

Internet Engineering Task Force
Audio Video Transport WG
INTERNET-DRAFT
February 23, 2001
Expires: August 23, 2001

Ari Lakaniemi, Nokia
Pasi Ojala, Nokia
Johan Sjöberg, Ericsson
Magnus Westerlund, Ericsson

RTP payload format for AMR-WB
<[draft-lakaniemi-avt-amrwb-00.txt](#)>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC 2026](#).

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 for Adaptive Multi-Rate Wideband (AMR-WB) speech encoded signals. The AMR-WB payload format is designed to be able to interoperate with existing AMR-WB transport formats. This document also includes a MIME type registration for AMR-WB.

1. Introduction

The Adaptive Multi-Rate Wideband (AMR-WB) speech codec [1] was originally developed by the Third Generation Partnership Project (3GPP) to be used in GSM and 3G systems. I.e. the AMR-WB codec will be widely used in cellular systems. The AMR-WB codec is developed to preserve high speech quality under a wide range of transmission conditions.

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. Mode adaptation functionality is one valuable aspect of the AMR-WB operation. In mobile radio systems (GSM) it 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, AMR-WB mode adaptation can be also utilized to adapt to the varying available transmission bandwidth. Basically the mode change can occur to any mode at any time.

The name and operational principles of the AMR-WB codec largely resemble those of the Adaptive Multi-Rate (AMR-NB) codec [2,12]. However, these are two separate speech codecs, the principal difference being that AMR-NB is so-called narrow band speech coding, using 8000 Hz sampling frequency, compared to 16000 Hz of the AMR-WB.

The AMR-WB codec is designed with a voice activity detector (VAD) [6] and generation of comfort noise (CN) parameters during silence periods [5]. Hence, the AMR-WB codec can reduce the number of transmitted bits and packets during silence periods to a minimum. The operation to send silence descriptor (SID) frames containing CN parameters at regular intervals non-speech periods is usually called discontinuous transmission (DTX) or source controlled rate (SCR) operation [4].

AMR-WB implementations must support all 9 speech coding modes. AMR-WB mode switching can occur between any speech frames, and current mode must be indicated by transmitting the mode information together with the speech encoded bits. The objective of AMR-WB design has been to enable highest possible speech quality under a variety of transmission channel conditions. To realize the mode adaptation the receiver needs to signal the AMR-WB mode it prefers to receive to the transmitter.

Due to the flexibility and robustness of AMR-WB, it is 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. class A bits. 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 of the damaged data. 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 [14], bit error tolerant AMR-WB traffic could achieve better performance over these types of links.

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

- 1) Utilizing the a partial checksum to cover headers and the most important AMR-WB speech bits of the payload. It is recommended that at least all class A bits are covered by the checksum.
- 2) Utilizing the a partial checksum to only cover headers, but a frame CRC to cover the class A bits of each AMR-WB 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 pays attention to the class B and C residual bit error rate. Though less sensitive to error than class A bits, class B and C 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-WB in UMTS can be found in [17].

Approach 1 is 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 CRCs 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.

To achieve better robustness against packet loss the payload supports Forward Error Correction (FEC). 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. [RFC 2733](#), 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 [15].

Yet another mechanism to enhance error robustness is the interleaving of AMR-WB speech frames. Sometimes several frames can be encapsulated into 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. The 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 usage scenarios. However, e.g. streaming applications are likely to be able to exploit interleaving to improve speech quality in lossy transmission conditions.

2. Requirements

The AMR-WB RTP payload format was designed to meet the following requirements:

- o Different levels of robustness must be supported, from no redundant data to extreme robustness capable of handling very high packet loss rates with no or small speech quality degradation.
- o Fast, bandwidth efficient, frame-wise AMR-WB mode adaptation must be supported. This means that it must be possible to send Codec Mode Requests back from the receiving side to the transmitting side with information on the preferred mode.

- o Source controlled rate operation (SCR) (also called DTX) and comfort noise parameter (CN) transmission defined in AMR-WB must be supported.

3. 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 [RFC 2119](#) [9].

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

The AMR-WB payload format consists of one payload header, a table of content, optionally one CRC per payload frame, and zero or more AMR-WB payload frames. The payload format is made as bandwidth efficient as possible by not using octet alignment for the payload header, table of content or the payload frames. However, the full payload is octet aligned. Therefore any unused bits in the last octet MUST be padded with zeros.

