RTP payload format for UEMCLIP speech codec
draft-ietf-avt-rtp-uemclip-01

Status of this Memo

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 becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

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.

This Internet-Draft will expire on March 24, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2007).
Abstract

This document describes the RTP payload format of UEMCLIP, an enhanced speech codec of ITU-T G.711. The bitstream has a scalable structure with an embedded u-law bitstream, also known as PCMU, thus providing a handy transcoding operation between narrowband and wideband speech.

Table of Contents

1. Introduction .................................................. 3
   1.1. Terminology ............................................. 3
2. Media Format Background ....................................... 4
3. Payload Format .................................................. 6
   3.1. RTP Header Usage ....................................... 6
   3.2. Multiple frames in an RTP packet ......................... 7
   3.3. Payload Data ............................................... 7
      3.3.1. Main Header ......................................... 7
      3.3.2. Sub-layer ............................................. 11
4. G.711 interoperability ......................................... 13
5. Congestion Control Considerations ............................. 14
6. Payload Format Parameters ..................................... 15
   6.1. Media type registration .................................. 15
   6.2. Mapping to SDP Parameters ............................... 16
      6.2.1. Dynamic transmission specification .................. 17
   6.3. Offer-Answer Model Considerations ....................... 18
      6.3.1. Offer-Answer Guidelines ............................ 18
      6.3.2. Examples ............................................. 19
7. Security Considerations ....................................... 21
8. IANA Considerations ........................................... 22
9. Normative References .......................................... 23
Authors’ Addresses ............................................... 24
Intellectual Property and Copyright Statements .................. 25
1. Introduction

This document specifies the payload format for sending UEMCLIP encoded speech using the Real-time Transport Protocol (RTP) [3]. UEMCLIP is an enhanced version of u-law ITU-T G.711, and designed to help the market for smooth transition towards the forthcoming wideband communication environment and while maintaining the interoperability and less transcoding load with the existing terminals, in which the implementation of G.711 is mandatory.

The background and the basic idea of the media format is described in Section 2. The details of the payload format is given in Section 3. The interoperability with G.711 issues are discussed in Section 4, and the consideration for congestion control is in Section 5. In Section 6.1, a media type registration for UEMCLIP RTP payload format and SDP mappings are provided.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [1].
2. Media Format Background

UEMCLIP stands for "U-law EMbedded Coder for Low-delay IP communication", and is basically an enhanced version of u-law ITU-T G.711, otherwise known as PCMU [8]. It is developed for VoIP (Voice over Internet Protocol) applications, and is especially suitable for wideband multi-point conferencing. The main goal of this codec is to provide a wideband communication platform that is highly interoperable with existing terminals equipped with G.711, and to stimulate the market to gradually shift to the wideband communication. Because the G.711 bitstream is embedded in the bitstream, costly transcoding can be avoided especially when interoperating with narrowband terminals.

This document does not discuss the implementation details of the encoder and decoder, but only describes the bitstream format. The implementation details will be available by other means.

Because of its scalable nature, there are a number of sub-bitstreams (sub-layer) within a UEMCLIP bitstream. By choosing appropriate sub-layers, the codec can adapt to the following requirements:

- Sampling frequency,
- Number of channels,
- Speech quality, and
- Bit-rate.

The current implementation of UEMCLIP codec includes three sub-coders, as shown in Table 1. The core layer is G.711 core, and other two are quality and bandwidth enhancement layers with bit-rate of 16 kbit/s each.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
<th>Bit-rate</th>
<th>Coding algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>G.711 core</td>
<td>64</td>
<td>u-law PCM</td>
</tr>
<tr>
<td>b</td>
<td>Lower-band</td>
<td>16</td>
<td>Time domain block</td>
</tr>
<tr>
<td></td>
<td>enhancement</td>
<td></td>
<td>quantization</td>
</tr>
<tr>
<td>c</td>
<td>Higher-band</td>
<td>16</td>
<td>MDCT block quantization</td>
</tr>
</tbody>
</table>

Table 1: Sub-layer description
Based on these sub-layers, UEMCLIP codec operates in four modes as shown in Table 2. Here, "Fs" is the sampling frequency in kHz. The absent Modes 2 and 5 are reserved for future extension to 32 kHz sampling modes. As the mode definition is expected to grow, any other modes not defined in this table MUST NOT be used for compatibility and interoperability reasons.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Ch</th>
<th>Fs</th>
<th>Layer a</th>
<th>Layer b</th>
<th>Layer c</th>
<th>Bit-rate w/o headers [kbps]</th>
<th>Total bit-rate [kbps]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>8</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>64</td>
<td>68.8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>16</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>80</td>
<td>85.6</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>8</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>80</td>
<td>85.6</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>16</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>96</td>
<td>102.4</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Mode description

UEMCLIP bitstream contains internal headers and other side-information apart from the layer data. This results in total bit-rate larger than the sum of the layers shown in the above table. The detail of the internal headers and auxiliary information are described in Section 3.3.1.

