LINK-LEVEL UNEQUAL ERROR DETECTION FOR SPEECH TRANSMISSION OVER 802.11 NETWORKS

Antonio Servetti, Juan Carlos De Martin[1]

Dipartimento di Automatica e Informatica / IEIIT-CNR[1]
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: [servetti|demartin]@polito.it

ABSTRACT

We present a technique to enhance the 802.11 link layer effectiveness by changing the scope of the standard 802.11 error detection step for speech packets. More specifically the proposed technique instructs the link layer to retransmit speech packets only if errors are detected in the most perceptually sensitive bit class, instead of plainly retransmitting all corrupt packets irrespective of error position. Experiments performed using the GSM-AMR speech coding standard showed that end-to-end delays are considerably reduced with respect to the reference case, while speech quality improves since the negative effect of errors in the less perceptually important bits is counterbalanced by the lower number of speech packets discarded because of retransmission limits.

1. INTRODUCTION

In recent years the IEEE 802.11 wireless standard has been adopted in an increasing number of places: many shopping malls, airports, train stations, universities, have now wireless infrastructures to provide people with tetherless access to the Internet. At the same time several portable devices, laptops, and even motor vehicles come with wireless support to benefit from the communication opportunities offered by ubiquitous network accessibility. This emerging scenario is creating the basis a large deployment of wireless multimedia applications.

So far, wireless LAN's have been mainly designed and developed to provide Internet based services. A strong interest is, however, emerging towards multimedia applications, and in particular towards interactive voice applications. WLAN-based telephony, in fact, can bring to the wireless environment all the typical advantages of Voice over IP, including the adotion of a single infrastructure for both data and voice traffic.

This work was supported in part by MIUR, Project FIRB-PRIMO, http://primo.ismb.it.

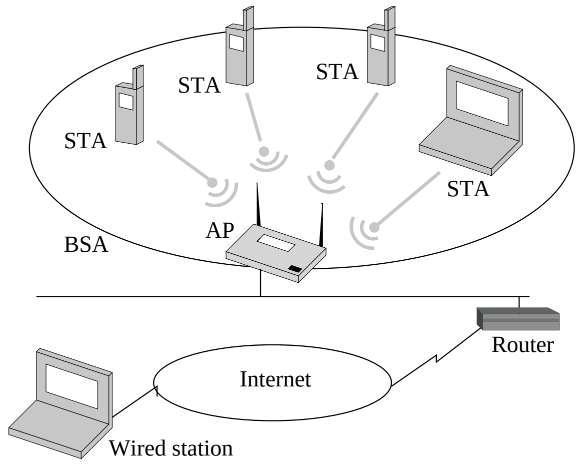


Fig. 1. IEEE 802.11-based network communications scenario.

However several challenges need to be addressed to provide successful interactive multimedia applications over a network originally designed for generic data traffic and characterized by potentially high error rates. Multimedia data require a completely different network behavior: throughput is no more the only parameter for measuring network performance, and packet losses can be –to some extent– tolerated if counterbalanced by timely packet delivery. Multimedia communications have strictly bounded quality of service requirements in terms of packet losses, end-to-end delays and jitter [1]. For example interactive voice communications require an end-to-end delay not greater then 150 ms, thus heavily restricting the possibility to retransmit lost or corrupted packets.

Moreover, in a wireless scenario, where packet losses are above all due to channel errors, audio and video applications prefer damaged packets over lost packets, i.e., multimedia compression algorithms provide a certain degree of error resilience so that the decoder can tolerate corrupted packets. Traditional cellular voice applications already benefit from Unequal Error Protection (UEP) schemes: speech data bits are classified by their importance in different groups and only the most valuable ones are protected by Forward Error Correction codes [2]. Current WLAN's protocols do not support this characteristic and packet containing only a small part of corrupted data are often discarded.

