RTP Payload for AMR-WB+ audio codec

draft-ietf-avt-rtp-amrwbplus-02.txt

Johan Sjöberg, Ericsson

Magnus Westerlund, Ericsson

Ari Lakaniemi, Nokia

IPR notice

- Nokia believes there is an unpublished Nokia patent application that may be relevant to this draft

- See
  http://www.ietf.org/ietf/IPR/nokia-ipr-draft-ietf-avt-rtp-amrwbplus-01.txt
  http://www.ietf.org/ietf/IPR/NOKIA

3GPP TSG SA WG4 news

- AMR-WB+ is one of the recommended codecs for 3GPP Packet Switched Streaming services (PSS) and for Multimedia Messaging Service (MMS).
- Final decision/approval by TSG SA plenary in September.
- AMR-WB+ technical specifications are ready and published.
- AMR-WB+ is one of the candidates to become a mandatory codec for Multimedia Broadcast/Multicast Service (MBMS).

Problems with previous solutions

The main problem for the payload format described in `draft-ietf-avt-rtp-amrwbplus-02.txt` is the overhead. The flexibility of the `-02` proposal costs quite some overhead, which is proportional to the number of frames per packet. Asymptotically, assuming many frames per packet, this overhead is 2 bytes/frame for basic mode and 4 bytes per frame for interleaved mode, leading to 800 bps or, resp., 1600 bps overhead if AMR-WB+ is operated with overclocking factor OC=1. Considering the fact that FT and ISF switching can be considered rare cases and also the TFI evolves predictable in time, we are wasting the overhead for transmission of redundancy.

New proposal - Principal

The new proposal takes into account the fact that FT switching is seldom and that ISF switching is even more seldom. For this reason it appears possible without big loss to constraint that the ISF must be CONSTANT throughout a packet. With the FT, there are less restrictions. It is rather assumed that the FT is constant for certain contiguous groups of frames. The TFI is regarded required information only for the first frame in the packet. For all other frames, the TFI can easily be deduced as long as the sequence number, or better the time difference (counted in frames) for all frames in the packet is known.

New Proposal – Description

This let's us define a payload header as follows:

```
+----------------+----------------+- ... -+----------------+
| Global Header  |  Toc entry #1  |       |  ToC entry #N  |
+----------------+----------------+- ... -+----------------+
```

The global header field which is required once per packet and contains the following information (1 byte):

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|  ISF    |TFI|L|
+-+-+-+-+-+-+-+-+
```

ISF: Indicates the internal sampling frequency employed for the corresponding frame. The index values correspond to internal sampling frequency as specified in Table 24 in [1]. This field SHALL be set to 0 for Frame Type values 0-13.

TFI (2 bits): An index from 0 (first) to 3 (last) indicating the position of the first audio frame of the payload in the AMR-WB+ superframe. This Field needs to be calculated also for frame type 0-9, but the value SHALL be ignored.

L (1 bit): Long displacement fields, indicates if the displacement field is 4 or 8 bits.

New Proposal – Description

Each ToC entry is of the following structure in basic (non-interleaved) mode:

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F| Frame Type  |   #frames     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

F (1 bit): If set to 1, indicates that this ToC is followed by another ToC; if set to 0, indicates that this ToC is the last in this payload header.

Frame Type (7 bit): Indicates the audio codec frame type used for the corresponding frame. Indicates the combination of AMR-WB+ core and stereo rate, special AMR-WB+ frame types, the AMR-WB rate, or comfort noise, as specified by Table 25 in [1].

#frames (8 bit): This field indicates the number of audio frames corresponding to the ToC entry. The number of frames may be between 1 and 256, indicated by entries from 0 to 255.

New Proposal – Description

and in interleaved mode with additional Displacement entries:

```
 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| Frame Type  |    #frames    | DIS1  |  ...  | DISi  |  ...  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ...  |  ...  | DISm  | padd  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The following fields are optional and shall only be present in case interleaving is used:

DIS1 .. DISn (4 or 8 bits): A list of m (m=#frames) displacement fields indicating the displacement of the i-th (i=1..#frames) audio frame relative to the preceding audio frame in the payload. Unlike in the existing draft the displacement is NOT any longer a time stamp offset. Rather it is specified in number of audio frames. The corresponding time stamp offset is easily derived from the audio frame length, which is constant throughout the packet. The displacement values may be between 0 and 15 encoding the number of audio frames between the i-th and the (i-1)-th frame in the payload.

Padd: In case the m-th displacement field is not aligned with the 4 LSB of the byte carrying it, padding bits shall be added in order to fill up the byte. All padding bits shall be zero.

New Proposal – Trade Off

- Reduction in overhead. Overhead for basic mode without switching costs 3 bytes per packet, e.g. for 5 frames per packet OC=1, overhead becomes 240 bps. Interleaving add 200 or 400 bps.
- In Basic payload format configuration (aggregation) ISF switching requires separate packets for the frames belong to the previous ISF value, and the ones with the new ISF value.
- In Interleaved payload format configuration an ISF switching requires ending of the previous interleaving pattern, and restarting it for the new one.

Next steps

- New draft `draft-ietf-avt-rtp-amrwbplus-03.txt` to be published soon including the new proposal.
- Your feedback
  - Questions?
  - Comments?
  - Suggestions?