Defining the sampling frequency and the number of channels does not result in a singular mode, i.e., there can be multiple modes for the same sampling frequency or number of channels. The supported modes would differ from the implementations, thus the sender and the receiver must exchange what mode to use for transmission.
3. Payload Format

As an RTP payload, UEMCLIP bitstream can contain one or more frames as shown in Figure 1.

```
+-----------------------------+-----------------------------+
| RTP Header                  | one or more frames of UEMCLIP|
| +++++++++++++++++++++++++++| ++++++++++++++++++++++++++++|
```

Figure 1: RTP payload format

UEMCLIP bitstream has a scalable structure, thus it is possible to reconstruct the signal by decoding a part of it. A UEMCLIP frame is composed of a main header (MH) followed by one or more (up to three) sub-layers (SL) as shown in Figure 2.

```
+-------------+-------------+-+
| MH | SL #1 | ... |
|     |       |   |
```

Figure 2: A UEMCLIP frame (bitstream format)

As a sub-layer, the core layer, i.e., "Layer a", MUST always be included. It should be noted that the location of the core layer may not be located at the top. The decoder MUST always refer to the layer ID for proper decoding.

The UEMCLIP bitstream does not include the following information: Mode and sampling frequency (Fs). As described before, this information SHOULD be exchanged while establishing a connection, for example, by means of SDP.

3.1. RTP Header Usage

Each RTP packet starts with a fixed RTP header, as explained in [3]. The following fields of the RTP fixed header used specifically for UEMCLIP streams are emphasized:
Payload type: The assignment of an RTP payload type for this packet format is outside the scope of this document, however, it is expected that a payload type in the dynamic range shall be assigned.

Timestamp: This encodes the sampling instant of the first speech signal sample in the RTP data packet. For UEMCLIP streams, the RTP timestamp MUST advance based on a multiple of 8 kHz, and in case the sampling rate can change during a session, this figure should equal to the maximum rate (in Hz) given in the mode range (see Section 6.2.1). For example, during a 8 kHz session, if a transition to a 16 kHz mode is allowed, the time stamp SHOULD advance using 16 kHz clock rate. For fixed modes, it should be either 8 or 16 kHz, based on the sampling rate.

Marker bit: If the codec is used for applications with discontinuous transmission (DTX, or silence compression), the first packet after a silence period during which packets have not been transmitted contiguously SHOULD have the marker bit in the RTP data header set to one. The marker bit in all other packets MUST be zero. Applications without DTX MUST set the marker bit to zero.

3.2. Multiple frames in an RTP packet

More than one UEMCLIP frame may be included in a single RTP packet by a sender. However, senders have the following additional restrictions:

- A single RTP packet SHOULD NOT include more UEMCLIP frames than will fit in the MTU of the RTP transport protocol.

- All frames contained in a single RTP packet MUST be of the same mode.

- Frames MUST NOT be split between RTP packets.

It is RECOMMENDED that the number of frames contained within an RTP packet be consistent with the application. Since UEMCLIP is designed for telephony application where delay has a great impact on the quality, then fewer frames per packet for lower delay, is preferable.

3.3. Payload Data

3.3.1. Main Header

The main header (MH) is placed at the top of a frame and has size of 10 bytes with additional optional enhanced header size. The content of the main header is defined in Figure 3.
Identification (ID): 8 bits

The value should be "0x95".

Byte size (BS): 16 bits

 Indicates the size in bytes of the rest of the UEMCLIP frame, i.e., the frame size minus 3 bytes (of ID and BS). It is encoded in network byte-order.

Mixing information (MX): 8 bits

Mixing information field.

Packet-loss Concealment information (PC): 40 bits

Packet-loss concealment (PLC) information field.

