RTP payload format and file storage format for AMR and AMR-WB audio

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

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 cite them other than as "work in progress".

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/lid-abstracts.txt

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

This document is an individual submission to the IETF. Comments should be directed to the authors.

Abstract

This document specifies a real-time transport protocol (RTP) payload format to be used for AMR and AMR-WB speech encoded signals. The payload format is designed to be able to interoperate with existing AMR and AMR-WB transport formats. Furthermore, a file format for storage of AMR and AMR-WB speech data is specified. Two separate MIME type registrations, one for AMR and one for AMR-WB, describing both RTP payload format and storage format are included.
1. Introduction

This payload description applies to the packetization of data from two different codecs, the Adaptive Multi-Rate (AMR) codec and the Adaptive Multi-Rate Wideband (AMR-WB) codec. It is important to remember that these are different codecs and they MUST always be handled as different payload types in RTP.

1.1. The Adaptive Multi-Rate speech codec

The adaptive multi-rate (AMR) speech codec [1] was developed by the European Telecommunications Standards institute (ETSI). The AMR codec is standardized for GSM, and is also chosen by the Third Generation Partnership Project (3GPP) as the mandatory codec for third generation systems. The AMR codec will be widely used in cellular systems.

The AMR codec is a multi-mode codec with 8 narrow band speech modes with bit rates between 4.75 and 12.2 kbps. The sampling frequency is 8000 Hz and processing is done on 20 ms frames, i.e. 160 samples per frame. The AMR modes are closely related to each other and use the same coding framework. Three of the AMR modes are already adopted standards of their own, the 6.7 kbps mode as PDC-EFR [10], the 7.4 kbps mode as IS-641 codec in TDMA [9], and the 12.2 kbps mode as GSM-EFR [8].

1.2. The Adaptive Multi-Rate Wideband speech codec

The Adaptive Multi-Rate Wideband (AMR-WB) speech codec [3] was originally developed by 3GPP to be used in GSM and 3G systems. The AMR-WB codec will be widely used in cellular systems.

The AMR-WB codec is a multi-mode speech codec with 9 wideband speech coding modes with bit-rates between 6.6 and 23.85 kbps. The sampling frequency is 16000 Hz and processing is performed on 20 ms frames, i.e. 320 speech samples per frame. The AMR-WB modes are closely related to each other and employ the same coding framework.

1.3. Common Characteristics for AMR and AMR-WB

The multi-mode feature is used to preserve high speech quality under a wide range of transmission conditions. In mobile radio systems (e.g. GSM) mode adaptation allows the system to adapt the balance between speech coding and error protection to enable best possible speech quality in prevailing transmission conditions. On the other hand, mode adaptation can be also utilized to adapt to the varying available transmission bandwidth. Codec implementations must support all specified speech coding modes, and mode switching can occur to
any mode at any time. The mode information must therefore be transmitted together with the speech encoded bits, to indicate the mode. To realize rate adaptation the decoder needs to signal the mode it prefers to receive to the encoder.

Both codecs include voice activity detection (VAD) and generation of comfort noise (CN) parameters during silence periods. Hence, the codecs can reduce the number of transmitted bits and packets during silence periods to a minimum. The operation to send CN parameters at regular intervals during silence periods is usually called discontinuous transmission (DTX) or source controlled rate (SCR) operation. The frames containing CN parameters are called Silence Indicator (SID) frames.

Due to the flexibility and robustness of these codecs, they are suitable also for other purposes than circuit switched cellular systems. Other suitable applications are real-time services over packet switched networks. The payload format should be designed for robustness against both bit errors and packet loss. The speech encoded bits have different perceptual sensitivity to bit errors and cellular systems exploit this by using unequal error protection and detection (UEP and UED).

The UED/UEP mechanism focus the correction and detection of corrupted bits to the perceptually most sensitive bits. A speech frame is only declared damaged if there are bit errors in the most sensitive bits, i.e. the class A bits see [2] and [4]. It is acceptable to have some bit errors in the other bits, i.e. class B and C. Also a damaged frame is still useful for error concealment in the decoding, which uses some of the less sensitive bits. This improves the speech quality compared to discarding the data.

Today there exist some link layers that do not discard packets with bit errors, e.g. SLIP and some wireless links. With the Internet traffic pattern shifting towards a more media-centric one, more link layers of such nature may emerge in the future. With transport layer support for partial checksums, for example those supported by UDP-Lite [13] (work in progress), bit error tolerant AMR and AMR-WB traffic could achieve better performance over these types of links.

There are at least two basic approaches for carrying AMR and AMR-WB traffic over bit error tolerant networks:

1) Utilizing a partial checksum to cover headers and the most important speech bits of the payload. It is recommended that at least all class A bits are covered by the checksum.

2) Utilizing a partial checksum to only cover headers, but a frame CRC to cover the class A bits of each speech frame in the payload.
In either approach, at least part of the class B/C bits are left without error-check and thus bit error tolerance is achieved.