If the option to transmit a robust sorted payload is enabled by the receiver, the transmitter may choose to sort the bits in the payload according to descending bit error sensitivity in order to enable UEP/UED outside RTP (e.g. UDP-lite). The sensitivity order for AMR-WB encoded speech bits for each mode is defined in Annex B of [3], the original bit order being as delivered by the AMR-WB speech encoder [1]. The AMR-WB frame types, or modes, are defined in [3].

Robustness against packet loss can be accomplished by using the possibility to retransmit previously transmitted frames together with the current (new) frame or frames. Another approach is using interleaving to reduced the speech quality effect of packet losses. Note that the usage of these options can be restricted by the MIME parameters during the session set-up. The AMR-WB performance over error tolerant links can be improved by delivering also the speech frames that have been corrupted with bit errors. However, UEP/UED MUST be used in such a way that the bit errors are allowed only in the least error sensitive bits. Bit errors in class A bits MUST NOT be allowed in any circumstances. This payload format provides two alternative methods to implement UED:

A. CRC calculation over the class A speech bits

If several consecutive speech frames are encapsulated into each payload, the optional CRC may be used to protect the class A speech bits of each frame, see table 1. The number of class A bits is

specified as informative in [3] and therefore copied into table 1 as normative for this payload format. Speech frames with errors in

class A bits MUST be marked with SPEECH_BAD for corrupted speech frames (FT=0..8) or SID_BAD for corrupted SID frames (FT=9), and be sent to the speech decoder to assist error concealment, see [7]. In this case the RTP header, payload header, and table of content should be covered by a transport layer CRC, e.g. UDP-lite. A packet MUST be discarded if the transport layer CRC detects errors in these bits.

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 [15]). Note that payloads containing a single frame are sorted in the same robust way regardless of the use of simple or robust sorting. The UED and/or UEP is recommended to cover at least the RTP header, payload header, table of content and all class A bits from all frames in the payload.

Support for unequal error detection is OPTIONAL. If either scheme is to be used, it MUST be signaled out of band (see [section 8](#)).

Index	Mode	Class A bits	total speech bits
0	AMR-WB 6.6	54	78
1	AMR-WB 8.85	64	113
2	AMR-WB 12.65	72	181
3	AMR-WB 14.25	72	213
4	AMR-WB 15.85	72	245
5	AMR-WB 18.25	72	293
6	AMR-WB 19.85	72	325
7	AMR-WB 23.05	72	389
8	AMR-WB 23.85	72	405
9	AMR-WB SID	40	40

Table 1. Specification of the number of class A bits for AMR-WB.

The speech quality in channel error conditions can be improved by delivering also the frames corrupted e.g. in transmission over a radio link to the receiver. Despite the bit-errors, providing damaged frames to the error concealment unit can improve the speech quality compared to case where corrupted frames are dropped. However, to accomplish this, a frame quality indicator is needed to mark the corrupted frames for the decoder. In many communication scenarios the AMR-WB frames 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.

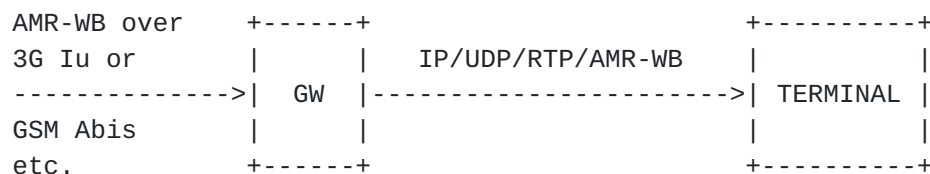


Figure 1: GW to VoIP terminal scenario.

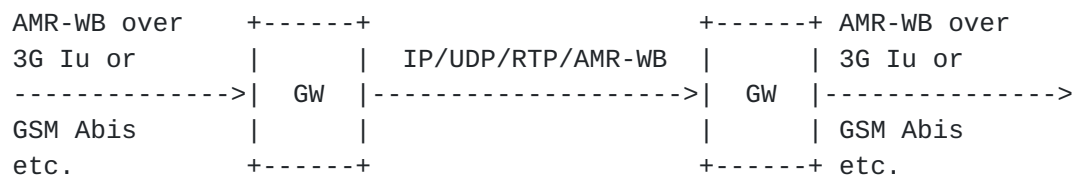


Figure 2. GW to GW scenario.