In this paper we propose to modify the link layer of IEEE 802.11 networks to better support voice communications allowing partially corrupted packets to be forwarded (and not discarded) without requiring additional retransmissions. IEEE 802.11 MAC layer provides a checksum to prevent forwarding of erroneous frames, if a bit or more are corrupted the packet is discarded and the sender will retransmit the data until a maximum retransmission limit is reached [3]. Since speech data bits are known to have different perceptual importance, they can be packed in sensitivity order and a checksum can be applied only to the most important subset. Partial checksum will prevent useful frames to be dropped and will reduce the number of retransmissions, thus reducing the network load and delay when the error probability is high.

Previous work addressed the problem of multimedia transmission over lossy networks suggesting to apply partial checksums to the UDP transport protocol [4]: if the packet is received with errors, it is delivered to the application only if the checksummed bits are correct, otherwise the packet is dropped. This solution manages to preserve useful data for use by the application layer, but even higher gains can be achieved if the idea is implemented at the link layer (as already mentioned in [5]): hop–by–hop retransmission, in the case of wrong checksum, is very fast, and the delivery of an "acceptable" packet can be pursued with lower delay.

The paper is organized as follows. In Section 2, we introduce the wireless Voice over IP scenario and the used speech coder. In Section 3, we describe the proposed partial checksum scheme. Results and conclusions are presented in Section 4 and 5, respectively.

2. VOICE COMMUNICATIONS OVER 802.11 WLAN'S

Voice over IP over wireless packet networks is becoming increasingly attractive. However two-way conversational applications are characterized by stringent requirements on end-to-end delay and packet losses can not increase too much to prevent significant perceptual degradation.

The WLAN environment is quite challenging on two counts: the wireless link is inherently noisy, due to fading and interference; the contention-based Medium Access Control (MAC) layer and the retransmission-based error control scheme may introduce strong delays.

Efficient WLAN-based telephony systems must thus be designed carefully to overcome the difficulties of the environment if toll quality service is to be delivered.

2.1. IEEE 802.11 Wireless LAN's

Users may conveniently access the Internet via Wireless LAN technology. Bridging functionality is provided by access points that interconnect wireless nodes to the wired infrastructure, i.e. the IEEE 802.11 WLAN in infrastructure mode.

The IEEE 802.11b physical layer describes a Direct Sequence Spread Spectrum (DSSS) system with an 11 Mbps bit-rate [3]. The MAC sublayer is responsible for the channel allocation procedures, frame formatting, error checking, fragmentation and reassembly. The fundamental transmission medium defined to support asynchronous data transfer on a best effort basis is called Distributed Coordination Function (DCF). It operates in a contention mode requiring all stations to contend for access to the channel for each packet transmitted. Contention services promote fair access to the channel for all stations.

In the IEEE 802.11 MAC, each data-type frame consists of the following basic components: a MAC header, a variable length information frame body, and a Frame Check Sequence (FCS). This last field is a 32-bit checksum used to verify the integrity of the data-unit. Upon packet transmission the destination station compares the packet FCS with a new one computed over all the received MAC bits. Only if all the bits are correct the two FCS's match each other and the packet is positively acknowledged by sending an ACK frame back to the source station. When, after a network error, an ACK is not received, the source station contends again for the channel to transmit the unacknowledged packet and, in case of further error, retries until a maximum retry limit is reached.

2.2. The GSM AMR-WB Speech Coding Standard

The GSM Adaptive Multi-Rate Wideband (AMR-WB) standard [6] is a state-of-the-art ACELP coder with 9 modes operating at bitrates from 6.6 kb/s to 23.85 kb/s. The coder modes are integrated in a common structure, where the bitrate scalability is obtained by adjusting the quantization schemes for the different parameters. The frame size is 20 ms, consisting of 4 sub-frames of 5 ms each.

Before transmission the speech encoder output bits are ordered according to their subjective importance. The reordered bits are further divided into three classes according to their importance: Class A, Class B, and Class C [2]. Class A contains the bits most sensitive to errors and any error in these bits typically results in a corrupted speech frame which should not be decoded without applying appropriate error concealment. Class B and C contain bits where increasing error rates gradually reduce the speech quality, but decoding of an erroneous speech frame is usually possible without annoying artifacts.