It is still important that the network designer pay attention to the class B and C residual bit error rate. Though less sensitive to errors than class A bits, class B bits are not insignificant and undetected errors in these bits cause degradation in speech quality. An example of residual error rates considered acceptable for AMR in UMTS can be found in [21] and for AMR-WB in [22].

Approach 1 is a bit efficient, flexible and simple way, but comes with two disadvantages, namely, a) bit errors in protected speech bits will cause the payload to be discarded, and b) when transporting multiple frames in a payload there is the possibility that a single bit error in protected bits gets all the frames discarded.

These disadvantages can be avoided if needed, with some overhead in the form of a frame-wise CRC (Approach 2). In problem a), the CRC makes it possible to detect bit errors in class A bits and use the frame for error concealment, which gives a small improvement in speech quality. Secondly (b), when transporting multiple frames in a payload the CRC’s remove the possibility that a single bit error in a class A bit gets all the frames discarded. Avoiding that gives an improvement in speech quality when transporting multiple frames and subject to bit errors.

The choice between the two approaches must be made based on the available bandwidth, and desired tolerance to bit errors. Neither solution is appropriate to all cases.

The payload format supports several means to increase robustness against packet loss. The simple scheme of repetition of previously sent data is one possibility. Another possible scheme which is more bandwidth efficient is to use payload external FEC, e.g. RFC2733 [20], which generates extra packets containing repair data. The whole payload can also be sorted in sensitivity order to support external FEC schemes using UEP. There is work in progress on a generic version of such a scheme [19].

Several frames can be encapsulated into a single RTP packet to decrease protocol overhead. One of the drawbacks of such approach is that in case of packet loss this means loss of several consecutive speech frames, which usually causes clearly audible distortion in reconstructed speech. Interleaving of frames can improve the speech quality in such cases by distributing the consecutive losses into series of single frame losses. However, interleaving and bundling several frames per payload will also increase end-to-end delay and is therefore not applicable to all types of applications. However, streaming applications are likely to be able to exploit interleaving to improve speech quality in lossy transmission conditions.
2. Payload format

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 RFC2119 [5].

The AMR and AMR-WB payload format supports transmission of multiple frames per payload, the use of fast codec mode adaptation, and robustness against packet loss and bit errors.

The payload format consists of one payload header with an optional interleaving extension, a table of contents, optionally one CRC per payload frame and zero or more payload frames.

The payload format is either bandwidth efficient or octet aligned, which mode of operation to use has to be signalled at session establishment. Only the octet aligned format has the possibility to use the robust sorting, interleaving and CRC to make it robust to packet loss and bit errors. In the octet aligned format the payload header, table of contents entries and the payload frames are individually octet aligned to make implementations efficient, but in the bandwidth efficient format only the full payload is octet aligned. If the option to transmit a robust sorted payload is enabled and employed, the full payload SHALL finally be ordered in descending bit error sensitivity order to be prepared for unequal error protection or unequal error detection schemes. The encoded bit streams are defined in sensitivity order in Annex B of [2] and [4], the original order as delivered from the speech encoder is defined in [1] and [3].

Octet alignment of a field or payload means that the last octet MUST be padded with zeroes at the end to fill the the octet.

The AMR frame types, or modes, are defined in [2] and the corresponding description for AMR-WB is found in [4]. Frame type 14 (only available for AMR-WB), SPEECH_LOST, and 15, NO_DATA, are needed to indicate not transmitted frames or lost frames. NO_DATA could mean both no data produced by the speech encoder for this frame or no data transmitted in this payload, i.e. valid data for this frame could be sent in an earlier or following packets. For example, when multiple frames are sent in each payload and comfort noise starts. A frame type sequence in a payload with 8 speech frames using AMR mode 7 is interrupted by DTX operation in the fifth frame, this looks like: {7,7,7,7,8,15,15,8}. The AMR SCR/DTX is described in [6] and AMR-WB SCR/DTX in [7].

Robustness against packet loss can be accomplished by using the possibility to retransmit previously transmitted frames together with the current frame or frames. Another approach is to use interleaving to reduce the speech quality effect of packet losses. The speech quality in case of packet losses when transmitting several frames per
packet can be improved by using OPTIONAL frame interleaving. The interleaving improves perceived speech quality since it introduces single frame errors instead of several consecutive frame errors. Note that interleaving can be applied only if the receiver has signaled support for it in capability description.

The AMR performance over error tolerant links can be improved by delivering also speech frames with bit errors. Unequal error detection is needed since bit errors SHOULD only be allowed in the least error sensitive bits. This payload format provides two alternative methods to implement unequal error detection:

A. CRC calculation over the class A speech bits

The optional CRC MAY be used to protect the class A speech bits. The number of class A bits is specified as informative for AMR in [2] and therefore copied into table 1 as normative for this payload format. The number of class A bits for AMR-WB are specified as normative in table 2 in [4] and these numbers MUST be used also for this payload format. Speech frames with errors in class A bits MUST be marked with SPEECH_BAD for corrupted speech frames (FT=0..7 for AMR and FT=0..8 for AMR-WB) or SID_BAD for corrupted SID frames (FT=8 for AMR and FT=9 for AMR-WB) and be sent to the speech decoder, see [6] and [7]. In this case the RTP header, payload header and table of contents should be covered by a transport layer checksum, e.g. UDP-lite [13]. Packets MUST be discarded if the transport layer checksum detects errors.