Enhanced-header Size (ES): 8 bits

Size of EH (enhanced header) in bytes.

Enhanced header (EH): 8*ES bits

Content of the enhanced header. When ES is 0, the enhanced header is non-existent.
3.3.1.1. Mixing information field

```
0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+-+-+-+
|C|R|V|   PW1   |
+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Mixing information field (MX)

Check bit #1 (C1): 1 bit

Validity flag of V1 and PW1. This bit being "1" indicates that both parameters are valid, and "0" indicates that the parameters should be ignored.

Reserved bit #1 (R1): 1 bit

This bit should be ignored.

VAD flag #1 (V1): 1 bit

Voice activity detection flag of the current frame. This flag being "1" indicates that the frame is an active (voice) segment, and "0" indicates that it is an inactive (non-voice) or a silent segment. This flag is specifically designed for mixing information and DTX judgement based this flag is not recommended.

Power #1 (PW1): 5 bits

Signal power code of the current frame.

3.3.1.2. PLC information field

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
|C|C|K|P1|   R|   P2|   PW2|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-
```

```
| 2|3|2|2|6|3|4|4|6|6|6|6|6|6|6|
```

```
| 0 1 2 3| 0 1 2 3 4 5 6| 0 1 2 3 4 5 6| 0 1 2 3 4 5 6 7|
```

```
|   R5   |
```

```
| 0 1 2 3 4 5 6 7|
```

Figure 5: PLC information field (PC)
Check bit #2 (C2): 1 bit

Validity flag of V2, K, P1, P2, and PW2. If the flag is "1", it means that all these parameters are valid, and "0" means that the parameters should be ignored. If any of these parameters is invalid, C1 should be set to "0".

Check bit #3 (C3): 1 bit

Payload validity indicator. This flag is normally set to "0". If a received packet has this flag set to "1", the payload data should be ignored and packet-loss concealment should be performed by the receiver. This flag is used in case of a multi-point conferencing, where the upstream packet was lost and the mixing server did not execute packet-loss concealment.

Reserved bit #2 (R2): 1 bit

This bit should be ignored.

VAD flag #2 (V2): 1 bit

Voice activity detection flag of the current frame. This may be as same as V1 in the mixing information, and may not be synchronous to the marker bit in the RTP header. This flag is specifically designed for packet-loss concealment and DTX judgement based this flag is not recommended.

Frame indicator (K): 4 bits

This value indicates the frame offset of P2 and PW2. Since it is a better idea to carry the pitch and power parameters as PLC information in a different frame, this frame offset value gives which frame the parameters are to be associated with. Since there are 4 bits allocated, it ranges between "0" and "15".

Reserved bit #3 (R3): 1 bit

This bit should be ignored.

Pitch lag #1 (P1): 7 bits

Pitch code of the current frame. The actual pitch lag is calculated as P1+20 samples in 8-kHz sampling rate. Pitch lag must be 20 <= pitch length <= 120. Codes ranging between "0x65" and "0x7F" are not used.
Reserved bit #4 (R4): 1 bit

This bit should be ignored.

Pitch lag #2 (P2): 7 bits

Pitch code of the offset frame. The actual pitch lag is calculated as P2+20 samples in 8-kHz sampling rate. Pitch lag must be 20 <= pitch length <= 120. Codes ranging between "0x65" and "0x7F" are not used. The offset value is defined as K.

Power #2 (PW2): 8 bits

Signal power code of the offset frame. The offset value is defined as K.

Reserved bits #5 (R5): 8 bits

These bits should be ignored.

3.3.2. Sub-layer

Sub-layer (SL) is a sub-header followed by layer bitstreams, as shown in Figure 6. The sub-header indicates the layer location and the number of bytes.

<table>
<thead>
<tr>
<th>CI</th>
<th>FI</th>
<th>QI</th>
<th>R6</th>
<th>SB</th>
<th>LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>2 3 4 5 6 7</td>
</tr>
</tbody>
</table>

Figure 6: Sub-layer format (SL)

Channel index (CI): 2 bits

Indicates the channel number. For all modes given in Table 2, this should be "0x1". The detail is given in Table 3.

Frequency index (FI): 2 bits

Indicates the frequency number. "0" means that the layer is in the base frequency band, higher number means that the layer is in respective frequency band. The detail is given in Table 3.
Quality index (QI): 2 bits

Indicates the quality layer number. "0" means that the layer is in the base layer, and higher number means that the layer is in respective quality layer. The detail is given in Table 3.

