at a gateway. The named telephone events payload type can be considered to be a very highly-compressed audio codec, and is treated the same as other codecs.
2.2 Use of RTP Header Fields

2.2.1 Timestamp

The RTP timestamp reflects the measurement point for the current packet. The event duration described in section 2.5 extends forwards from that time. For events that span multiple RTP packets, the RTP timestamp identifies the beginning of the event, i.e., several RTP packets may carry the same timestamp. For long-lasting events that have to be split into subevents (see below, section 2.5.1.3), the timestamp indicates the beginning of the subevent.

2.2.2 Marker Bit

The RTP marker bit indicates the beginning of a new event. For long-lasting events that have to be split into subevents (see below, section 2.5.1.3), only the first subevent will have the marker bit set.

2.3 Payload Format

The payload format for named telephone events is shown in Figure 1.

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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     event     |E|R| volume    |          duration             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Figure 1: Payload Format for Named Events

2.3.1 Event Field

The event field is a number between 0 and 255 identifying a specific telephony event. An IANA registry of codepoints for this field has been established (see IANA Considerations, section 8). The initial content of this registry consists of the events defined in section 3.

2.3.2 E ("End") Bit

If set to a value of one, the "end" bit indicates that this packet contains the end of the event. For long-lasting events that have to be split into subevents (see below, section 2.5.1.3), only the final packet for the final subevent will have the "E" bit set.

2.3.3 R Bit

This field is reserved for future use. The sender MUST set it to zero, the receiver MUST ignore it.
2.3.4 Volume Field

For DTMF digits and other events representable as tones, this field describes the power level of the tone, expressed in dBm₀ after dropping the sign. Power levels range from 0 to -63 dBm₀. Thus, larger values denote lower volume. This value is defined only for events for which the documentation indicates that volume is applicable. For other events, the sender MUST set volume to zero and the receiver MUST ignore the value.

2.3.5 Duration Field

The duration field indicates the duration of the event or subevent being reported, in timestamp units, expressed as an unsigned integer. For a non-zero value, the event or subevent began at the instant identified by the RTP timestamp and has so far lasted as long as indicated by this parameter. The event may or may not have ended. If the event duration exceeds the maximum representable by the duration field, the event is split into several contiguous subevents as described below (section 2.5.1.3).

The special duration value of zero is reserved to indicate that the event lasts "forever", i.e., is a state and is considered to be effective until updated. A sender MUST NOT transmit a zero duration for events other than those defined as states. The receiver SHOULD ignore an event report with zero duration if the event is not a state.

Events defined as states MAY contain a non-zero duration, indicating that the sender intends to refresh the state before the time duration has elapsed ("soft state").

For a sampling rate of 8000 Hz, the duration field is sufficient to express event durations of up to approximately 8 seconds.

2.4 Optional MIME Parameters

As indicated in the MIME registration for named events in section 6.1, the telephone-event MIME type supports two optional parameters: the "events" parameter, and the "rate" parameter.

The "events" parameter lists the events supported by the implementation. Events are listed as one or more comma-separated elements. Each element can either be a single integer or an integer followed by a hyphen and a larger integer, representing a range of consecutive event codepoints. No white space is allowed in the argument. The integers designate the event numbers supported by the implementation.
The "rate" parameter describes the sampling rate, in Hertz, and hence the units for the RTP timestamp and event duration fields. The number is written as a floating point number or as an integer. If omitted, the default value is 8000 Hz.

2.4.1 Relationship to SDP

The recommended mapping of MIME optional parameters to SDP is given in section 3 of RFC 3555 [N-5]. The "rate" MIME parameter for the named event payload type follows this convention: it is expressed as usual as the <clock rate> component of the a=rtpmap: attribute line.

The "events" MIME parameter deviates from the convention suggested in RFC 3555 because it omits the string "events=" before the list of supported events.

a=fmtp:<format> <list of values>

The list of values has the format described above for the MIME parameter. The list does not have to be sorted.

For example, if the payload format uses the payload type number 100, and the implementation can handle the DTMF tones (events 0 through 15) and the dial and ringing tones, it would include the following description in its SDP message:

m=audio 12345 RTP/AVP 100
a=rtpmap:100 telephone-event/8000
a=fmtp:100 0-15,66,70

The following sample media type definition corresponds to the SDP example above:

audio/telephone-event;events="0-15,66,67";rate="8000"

2.5 Procedures

This section defines the procedures associated with the named event payload type. Additional procedures may be specified in the documentation associated with specific event codepoints.

2.5.1 Sending Procedures

2.5.1.1 Negotiation of Payloads

Negotiation of payloads between sender and receiver is achieved by out-of-band means, using SDP, for example.
The sender SHOULD indicate what events it supports, using the optional "events" parameter associated with the telephone-events MIME type. If the sender receives an "events" parameter from the receiver, it MUST restrict the set of events it sends to those listed in the received "events" parameter. For backward compatibility, if no "events" parameter is received, the sender SHOULD assume support for the DTMF events 0-15 but for no other events.

2.5.1.2 Transmission of Event Packets

DTMF digits and named telephone events are carried as part of the audio stream, and MUST use the same sequence number and time-stamp base as the regular audio channel to simplify the generation of audio waveforms at a gateway.

An audio source SHOULD start transmitting event packets as soon as it recognizes an event, and continue to send updates until the event has ended. The update packet MUST have the same RTP timestamp value as the initial packet for the event, but the duration MUST be increased to reflect the total cumulative duration since the beginning of the event.

The first packet for an event MUST have the "M" bit set. The final packet for an event MUST have the "E" bit set, but setting of the "E" bit MAY be deferred until the final packet is retransmitted (see section 2.5.1.4). Intermediate packets for an event MUST NOT have either the "M" bit or the "E" bit set.

Sending of a packet with the "E" bit set is OPTIONAL if the packet reports two events which are defined as mutually exclusive states, or if the final packet for one state is immediately followed by a packet reporting a mutually exclusive state. (For events defined as states, the appearance of a mutually exclusive state implies the end of the previous state.)

A source has wide latitude as to how often it sends event updates. A natural interval is the spacing between non-event audio packets. (Recall that a single RTP packet can contain multiple audio frames for frame-based codecs and that the packet interval can vary during a session.) Alternatively, a source MAY decide to use a different spacing for event updates, called an event period, with a value of 50 ms RECOMMENDED.

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

For robustness, the sender SHOULD retransmit "state" events periodically.

Timing information is contained in the RTP timestamp, allowing precise recovery of inter-event times. Thus, the sender does not need to maintain precise or consistent time intervals between event packets.

### 2.5.1.3 Long Duration Events

If an event persists beyond the maximum duration expressible in the duration field (0xFFFF), the sender MUST send a packet reporting this maximum duration but MUST NOT set the "E" bit in this packet. The sender MUST then begin reporting a new "subevent" with the RTP timestamp set to the time at which the previous subevent ended and the duration set to the cumulative duration of the new subevent. The "M" bit of the first packet reporting the new subevent MUST NOT be set. The sender MUST repeat this procedure as required until the end of the complete event has been reached. The final packet for the complete event MUST have the "E" bit set (either on initial transmission or on retransmission as described below).