The speech quality in case of packet losses when transmitting several AMR-WB 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.

3.1. The payload header

The length of the payload header is either 7 or 15 bits, depending on whether the interleaving is used or not. Figures 3a and 3b illustrate the header structure. Header bits are specified in following two subclauses.

3.1.1. Required fields of the payload header

S (1 bit): Indicates, if set, that the bits in the payload is robust sorted. If not set, simple payload sorting is employed. Note that this bit can be set only if the receiver has signaled support for the OPTIONAL robust payload sorting.

C (1 bit): Indicates the existence of OPTIONAL CRC fields in the payload table of content. Note that this bit can be set only if the receiver has signaled support for the OPTIONAL CRC.

I (1 bit): Indicates, if set, that frames in this payload are interleaved, and that ILL and ILP fields are present in the payload header. If not set, frames in this payload are successive frames and

ILL and ILP fields are not present in the payload header. Note that

this bit can be set only if the receiver has signaled support for interleaving.

CMR (4 bits): Indicates Codec Mode Requested for the other communication direction. It is only allowed to request one of the AMR-WB speech modes (frame type index 0...8, see Table 1a in [3]). CMR value 15 indicates that no mode request is present, other values are for future use.

3.1.2. Optional fields of the payload header

ILL (4 bits): OPTIONAL field that is present only if I=1. 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 I=1. 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 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: $n+1$, $n+1+(L+1)$, $n+1+2*(L+1)$, ..., $n+(N-1)*(L+1)$

...

The last packet of an interleave group: $ILL=L$, $ILP=L$
Payload: $s+L$

Frames: $n+L$, $n+L+(L+1)$, $n+L+2*(L+1)$, ..., $n+L+(N-1)*(L+1)$

Interleaved frames MUST be stored in the payload in timestamp-increasing order. Furthermore, the interleaved payloads within an interleave group MUST be sent according to increasing order of ILP field, and each payload of an interleave group MUST contain equal number of frames. It is RECOMMENDED that ILL remains constant throughout the session. If ILL is to be changed, the change SHOULD be done between interleaving groups, i.e. the ILP of the previous packet was L. Furthermore, because of the inter-frame dependent nature of AMR-WB coding, it is RECOMMENDED that ILL values greater than or equal to 2 are used to enable better error recovery in the decoder in case of lost interleaved payload. Note also that using value ILL=0 or using interleaving for payload carrying only one frame is not meaningful.

```

0
0 1 2 3 4 5 6
+--+--+--+--+--+
|S|C|I| CMR |
+--+--+--+--+--+

```

Figure 3a: AMR-WB payload header, I=0.

```

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

```

Figure 3b: AMR-WB payload header, I=1.

3.2. The payload table of content and CRCs

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

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

FT (4 bits): Frame type indicator, indicating the AMR-WB speech coding mode or comfort noise (CN) mode. The mapping of AMR-WB modes to FT is given in Table 1a in [3]. If FT=14 (lost frame) or FT=15 (no transmission/no reception), no CRC or payload frame is present.

Q (1 bit): The frame quality bit indicates, if not set, that the payload is corrupted and the receiver should set the RX_TYPE (see [4]) to SPEECH_BAD or SID_BAD depending on the frame type (FT).


```

0
0 1 2 3 4 5
+--+--+--+--+
|F|  FT  |Q|
+--+--+--+--+

```

Figure 4: Table of content (ToC) entry field.

CRC (8 bits): OPTIONAL field, exists if the payload header bit C is set (C=1). The 8 bit CRC is used for error detection. These 8 parity bits are generated according to section 4.1.4 in [3].

```

0
0 1 2 3 4 5 6 7
+--+--+--+--+--+--+
|      CRC      |
+--+--+--+--+--+--+

```

Figure 5: CRC field.

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

```

0          1          2          3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|F| FT  |Q|F| FT  |Q|F| FT  |Q|      CRC      |      CRC  |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|  |      CRC  |
+--+--+--+--+--+--+--+--+--+

```

Figure 6: The ToC and CRCs for a payload with three speech frames.

