Technical Specification Group Services and System Aspects **TSGS#12(01)0309**
Meeting #12, Stockholm, Sweden, 18-21 June 2001

Source: **TSG-SA WG4**

Title: **CRs to TS 26.235 on Update of AMR-NB and AMR-WB RTP payload (Release 4)**

Document for: **Approval**

Agenda Item: **7.4.3**

The following CR, agreed at the TSG-SA WG4 meeting #17, is presented to TSG SA #12 for approval.

Spec	CR	Rev	Phase	Subject	Cat	Vers	WG	Meeting	S4 doc
26.235	001		REL-4	Update of AMR-NB and AMR-WB RTP payload	F	4.0.0	S4	TSG-SA WG4#17	S4-010343

3GPP TSG- SA4 Meeting #17
Naantali, Finland, June 4 – 8

Tdoc S4 01 0343

CR-Form-v3

CHANGE REQUEST

⌘	**26.235**	**CR**	**001**	⌘	rev	**-**	⌘	Current version:	**4.0.0**	⌘

*For **HELP** on using this form, see bottom of this page or look at the pop-up text over the ⌘ symbols.*

Proposed change affects: ⌘ (U)SIM [ ] ME/UE [X] Radio Access Network [ ] Core Network [ ]

Title:	⌘	Update of AMR-NB and AMR WB RTP payload
Source:	⌘	TSG-SA WG4
Work item code:	⌘	IMS-CODEC
Date:	⌘	June 8, 2001
Category:	⌘	**F**
Release:	⌘	REL-4

Use one of the following categories:
- ***F*** *(essential correction)*
- ***A*** *(corresponds to a correction in an earlier release)*
- ***B*** *(Addition of feature),*
- ***C*** *(Functional modification of feature)*
- ***D*** *(Editorial modification)*

Detailed explanations of the above categories can be found in 3GPP TR 21.900.

Use one of the following releases:
- *2 (GSM Phase 2)*
- *R96 (Release 1996)*
- *R97 (Release 1997)*
- *R98 (Release 1998)*
- *R99 (Release 1999)*
- *REL-4 (Release 4)*
- *REL-5 (Release 5)*

Reason for change:	⌘	AMR and AMR WB RTP payload and MIME type registration has changed in IETF after the TS 26.235 specification was approved
Summary of change:	⌘	AMR and AMR WB RTP payload and MIME type registration are updated.
Consequences if not approved:	⌘	AMR and AMR WB RTP payload and MIME type registration are misaligned in IETF and 3GPP.

Clauses affected:	⌘	Section 9 is corrected, normative annex B is changed and normative annex D is removed
Other specs affected:	⌘	[ ] Other core specifications ⌘ [ ] Test specifications [ ] O&M Specifications
Other comments:	⌘	

9 Multimedia stream encapsulation

9.1 MIME media types

The terminal shall declare the mandatory and any optional media streams using the codec specific MIME media types in the associated SDP syntax. The MIME media types for the mandatory and optional codecs shall be according to the corresponding types registered by IANA.

- AMR narrowband speech codec MIME media type as specified in Annex B.
- AMR wideband speech codec MIME media type is specified in Annex ~~D~~B.
- H.263 video codec MIME media type is specified in Annex C.
- MPEG-4 visual simple profile level 0 MIME media type as specified in RFC 3016 [5].
- T.140 Text Conversation MIME media type as specified by RFC 2793 [24].

9.2 RTP payload

RTP payload formats specified by IETF shall be used for real time media streams.

RTP payload format for the AMR narrowband speech codec is specified in Annex B.

RTP payload format for the AMR wideband speech codec is specified in Annex ~~D~~B.

RTP payload format for the ITU-T H.263 video codec is specified in IETF RFC 2429 [3].

RTP payload format for the MPEG-4 visual simple profile level 0 is specified in IETF RFC 3016 [5].