B. Robust sorting of payload bits

Robust behavior can also be accomplished by robust sorting of the payload. This enables the use of UED (e.g. UDP-lite) and UEP (e.g. ULP [19]). The UED and/or UEP is recommended to cover at least the RTP header, payload header, table of contents and class A bits.

Support for unequal error detection is OPTIONAL. If either scheme is to be used, it MUST be signaled out of band (see section 7).
Table 1. The number of class A bits for the AMR codec.

A frame quality indicator is included for interoperability with the ATM payload format described in ITU-T I.366.2, the UMTS Iu interface [17] and other transport formats. The speech quality is increased if damaged frames are forwarded to the speech decoder error concealment unit and not dropped. In many communication scenarios the AMR encoded bits will be transmitted from one IP/UDP/RTP terminal to a terminal in a system with another transport format and/or vice versa. The transport format transcoding will be done in a gateway. A second likely scenario is that IP/UDP/RTP is used as transport between other systems, i.e. IP is originated and terminated in gateways on both sides of the IP transport.

2.1. The payload header
The length of the payload header is either 4 or 8 bits plus optionally an 8 bit interleaving header. The bits in the header are specified as follows:

CMR (4 bits): Indicates Codec Mode Requested for the other communication direction. It is only allowed to request one of the speech modes of the used codec, frame type index 0..7 for AMR, see Table 1a in [2] or frame type index 0..8 for AMR-WB, see Table 1a in [4]. CMR value 15 indicates that no mode request is present, other values are for future use.

P: Is a padding bit, always set to zero.

```
0 0 1 2 3
+++++++|
CMR |
+++++++```

Figure 3: Payload header for bandwidth efficient operation.

```
0 0 1 2 3 4 5 6 7
+++++++|
CMR |P|P|P|P|
+++++++```

Figure 4: Payload header for octet aligned operation.

If the use of interleaving is signaled out of band at session set up, and if octet aligned operation is signaled interleaving is used and the payload header is extended with two 4 bit fields, ILL and ILP, used to describe the interleaving scheme.

ILL (4 bits): OPTIONAL field that is present only if interleaving is signaled. The value of this field specifies the interleaving length used for frames in this payload.

ILP (4 bits): OPTIONAL field that is present only if interleaving is signaled. The value of this field indicates the interleaving index for frames in this payload. The value of ILP MUST be smaller than or equal to the value of ILL. Erroneous value of ILP SHOULD cause the payload to be discarded.

The value of the ILL field defines the length of an interleave group: ILL=L implies that frames in (L+1)-frame intervals are picked into the same interleaved payload, and the interleave group consists of L+1 payloads. The size of the interleaving group is the N*(L+1), if N is the number of frames per payload. The value of ILP=p in payloads belonging to the same group runs from 0 to L. The interleaving is meaningful only when number of frames per payload N is greater than...
or equal to 2. Thus, when N frames are transmitted in each payload of a group, the interleave group consists of payloads with sequence numbers s...s+L, and frames encapsulated into these payloads are f...f+N*(L+1)-1.

To put this in a form of an equation, let’s assume that the first frame of an interleave group is n, the first payload of the group is s, number of frames per payload is N, ILL=L and ILP=p (p in range 0...L), the frames contained by the payload s+p are n + p + k*(L+1), where k runs from 0 to N-1. I.e.

The first packet of an interleave group: ILL=L, ILP=0
Payload: s
Frames: n, n+(L+1), n+2*(L+1), ..., n+(N-1)*(L+1)

The second packet of an interleave group: ILL=L, ILP=1
Payload: s+1
Frames: n+1, n+1+(L+1), n+1+2*(L+1), ..., n+1+(N-1)*(L+1)

... 

The last packet of an interleave group: ILL=L, ILP=L
Payload: s+L
Frames: n+L, n+L+(L+1), n+L+2*(L+1), ..., n+L+(N-1)*(L+1)

0                  1
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  CMR  |P|P|P|P|  ILL  |  ILP  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Figure 5: Octet aligned operation payload header with interleaving extension.

2.2. The payload table of contents and CRCs

The table of contents (ToC) consists of one entry for each speech frame in the payload. A table of contents entry includes several specified fields as follows:

F (1 bit): Indicates if this frame is followed by further frames. F=1 further frames follow, F=0 last frame.

FT (4 bits): Frame type indicator, indicating the AMR speech coding mode or comfort noise (SID) mode. The mapping of existing AMR modes to FT is given in Table 1a in [2] for AMR and in Table 1a in [4] for AMR-WB. If FT=14 (speech lost, available only in AMR-WB) or FT=15 (No transmission/no reception) no CRC or payload frame is present.
Q (1 bit): The payload quality bit indicates, if not set, that the payload is severely damaged and the receiver should set the RX_TYPE, see [6], to SPEECH_BAD or SID_BAD depending on the frame type (FT).