3.3. AMR-WB speech frame

An AMR-WB speech frame represents one encoded speech frame encoded using the mode according to the FT field in ToC entry corresponding to this frame. The length of this field is implicitly defined by the AMR-WB mode in the FT field. The AMR-WB speech bits SHALL be sorted according to [Appendix B](#) of [3].

3.4. Compound AMR-WB payload

The compound AMR-WB payload consists of one AMR-WB payload header, the table of content, and one or more AMR-WB payload frames, see [section 3.1](#), 3.2 and 3.3. These can be combined either by using robust or simple payload sorting. The S-bit in the AMR-WB payload

header indicates which method is used.

Definitions for describing the compound AMR-WB payload:

b(m) - bit m of the compound AMR-WB payload
t(n,m) - bit m in the table of content entry for speech frame n
p(n,m) - bit m in the CRC for speech frame n
f(n,m) - bit m in speech frame n
F(n) - number of bits in speech frame n, defined by FT
h(m) - bit m of payload header
H - number of bits in payload header, 7 or 15 bits
C - number of CRC bits, 0 or 8 bits
N - number of payload frames in the payload
S - number of unused bits in the last octet of the payload

Payload frames f(n,m) are ordered in the order they are delivered by the AMR-WB speech encoder, i.e. frame n is preceding frame n+1. All frames between the oldest one and the most recent one MUST be present in the payload, the only exception is interleaving, when the frame order are defined in [section 3.1.2](#). If some of the frames are not available because of a frame loss or they are not transmitted, e.g. due to DTX, those MUST be replaced by lost speech or by no transmission/no reception type frames, respectively.

3.4.1. Robust payload sorting

As described earlier, a bit error in a more sensitive bit is subjectively more annoying than in a less sensitive bit. Therefore, to enable protection of only the most sensitive bits of a payload with a forward error detection code, e.g. a CRC outside RTP, the bits inside a payload can be ordered into sensitivity order. The protection SHOULD cover an appropriate number of octets from the beginning of the payload, covering at least the AMR-WB payload header, ToC, and class A bits (see Table 1). Exactly how many octets that needs protection depends on the network and application. To maintain sensitivity ordering inside the AMR-WB payload, when more than one speech frame is transmitted in one payload, reordering of the bits in the payload is needed.

The AMR-WB payload header, ToC and CRCs SHALL still be placed unchanged in the beginning of the robust sorted payload. Thereafter, the payload frames are sorted with one bit alternating from each AMR-WB payload frame.

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

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

```
}  
/* table of content */  
for (j = 0; j < N; j++){
```

```

    for (i = 0; i < 6; i++){
        b(k++) = t(j,i);
    }
}
/* CRCs */
for (j = 0; j < N; j++){
    for (i = 0; i < C; i++){
        b(k++) = p(j,i);
    }
}
/* payload frames */
max = max(F(0),...,F(N-1));
for (i = 0; i < max; i++){
    for (j = 0; j < N; j++){
        if (i < F(j)){
            b(k++) = f(j,i);
        }
    }
}
/* padding */
S = 8 - k%8;
if (S < 8){
    for (i = 0; i < S; i++){
        b(k++) = 0;
    }
}

```

3.4.2. Simple payload sorting

If multiple frames are encapsulated into the payload and robust payload sorting is not used, the payload is formed as concatenation of the AMR-WB payload header, ToC, possibly optional CRC fields, and the AMR-WB speech frames. However, the bits inside each AMR-WB payload frame are ordered into sensitivity order as defined in Annex B of [3].

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

```

/* payload header */
k=0;
for (i = 0; i < H; i++){
    b(k++) = h(i);
}
/* table of content */
for (j = 0; j < N; j++){
    for (i = 0; i < 6; i++){
        b(k++) = t(j,i);
    }
}

```

```
}  
/* CRCs */  
for (j = 0; j < N; j++){
```

```
    for (i = 0; i < C; 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 */
S = 8 - k%8;
if (S < 8){
    for (i = 0; i < S; i++){
        b(k++) = 0;
    }
}
```

3.5. Decoding security consideration

If the payload length calculation based on C, I, F and FT fields does not indicate the same length as the actually received payload size, the payload should be dropped as erroneous. Decoding AMR-WB frames that are parsed based on erroneous header information could severely degrade the speech quality.