### 2.5.1.4 Retransmission of Final Packet

The final packet for each event and for each subevent SHOULD be sent a total of three times at the interval used by the source for updates. (If a new event is recognized during the retransmissions and RFC 2198 [N-2] is in use, the old event will be part of the redundancy in the RFC 2198 payloads.) This ensures that the duration of the event or subevent can be recognized correctly even if an instance of the last packet is lost.

A sender MAY delay setting the "E" bit until retransmitting the last packet for a tone, rather than setting the bit on its first transmission. This avoids having to wait to detect whether the tone has indeed ended. Once the sender has set the "E" bit for a packet, it MUST continue to set the "E" bit for any further retransmissions of that packet.

### 2.5.1.5 Packing Multiple Events Into One Packet

Multiple named events can be packed into a single RTP packet if and only if the events are consecutive and contiguous, i.e., occur without overlap and without pause between them, and if the last event packed into a packet occurs quickly enough to avoid excessive delays at the receiver.
This approach is similar to having multiple frames of frame-based audio in one RTP packet.

The constraint that packed events not overlap implies that events designated as states can be followed in a packet only by other state events which are mutually exclusive to them. The constraint itself is needed so that the beginning time of each event can be calculated at the receiver.

In a packet containing events packed in this way, the RTP timestamp MUST identify the beginning of the first event or subevent in the packet. The "M" bit MUST be set (since the packet records the beginning of at least one event). The "E" bit and duration for each event in the packet MUST be set using the same rules as if that event were the only event contained in the packet.

For events with a duration shorter than a typical packet interval, for example, V.21 bits, it is RECOMMENDED that multiple events are represented by a single RFC 2198 [N-2] packet, as described in section 5.

2.5.1.6 RTP Sequence Number

The RTP sequence number MUST be incremented by one in each successive RTP packet sent. Incrementing applies to retransmitted as well as initial instances of event reports, to permit the receiver to detect lost packets for RTCP receiver reports.

2.5.2 Receiving Procedures

2.5.2.1 Indication of Receiver Capabilities using SDP

Receivers can indicate which named events they can handle, for example, by using the Session Description Protocol (RFC 2327 [N-3]). SDP descriptions using the event payload MUST contain an fntp format attribute that lists the event values that the receiver can process.

2.5.2.2 Playout of Tone Events

In the gateway scenario, an Internet telephony gateway connecting a packet voice network to the PSTN recreates the DTMF or other tones and injects them into the PSTN. Since, for example, DTMF digit recognition takes several tens of milliseconds, the first few milliseconds of a digit will arrive as regular audio packets. Thus, careful time and power (volume) alignment between the audio samples and the events is needed to avoid generating spurious digits at the receiver. Playout when audio packets continue to arrive as the event proceeds is discussed further in section 5.2 below.
Receiver implementations MAY use different algorithms to create tones, including the two described here. Note that not all implementations have the need to recreate a tone; some may only care about recognizing the events.

In the first algorithm, the receiver simply places a tone of the given duration in the audio playout buffer at the location indicated by the timestamp. As additional packets are received that extend the same tone, the waveform in the playout buffer is extended accordingly. (Care has to be taken if audio is mixed, i.e., summed, in the playout buffer rather than simply copied.) Thus, if a packet in a tone lasting longer than the packet interarrival time gets lost and the playout delay is short, a gap in the tone may occur.

Alternatively, the receiver can start a tone and play it until it receives a packet with the "E" bit set, the next tone, distinguished by a different timestamp value or a given time period elapses. This is more robust against packet loss, but may extend the tone beyond its original duration if all retransmissions of the last packet in an event are lost. Limiting the time period of extending the tone is necessary to avoid that a tone "gets stuck". This algorithm is not a license for senders to set the duration field to zero; it MUST be set to the current duration as described, since this is needed to create accurate events if the first event packet is lost, among other reasons.

Regardless of the algorithm used, the tone SHOULD NOT be extended by more than three packet interarrival times. A slight extension of tone durations and shortening of pauses is generally harmless.

If a receiver has extended a tone by the maximum extension duration and started playing silence, it MUST NOT resume playing the tone when later packets for that event arrive, as this would cause spurious events to be detected downstream.

If a receiver receives an event packet for an event which it is not currently playing out and the packet does not have the "M" bit set, earlier packets for that event have evidently been lost. This can also be determined by gaps in the RTP sequence number. The receiver MAY determine on the basis of retained history and the timestamp and event code of the current packet that it corresponds to an event already played out and lapsed. In that case further reports for the event MUST be ignored, as indicated in the previous paragraph. If this is not so, the receiver MAY attempt to play the event out to the complete duration indicated in the event report. The appropriate behaviour will depend on the event type concerned, and requires consideration of the relationship of the event to audio media flows.
and whether correct event duration is essential to the correct operation of the media session.

A receiver SHOULD NOT rely on a particular event packet spacing, but instead MUST use the event timestamps and durations to determine timing and duration of playout.

The receiver MUST calculate jitter for RTCP receiver reports based on all packets with a given timestamp. Note: The jitter value should primarily be used as a means for comparing the reception quality between two users or two time-periods, not as an absolute measure.

If a zero volume is indicated for an event for which the volume field is defined, then the receiver MAY reconstruct the volume from the volume of non-event audio or MAY use the nominal value specified by the ITU Recommendation or other document defining the tone. This ensures backwards compatibility with RFC 2833, where the volume field was defined only for DTMF events.

2.5.2.3 Long Duration Events

If an event report is received with duration equal to the maximum duration expressible in the duration field (0xFFFF) and the "E" bit for the report is not set, the event report may mark the end of a subevent generated according to the procedures of section 2.5.1.3. If another report for the same event type is received, the receiver MUST compare the RTP timestamp for the new event with the sum of the RTP timestamp of the previous report plus the duration (0xFFFF). The receiver uses the absence of a gap between the events to detect that it is receiving a single long-duration event.

The total duration of a long duration event is (obviously) the sum of the durations of the subevents used to report it. This is equal to the duration of the final subevent (as indicated in the final packet for that subevent), plus 0xFFFF multiplied by the number of subevents preceding the final subevent.

2.5.2.4 Multiple Events In a Packet

The procedures of section 2.5.1.5 require that if multiple events are reported in the same packet, they are contiguous and non-overlapping. As a result, it is not strictly necessary for the receiver to know the start times of the events following the first one in order to play them out -- it needs only to respect the duration reported for each event. Nevertheless, if knowledge of the start time for a given event after the first one is required, it is equal to the sum of the start time of the preceding event plus the duration of the preceding event.
2.5.2.5 Soft States

If the duration of a soft state event expires, the receiver SHOULD consider the value of the state to be "unknown" unless otherwise indicated in the event documentation (e.g., in section 3).

2.6 Reliability

The named event mechanism uses three complementary redundancy mechanisms to deal with lost packets:

Intra-event updates:

Events that last longer than one event period (e.g., 50 ms) are updated periodically, so that the receiver can reconstruct the event and its duration if it receives any of the update packets, albeit with delay. This mechanism is described in section 2.6.1 and is most helpful for longer events.