RTP payload format for the ITU-T T.140 text conversation coding is specified in IETF RFC 2793 [24].

Annex B (normative): AMR and AMR-WB RTP payload and MIME type registration

This section specifies the AMR and AMR-WB speech codec RTP payload and MIME type registration.

Note: The intention is to replace this normative annex with the IETF RFC defining the AMR and AMR-WB RTP payload and MIME media type registration when the RFC is available.

B.1 AMR and AMR WB RTP payload

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, the 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 signaled 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 octet. Note that this padding is separate from padding indicated by the P bit in the RTP header.

The AMR frame types, or modes, are defined in [2] and the corresponding description for AMR-WB is found in [4]. The extra comfort noise types specified in table 1a in [2], i.e. frame type 9-11 GSM-EFR CN, IS-641 CN and PDC-EFR CN, MUST NOT be used in this payload format. 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, looks like: {7,7,7,7,8,15,15,8}. Note that packets containing only NO_DATA frames SHOULD not be transmitted. Also, NO_DATA frames at the end of a packet SHOULD NOT be transmitted, except in the case of interleaving. 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. This is done by using a sliding window to group the speech frames to send in each payload, see figure 1. A packet containing redundant frames will not look different from a packet with only new frames. The receiver may receive multiple copies or versions (encoded with different modes) of a frame for a certain timestamp if no packet losses are experienced. If multiple versions of a speech frame is received, it is RECOMMENDED that the mode with the highest rate is used by the speech decoder.

```
--+---------+---------+---------+---------+---------+---------+---------+--
  | f(n-2)  | f(n-1)  |  f(n)   | f(n+1)  | f(n+2)  | f(n+3)  | f(n+4)  |
--+---------+---------+---------+---------+---------+---------+---------+--
   <-     p(n-2)      ->
             <-     p(n-1)      ->
                       <-      p(n)       ->
                                 <-     p(n+1)      ->
                                           <-     p(n+2)      ->
                                                     <-     p(n+3)      ->
```

Figure 1: An example of retransmission where each frame is retransmitted one time in the following payload. f(n-2)..f(n+4) denotes a sequence of speech frames and p(n-2)..p(n+3) a sequence of payloads.

The sender is responsible for selecting an appropriate amount of redundancy based on feedback about the channel, e.g. RTCP receiver reports. To avoid congestion problems, congestion control MUST be considered, see also section B.2. With AMR it is possible to add redundancy with little or no extra bandwidth by switching to an AMR mode with lower rate.

Another approach to increase robustness against packet loss is to use the OPTIONAL frame interleaving to reduce the speech quality effect of packet losses. 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 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 SHOULD 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.

```
                     Class A   total speech
Index    Mode          bits         bits
----------------------------------------
  0      AMR 4.75      42            95
  1      AMR 5.15      49           103
  2      AMR 5.9       55           118
  3      AMR 6.7       58           134
  4      AMR 7.4       61           148
  5      AMR 7.95      75           159
  6      AMR 10.2      65           204
  7      AMR 12.2      81           244
  8      AMR SID       39            39
```

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 improved if damaged frames are forwarded to the speech decoder error concealment unit and not dropped. In many communication scenarios the AMR or AMR-WB 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.

```
AMR or AMR-WB over
I.366.{2,3} or +------+                        +----------+
3G Iu or       |      |      IP/UDP/RTP/AMR    |          |
-------------->|  GW  |----------------------->| TERMINAL |
```

```
GSM Abis          |       |                        |          |
etc.              +-------+                        +----------+
```

Figure 2: GW to VoIP terminal scenario

```
AMR or AMR-WB                                                    AMR or AMR-WB
   over                                                             over
I.366.{2,3} or +-------+                      +-------+ I.366.{2,3} or
3G Iu or       |       |  IP/UDP/RTP/AMR or   |       | 3G Iu or
-------------->|  GW   |--------------------->|  GW   |--------------->
GSM Abis       |       |  IP/UDP/RTP/AMR-WB   |       | GSM Abis
etc.           +-------+                      +-------+ etc.
```

Figure 3: GW to GW scenario

The complete payload consists of one payload header (section B.1.2) a table of contents (section B.1.3) and one or more speech frames (section B.1.4) sorted in either simple or robust order. The process by which the complete payload is assembled is described in section B.1.5.