P: Is a padding bit, always set to zero.

Figure 6: Table of contents entry field for bandwidth efficient operation.

Figure 7: Table of contents entry field for octet aligned operation.

CRC (8 bits): OPTIONAL field, exists if the use of CRC is signaled at session set up. The 8 bit CRC is used for error detection. The algorithm to generate these 8 parity bits are defined in section 4.1.4 in [2].

Figure 8: CRC field

The ToC and CRCs are arranged with all table of contents entries fields first followed by all CRC fields. The ToC starts with the frame data belonging to the oldest speech frame.

Figure 9: The ToC and CRCs for a payload with three speech frames
2.3. Speech frame

A speech frame represents one frame encoded with the mode according to the ToC field FT. The length of this field is implicitly defined by the AMR mode in the FT field. The bits SHALL be sorted according to Appendix B of [2] for AMR and Appendix B of [4] for AMR-WB.

If octet aligned operation is used, the last octet of each speech frame MUST be padded with zeroes at the end if not all bits are used.

2.4. Compound payload

The compound payload consists of one AMR payload header, the table of contents and one or more speech frames, see section 2.1, 2.2 and 2.3. These elements SHALL be put together to form a payload with either simple or robust sorting. If the bandwidth efficient operation is used only simple sorting MUST be used.

Definitions for describing the compound AMR payload:

\[ b(m) \quad \text{- bit } m \text{ of the compound AMR payload, octet aligned} \]
\[ o(n,m) \quad \text{- bit } m \text{ of octet } n \text{ in the octet description of the compound AMR payload, bit 0 is MSB} \]
\[ t(n,m) \quad \text{- bit } m \text{ in the table of contents entry for speech frame } n \]
\[ p(n,m) \quad \text{- bit } m \text{ in the CRC for speech frame } n \]
\[ f(n) \quad \text{- number of bits in speech frame } n, \text{ defined by FT} \]
\[ h(m) \quad \text{- bit } m \text{ of payload header} \]
\[ C(n) \quad \text{- number of CRC bits for speech frame } n, \text{ 0 or 8 bits} \]
\[ N \quad \text{- number of payload frames in the payload} \]
\[ S \quad \text{- number of unused bits} \]

Payload frames \( f(n,m) \) are ordered in consecutive order, where frame \( n=1 \) is preceding frame \( n=2 \). Within one payload all frames between the oldest and most recent MUST be present, if not interleaving is used then the interleaving rules defined in section 2.1 applies. If speech data is missing for one or more frames in the sequence of frames in the payload, due to e.g. DTX, send the NO_DATA frame type for these frames. This does not mean that all frames must be sent, only that the sequence of frames in one payload MUST indicate missing frames.

The compound AMR payload, \( b \), is mapped into octets, \( o \), where bit 0 is MSB.

2.4.1. Simple payload sorting

If multiple new frames are encapsulated into the payload and robust payload sorting is not used, the payload is formed by concatenating the payload header, the ToC, optional CRC fields and the speech
frames in the payload. However, the bits inside a frame are ordered into sensitivity order as defined in [2] for AMR and [4] for AMR-WB.

### 2.4.1.1. Simple payload sorting for bandwidth efficient operation

The simple payload sorting algorithm is defined in C-style as:

```c
/* payload header */
k=0; H=4;
for (i = 0; i < H; i++){
    b(k++) = h(i);
}

/* table of contents */
T=6;
for (j = 0; j < N; j++){
    for (i = 0; i < T; i++){
        b(k++) = t(j,i);
    }
}

/* payload frames */
for (j = 0; j < N; j++){
    for (i = 0; i < F(j); i++){
        b(k++) = f(j,i);
    }
}

/* padding */
S = (k%8 == 0) ? 0 : 8 - k%8;
for (i = 0; i < S; i++){
    b(k++) = 0;
}

/* map into octets */
for (i = 0; i < k; i++){
    o(i/8,i%8)=b(i)
}
```

### 2.4.1.2. Simple payload sorting for octet aligned operation

In octet aligned operation is the simple payload sorting algorithm defined in C-style as:

```c
/* payload header */
k=0; H=8;
if (interleaving){
    H+=8;       /* Interleaving extension */
}
for (i = 0; i < H; i++){
    b(k++) = h(i);
}
```

Sjoberg et al.
/* table of contents */
T=8;
for (j = 0; j < N; j++){
    for (i = 0; i < T; i++){
        b(k++) = t(j,i);
    }
}
/* CRCs, only if signaled */
if (crc) {
    for (j = 0; j < N; j++){
        for (i = 0; i < C(j); i++){
            b(k++) = p(j,i);
        }
    }
}
/* payload frames */
for (j = 0; j < N; j++){
    for (i = 0; i < F(j); i++){
        b(k++) = f(j,i);
    }
    /* padding of each speech frame */
    S = (k%8 == 0) ? 0 : 8 - k%8;
    for (i = 0; i < S; i++){
        b(k++) = 0;
    }
    /* map into octets */
    for (i = 0; i < k; i++){
        o(i/8,i%8)=b(i)
    }
}

2.4.2. Robust payload sorting

Robust payload sorting is only supported in octet aligned operation and must be signaled at session set up.