Repeat last event packet:

As described in section 2.5.1.4, the last event packet is transmitted a total of three times if there is no subsequent event. This mechanism is applicable for widely-spaced events.

Multi-event redundancy:

Section 2.6.2 describes how a summary of earlier events MAY be carried in RFC 2198 redundancy payloads. This is particularly useful for sequences of short events, e.g., digits dialed by a modem or autodialer or in-band tone signaling sequences (section 3.2 or 3.5).

2.6.1 Intra-Event Updates

During an event, the RTP event payload format provides incremental updates on the event. The error resiliency afforded by this mechanism depends on whether the first or second algorithm in section 2.5.2.2 is used and on the playout delay at the receiver. For example, if the receiver uses the first algorithm and only places the current duration of tone signal in the playout buffer, for a playout delay of 120 ms and a packet gap of 50 ms, two packets in a row can get lost without causing a premature end of the tone generated.

2.6.2 Multi-Event Redundancy

The audio redundancy mechanism described in RFC 2198 [N-2] MAY be used to recover from packet loss across events. For the suggested packet gap of 50 ms, the effective data rate is r times 64 bits.
The timestamp offset in this redundancy scheme has 14 bits, so that it allows a single packet to "cover" 2.048 seconds of telephone events at a sampling rate of 8000 Hz. Including the starting time of previous events 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 event payload as a highly-compressed version of the current audio frame. In that mode, each RTP packet during an event would contain the current audio codec rendition (say, G.723.1 [I-4] or G.729 [I-5] of this digit as well as the representation described in section 2, plus any previous events seen earlier.

This approach allows dumb gateways that do not understand this format to function. See also the discussion in section 1.

The payload format described here achieves a higher redundancy even in the case of sustained packet loss than the method proposed for the Voice over Frame Relay Implementation Agreement [I-18]. In short, senders generate updates at regular intervals, thus ensuring that each event is transmitted multiple times. RFC 2198 [N-2] is used to recover events where all packets sent during the event have been lost.

3. Specification of Codepoints For Telephone Events

This document defines two classes of named events:

1) DTMF tones (section 3.1);

2) data and fax-related tones (section 3.2);

It is intended that other RFCs define additional events, and in particular define and update the events present in RFC 2833 but not documented here.

The tables listing the event codepoints for each class indicate whether the respective events are states, tones, or other. For tone
events, the tables indicate whether the volume field is applicable or must be set to 0.

3.1 DTMF Events

DTMF signalling [N-8] is typically generated by a telephone set or possibly by a PBX. DTMF digits may be consumed by entities such as gateways or application servers in the IP network, or by entities such as telephone switches or IVRs in the circuit switched network.

The DTMF events support two possible applications at the sending end, and two at the receiving end. In the first application at the sending end, the Internet telephony gateway detects DTMF on the incoming circuits and sends the RTP payload described here instead of regular audio packets. The gateway likely has the necessary digital signal processors and algorithms, as it often needs to detect DTMF, e.g., for two-stage dialing. Having the gateway detect tones relieves the receiving Internet end system from having to do this work and also avoids having low bit-rate codecs like G.723.1 [I-4] render DTMF tones unintelligible. In the second application, an Internet end system such as an "Internet phone" can emulate DTMF functionality without concerning itself with generating precise tone pairs and without imposing the burden of tone recognition on the receiver.

A similar distinction occurs at the receiving end. In the gateway scenario, an Internet telephony gateway connecting a packet voice network to the PSTN recreates the DTMF tones or other telephony events and injects them into the PSTN. In the end system scenario, the DTMF events are consumed by the receiving entity itself.

Table 1 shows the DTMF-related named event codepoints within the telephone-event payload format. The DTMF digits 0-9 and * and # are commonly supported. DTMF digits A through D are less frequently encountered, typically in special applications such as military networks.

ITU-T Recommendation Q.24 [N-9], Table A-1, indicates that the legacy switching equipment in the countries surveyed expects a minimum recognizable signal duration of 40 ms, a minimum pause between signals of 40 ms, and a maximum signalling rate of 8 to 10 digits per second depending on the country.
### 3.2 Data Modem and Fax Events

This section defines a few of the control events and tones that can appear on a subscriber line serving a fax machine or modem. Their purpose is to support negotiation, start-up and takedown of FAX and modem sessions and transitions between operating modes. The actual FAX and modem content are carried by other payload types (e.g., G.711 [I-3], T.38 [I-8], or, in specific circumstances, V.150.1 [I-17] modem relay, RFC 2793 [I-1], or CLEARMODE [I-2]. The events are organized into several groups, corresponding to the ITU-T Recommendation in which they are defined.

NOTE: implementors SHOULD NOT rely on the descriptions of the various modem protocols described below without consulting the original references (generally ITU-T Recommendations). The descriptions are provided in this document to give a context for the use of the events defined here. They frequently omit important details needed for implementation.

The typical application of these events is to allow the Internet to serve as a bridge between terminals operating on the PSTN. This application is characterized as follows:

- each gateway will act both as sender and as receiver;
- time constraints apply to the exchange of signals, making the early identification and reporting of events desirable so that receiver playout can proceed in timely fashion;
- transfer of the events must be reliable.
In some cases, an implementation may simply ignore certain events, such as fax tones, that do not make sense in a particular environment. Section 2.4.1 specifies how an implementation can use the SDP "fmtp" parameter within an SDP description to indicate its inability to understand a particular event or range of events.

Regardless of which events they support, implementations MUST be prepared to send and receive data signals using payload types other than telephone-event, simultaneously with the use of the latter. This is discussed further in section 5.3.

A further word on time constraints is in order. Time constraints governing the duration of tones do not pose a problem when using the telephone-events payload type: the payload specifies the duration and the receiving gateway can play out the tones accordingly. Problems come when time constraints are specified for the duration of silence between tones. A silent period of "at least x ms" is not a problem -- event notifications can be received late, but they can still be played out at their specified durations.

The problem arises with requirements of silence for "exactly" some period or for "at most" some period. The most general constraint of the latter type has to do with the operation of echo suppressors (ITU-T Rec. G.164 [N-6] and echo cancellers (ITU-T Rec. G.165 [N-7]). These devices may re-activate after as little as 100 ms of no signal on the line. As a result, in any situation where echo suppressors or cancellers must be disabled for signalling to work, tone events must be reported quickly enough to ensure that these devices do not become reenabled. This principle is reflected in the succeeding sections.

3.2.1 V.21 Events

V.21 [N-12] is a modem protocol offering data transmission at a maximum rate of 300 bits/s. Two channels are defined, supporting full duplex data transmission if required. One channel uses frequencies 980 Hz for "1" and 1180 Hz for "0"; the other channel uses frequencies 1650 Hz for "1" and 1850 Hz for "0". The modem can operate synchronously or asynchronously.

V.21 is used by other protocols (e.g., V.8bis, V.18, T.30) for transmission of control data, and is also used in its own right between text terminals. The telephone-events payload type SHOULD NOT be used to carry user data as opposed to control data -- other payload types such as G.711 [I-3], RFC 2793 [I-1], or V.150.1 [I-17] modem relay are more suitable for that purpose. The V.21 events are summarized in Table 3.