4. RTP header usage

The RTP header marker bit (M) is used to mark (M=1) the payloads containing the first speech frame after a CN period. For all other payloads the marker bit is set to 0 (M=0).

The timestamp corresponds to the sampling time of the first sample of the first encoded AMR-WB frame in the payload. A frame can either be encoded speech, comfort noise parameters, LOST_FRAME, or NO_TRANSMISSION. The unit used to compute timestamp is one sample. The duration of one AMR-WB speech frame is 20 ms and the sampling frequency is 16 kHz, corresponding to 320 speech samples per frame. Thus, the timestamp is increased by 320 for each consecutive frame. If the optional interleaving functionality is not used, all frames in a packet MUST be successive frames, stored in the same order as delivered by the AMR-WB speech encoder. If the interleaving is employed, the frames encapsulated into a payload MUST be picked as defined in [section 3.1.2](#).

5. Congestion Control

The need of congestion control for data transported with RTP has to be considered. 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-WB are 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 that the FEC itself does not worsen the problem. Therefore, it is RECOMMENDED that applications using this payload implements congestion control. The actual mechanism for congestion control is not specified but should be suitable for real-time flows, e.g. [16].

6. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [10]. 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 data, the main security issues are confidentiality and authentication of the speech itself. Some other smaller issues also exist. The payload format itself does not have any support for security. These issues have to be solved by a payload external mechanism.

6.1. 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 frame header as they only carry information about the requested AMR-WB mode, AMR-WB 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 robustness against bit errors will be non, unless an encryption method without error-propagation is used, e.g. a stream cipher. This is only an issue when using UEP/UED, when bit errors can be accepted in some part of the payload.

6.2. Authentication

To authenticate the sender of the speech an external mechanism have 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 to tamper with the packetization of the speech data, some extra data could be protected. The data is: RTP timestamp, RTP sequence number, RTP marker bit. Tampering could result in erroneous decapsulation/decoding that could lower speech quality. Tampering with the AMR-WB mode request field can result in that the sender receives speech in a different quality than desired.

7. Examples

7.1. Simple example

In the simple example one AMR-WB frame is encapsulated into the payload. Simple payload sorting is used ($S=0$), no CRC fields are present ($C=0$), and interleaving is not used ($I=0$). A 23.05 kbps mode is requested for the reverse link ($CMR=7$), and the payload was not damaged at IP origin ($Q=1$). The AMR-WB mode is the 12.65 kbps mode ($FT=2$). The speech encoded bits are put into $f(0...252)$ in descending sensitivity order according to [3].

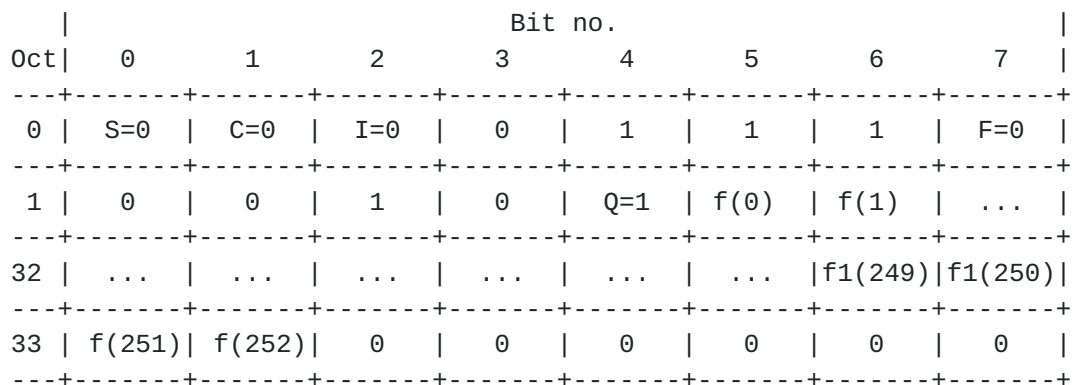


Figure 7: One AMR-WB frame per payload example.

