### Source: Nokia

### Title: AMR-WB+ configurations in characterisation test phase 2

**Document for: Discussion**

**Agenda Item: 8**

# Introduction

This document introduces the coding modes and packetisation parameters used for the Extended Adaptive Multi-Rate (AMR-WB+) codec in the 3GPP PSS/MMS/MBMS audio codec characterisation tests.

# Encoding and packetisation

The factors having an effect on the overall bit-rate employed by the encoded audio stream are the codec source bit-rate, number of audio frames per packet, and settings of the RTP payload format. Furthermore, in the characterisation tests also the compressed IP/UDP/RTP header was counted in as part of the audio bit-rate. Additional factor regarding packetisation strategy is the buffering delay required in the receiver for correct reconstruction of the received audio stream in a timely manner, especially if the frame-level interleaving is used.

Selecting the optimal packetisation strategy is inevitably a trade-off between audio quality in error-free channel and robustness against packet losses: by increasing the number of audio frames per packet we can reduce overhead and thereby use higher source coding bit-rate, but at the same time increased number of frames per packet implies larger block of missing data in case a packet loss occurs. Although frame-level interleaving can be used to reduce effect of a packet loss in case we carry several frames per packet, interleaving also introduces additional delay, which needs to be limited for most real-life applications. Thus, the optimal choice is therefore heavily dependent on the range of transmission conditions we want to prepare to, as well as the delay requirements set by the application.

## SDU size for AMR-WB+

The assumption was that the IP/UDP/RTP header compressed by ROHC algorithm occupies on average 8 bytes per SDU, and the rest of the available bandwidth is used by the audio (RTP) payload. Furthermore, we assumed usage of constant codec bit-rate within an experiment. With this choice the AMR-WB+ payload structure introduces 3 bytes of payload header information for each SDU when basic mode payload is employed. If the interleaved mode of the payload is used, there is one additional byte of information for each audio frame carried in the payload. Hence, we have the following formulas for computing the SDU size (as bytes per packet) including the ROHC information:

SDU\_size = 8 + 3 + n \* Frame\_size (basic mode payload)

and

SDU\_size = 8 + 3 + n \* (1 + Frame\_size) (interleaved payload)

In above formulas n stands for number for audio frames per packet.

The overall bandwidth (as bytes) can be computed by taking into account the SDU transmission rate (as packets per second):

BW = SDU\_size \* Packet\_rate

## Buffering delay in the receiver

Since in a streaming scenario it is reasonable to assume that we are transmitting pre-encoded and pre-stored audio, there is no delay component due to algorithmic or processing delays in the encoder side – i.e. we assume that all frames belonging to a stream are immediately available for transmission. Thus, the main delay component we need to take into account is the buffering delay required in the receiver to allow full reconstruction of received audio stream in timely manner. In practice a receiver cannot know the buffering delay required for correct and uninterrupted reconstruction of the audio stream in case of interleaving without the sender explicitly telling it. On the other hand, a receiver does not need to know the exact interleaving pattern – it is enough to know the buffering requirements (as number of frames or bytes) and required initial buffering time to ensure correct reconstruction.

However, for these experiments we have chosen a regular interleaving pattern(s) for simplicity. In our approach we pick frames at d+1 frame intervals, and encapsulate n frames per packet, which results in d+1 packets per interleaving block. The resulting block size (as number of frames) is

N = n \* (d + 1)

spanning over time

Dur = N \* Frame\_duration.

Figure 1 shows an example of an interleaving block with d = 4 and n = 2.



Figure 1: An example of interleaving with regular pattern.

Furthermore, we have assumed buffering time covering a full interleaving block, which would actually ensure uninterrupted flow of frames even when using an irregular interleaving pattern within a block.

## Experiment 2-1

The overall bit-rate available for audio stream in this experiment is 16 kbit/s = 2000 bytes per second. The input audio signal was encoded using coding mode 17 (see Table 25 in 3GPP TS 26.290 for full list of mode index specifications) requiring 30 bytes per frame (= 12 kbit/s) and encapsulated two audio frames in each payload. Furthermore, the interleaving depth of 48 was used. This implies SDU size

SDU\_size = 8 + 3 + 2 \* (1 + 30) = 73 bytes / packet.