A bit error in a more sensitive bit is subjectively more annoying than in a less sensitive bit. Therefore, to be able to protect only the most sensitive bits in a payload packet with a forward error detection or correction code, e.g. a checksum outside RTP or ULP [19], the bits inside a frame are ordered into sensitivity order. The protection SHOULD cover an appropriate number of octets from the beginning of the payload, covering at least the AMR payload header, ToC and class A bits (see [2]). If CRCs are used together with robust sorting only the payload header and the ToC should be covered by the transport checksum. Exactly how many octets need protection depends on the network and application. To maintain sensitivity ordering inside the AMR payload, when more than one speech frame is transmitted in one payload, reordering of the data is needed.
When robust sorting mode is used, the reordering to maintain the sensitivity ordered AMR payload SHALL be performed on octet level. The AMR payload header, ToC and CRCs SHALL still be placed unchanged in the beginning of the payload. Thereafter, the payload frames are sorted with one octet alternating from each payload frame.

The robust payload sorting algorithm is defined in C-style as:

```c
/* payload header */
k=0; H=8;
if (interleaving){
    H += 8;       /* interleaving extension */
}
for (i = 0; i < H; i++){
    b(k++) = h(i);
}
/* table of contents */
for (j = 0; j < N; j++){
    for (i = 0; i < 8; i++){
        b(k++) = t(j,i);
    }
}
/* CRCs */
if (crc){
    for (j = 0; j < N; j++){
        for (i = 0; i < C(j); i++){
            b(k++) = p(j,i);
        }
    }
}
/* payload frames */
for (j = 0; j < N; j++){
    P(j) = F(j)%8 == 0 ? 0 : 8 - F(j)%8;
    max = (max(F(0),..,F(N-1))-1)/8 +1;
    for (i = 0; i < max; i++){
        for (j = 0; j < N; j++){
            for (l = 0; l < 8; l++){
                if (i < F(j)+P(j)){
                    if (i < F(j)){
                        b(k++) = f(j,i);
                    }else{
                        b(k++) = 0;
                    }
                }
            }
        }
    }
}
```
/* map into octets */
for (i = 0; i < k; i++)
  o(i/8,i%8)=b(i)
}

2.5. Decoding security consideration

If the payload length calculation, using the information from
signaling plus the F and FT fields, does not indicate the same length
as the size of the payload actually received, the payload should be
dropped. Decoding a packet that has errors in length indicator bits
could severely degrade the speech quality.

2.6. Implementation considerations

Implementations SHOULD include both bandwidth efficient and octet
aligned operation to give a high possibility of interoperability. The
implementation of robust sorting, interleaving and CRCs are OPTIONAL.

3. RTP header usage

The RTP header marker bit (M) is used to mark (M=1) the packages
containing as their first frame the first speech frame after a
comfort noise period in DTX operation. For all other packages the
marker bit is set to 0 (M=0).

The timestamp corresponds to the sampling instant of the first sample
encoded for the first frame in the packet. A frame can be either
encoded speech, comfort noise parameters, NO_DATA, or SPEECH_LOST
(only for AMR-WB). The timestamp unit is in samples. The duration of
one speech frame is 20 ms and the sampling frequency is 8 kHz,
corresponding to 160 encoded speech samples per frame for AMR and 16
kHz corresponding to 320 samples per frame in AMR-WB. Thus, the
timestamp is increased by 160 for AMR and 320 for AMR-WB for each
consecutive frame. All frames in a packet MUST be successive 20 ms
frames as delivered by the speech encoder except if interleaving is
employed, then frames encapsulated into a payload MUST be picked as
defined in section 2.1.

4. Congestion Control

The need of congestion control for data transported with RTP has to
be considered. AMR and AMR-WB speech data have some elastic
properties due to the different bandwidth demand for each mode.
Another parameter that can reduce the bandwidth demand for AMR and
AMR-WB is how many frames of speech data that are encapsulated in
each payload. This will reduce the number of packets and the overhead
from IP/UDP/RTP headers. If using forward error correction (FEC) there is also the need to regulate the amount, so the FEC itself does not worsen the problem. Therefore, it is RECOMMENDED that applications using this payload implement congestion control. The actual mechanism for congestion control is not specified but should be suitable for real-time flows, e.g. "Equation-Based Congestion Control for Unicast Applications" [18].

5. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [11]. This implies that confidentiality of the media streams is achieved by encryption. Because the payload format is arranged end-to-end, encryption MAY be performed after encapsulation 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.

As this format transports encoded speech, the main security issues are decoding security (see section 2.5), confidentiality and authentication of the speech itself. The payload format itself does not have any support for security. These issues have to be solved by a payload external mechanism.