B.1.1. 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 packets the marker bit is set to zero (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 except if interleaving is employed, then frames encapsulated into a payload MUST be picked as defined in section B.1.2.

The payload MAY be padded using P bit in the RTP header. The assignment of an RTP payload type for this new packet format is outside the scope of this document, and will not be specified here. It is expected that the RTP profile for a particular class of applications will assign a payload type for this encoding, or if that is not done then a payload type in the dynamic range SHOULD be chosen.

B.1.2. The payload header

The payload header consists of a 4 bit codec mode request. If octet aligned operation is used the payload header is padded to fill an octet and optionally an 8 bit interleaving header may extend the payload 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. It is RECOMMENDED that the encoder follows a received mode request, but if the encoder has reason for not follow the mode request, e.g. congestion control, it MAY use another mode. The codec mode request (CMR) MUST be set to 15 for packets sent to a multicast group. The encoder in the sender SHOULD ignore mode requests when sending to a multicast session but MAY use RTCP feedback information as a hint that a mode change is needed. The codec mode selection MAY be restricted by the mode set definition at session set up. If so, the selected codec mode MUST be in the signaled mode set.

R: Is a reserved bit that MUST be set to zero. All R bits MUST be ignored by the receiver.

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

Figure 4: Payload header for bandwidth efficient operation.

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

Figure 5: Payload header for octet aligned operation.

If the use of interleaving is signaled out of band at session set up, octet aligned operation MUST be used. When interleaving is used 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 ILL MUST only be changed between interleave groups. The value of ILP=p in payloads belonging to the same group runs from 0 to L. The interleaving is meaningful only when the number of frames per payload (N) is greater than or equal to 2. All payloads in an interleave group MUST contain equally many speech frames. 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, 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  |R|R|R|R|  ILL  |  ILP  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: Octet aligned operation payload header with interleaving extension.

B.1.3. 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 speech frames in this payload or not. F=1 further frames follow, F=0 last frame.

FT (4 bits): Frame type indicator, indicating the AMR or AMR-WB speech coding mode or comfort noise (SID) mode. The mapping of existing 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, MUST be set to zero.

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

Figure 7: Table of contents entry field for bandwidth efficient operation.

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

Figure 8: An example of a ToC when using bandwidth efficient operation.

```
    0
    0 1 2 3 4 5 6 7
   +-+-+-+-+-+-+-+-+
   |F|  FT   |Q|P|P|
   +-+-+-+-+-+-+-+-+
```

Figure 9: 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 and SHALL only be used in octet aligned operation. 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].

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

Figure 10: 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.

```
    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|P|P|F|  FT   |Q|P|P|F|  FT   |Q|P|P|      CRC      |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
    |        CRC       |        CRC       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 11: The ToC and CRCs for a payload with three speech frames when using octet aligned operation.

B.1.4. 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 mode in the FT field. The bits SHALL be sorted according to Annex B of [2] for AMR and Annex 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.

B.1.5. Compound payload

The compound payload consists of one payload header, the table of contents and one or more speech frames, see section B.1.2, B.1.3 and B.1.4. These elements SHALL be put together to form a payload with either simple or robust sorting. If the bandwidth efficient operation is used, simple sorting MUST be used.

Definitions for describing the compound payload:

b(m)   - bit m of the compound payload, octet aligned
o(n,m) - bit m of octet n in the octet description of the compound payload, bit 0 is MSB
t(n,m) - bit m in the table of contents 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
C(n)   - number of CRC bits for speech frame n, 0 or 8 bits
P(n)   - number of padding bits for speech frame n
N   - number of payload frames in the payload
S   - number of unused bits

Payload frames f(n,m) are ordered in consecutive order, where frame n is preceding frame n+1. Within one payload with multiple speech frames the sequence of speech frames MUST contain all speech frames in the sequence. If interleaving is used the interleaving rules defined in section B.1.2 applies for which frames that are contained in the payload. 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 in the ToC 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. Payloads containing only NO_DATA frames SHOULD NOT be transmitted.

The compound payload, b, is mapped into octets, o, where bit 0 is MSB.

B.1.5.1. 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.

B.1.5.1.1. Simple payload sorting for bandwidth efficient operation

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