Sending implementations MUST report a completed event for every bit transmitted (i.e., rather than at transitions between "0" and "1").
Implementations SHOULD pack multiple events into one packet, using the procedures of section 2.5.1.5. Eight to ten bits is a reasonable packetization interval.

Reliable transmission of V.21 events is important, to prevent data corruption. Reporting an event per bit rather than per transition increases reporting redundancy and thus reporting reliability, since each event completion is retransmitted three times as described in section 2.5.1.4. To reduce the number of packets required for reporting, implementations SHOULD carry the retransmitted events using RFC 2198 [N-2] redundancy encoding.

<table>
<thead>
<tr>
<th>Event</th>
<th>Frequency (Hz)</th>
<th>Encoding (decimal)</th>
<th>Type</th>
<th>Volume?</th>
</tr>
</thead>
<tbody>
<tr>
<td>V.21 channel 1, &quot;0&quot; bit</td>
<td>1180</td>
<td>37</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>V.21 channel 1, &quot;1&quot; bit</td>
<td>980</td>
<td>38</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>V.21 channel 2, &quot;0&quot; bit</td>
<td>1850</td>
<td>39</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>V.21 channel 2, &quot;1&quot; bit</td>
<td>1650</td>
<td>40</td>
<td>tone</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 2: Events for V.21 signals

3.2.2 V.8 Events

V.8 [N-11] is an older general negotiation and control protocol, supporting startup for the following terminals: H.324 [I-6] multimedia, V.18 [I-21] text, T.101 [I-9] videotext, T.30 [N-10] send or receive FAX, and a long list of V-series modems including V.34 [I-13], V.90 [I-14], V.91 [I-15], and V.92 [I-16]. In contrast to V.8bis [I-19], in V.8 only the calling terminal can determine the operating mode.

V.8 defines four signals which consist of bits transferred by V.21 [N-12] at 300 bits/s: the call indicator signal (CI), the call menu signal (CM), the CM terminator (CJ), and the joint menu signal (JM). In addition, it uses tones defined in V.25 [N-13] and T.30 [N-10] (described below), and one tone (ANSam) defined in V.8 itself. The calling terminal sends using the V.21 low channel; the answering terminal uses the high channel.
The basic protocol sequence is subject to a number of variations to accommodate different terminal types. A pure V.8 sequence is as follows:

1) After an initial period of silence, the calling terminal transmits the V.8 CI signal. It repeats CI at least three times, continuing with occasional pauses until it detects ANSam tone. The CI indicates whether the calling terminal wants to function as H.324, V.18, T.30 send, T.30 receive, or a V-series modem.

2) The answering terminal transmits ANSam after detecting CI. ANSam will disable any G.164 [N-6] echo suppressors on the circuit after 400 ms and any G.165 [N-7] echo cancellors after one second of ANSam playout.

3) On detecting ANSam, the calling terminal pauses at least half a second, then begins transmitting CM to indicate detailed capabilities within the chosen mode.

4) After detecting at least two identical sequences of CM, the answering terminal begins to transmit JM, indicating its own capabilities (or offering an alternative terminal type if it cannot support the one requested).

5) After detecting at least two identical sequences of JM, the calling terminal completes the current octet of CM, then transmits CJ to acknowledge the JM signal. It pauses exactly 75 ms, then starts operating in the selected mode.

6) The answering terminal transmits JM until it has detected CJ. At that point it stops transmitting JM immediately, pauses exactly 75 ms, then starts operating in the selected mode.

The CI, CM, and JM signals all consist of a fixed sequence of ten "1" bits followed by a signal-dependent pattern of ten synchronization bits, followed by one or more octets of variable information. Each octet is preceded by a "0" start bit and followed by a "1" stop bit. The combination of the synchronization pattern and V.21 channel uniquely identifies the message type. The CJ signal consists of three successive octets of all zeros with stop and start bits but without the preceding "1"s and synchronizing pattern of the other signals.

If both gateways support V.21 bit events (section 3.2.2), the sending gateway for a given message MUST report each instance of a CM, JM, CI, and CJ signal respectively as a series of V.21 bit events. A packetization interval of 10 events per packet is suggested, since V.8 signals are organized in this way.
The overlapping nature of V.8 signalling means that there is no risk of silence exceeding 100 ms once ANSam has disabled any echo control circuitry. However, the 75 ms pause before entering operation in the selected data mode will require both the calling and the answering gateways to recognize the completion of CJ, so they can change from playout of telephone-events to playout of the data-bearing payload after the 75 ms period.

<table>
<thead>
<tr>
<th>Event</th>
<th>Frequency Hz</th>
<th>Encoding (decimal)</th>
<th>Type</th>
<th>Volume?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANSam</td>
<td>2100 x 15</td>
<td>34</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>/ANSam</td>
<td>2100 x 15</td>
<td>35</td>
<td>tone</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 3: Events for V.8 signals

Modified answer tone ANSam consists of a sinewave signal at 2100 Hz with phase reversals at an interval of 450 ms, amplitude-modulated by a sine wave at 15 Hz. The modulated envelope ranges in amplitude between 0.8 and 1.2 times its average amplitude. The average transmitted power is governed by national regulations. Thus it makes sense to indicate the volume of the signal. The ANSam phase reversals are allowed only if echo canceller disabling is required.

The sender MUST report ANSam as soon as it is recognized, providing updates at reasonable intervals as it continues. However, an ANSam event packet SHOULD NOT be sent until it is possible to discriminate between an ANSam event and an ANS event (see V.25 events, below). If a phase reversal is detected, the sender MUST report completion of the ANSam event and beginning of the /ANSam event at the time that the reversal was detected. If another phase reversal is detected, the sender MUST report the end of the /ANSam event and the beginning of an ANSam event, continuing in this way until the tone is removed.

3.2.3 V.25 Events

V.25 [N-13] is a start-up protocol antedating V.8 [N-11] and V.8bis [I-19]. It specifies the exchange of two tone signals:

CT:

"The calling tone consists of a series of interrupted bursts of 1300 hz tone, on for a duration of not less than 0.5 s and not more than 0.7 s and off for a duration of not less than 1.5 s and..."
not more than 2.0 s." [N-13]. Modems not starting with the V.8 call initiation signal often use this tone.

ANS:

Answering tone. This 2100 Hz tone is used to disable echo suppression for data transmission [N-13], [N-10]. For fax machines, Recommendation T.30 [N-10] refers to this tone as called terminal identification (CED) answer tone. ANS differs from V.8 ANSam in that ANSam varies in amplitude due to modulation by a 15 Hz signal.

V.25 specifically includes procedures for disabling echo suppressors as defined by ITU-T Rec. G.164 [N-6]. However, G.164 echo suppressors have now for the most part been replaced by G.165 [N-7] echo cancellers, which require phase reversals in the disabling tone (see ANSam above). As a result, V.25 was modified in July, 2001 to say that phase reversal in the ANS tone is required if echo cancellers are to be disabled.

One possible V.25 sequence is as follows:

1) The calling terminal starts generating CT as soon as the call is connected.

