Reducing packet loss by employing damaged speech data

Florian Hammer, Tomas Nordström, Peter Reichl, Gernot Kubin

Abstract

In state-of-the-art access networks like xDSL or UMTS, lots of potentially useful speech data packets are dropped due to bit errors. Many of these packet losses could be avoided by the use of UDPlite, a modified UDP protocol, which allows for unequal error protection of the voice data. While the classic UDP calculates the checksum over the entire payload, and drops the packet if in error, UDPlite allows checksum coverage of only an arbitrary part of the packet. The paper starts by pointing out the close relationship between this new transport protocol proposal and recent speech codec developments, like the 3GPP Adaptive MultiRate (AMR) codec, which provides "uneven level protection" (ULP) by ordering the speech coder output bits by their subjective importance.

The rest of the paper focuses on using UDPlite and the ULP of the AMR-speech-codec (at 12.2 kb/s) to simulate RTP/UDPlite/IP voice transport over links with bit errors ranging from 10^{-5} to 10^{-3}. To this end, three transmission scenarios are distinguished:

1. Traditional IP-transport. The UDPlite checksum covers the entire packet, thus the packet is dropped if a bit error occurs anywhere in the packet.
2. The UDPlite checksum covers the header and the perceptually important bits. This prevents packet loss in case of bit errors in perceptually low important parts of the speech data.
3. The UDPlite checksum covers the packet header only and keeps all damaged speech data.

In any case, whenever a packet is dropped, the packet loss concealment of the AMR codec is used.

The resulting speech samples have been evaluated using the PESQ-algorithm, an ITU-T standard (Rec. P.862) for “Perceptual Evaluation of Speech Quality”. The evaluation indicates that the use of as much damaged speech data as possible is beneficial with respect to the observed speech quality. Thus, our results indicate that the UDPlite checksum should only cover the IP/UDP/RTP header and none of the speech data.