```
/* 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)
    }
```

B.1.5.1.2. Simple payload sorting for octet aligned operation

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

```
/* 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 */
    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)
    }
```

B.1.5.2. 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 payload header, ToC and class A bits, see table 1 (AMR) and [4] (AMR-WB). 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 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 payload SHALL be performed on octet level. The 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:

```
/* 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));
    for (i = 0; i < max; i+=8){
      for (j = 0; j < N; j++){
        for (l = 0; l < 8; l++){
          if (i+l < F(j)+P(j)){
            if (i+l< F(j)){
              b(k++) = f(j,i+l);
```

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

B.1.6. 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. Furthermore, all receivers MUST be able to receive any speech frame multiple times, both exact duplicates and in different AMR modes.

B.1.7. 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.

B.2. 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].

B.3. 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 B.1.6), 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, e.g. SRTP [23].

Interleaving MAY affect encryption. Depending on the used encryption scheme there MAY be restrictions on for example the time when keys can be changed.

B.3.1. 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.

B.3.2. 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 codec mode request field can result in that the sender must receive speech in a different quality than desired.

B.4. Examples

B.4.1. Bandwidth efficient examples

B.4.1.1. 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 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  CMR  |F|  FT   |Q|f(0)                                         |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                   f(147)|P|P|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 12: One frame per packet example.

B.4.1.2. 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 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  CMR  |F|  FT   |Q|F|  FT   |Q|f(0)                           |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
```

```
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                 f(131)|g(0)                   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |   g(176)|P|P|P|
 +-+-+-+-+-+-+-+-+
```

Figure 13: Two frame per packet example.

B.4.2. 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.

```
 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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  CMR  |R|R|R|R|  ILL  |  ILP  |F|  FT   |Q|P|P|F|  FT   |Q|P|P|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     CRC1      |     CRC3      |    f1(0..7)   |    f3(0..7)   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |   f1(8..15)   |   f3(8..15)   |  f1(16..23)   |  f3(16..23)   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 : ...                                                           :
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |f1(152..158) |P|f3(152..158) |P|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 14: Example with CRCs, interleaving and robust sorting.

B.5. 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.

B.5.1. 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 be used if the receiver has signaled the use of 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 SHALL be used if the receiver has signaled the use of 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 be applied if the receiver has signaled the use of it in the capability description, and 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 B.1.2 for details on interleaving).

B.5.2. Storage mode

The storage mode is used for storing speech frames, e.g. as a file or e-mail attachment. The file begins with a magic mumber to identify that it is an AMR or AMR-WB file. AMR and AMR-WB have different magic numbers. The magic number for AMR corresponds to the ASCII character string "#!AMR\n" and for AMR-WB "#!AMR-WB\n", i.e. 0x2321414d520a and 0x2321414d522d57420a.

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 B.1.3)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 9. Following this first octet comes the encoded speech frames bits (see section B.1.4). The last octet of each frame is padded with zeroes, if needed, to achieve octet alignment. An example is given in figure 15.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|P|  FT   |Q|P|P|                                               |
+-+-+-+-+-+-+-+-+                                               +
|                                                               |
+                Speech bits for frame n                        +
|                                                               |
+                                                           +-+-+
|                                                           |P|P|
```

+-+

Figure 15: An example of storage format with one AMR 5.9 kbit/s frames (118 speech bits). Note that bits marked with P, "padding" MUST be set to zero.