7.2. Example with CRCs

In this example two frames are transmitted in one payload. Simple payload sorting is used ($S=0$), CRC fields are present ($C=1$), and interleaving is not used ($I=0$). No mode request is sent ($CMR=15$), and neither of the frames is corrupted ($Q=1$). The payload contains one frame at 14.25 kbps mode ($FT=3$) and one frame at 15.85 kbps mode ($FT=4$). Bits $p1(0...7)$ and $p2(0...7)$ mark the CRC checksum for the first and second frames, respectively. The bits of the first frame

are denoted by $f_1(0 \dots 284)$, and bits of the second frame are marked by $f_2(0 \dots 316)$.

	Bit no.							
Oct	0	1	2	3	4	5	6	7
0	S=0	C=1	I=0	1	1	1	1	F=1
1	0	0	1	1	Q=1	F=0	0	1
2	0	0	Q=1	p1(0)	p1(1)	p1(2)	p1(3)	p1(4)
3	p1(5)	p1(6)	p1(7)	p2(0)	p2(1)	p2(2)	p2(3)	p2(4)
4	p2(5)	p2(6)	p2(7)	f1(0)	f1(1)	...	...	...
40	...	...	...	...	...	...	f1(283)	f1(284)
41	f2(0)	f2(1)	...	...	...	...	...	...
80	...	...	...	f2(315)	f2(316)	0	0	0

Figure 8: Example with two AMR-WB frames and CRCs.

7.3. Example with multiple frames per payload and robust sorting

In this example two frames are transmitted in one payload with robust sorting ($S=1$). No CRC is used ($C=0$), interleaving is not used ($I=0$), and 8.85 kbps mode frame is requested from the reverse link ($CMR=1$). Both frames are undamaged ($Q=1$), and the two frames in the payload are encoded at 14.25 kbps ($FT=3$) and 15.85 kbps ($FT=4$) modes. The first frame is represented by $f1(0...284)$ and the subsequent frame by $f2(0...316)$.

	Bit no.							
Oct	0	1	2	3	4	5	6	7
0	S=1	C=0	I=0	0	0	0	1	F=1
1	0	0	1	1	Q=1	F=0	0	1
2	0	0	Q=1	f1(0)	f2(0)	f1(1)	f2(1)	...
74	...	f1(283)	f2(283)	f1(284)	f2(284)	f2(285)	f2(286)	...
78	...	...	...	f2(316)	f2(317)	0	0	0

Figure 9: Example with two AMR-WB frames per payload and robust

sorting.

8. The AMR-WB MIME type registration

This chapter defines the MIME type for the Adaptive Multi-Rate Wideband (AMR-WB) speech codec. AMR-WB implementations according to [1] MUST support all nine coding modes. The fast mode adaptation is supported by transmitting the mode information in-band together with encoded speech data to allow mode change without any additional signaling. Furthermore, fast mode adaptation requires transmission of codec mode request inside payload.

In addition to the speech codec, AMR-WB specifications also include Discontinuous Transmission / comfort noise (DTX/CN) functionality [4]. The DTX/CN switches the transmission off during silent periods of the speech and only SID frames containing CN parameter updates are sent at regular intervals. Also the AMR-WB DTX/CN MUST be supported.

It is possible that the receiver may only want to receive a certain AMR-WB mode or a subset of AMR-WB modes, due to link limitations in some cellular systems, e.g. the GSM/GERAN radio link can require that only a subset of AMR-WB modes is used. Therefore, it is possible to request a specific set of AMR-WB modes in capability description and the encoder MUST abide this request. If the request for mode set is not given, any mode may be used or requested.

The AMR-WB 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 possibility to define the minimum number of frames between mode changes and to limit the mode change to happen into neighboring modes only.

The receiver can limit the number of AMR-WB frames encapsulated into one RTP packet, and if maximum number of frames per packet is given in capability description, the transmitter MUST comply with this limitation. This is an OPTIONAL feature and if no parameter is given in capability description, the transmitter can encapsulate any number of AMR-WB 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 enable unequal error protection and/or detection outside RTP, 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 AMR-WB frames per packet can be improved by using OPTIONAL frame

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 specifying `maxframes=N` and `interleaving=L`, the maximum size of an interleave group would be $N*(L+1)$ (see [section 3.1.2](#) for details on interleaving).

8.1. 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.