5.1. Confidentiality

To achieve confidentiality of the encoded speech all speech data bits must be encrypted. There is less need to encrypt the payload header or the table of contents as they only carry information about the requested speech mode, frame type and frame quality. This information could be useful to some third party, e.g. quality monitoring. The type of encryption used can not only have impact on the confidentiality but also on error robustness. The error robustness against bit errors will be none, unless an encryption method without error-propagation is used, e.g. a stream cipher. This is only an issue when using UEP/D, when bit errors can be accepted in some part of the payload.

5.2. Authentication

To authenticate the sender of the speech an external mechanism has to be added. It is RECOMMENDED that such a mechanism protects all the speech data bits. Note that the use of UED/UEP is difficult to combine with authentication. To prevent a man in the middle from tampering with the packetization of the speech data, some extra data SHOULD be protected. The data is: the payload header, ToC, CRCs, RTP timestamp, RTP sequence number, and the RTP marker bit. Tampering
could result in erroneous depacketization/decoding that could lower speech quality. Tampering with the AMR mode request field can result in that the sender must receive speech in a different quality than desired.

6. Examples

6.1. Bandwidth efficient examples

6.1.1. Single frame example

The bandwidth efficient single frame per payload example is employing AMR, no valid Codec Mode Request CMR is sent (CMR=15), the payload was not damaged at IP origin (Q=1). The mode is AMR 7.4 kbps (FT=4). The speech encoded bits are put into f(0) to f(147) in descending sensitivity order according to [2].

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  CMR |F|  FT   |Q|f(0)                                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         f(147)|P|P|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 10: One frame per packet example.

6.1.2. Multi frame example

The bandwidth efficient multiple frame per payload example is employing AMR-WB, a Codec Mode Request CMR for the AMR-WB 8.85 kbps mode is sent (CMR=1), the payloads were not damaged at IP origin (Q=1). The mode is AMR-WB 6.6 kbps (FT=0) for the first frame, f(0) to f(131), and AMR-WB 8.85 kbps (FT=1) for the second frame, g(0) to g(176). The speech encoded bits are put into f(0) to f(131) and g(0) to g(176) in descending sensitivity order according to [4].

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  CMR |F|  FT   |Q|f(0)                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       f(147)|P|P|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Sjoeborg et al.                                                 [Page 17]
6.2. Octet aligned operation examples

In this example octet aligned operation of the payload format is used. Two AMR frames with 7.95 kbps mode (FT=5) are sent in the payload. A mode request is sent, requesting the 10.2 kbps mode for the other link (CMR=6). CRC is used. Interleaving is used with depth ILL=1 and index ILP=0. The first frame is frame 1, f1(0..158), and the second frame in the payload is frame 3 due to interleaving, f3(0..158). For each payload frame a CRC is calculated CRC1(0..7) for frame 1 and CRC3(0..7) for frame 3. Robust payload sorting is used.

Figure 11: Two frame per packet example.

Figure 12: Example with CRCs, interleaving and robust sorting.
7. MIME type registration

This chapter defines the MIME types for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) speech codecs, [1] and [3], respectively. To distinguish between the two codecs and emphasize that seamless switching is possible only within each of these two codecs the MIME types are kept separate although they are very similar. The data format and parameters are specified for both real-time transport and for storage type applications (e.g. e-mail attachment, multimedia messaging). The former is referred to as RTP mode and the latter as storage mode.

Implementations according to [1] and [3] MUST support all eight coding modes for AMR and all nine coding modes for AMR-WB. The mode change within each codec can occur at any time during operation and therefore the mode information is transmitted in-band together with speech bits to allow mode change without any additional signaling.

In addition to the speech codec, AMR and AMR-WB specifications also include Discontinuous Transmission / comfort noise (DTX/CN) functionality [14] and [15]. The DTX/CN switches the transmission off during silent parts of the speech and only CN parameter updates, SID frames, are sent at regular intervals.

7.1. RTP mode

It is possible that the decoder may want to receive a certain speech mode or a subset of modes, due to link limitations in some cellular systems, e.g. the GSM radio link can only use a subset of at most four modes. A GSM subset can consist of any combination of the 8 AMR modes or 9 AMR-WB modes. Therefore, it is possible to request a specific set of speech modes in capability description and the encoder MUST abide by this request. If the request for mode set is not given any mode may be used or requested.

The codec can in principle perform a mode change at any time between any two modes. To support interoperability with GSM through a gateway it is possible to set limitations for mode changes. The decoder has the possibility to define the minimum number of frames between mode changes and to limit the mode change to transition into neighboring modes only.

It is also possible to limit the number of speech frames encapsulated into one RTP packet. This is an optional feature and if no parameter is given in the capability description, the transmitter MAY encapsulate any number of speech frames into one RTP packet.

The payload CRC UED MUST only be used if the receiver has signaled support for this functionality in the capability description.
To support unequal error protection and/or detection the payload format supports robust payload sorting. The robust payload sorting is an OPTIONAL feature and MUST only be used if the receiver has signaled support for this functionality in the capability description.