2) The called terminal waits in silence for 1.8 to 2.5 s after answer, then begins to transmit ANS continuously. If echo cancellers are on the line the phase of the ANS signal is reversed every 450 ms. ANS will not reach the calling terminal until the echo control equipment has been disabled. Since this takes about a second it can only happen in the gap between one burst of CT and the next.

3) Following detection of ANS, the calling terminal may stop generating CT immediately or wait until the end of the current burst to stop. In any event, it must wait at least 400 ms (at least 1 s if phase reversal of ANS is being used to disable echo cancellers) after stopping CT before it can generate the calling station response tone. This tone is modem-specific, not specified in V.25.

4) The called terminal plays out ANS for 2.6 to 4.0 seconds or until it has detected calling station response for 100 ms. It waits 55-95 ms (nominal 75 ms) in silence. (Note that the upper limit of 95 ms is rather close to the point at which echo control may reestablish itself.) If the reason for ANS termination was timeout rather than detection of calling station response, the called terminal begins to play out ANS again to maintain disabling of echo control until the calling station responds.
The events defined for V.25 signalling are shown in Table 5. The gateway at the calling end SHOULD use a packetization interval smaller than the nominal duration of a CT burst, to ensure that CT playout at the called end precedes the sending of ANS from that end.

The gateway at the called end MUST report ANS as soon as it is recognized, providing updates at reasonable intervals as it continues. However, an ANS event packet SHOULD NOT be sent until it is possible to discriminate between an ANS event and an ANSam event (see V.8 events, above). If a phase reversal is detected, the sender MUST report completion of the ANS event and beginning of the /ANS event at the time that the reversal was detected. If another phase reversal is detected, the sender MUST report the end of the /ANS event and the beginning of an ANS event, continuing in this way until the tone is removed.

<table>
<thead>
<tr>
<th>Event</th>
<th>Frequency Hz</th>
<th>Encoding (decimal)</th>
<th>Type</th>
<th>Volume?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer tone (ANS)</td>
<td>2100</td>
<td>32</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>/ANS</td>
<td>2100 rev</td>
<td>33</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>CT</td>
<td>1300</td>
<td>49</td>
<td>tone</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 4: Events for V.25 signals

3.2.4 T.30 Events

ITU-T Recommendation T.30 [N-10] defines the procedures used by Group III FAX terminals. The pre-message procedures for which the events of this section are defined are used to identify terminal capabilities at each end and negotiate operating mode. Post-message procedures are also included, to handle cases such as multiple document transmission. FAX terminals support a wide variety of protocol stacks, so T.30 has a number of options for control protocols and sequences.

T.30 defines two tone signals used at the beginning of a call. The CNG signal is sent by the calling terminal. It is a pure 1100 Hz tone played in bursts: 0.5 s on, 3 s off. It continues until timeout or until the calling terminal detects a response.

The called terminal waits in silence for at least 200 ms. It then may return CED tone, which is identical to V.25 ANS, or else V.8 ANSam if it has V.8 capability. If ANSam is returned and the calling
terminal has V.8 capability, it transmits CI to begin a V.8 negotiation. Otherwise, the calling and called terminals enter the T.30 negotiation phase.

In the negotiation phase the terminals exchange binary messages using V.21 signals, high channel frequencies only. Each message is preceded by a one-second (nominal) preamble consisting entirely of HDLC flag octets (0x7E). This flag has the function of preparing echo control equipment for the message which follows.

The pre-transfer messages exchanged using the V.21 coding are:

Digital Identification Signal (DIS):

Characterizes the standard ITU-T capabilities of the called terminal.

Digital Transmit Command (DTC):

The digital command response to the standard capabilities identified by the DIS signal.

Digital Command Signal (DCS):

The digital set-up command responding to the standard capabilities identified by the DIS signal.

Confirmation To Receive (CFR):

A digital response confirming that the entire pre-message procedure has been completed and the message transmissions may commence.

If the calling terminal wishes to transmit a document, the three messages exchanged are DIS (from the called terminal), DCS, and CFR. If it wishes to receive, the sequence changes to DIS, DTC, DCS, and CFR. Each message may consist of multiple frames, each bounded by HDLC flags. The messages are organized as a series of octets, but T.30 calls for the insertion of extra "0" bits to prevent spurious recognition of HDLC flags.

T.30 also provides for the transmission of control messages after document transmission has completed (e.g., to support transmission of multiple documents). The transition back from the modem used for document transmission (V.17 [I-10], V.27ter [I-11], V.29 [I-12], V.34 [I-13]) to V.21 signalling is preceded by 75 ms (nominal) of silence). Control message transmission is preceded by the preamble described above.
Before CFR the transmitting terminal sends a training signal consisting of a steady string of V.21 high channel zeros (1850 Hz tones) for 1.5 s. The sender MAY report this training signal either as a single extended V.21 upper channel "0" event, or as a series of "0" events of normal duration. The event(s) MUST be reported as soon as the training signal is recognized, with updates at reasonable intervals thereafter.

Applications supporting T.30 signalling using the telephone-events payload MUST transfer T.30 messages in the form of sequences of bits, using the V.21 bit events defined in section 3.2.2. The transmitted information MUST include the complete contents of the message: the initial HDLC flags, the information field, the checksum, and the terminating HDLC flags.

Transmission MUST also include the extra "0" bits added to prevent false recognition of HDLC flags at the receiver. Implementors should note that these extra "0" bits mean that in general T.30 messages as transmitted on the wire will not come out to an even multiple of octets. Sending implementations MAY choose to vary the packetization interval to include exactly one octet of information plus any extra "0" bits inserted into that octet.

The events defined for T.30 signalling are shown in Table 6. The CED and /CED events represent exactly the same tone signals as V.8 ANS and /ANS, and are given the same codepoints; they are reproduced here only for convenience. For reporting of CNG, the gateway at the calling end SHOULD use a packetization interval smaller than the nominal duration of a CNG burst, to ensure that CED has time to disable echo control before it times out.

The gateway at the called end MUST report CED as soon as it is recognized, providing updates at reasonable intervals as it continues. However, a CED event packet SHOULD NOT be sent until it is possible to discriminate between a CED event and an ANSam event (see V.8 events, above). If a phase reversal is detected, the sender MUST report completion of the CED event and beginning of the /CED event at the time that the reversal was detected. If another phase reversal is detected, the sender MUST report the end of the /CED event and the beginning of an CED event, continuing in this way until the tone is removed.

<table>
<thead>
<tr>
<th>Event</th>
<th>Frequency Hz</th>
<th>Encoding (decimal)</th>
<th>Type</th>
<th>Volume?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNG (Calling tone)</td>
<td>1100</td>
<td>36</td>
<td>tone</td>
<td>yes</td>
</tr>
</tbody>
</table>
4. RTP Payload Format for Telephony Tones

4.1 Introduction

As an alternative to describing tones and events by name, as described in section 2, it is sometimes preferable to describe them by their waveform properties. In particular, recognition is faster than for naming signals since it does not depend on recognizing durations or pauses.