Reserved #6 (R6): 2 bits

Not used (reserved). The value must be "0".

Sub-layer Size (SB): 8 bits

Indicates the byte size of the following sub-layer data.

Layer Data (LD): SB*8 bits

The actual sub-layer data.

### 3.3.2.1. Layer index encoding

The layer index is encoded using values of channel number, quality number, and frequency-band number encoded with 2-bits each, in the appearing order. The last 2 bits are reserved for future use, and all implementation should ignore this field. For all the layers shown in Table 1, the layer indices are shown in Table 3.

<table>
<thead>
<tr>
<th>Layer</th>
<th>CI</th>
<th>FI</th>
<th>QI</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3: Layer indices
4. G.711 interoperability

As given in Section 2, u-law encoded G.711 bitstream (Layer a) is the core layer of a UEMCLIP bitstream, and is always embedded. This means that transcoding from UEMCLIP bitstream to G.711 does not have to undergo decoding and re-encoding procedures, but simple extraction would only suffice. However, this does not apply for the reverse procedure, i.e., transcoding from G.711 to UEMCLIP, because the side information in the main header must be assigned separately.

The transcoding from UEMCLIP to u-law G.711 can be done easily by finding an appropriate sub-layer. Within a frame, the transcoder should look for a sub-layer with layer index "0x00", and subsequent LD which has size of SB*8 bits (UEMCLIP has a 20-ms frame thus, SB=160) are the actual G.711 bitstream data. It should be noted that transcoder should not always expect the core layer to be located right after the main header.

On the other hand, the transcoding from G.711 to UEMCLIP is not entirely straight-forward. Since there are no means to generate enhancement sub-layers, a G.711 bitstream can only be converted to UEMCLIP Mode 0 bitstream. If the original G.711 bitstream is encoded in A-law, it should first be converted to u-law to become the core layer. Because a UEMCLIP frame size is 20 ms, u-law encoded G.711 bitstream MUST be a 160-sample chunk to become a core layer. For the main header contents, when the UEMCLIP encoder is not available, it should follow the following guidelines.

- ID must be set "0x95".
- Byte size (BS) should be set 7 bytes of the main header, plus sub-header size (2) added with number of samples in G.711 (SB).
- The enhanced-header size (ES) set to "0x00".
- The check bit for mixing and PLC (C1 and C2) should be set 0.
- The payload validity indicator (C3) should be set 0.

For the core layer (i.e., u-law G.711 bitstream), it should have the following sub-layer header:

- All CI, FI, QI, R6 MUST be 0.
- Sub-layer size (SB) MUST be 160 for 20 ms frame.
5. Congestion Control Considerations

The general congestion control considerations for transporting RTP data apply to UEMCLIP over RTP [3] as well as any applicable RTP profile like AVP [4]. UEMCLIP does not have any built-in mechanism for reducing the bandwidth. Packing more frames in each RTP payload can reduce the number of packets sent, and hence the overhead from IP/UDP/RTP headers, at the expense of increased delay and reduced error robustness against packet losses. It should be treated with care because increased delay means reduced quality.
6. Payload Format Parameters

6.1. Media type registration

This registration is done using the template defined in [5] and following [7].

Media type name: audio

Media subtype name: UEMCLIP

Required parameters:

Mode: This defines bit-rate, sampling frequency and layer structure of the bitstream. This parameter is necessary because this is not signaled within the bitstream. Allowed values are 0, 1, 3, and 4.

Optional parameters:

ptime: See RFC 4566 [6].

maxptime: See RFC 4566 [6].

dynmode: Indicates dynamic mode change is allowed. Possible values are comma separated list of modes from the supported mode set: 0, 1, 3, and 4. This option MUST be exclusively used with respect to "fixmode". When not specified, the mode transmission defaults to "fixmode" with the default modes specified in Table 4. See Section 6.2.1 "Dynamic transmission specification" for details.

fixmode: Indicates dynamic mode change is prohibited. Possible values are comma separated list of modes from the supported mode set: 0, 1, 3, and 4. This option MUST be exclusively used with respect to "dynmode". See Section 6.2.1 "Dynamic transmission specification" for details.