The speech quality in case of packet losses when transmitting several speech frames per packet can be improved by using the OPTIONAL frame level interleaving. The interleaving improves perceived speech quality since it introduces series of single frame errors instead of several consecutive frame errors. Interleaving MUST only be applied if the receiver has signaled support for it, and if used, the interleaving length MUST NOT exceed the limitation given in capability description. Note that the receiver can use the MIME parameters to limit increased buffering requirements caused by the interleaving. For example, interleaving=I defines the maximum size of an interleave group to I=N*(L+1) (see section 2.1 for details on interleaving).

7.2. Storage mode

The storage mode is used for storing speech frames, e.g. as a file or e-mail attachment.

The first octet of the file is the storage header and indicates with its first bit if the file contains AMR or AMR-WB speech. The rest of the header octet is reserved for future use.

```
0 1 2 3 4 5 6 7
+-------------------+
|C|Reserved          |
+-------------------+
```

Figure 13: Storage header, C=0 indicates AMR and C=1 AMR-WB.

The speech frames are stored in consecutive order in octet aligned manner. This implies that the first octet after the last octet of frame n must be the first octet of frame n+1. The first octet of each stored speech frame consists of a 4-bit FT field (see definition in section 2.2) and a Q bit. The positions of the fields correspond to the positions of the corresponding fields of an octet aligned table of contents entry, see figure 7. Following this first octet comes the encoded speech frames bits (see section 2.3). The last octet of each frame is padded with zeroes, if needed, to achieve octet alignment. An example is given in figure 14.
Speech frames lost in transmission and non-received frames between SID updates during non-speech period MUST be stored as NO_DATA frames (frame type 15, see definition in [2] and [4]) or SPEECH_LOST (only available for AMR-WB) to keep synchronization with the original media.

7.3. AMR MIME Registration

MIME-name for the AMR codec is allocated from IETF tree since AMR is expected to be widely used speech codec in VoIP applications. Some parts of this chapter will distinguish between RTP and storage modes.

Media Type name: audio

Media subtype name: AMR

Required parameters: none

Optional parameters for RTP mode:

octet-align: If present, octet aligned operation SHALL be used. If not present, band width efficient operation is employed.

mode-set: Requested AMR mode set. Restricts the active codec mode set to a subset of all modes. Possible values are comma separated list of modes: 0,...,7 (see Table 1a [2] an example is given in section 7.5). If not present, all speech modes are available.

mode-change-period: Defines a number N which restricts the mode changes in such a way that mode changes are only allowed on multiples of N, initial state of the phase is arbitrary. If this parameter is not present, mode change can happen at any time.

mode-change-neighbor: If present, mode changes SHALL only be made to neighboring modes in the active codec mode set. Neighboring modes are the ones closest in bit rate to the
current mode, both higher and lower rate included. If not present, change between any two modes in the active codec mode set is allowed.

maxframes: Maximum number of speech frames in one RTP packet. The receiver may set this parameter in order to limit the buffering requirements or delay.

crc: If present, CRCs SHALL be included in the payload, otherwise not. Requires also the octet-align parameter to be sent.

robust-sorting: If present, the payload SHALL employ robust payload sorting. If not present simple payload sorting SHALL be used. Requires also the octet-align parameter to be sent.

interleaving: Indicates that frame level interleaving SHALL be used and its value defines a maximum number of frames in the interleaving group (see section 2.1). If this parameter is not present, interleaving SHALL not be used. Requires also the octet-align parameter to be sent.

Optional parameters for storage mode: none

Encoding considerations for RTP mode: See chapter 2.

Encoding considerations for storage mode: See section 7.2.

Security considerations: see chapter 5 "Security".

Public specification: please refer to chapter 8 "References".

Additional information for storage mode:
Magic number: none
File extensions: amr, AMR
Macintosh file type code: none
Object identifier or OID: none

Person & email address to contact for further information:
  johan.sjoberg@ericsson.com
  ari.lakaniemi@nokia.com

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

Author/Change controller:
  johan.sjoberg@ericsson.com
  ari.lakaniemi@nokia.com

7.4. AMR-WB MIME Registration

MIME-name for the AMR-WB codec is allocated from IETF tree since AMR-WB is expected to be widely used speech codec in VoIP applications.
Some parts of this chapter will distinguish between RTP and storage modes.

Media Type name: audio

Media subtype name: AMR-WB

Required parameters: none

Optional parameters for RTP mode:
- **octet-align**: If present, octet aligned operation SHALL be used. If not present, band width efficient operation is employed.
- **mode-set**: Requested AMR-WB mode set. Restricts the active codec mode set to a subset of all modes. Possible values are comma separated list of modes: 0,...,8 (see Table 1a [4]). If not present, all speech modes are available.
- **mode-change-period**: Defines a number N which restricts the mode changes in such a way that mode changes are only allowed on multiples of N, initial state of the phase is arbitrary. If this parameter is not present, mode change can happen at any time.
- **mode-change-neighbor**: If present, mode changes SHALL only be made to neighboring modes in the active codec mode set. Neighboring modes are the ones closest in bit rate to the current mode, both higher and lower rate included. If not present, change between any two modes in the active codec mode set is allowed.
- **maxframes**: Maximum number of speech frames in one RTP packet. The receiver may set this parameter in order to limit the buffering requirements or delay.
- **crc**: If present, CRCs SHALL be included in the payload, otherwise not. Requires also the octet-align parameter to be sent.
- **robust-sorting**: If present, the payload SHALL employ robust payload sorting. If not present simple payload sorting SHALL be used. Requires also the octet-align parameter to be sent.
- **interleaving**: Indicates that frame level interleaving SHALL be used and its value defines a maximum number of frames in the interleaving group (see section 2.1). If this parameter is not present, interleaving SHALL not be used. Requires also the octet-align parameter to be sent.