There is no single international standard for telephone tones such as dial tone, ringing (ringback), busy, congestion ("fast-busy"), special announcement tones or some of the other special tones, such as payphone recognition, call waiting or record tone. However, ITU-T Recommendation E.180 [I-20] notes that across all countries, these tones share a number of characteristics:

- Telephony tones consist of either a single tone, the addition of two or three tones or the modulation of two tones. (Almost all tones use two frequencies; only the Hungarian "special dial tone" has three.) Tones that are mixed have the same amplitude and do not decay.

- In-band tones for telephony events are in the range of 25 Hz (ringing tone in Angola) to 2600 Hz (the tone used for line signalling in SS No. 5 and R1). The in-band telephone frequency range is limited to 3400 Hz. R2 defines a 3825 Hz out-of-band tone for line signalling on analogue trunks. (The piano has a range from 27.5 to 4186 Hz.)

- Modulation frequencies range between 15 (ANSam tone) to 480 Hz (Jamaica). Non-integer frequencies are used only for frequencies of 16 2/3 and 33 1/3 Hz. (These fractional frequencies appear to be derived from AC power grid frequencies.)

### Table 5: Events for T.30 signals

<table>
<thead>
<tr>
<th>Event</th>
<th>Code</th>
<th>Duration</th>
<th>Type</th>
<th>Mod.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CED</td>
<td>2100</td>
<td>32</td>
<td>tone</td>
<td>yes</td>
</tr>
<tr>
<td>/CED</td>
<td>2100</td>
<td>33</td>
<td>tone</td>
<td>yes</td>
</tr>
</tbody>
</table>

CED (Called tone)

/CED (ph. rev.)
- Tones that are not continuous have durations of less than four seconds.

- ITU Recommendation E.180 [I-20] notes that different telephone companies require a tone accuracy of between 0.5 and 1.5%. The Recommendation suggests a frequency tolerance of 1%.

### 4.2 Examples of Common Telephone Tone Signals

As an aid to the implementor, Table 15 summarizes some common tones. The rows labeled "ITU ..." refer to ITU-T Recommendation E.180 [I-20]. Note that there are no specific guidelines for these tones. In the table, the symbol "*" indicates addition of the tones, without modulation, while "**" indicates amplitude modulation. The meaning of these tones is described in section 3.3.

<table>
<thead>
<tr>
<th>Tone Name</th>
<th>Frequency</th>
<th>On Period (s)</th>
<th>Off Period (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNG</td>
<td>1100</td>
<td>0.5</td>
<td>3.0</td>
</tr>
<tr>
<td>V.25 CT</td>
<td>1300</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>CED</td>
<td>2100</td>
<td>3.3</td>
<td>--</td>
</tr>
<tr>
<td>ANS</td>
<td>2100</td>
<td>3.3</td>
<td>--</td>
</tr>
<tr>
<td>ANSam</td>
<td>2100*15</td>
<td>3.3</td>
<td>--</td>
</tr>
<tr>
<td>V.21 &quot;0&quot; bit,</td>
<td>1180</td>
<td>0.00333</td>
<td>--</td>
</tr>
<tr>
<td>channel 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V.21 &quot;1&quot; bit,</td>
<td>980</td>
<td>0.00333</td>
<td>--</td>
</tr>
<tr>
<td>channel 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V.21 &quot;0&quot; bit,</td>
<td>1850</td>
<td>0.00333</td>
<td>--</td>
</tr>
<tr>
<td>channel 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V.21 <em>&quot;1&quot;</em> bit,</td>
<td>1650</td>
<td>0.00333</td>
<td>--</td>
</tr>
<tr>
<td>channel 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITU dial tone</td>
<td>425</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>U.S. dial</td>
<td>350+440</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>tone</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ITU ringing tone 425 0.67-1.5 3-5

U.S. ringing tone 440+480 2.0 4.0

ITU busy tone 425

U.S. busy tone 480+620 0.5 0.5

ITU congestion tone 425

U.S. congestion tone 480+620 0.25 0.25

Table 6: Examples of telephony tones

4.3 Use of RTP Header Fields

4.3.1 Timestamp

The RTP timestamp reflects the measurement point for the current packet. The event duration described in section 4.3.3 extends forwards from that time.

4.3.2 Marker Bit

The tones payload type uses the marker bit to distinguish the first RTP packet reporting a given instance of a tone from succeeding packets for that tone. The marker bit SHOULD be set to 1 for the first packet, and to 0 for all succeeding packets relating to the same tone.

4.3.3 Payload Format

Based on the characteristics described above, this document defines an RTP payload format called "tone" that can represent tones consisting of one or more frequencies. (The corresponding MIME type is "audio/tone"). The default timestamp rate is 8000 Hz, but other rates may be defined. Note that the timestamp rate does not affect the interpretation of the frequency, just the durations.
In accordance with current practice, this payload format does not have a static payload type number, but uses a RTP payload type number established dynamically and out-of-band.

The payload format is shown in Figure 2.

```
+----------------+-----------------+-----------------+
| modulation     | T               | volume          |
| duration       |                 |                 |
+----------------+-----------------+-----------------+
| R R R R | frequency | R R R R | frequency |
+----------------+-----------------+-----------------+
| R R R R | frequency | R R R R | frequency |
```

Figure 2: Payload Format for Tones

The payload contains the following fields:

**modulation:**

The modulation frequency, in Hz. The field is a 9-bit unsigned integer, allowing modulation frequencies up to 511 Hz. If there is no modulation, this field has a value of zero.

**T:**

If the "T" bit is set (one), the modulation frequency is to be divided by three. Otherwise, the modulation frequency is taken as is.

This bit allows frequencies accurate to 1/3 Hz, since modulation frequencies such as 16 2/3 Hz are in practical use.

**volume:**

The power level of the tone, expressed in dBm0 after dropping the sign, with range from 0 to -63 dBm0. (Note: A preferred level range for digital tone generators is -8 dBm0 to -3 dBm0.)
duration:

The duration of the tone, measured in timestamp units. The tone begins at the instant identified by the RTP timestamp and lasts for the duration value. The value of zero is not permitted and tones with such a duration SHOULD be ignored.

The definition of duration corresponds to that for sample-based codecs, where the timestamp represents the sampling point for the first sample.

frequency:

The frequencies of the tones to be added, measured in Hz and represented as a 12-bit unsigned integer. The field size is sufficient to represent frequencies up to 4095 Hz, which exceeds the range of telephone systems. A value of zero indicates silence. A single tone can contain any number of frequencies. If the number of frequencies it contains is odd, padding SHALL be added to bring the packet to a 32-bit boundary. ([RFC 3550][1] requires that padding be set to all zeroes.)

R:

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

4.3.4 Optional MIME Parameters

The "rate" parameter describes the sampling rate, in Hertz. The number is written as a floating point number or as an integer. If omitted, the default value is 8000 Hz.

4.4 Procedures

This section defines the procedures associated with the tones payload type.

4.4.1 Sending Procedures

As indicated by the examples in Table 15, the duration of an individual tone may range from a few milliseconds to a number of seconds. Timing considerations dictate some general guidelines for how these two extremes should be handled by the sender. For tones directed to human listeners, timing is not critical, within a tolerance of 100 ms or so at either beginning or end. For tones directed to remote equipment, the most critical aspect of timing is intra-stream time relationships -- that is, the individual tone durations and the interval between tones for a related sequence of
them. The timing of the start of playout of a related sequence is less critical within limits.