Media Type name: audio

Media subtype name: AMR-WB

Required parameters: none

Optional parameters:

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 [\[3\]](#), an example is given in [section 8.4](#)). 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. If not present, change between any two modes in the active codec mode set is allowed.

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

crc: If present, transmission of CRCs in the payload is supported, otherwise not supported.

robust-sorting: If present, robust payload sorting is supported, otherwise not supported and simple payload sorting **SHALL** be used.

interleaving: Indicates that the frame interleaving is supported and defines a maximum value for interleaving length field ILL (see [section 3.1.2](#)). If this parameter is not present, the interleaving is not supported.

Encoding considerations: See [section 3](#) in this document.

Security considerations: see chapter 6 "Security Consideration".

Public specification: please refer to chapter 9 "References".

Person & email address to contact for further information:

ari.lakaniemi@nokia.com

pasi.s.ojala@nokia.com

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

Author/Change controller:

ari.lakaniemi@nokia.com

pasi.s.ojala@nokia.com

8.2. Mapping to SDP Parameters

Parameters are mapped to SDP [11] as usual.

Example usage in SDP:

```
m=audio 49120 RTP/AVP 97
```

```
a=rtpmap:97 AMR-WB/16000
```

```
a=fmtp:97 mode-set=2,3,4,5,6; maxframes=1
```

9. References

- [1] 3GPP TS 26.190 "AMR Wideband speech codec; Transcoding functions".
- [2] 3GPP TS 26.090 "AMR speech codec; Transcoding functions".
- [3] 3GPP TS 26.201 "AMR Wideband speech codec; Frame Structure".
- [4] 3GPP TS 26.193 "AMR Wideband Speech Codec; Source Controlled Rate operation".
- [5] 3GPP TS 26.192 "AMR Wideband speech codec; Comfort Noise aspects".
- [6] 3GPP TS 26.194 "AMR Wideband speech codec; Voice Activity Detector (VAD)".
- [7] 3GPP TS 26.191 "AMR Wideband speech codec; Error concealment of lost frames".
- [8] 3GPP TS 25.415 "UTRAN Iu Interface User Plane Protocols".
- [9] IETF [RFC 2119](#), "Key words for use in RFCs to Indicate Requirement Levels".

[10]IETF [RFC 1889](#), "RTP: A Transport Protocol for Real-Time Applications".

- [11] IETF [RFC 2327](#) "SDP: Session Description Protocol", April 1998.
- [12] IETF [draft-ietf-avt-rtp-amr-03.txt](#), "RTP payload format for AMR", work in progress.
- [13] IETF [draft-westberg-realtime-cellular-01.txt](#), "Realtime Traffic over Cellular Access Networks", work in progress.
- [14] IETF [draft-larzon-udplite-03.txt](#), "The UDP Lite Protocol", work in progress.
- [15] IETF [draft-ietf-avt-ulp-00.txt](#), "An RTP Payload Format for Generic FEC with Uneven Level Protection", work in progress.
- [16] S. Floyd, M. Handley, J. Padhye, J. Widmer, "Equation-Based Congestion Control for Unicast Applications", ACM SIGCOMM 2000, Stockholm, Sweden.
- [17] 3GPP TS 26.202 "AMR Wideband speech codec; Interface to Iu and Uu".

10. Authors' addresses

Ari Lakaniemi
Nokia Research Center
P.O.Box 407
FIN-00045 Nokia Group
Finland
E-mail: ari.lakaniemi@nokia.com

Pasi Ojala
Nokia Research Center
P.O.Box 100
FIN-33721 Tampere
Finland
E-mail: pasi.s.ojala@nokia.com

Johan Sjöberg
Ericsson Research
Ericsson Radio System AB
Torshamsgatan 23
SE-164 80 Stockholm
SWEDEN
E-mail: johan.sjoberg@ericsson.com

Magnus Westerlund
Ericsson Research

Ericsson Radio System AB
Torshamsgatan 23

Lakaniemi/Ojala/Sjoberg/Westerlund

[Page 20]

INTERNET-DRAFT

RTP Payload Format for AMR-WB

February 23, 2001

SE-164 80 Stockholm

SWEDEN

E-mail: magnus.westerlund@ericsson.com

This Internet-Draft expires in August 23, 2001.