Encoding considerations: This type is defined for transferring UEMCLIP-encoded data via RTP using the payload format specified in Section 3 "Payload Format". This media type is framed and binary.

Security considerations: See Section 7 "Security Considerations".

Interoperability considerations: This media is interoperable with u-law encoded ITU-T G.711. see Section 4 "G.711 interoperability".
Published specification: RFC xxxx (This RFC)

Applications that use this media type: Audio and video streaming and conferencing tools.

Additional information: None

Intended usage: COMMON

Person & email address to contact for further information: Yusuke Hiwasaki <hiwasaki.yusuke@lab.ntt.co.jp>

Author: Yusuke Hiwasaki

Change Controller: IETF Audio/Video Transport Working Group delegated from the IESG

6.2. Mapping to SDP Parameters

The media types audio/UEMCLIP are mapped to fields in the Session Description Protocol (SDP) [6] as follows:

Payload type: Since it is not registered in [4], any RTP packets that carry UEMCLIP as payload type MUST be treated as a dynamic payload type.

Media name: The "m=" line of SDP MUST be audio.

Encoding name: Registered media subtype name should be used for the "a=rtpmap" line.

Sampling Frequency: Depending on the mode to communicate, clock rate (sampling frequency) specified in "a=rtpmap" MUST be selected from the ones defined in Table 2.

Encoding parameters: Since this is an audio stream, the encoding parameters indicate the number of audio channels, and this SHOULD default to "1", as selected from the ones defined in Table 2. This is OPTIONAL.

Packet time: A frame length of any UEMCLIP is 20 ms, thus the argument of "a=ptime" MUST be a multiple of "20". When not listed in SDP, it should also default to the minimum size: "20".
Bandwidth: As described in [6], bandwidth line is OPTIONAL. When there is no bandwidth restrictions, the numbers MUST be the largest value out of the Table 2, and the unit should be "kbit/s" with the fraction raised to the unit, including header overheads down to Layer 3. If any restrictions apply, then the value MUST be the largest of the Table 2 that satisfy the restriction, by the same calculation procedure. It MUST NOT encode with bit-rate larger than the answered bit-rate bandwidth.

UMECLIP specific: Any description specific to UEMCLIP are defined in the Format Specification Parameters ("a=fmtp"). Each parameters MUST be separated with ";", and if any attributes (value) exists, it MUST be defined with "+". For compatibility reasons, any application/terminal MUST ignore any parameters that does not appear below. This is to ensure the upper-compatibility with later added parameters for the future enhancements. The dynamic transmission specification parameters should be defined here (see Section 6.2.1).

6.2.1. Dynamic transmission specification

Since UEMCLIP codec can operate in number of modes, it is desirable to specify the range of modes that an encoder or a decoder can operate at.

UEMCLIP decoders SHOULD accept bitstreams in any modes. However, the implementation limitation may fail to adopt to the dynamic bit-rate change. Thus introduced here is two concepts: "dynamic mode" (denoted as "dynmode"), where the dynamic mode (bit-rate) change is allowed, and "fixed mode" (denoted as "fixmode"), where the change is not permitted. Both modes MUST be used exclusively.

"fixmode" is used to specify no modification of the operating mode (bit-rate) during the session. It MUST operate exclusively to "dynmode". It should specify the possible combination of mode numbers, delimited by commas ",". When offering a "fixmode", the offerer SHOULD list the mode numbers in descending priority order. The answerer MUST select a single suitable mode number and reply as "fixmode" with one argument.

On the other hand, "dynmode" is used to allow modification of the operating mode during the session. It MUST operate exclusively to "fixmode". The offerer should specify the possible combination of mode numbers, delimited by commas ",". The answerer can either select a number of suitable modes and reply as "dynmode" in the same manner, or select a single suitable mode number and reply as "fixmode" with one argument.
The mode numbers that can be specified as arguments to "fixmode" or "dynmode" are restricted by a combination of a sampling frequency and a number of audio channels, as shown in Table 2. This is because SDP binds a payload type to a combination of a sampling frequency and a number of audio channels. When a "fixmode" or "dynmode" is not given, it MUST be interpreted as being defaulting to the fixed mode ("fixmode") and MUST use the default value specified in Table 4.