In the case of longer-duration tones, implementations SHOULD expect to generate multiple RTP packets for the same tone instance. The considerations just enumerated suggest that a packetization interval in the order of 50 ms may be acceptable, in terms of the initial delay it imposes on remote playback. Implementations MAY adjust the packetization interval to suit the nature of the tones being played out. The packetization interval SHOULD remain constant until the tone ends in order not to distort playout times through buffer under-runs.

The RTP timestamp MUST be updated for each packet generated (in contrast, for instance, to the timestamp for packets carrying telephone-events). The first RTP packet for a tone SHOULD have the marker bit set to 1. Subsequent packets for the same tone SHOULD have the marker bit set to 0, and the RTP timestamp in each subsequent packet MUST equal the sum of the timestamp and the duration in the preceding packet. A final RTP packet SHOULD be generated as soon as the end of the tone is detected, without waiting for the latest packetization interval to elapse.

If the tones are meant for machine consumption, the intervals between them are potentially critical. Implementations may be aware of this situation, or may infer it from a heuristic such as that the tones are less than a second in duration. In this situation, it is RECOMMENDED that if a tone follows another tone within a period of 100 ms or less, the new tone should be reported as soon as it has been identified. The suggested 50 ms packetization interval should be applied to subsequent reports for the same tone.

The above advice applies to tones lasting in the order of 25 ms or more. Shorter tones, which are likely to be from modems, SHOULD be reported in batches. The tones payload format requires that each tone be reported in a separate RTP packet, but it is RECOMMENDED that multiple RTP packets be reported in the same UDP packet. Individual tones should be given their actual durations (i.e., from transition point to transition point) rather than reporting a new tone at each bit boundary.

4.4.2 Receiving Procedures

Receiving implementations play out the tones as received. When playing out successive tone reports for the same tone (marker bit is zero, the RTP timestamp is contiguous with that of the previous RTP packet, and payload content is identical), the receiving implementation SHOULD continue the tone without change or a break.
5. Application Considerations

5.1 Combining Tones and Named Events

Gateways which send signalling events via RTP MAY send both named signals (section named) and the tone representation (section tones) as a single RTP session, using the redundancy mechanism defined in RFC 2198 [N-2] to interleave the two representations. It is generally a good idea to send both, since it allows the receiver to choose the appropriate rendering.

If a gateway cannot present a tone representation, it SHOULD also send the audio tones as regular RTP audio packets using either the codec used for regular speech signals or a codec that is known to carry such signals successfully (e.g., PCMU).

Some low-rate codecs cannot accurately represent certain tones, such as DTMF.

5.2 Simultaneous Generation of Audio and Events

A source can choose between four approaches:

Events and audio:

The source sends events and encoded audio packets (e.g., PCMU or the codec used for speech signals) for the same time instant. In that mode, events are treated as redundant encodings for the encoded audio stream.

Events only:

The source does not send encoded audio while event tones are active and only sends named events, without any redundancy beyond the periodic updates of longer-lasting events.

Events only, with redundancy:

The source does not send encoded audio while event tones are active. It only sends named events, but uses RFC 2198 [N-2] redundancy, with named events as both primary and redundant encodings.

Events and audio, with redundancy:
During an event, the source sends both named events and audio, using RFC 2198 [N-2] to interleave audio data, current and redundant named events.

The choices above do not affect the event redundancy mechanism described in section 2.6.

Note that a period covered by a named event may overlap in time with a period of audio encoded by other means. This is likely to occur at the onset of a tone and is necessary to avoid possible errors in the interpretation of the reproduced tone at the remote end. Implementations supporting this payload format MUST be prepared to handle the overlap. It is RECOMMENDED that gateways only render the encoded tone since the audio may contain spurious tones introduced by the audio compression algorithm. However, it is anticipated that these extra tones in general should not interfere with recognition at the far end.

5.3 Strategies For Handling FAX and Modem Signals

As described in section 3.2, the typical data application involves a pair of gateways interposed between two terminals, where the terminals are in the PSTN. The gateways are likely to be serving a mixture of voice and data traffic, and need to adopt payload types appropriate to the media flows as they occur. If voice compression is in use for voice calls, this means that the gateways need the flexibility to switch to other payload types when data streams are recognized.

Within the established IETF framework, this implies that the gateways must negotiate the potential payloads (voice, telephone-events, tones, voice-band data, T.30 FAX [I-8], and possibly RFC 2793 [I-1] text and CLEARMODE [I-2] octet streams) as separate payload types. From a timing point of view, this is most easily done at the beginning of a call, but results in an over-allocation of resources at the gateways and in the intervening network.

One alternative is to use named events to buy time while out-of-band signals are exchanged to update to the new payload type applicable to the session. Thanks to the events defined in section 3.2, this is a viable approach for sessions beginning with V.8, V.8bis, T.30, or V.25 control sequences.

Named data-related events also allow gateways to optimize their operation when data signals are received in a relatively general form. One example is the use of V.8-related events to deduce that the voice-band data being sent in a G.711 payload comes from a higher-speed modem and therefore requires disabling of echo cancellors.
All of the control procedures described in section 3.2 eventually give way to data content. As mentioned above, this content will be carried by other payload types. Receiving gateways MUST be prepared to switch to the other payload type within the time constraints associated with the respective applications. (For several of the procedures documented below, the sender provides 75 ms of silence between the initial control signalling and the sending of data content.) In some cases (V.8bis [I-19], T.30 [N-10]), further control signalling may happen after the call has been established.

A possible strategy is to send both telephone-events and the data payload in an RFC 2198 redundancy arrangement. The receiving gateway then propagates the data payload whenever no event is in progress. For this to work, the data payload and events (when present) MUST cover exactly the same time period; otherwise spurious events will be detected downstream.

Note that there are a number of cases where no control sequence will precede the data content. This is true, for example, for a number of legacy text terminal types. In such instances, the events defined in section 3.2.6 in particular MAY be sent to help the remote gateway optimize its handling of the alternative payload.

5.4 Examples

5.4.1 Use of RFC 2198 Redundancy With Named Events

A typical RTP packet, where the user is just dialing the last digit of the DTMF sequence "911", is shown in Figure 3. The first digit was 200 ms long (1600 timestamp units) and started at time 0, the second digit lasted 250 ms (2000 timestamp units) and started at time 800 ms (6400 timestamp units), the third digit was pressed at time 1.4 s (11,200 timestamp units) and the packet shown was sent at 1.45 s (11,600 timestamp units). The frame duration is 50 ms. To make the parts recognizable, Figure 3 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 telephone event payload, respectively.
Table 16 shows all packets up to and including the packet shown in the figure. The last three columns describe the duration fields in the event payloads. The timestamp offset is not shown. We assume here that the digits happen to start on a 50 ms multiple, which is somewhat unlikely.
5.4.2 Combined Tone and Telephone-event Payloads

The payload formats in sections 2 and 4 can be combined into a single payload using the method specified in RFC 2198 [N-2]. Figure 4 shows an example. In that example, the RTP packet combines two "tone" and one "telephone-event" payloads. The payload types are chosen arbitrarily as 97 and 98, respectively, with a sample rate of 8000 Hz. Here, the redundancy format has the dynamic payload type 96.