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.

B.5.3. 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 and no other signal indicate octet aligned operation, bandwidth 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 B.5.5). 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. Implies automatically that octet-align operation is used.

robust-sorting: If present, the payload SHALL employ robust payload sorting. If not present simple payload sorting SHALL be used. Implies automatically that octet-align operation is used.

interleaving: Indicates that frame level interleaving SHALL be used and its value defines a maximum number of frames in the interleaving group (see section B.1.2). If this parameter is not present, interleaving SHALL not be used. Implies automatically that octet-align operation is used.

Optional parameters for storage mode: none

Encoding considerations for RTP mode: See chapter 2 of RFC XXXX.

Encoding considerations for storage mode: See section 6.2 of RFC XXXX.

Security considerations: see chapter 4 "Security" of RFC XXXX.

Public specification: please refer to chapter 7 "References" of RFC XXXX.

Additional information for storage mode:
Magic number: !#AMR\n
File extensions: amr, AMR
Macintosh file type code: none

Object identifier or OID: none
Intended usage: COMMON. It is expected that many VoIP applications (as well as mobile applications) will use this type.

B.5.4. 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 and no other signal indicate octet aligned operation, bandwidth 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. Implies automatically that octet-align operation is used.

robust-sorting: If present, the payload SHALL employ robust payload sorting. If not present simple payload sorting SHALL be used. Implies automatically that octet-align operation is used.

interleaving: Indicates that frame level interleaving SHALL be used and its value defines a maximum number of frames in the interleaving group (see section B.1.2). If this parameter is not present, interleaving SHALL not be used. Implies automatically that octet-align operation is used.

Optional parameters for storage mode: none

Encoding considerations for RTP mode: See chapter 2 of RFC XXXX.

Encoding considerations for storage mode: See section 6.2 of RFC XXXX.

Security considerations: see chapter 4 "Security" of RFC XXXX.

Public specification: please refer to chapter 7 "References" of RFC XXXX.

Additional information for storage mode:
Magic number: #!AMR-WB\n
File extensions: awb, AWB
Macintosh file type code: none
Object identifier or OID: none

B.5.5 Mapping to SDP Parameters

Please note that this chapter applies only 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 maxframes=3; interleaving=15

B.6. References

[1] 3G TS 26.090, "Adaptive Multi-Rate (AMR) speech transcoding".

[2] 3G TS 26.101, "AMR Speech Codec Frame Structure".

[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".

[9] TIA/EIA -136-Rev.A, part 410 - "TDMA Cellular/PCS - Radio Interface, Enhanced Full Rate Voice Codec (ACELP). Formerly IS-641. TIA published standard, 1998".

[10] ARIB, RCR STD-27H, "Personal Digital Cellular Telecommunication System RCR Standard".

[11] IETF RFC1889, "RTP: A Transport Protocol for Real-Time Applications".

[12] IETF draft-westberg-realtime-cellular-01.txt, "Realtime Traffic over Cellular Access Networks".

[13] IETF draft-larzon-udplite-04.txt, "The UDP Lite Protocol".

[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".

[16] M. Handley and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998

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

[18] S. Floyd, M. Handley, J. Padhye, J. Widmer, "Equation-Based Congestion Control for Unicast Applications", ACM SIGCOMM 2000, Stockholm, Sweden

[19] IETF draft-ietf-avt-ulp-00.txt, "An RTP Payload Format for Generic FEC with Uneven Level Protection ".

[20] IETF RFC2733, "An RTP Payload Format for Generic Forward Error Correction".

[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".

[23] draft-ietf-avt-srtp-00.txt, "The Secure Real Time Transport Protocol".