<table>
<thead>
<tr>
<th>Fs [Hz]</th>
<th>Channels</th>
<th>Selectable modes</th>
<th>Default mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>8000</td>
<td>1</td>
<td>0,3</td>
<td>0</td>
</tr>
<tr>
<td>16000</td>
<td>1</td>
<td>1,4</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: Default modes

6.3. Offer-Answer Model Considerations

6.3.1. Offer-Answer Guidelines

The procedures related to exchanging SDP messages MUST follow [2]. Other than that, followings are the guidelines for establishing a session using an offer-answer model.

- When multiple UEMCLIP dynamic payload type number is offered, an answerer SHOULD select a single payload type number, i.e., one sampling frequency and channel condition.

- An offerer SHOULD offer every possible combination of sampling frequency, channel number, and fmtp parameters including dynamic/fixed mode. When the transmission bandwidth is restricted, it MUST be offered in accordance to the restriction.

- When offering/answering SDP, any fmtp parameters which are undefined MUST be ignored. If any unknown/undefined parameters should be offered, an answerer MUST delete the entry from the answer message. In this case, the offerer MUST use the default value for any deleted parameters.

- If a dynamic mode ("dynmode") is offered, an answerer MUST select either "dynmode" or "fixmode", according to ones capabilities. When fixed mode ("fixmode") is offered, an answerer MUST only answer "fixmode". In the case of answering fixed mode ("fixmode"), answerer MUST select a single mode out of offered mode, regardless of dynamic/fixed mode specification. If a mode is not offered at all, the session MUST default to fixed mode, and
the default mode value, as shown in Table 4, MUST be used, based on the sampling frequency and number of channels specified elsewhere.

- When an offered condition does not fit an answerer’s capabilities, it naturally MUST not answer the conditions, and session MAY proceed to re-INVITE, if possible. If a condition (mode) is decided upon, an offerer and an answerer MUST transmit on this condition.

6.3.2. Examples

When an offerer indicates that he/she wishes to dynamically switch between modes (0,1,3, and 4) during a session, an example of an offered SDP can be:

```
m=audio 5004 RTP/AVP 96 97
a=rtpmap:96 UEMCLIP/16000/1
a= fmtp:96 dynmode=4,1
a=rtpmap:97 UEMCLIP/8000/1
a= fmtp:97 dynmode=3,0
```

When an answerer can only operate in modes 1 and 0, and cannot dynamically switch between modes during a session, an example of answer will be as follows:

```
m=audio 5004 RTP/AVP 96
a=ptime:20
a=rtpmap:96 UEMCLIP/16000/1
a= fmtp:96 fixmode=1
```

As a result, both will start communication with mode 1. It should be noted that mode change during this session MUST NOT be done, because the answerer’s response is "fixmode". If an offerer wants to stick to one single mode during a session but can receive either Modes 4 or 1 bitstreams, the SDP should be:

```
m=audio 5004 RTP/AVP 96
a=rtpmap:96 UEMCLIP/16000/1
a= fmtp:96 fixmode=4,1
```

The "ptime" attribute is used to denote the packetization interval. When not specified, it SHOULD default to 20. Since UEMCLIP uses 20 msec frames, ptimes values of multiples of 20 imply multiple frames per packet. In the example below, the ptime is set to 60, and this means that there are 3 frames in each packet. When ftmp line is not present, it should default to fixmode with Mode 0 (see Section 6.2.1).
m=audio 5004 RTP/AVP 96
a=rtpmap:96 UEMCLIP/16000/1
a=ptime:60
7. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [3] and any appropriate profiles. This implies that confidentiality of the media streams is achieved by encryption by other means.

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

Another potential threat are memory attacks by illegal layer indices or byte numbers. The implementer of the decoder should always be aware that the indicated numbers may be corrupted and does not point to the right sub-layer and may force reading beyond the bitstream boundaries.
8. IANA Considerations

It is requested that one new media subtype (audio/UEMCLIP) is registered by IANA. For details, see Section 6.1.
9. Normative References


Authors’ Addresses

Yusuke Hiwasaki
NTT Corporation
3-9-11 Midori-cho,
Musashino-shi
Tokyo  180-8585
Japan

Phone: +81(422)59-4815
Email: hiwasaki.yusuke@lab.ntt.co.jp

Hitoshi Ohmuro
NTT Corporation
3-9-11 Midori-cho,
Musashino-shi
Tokyo  180-8585
Japan

Phone: +81(422)59-2151
Email: ohmuro.hitoshi@lab.ntt.co.jp