The packet represents a snapshot of U.S. ringing tone, 1.5 seconds (12,000 timestamp units) into the second "on" part of the 2.0/4.0 second cadence, i.e., a total of 7.5 seconds (60,000 timestamp units) into the ring cycle. The 440 + 480 Hz tone of this second cadence started at RTP timestamp 48,000. Four seconds of silence preceded it, but since RFC 2198 only has a fourteen-bit offset, only 2.05 seconds (16383 timestamp units) can be represented. Even though the tone sequence is not complete, the sender was able to determine that this is indeed ringback, and thus includes the corresponding named event.
Figure 4: Combining tones and events in a single RTP packet
6. MIME Registration

6.1 audio/telephone-event

MIME media type name: audio
MIME subtype name: telephone-event
Required parameters: none.
Optional parameters:

The "events" parameter lists the events supported by the implementation. Events are listed as one or more comma-separated elements. Each element can either be a single integer or two integers separated by a hyphen. No white space is allowed in the argument. The integers designate the event numbers supported by the implementation.

The "rate" parameter describes the sampling rate, in Hertz. The number is written as a floating point number or as an integer. If omitted, the default value is 8000 Hz.

Encoding considerations:

This type is only defined for transfer via RTP [N-4].

Security considerations:

See the "Security Considerations" section (section 7) in this document.

Interoperability considerations: none

Published specification: This document.

Applications which use this media:

The telephone-event audio subtype supports the transport of events occurring in telephone systems over the Internet.

Additional information:

1. Magic number(s): N/A
2. File extension(s): N/A
3. Macintosh file type code: N/A
6.2 audio/tone

MIME media type name: audio

MIME subtype name: tone

Required parameters: none

Optional parameters:

The "rate" parameter describes the sampling rate, in Hertz. The number is written as a floating point number or as an integer. If omitted, the default value is 8000 Hz.

Encoding considerations:

This type is only defined for transfer via RTP [N-4].

audio/tone MIME body parts contain binary data. A content-transfer-encoding of "binary" is strongly encouraged for messaging environments which support binary transport. A content-transfer-encoding of base-64 (and the associated transformation) is strongly encouraged for messaging environments which do not support binary transfer.

Security considerations:

See the "Security Considerations" section (section 7) in this document.

Interoperability considerations: none

Published specification: This document.

Applications which use this media: The tone audio subtype supports the transport of pure composite tones, for example those commonly used in the current telephone system to signal call progress.

Additional information:

1. Magic number(s): N/A

2. File extension(s): N/A

3. Macintosh file type code: N/A
7. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification (RFC 3550 [N-4]), and any appropriate RTP profile (for example RFC 3551 [I-22]). 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.

This payload type does not exhibit any significant non-uniformity in the receiver side computational complexity for packet processing to cause a potential denial-of-service threat.

Additional security considerations are described in RFC 2198 [N-2].

A security review of this payload format found no additional considerations.

8. IANA Considerations

This document defines two new RTP payload formats, named telephone-event and tone, and associated Internet media (MIME) types, audio/telephone-event and audio/tone. It also defines a number of codepoints for events.

Within the audio/telephone-event type, events MUST be registered with IANA. Registrations are subject to approval by the current chair of the IETF audio/video transport working group, or by an expert designated by the transport area director if the AVT group has closed. The initial registry content is shown in the following table, and consists of the events defined in section 3 of this document.

The meaning of new events MUST be documented either as an RFC or an equivalent standards document produced by another standardization body, such as ITU-T.

<table>
<thead>
<tr>
<th>Event Code</th>
<th>Event Name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>DTMF digit &quot;0&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>RFC Reference</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>1</td>
<td>DTMF digit &quot;1&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>2</td>
<td>DTMF digit &quot;2&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>3</td>
<td>DTMF digit &quot;3&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>4</td>
<td>DTMF digit &quot;4&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>5</td>
<td>DTMF digit &quot;5&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>6</td>
<td>DTMF digit &quot;6&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>7</td>
<td>DTMF digit &quot;7&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>8</td>
<td>DTMF digit &quot;8&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>9</td>
<td>DTMF digit &quot;9&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>10</td>
<td>DTMF digit &quot;*&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>11</td>
<td>DTMF digit &quot;#&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>12</td>
<td>DTMF digit &quot;A&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>13</td>
<td>DTMF digit &quot;B&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>14</td>
<td>DTMF digit &quot;C&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>15</td>
<td>DTMF digit &quot;D&quot;</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>32</td>
<td>ANS (V.25 Answer tone)</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td></td>
<td>Also known as CED (T.30 Called tone)</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>/ANS (V.25 Answer tone with phase shift)</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td></td>
<td>Also known as /CED (T.30 Called tone with phase shift)</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>ANSam (V.8 amplitude modified Answer tone)</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>35</td>
<td>/ANSam (V.8 amplitude modified Answer tone with phase shift)</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>36</td>
<td>CNG (T.30 Calling tone)</td>
<td>&lt;This RFC&gt;</td>
</tr>
<tr>
<td>37</td>
<td>V.21 channel 1, &quot;0&quot; bit</td>
<td>&lt;This RFC&gt;</td>
</tr>
</tbody>
</table>
9. Acknowledgements

The suggestions of the Megaco working group are gratefully acknowledged. Detailed advice and comments were provided by Hisham Abdelhamid, Flemming Andreasen, Fred Burg, Steve Casner, Dan Deliberato, Fatih Erdin, Bill Foster, Mike Fox, Mehryar Garakani, Gunnar Hellstrom, Rajesh Kumar, Terry Lyons, Steve Magnell, Zarko Markov, Kai Miao, Satish Mundra, Kevin Noll, Vern Paxson, Oren Peleg, Colin Perkins, Raghavendra Prabhu, Moshe Samoha, Todd Sherer, Adrian Soncodi, Yaakov Stein, Mira Stevanovic, Alex Urquizu and Herb Wildfeur.

10. Authors

Henning Schulzrinne  
Dept. of Computer Science  
Columbia University  
1214 Amsterdam Avenue  
New York, NY 10027  
USA

electronic mail: schulzrinne@cs.columbia.edu

Scott Petrack  
eDial  
USA

electronic mail: scott.petrack@edial.com

Tom Taylor  
Nortel Networks  
1852 Lorraine Ave.  
Ottawa, Ontario
References

12.1 Normative References


Informative References


[I-19] International Telecommunication Union, "Procedures for the identification and selection of common modes of operation between data circuit-terminating equipments (DCEs) and between data terminal equipments (DTEs) over the public switched telephone network and on leased point-to-point telephone-type circuits", Recommendation V.8bis, ITU-T, Geneva, Switzerland, Nov. 2000.


Disclaimer of validity:

"The IETF 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 this 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. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR 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 this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Copyright Notice

Copyright (C) The Internet Society (2004). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Disclaimer

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY 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 HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.