Optional parameters for storage mode: none

Encoding considerations for RTP mode: See chapter 2.

Encoding considerations for storage mode: See section 7.2.

Security considerations: see chapter 5 "Security".
Public specification: please refer to chapter 8 "References".

Additional information for storage mode:
- Magic number: none
- File extensions: amr, AMR
- Macintosh file type code: none
- Object identifier or OID: none

Person & email address to contact for further information:
- johan.sjoberg@ericsson.com
- ari.lakaniemi@nokia.com

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

Author/Change controller:
- johan.sjoberg@ericsson.com
- ari.lakaniemi@nokia.com

7.5 Mapping to SDP Parameters

Please note that this chapter applies to the RTP mode.

Example of usage of AMR in SDP [16], possible GSM gateway scenario:
m=audio 49120 RTP/AVP 97
a=rtpmap:97 AMR/8000
a=fmtp:97 mode-set=0,2,5,7; mode-change-period=2; mode-change-neighbor; maxframes=1

Example of usage of AMR-WB in SDP [16], possible VoIP scenario:
m=audio 49120 RTP/AVP 98
a=rtpmap:98 AMR-WB/16000
a=fmtp:98 octet-align

Example of usage of AMR-WB in SDP [16], possible streaming scenario:
m=audio 49120 RTP/AVP 99
a=rtpmap:99 AMR-WB/16000
a=fmtp:99 octet-align; maxframes=3; interleaving=15

8. References

[1] 3G TS 26.090, "Adaptive Multi-Rate (AMR) speech transcoding".
[3] 3GPP TS 26.190 "AMR Wideband speech codec; Transcoding functions".
[4] 3GPP TS 26.201 "AMR Wideband speech codec; Frame Structure".
[5] IETF RFC 2119, "Key words for use in RFCs to Indicate Requirement Levels".

[6] 3G TS 26.093, "AMR Speech Codec; Source Controlled Rate operation".

[7] 3GPP TS 26.193 "AMR Wideband Speech Codec; Source Controlled Rate operation".

[8] GSM 06.60, "Enhanced Full Rate (EFR) speech transcoding".


[14] GSM 06.92, "Comfort noise aspects for Adaptive Multi-Rate (AMR) speech traffic channels".

[15] 3GPP TS 26.192 "AMR Wideband speech codec; Comfort Noise aspects".


[17] 3G TS 25.415 "UTRAN Iu Interface User Plane Protocols"


[21] 3G TS 26.102, "AMR speech codec interface to Iu and Uu".

[22] 3GPP TS 26.202 "AMR Wideband speech codec; Interface to Iu and Uu".
9. Authors’ addresses

Johan Sjoberg                  Tel:   +46 8 50878230
Ericsson Research              EMail: Johan.Sjoberg@ericsson.com
Ericsson Radio Systems AB      
Torshamnssgatan 23             
SE-164 80 Stockholm, SWEDEN

Magnus Westerlund              Tel:   +46 8 4048287
Ericsson Research              EMail: Magnus.Westerlund@ericsson.com
Ericsson Radio Systems AB      
Torshamnssgatan 23             
SE-164 80 Stockholm, SWEDEN

Ari Lakaniemi                  Tel:   +358 40 5276440
Nokia Research Center          EMail: ari.lakaniemi@nokia.com
P.O.Box 407                    
FIN-00045 Nokia Group, FINLAND

Petri Koskelainen              Email: petri.koskelainen@nokia.com
Nokia Research Center          
P.O.Box 100                    
FIN-33721 Tampere, FINLAND

Tim Fingscheidt                Tel:   +49 89 722 57658
Siemens AG, ICP CD             Fax:   +49 89 722 46489
Grillparzerstrasse 10-18       EMail: Tim.Fingscheidt@mch.siemens.de
D - 81675 Munich, GERMANY

Bernhard Wimmer                Tel:   +49 89 722 23247
Siemens AG, ICP CD             Fax:   +49 89 722 46489
Grillparzerstrasse 10-18       EMail: Bernhard.Wimmer@mch.siemens.de
D - 81675 Munich, GERMANY

Qiaobing Xie                   Tel:   +1-847-632-3028
Motorola, Inc.                 EMail: qxie1@email.mot.com
1501 W. Shure Drive, #2309     
Arlington Heights, IL 60004, USA

Sanjay Gupta                   Tel:   +1-847-435-0306
Motorola, Inc.                 EMail: QA4496@email.mot.com
1501 W. Shure Drive, #3205     
Arlington Heights, IL 60004, USA

This Internet-Draft expires September 30, 2001.