Using ISF=1 with two 20 ms frames per packet, we can transmit on average at 40 ms intervals to provide real-time service, which implies 25 packets per second. Combining these we will have overall bandwidth of

BW = 73 bytes / packet \* 25 packets / second = 1825 bytes / second,

which translates into 14.6 kbit/s.

The duration of an interleaving block using these packetisation parameters is

Block\_duration = 2 \* (48 + 1) \* 20 ms = 1.96 s

## Experiment 2-2

In this experiment the maximum bit-rate for the audio stream was 24 kbit/s. Encoding using mode 37 (53 bytes per frame = 21.2 kbit/s) and transmitting 2 frames per packet with interleaving depth 48 we will have SDU size and bandwidth requirements

SDU\_size = 8 + 3 + 2 \* (1 + 53) = 119 bytes / packet

BW = 119 bytes / packet \* 25 packets / second = 2975 bytes / second,

implying bit-rate of 23.8 kbit/s.

The number of packets and interleaving step are the same as in experiment 2-1, resulting in 1.96 second interleaving block duration.

## Experiment 2-3

For the 20 kbit/s audio channel we used encoding mode 31 (43 bytes per frame = 17.2 kbit/s) with 2 frames per payload packetisation, using interleaving depth 48. This implies SDU size and bandwidth

SDU\_size = 8 + 3 + 2 \* (1 + 43) = 99 bytes / packet

BW = 99 bytes / packet \* 25 packets / second = 2475 bytes / second,

leading to 19.8 kbit/s audio bit-rate.

Also in this case the duration of an interleaving block is 1.96 seconds.

## Experiment 2-4

For the 40 kbit/s case we encoded using mode 37 (53 bytes per frame) and internal sampling frequency (ISF) 38.4 kHz. Selected ISF changes the frame duration to 13.33 ms, which means 75 frames per second (= 31.8 kbit/s). In this scenario we encapsulated only single frame per packet, and interleaving was not used. The resulting SDU size and bandwidth are

SDU\_size = 8 + 3 + 1 \* 53 = 64 bytes / packet

BW = 64 bytes / packet \* 75 packets / second = 4800 bytes / second,

resulting in 38.4 kbit/s bit-rate.

Since in this case we have no interleaving, the packetisation and buffering delay is 20 ms.

# Simulation tools

This section briefly describes the processing chain prepared for channel error simulation.



## AMR-WB+ encoder and packetisation

The preprocessed and concatenated material was encoded using the floating-point AMR-WB+ encoder [1] in mode configurations mentioned in Section 2. A separate packetisation tool was prepared for creating AMR-WB+ RTP packets with selected parameters (number of frames per packet, interleaving settings). This tool naturally also provides an interface between the encoded AMR-WB+ audio frames and the network simulator used for modelling the network.

## Error insertion

The network simulator provided by Siemens performed the error insertion, i.e. removal of RTP packets from the .rtp file. The simulator was configured for each experiment by selecting the bearer and error pattern.

## RTP receiver and buffer

Another tool was prepared for parsing the RTP packets back to AMR-WB+ audio frames. The RTP packet stream modified by the network simulator was given as input to this RTP receiver tool, which parsed the AMR-WB+ RTP payload, performed appropriate buffering, and reconstructed the stream of audio frames based on timestamps. Frames carried by the packets that were lost due to error insertion were replaced by FRAME\_ERASURE type of frames to trigger error concealment in the decoder.

## AMR-WB+ decoder

The reconstructed audio stream was decoded using the fixed-point AMR-WB+ decoder [2], as specified in the test plan

# References

[1] TS 26.304 "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Floating-point ANSI-C code"

[2] TS 26.273 ANSI-C code for the fixed-point Extended Adaptive Multi-Rate - Wideband (AMR-WB+) speech codec