3. PARTIAL CHECKSUM PROTECTION FOR WIRELESS REAL-TIME MULTIMEDIA

Modern multimedia applications and coding standards can often deal with corrupted frames better than with lost ones. Error resilience and sophisticated error concealment techniques, in fact, can hide quite well the perceptual effect of a limited number of erroneous bits.

In wireless scenarios multimedia communications heavily suffer from the error prone nature of the radio channel. Bit error rates as bad as 10^{-1} can be experienced depending on the environment, moving speed, and interfering traffic. IEEE 802.11 protocol standard implementations prevent erroneous packets to be forwarded to the next hop or to be received by the application. Checksums on the MAC and on the UDP data units force corrupted packets to be dropped wherever the bit errors occur. Since multimedia data bits have a different error sensitivity and some errors can be tolerated or concealed more than others, a new protocol stack would be advisable to reduce or avoid unnecessary packet discarding when only unimportant bits are affected.

A modified version of UDP, called UDP Lite [4], proposes a partial checksum on the UDP data unit to protect only the UDP header and the sensitive part of the payload. If an error occurs in the checksummed bits, the receiver should drop the packet, otherwise it is passed to the application. The UDP lite header is the same as the UDP header except for the (redundant) UDP length field that has been replaced by the UDP Lite coverage field, where the sending application can specify how many bytes, starting from the first byte of the header, are sensitive to errors. Using this feature the application can then inform the network layers to accept errors in part of the data payload in order to reduce the number of unnecessarily discarded packets at the receiver.

As already proposed in [7] UDP Lite should be used as a transport protocol in error tolerant networks where the link layer does not check the integrity of its data units or where this feature can be disabled for UDP Lite packets. However for wireless networks, with periods of heavy fading and interference, the link layer cannot avoid checksumming its header to prevent misdelivered packets. Furthermore partial checksum policy in the IEEE 802.11 MAC protocol could be preferred because forces hop-by-hop retransmission of heavily corrupted packets and not only detection at the receiver as in the UDP case.

3.1. Link-Layer Partial Checksum

In the following we present a partial checksum implementation for the IEEE 802.11 MAC layer: the reduction of the amount of checksummed bits lessens the number of dropped packets and shortens the average end-to-end delay, and, even if more packets are received partially corrupted, it is showed to effectively increase the perceived quality of voice communications.

In order to deliver uncorrupted data over an error prone transmission channel, IEEE 802.11 standard allows retransmission of erroneous packets. A 32-bit checksum that covers the whole MAC data unit is used to verify the data integrity and a single bit error is sufficient to discard the packet. If an acknowledgment of successful reception is not sent back to the sender, the packet is transmitted again.

Mathematically the Packet Error Rate (PER) is expressed by the equation

$$PER = 1 - (1 - BER)^n \tag{1}$$

where n represents the number of checksummed bits per packet and BER is the Bit Error Probability. If a maximum of $N - 1$ retransmissions are allowed the Packet Loss Rate (PLR), that is the probability of N unsuccessful transmissions, is

$$PLR = PER^N. \tag{2}$$

A partial checksum solution for real-time multimedia data, where only $m < n$ bit of the payload are checked for integrity, reduces the Packet Loss Rate. For example, in a 23.85 kb/s GSM AMR-WB speech communication 72 bits out of 477 should be protected for each speech frame (no channel overhead is considered at the moment).

In a wireless network retransmission occurs for every hop that must be traversed from the source to the destination node. Each retransmission force the wireless node to contend again for the channel introducing variable and longer delay especially during high channel load. Moreover bad channel behavior increases the number of retransmissions and worsens channel congestion. If the Packet Error Rate for a partial checksummed packet is lower, the average number of retransmissions, given by

$$\overline{N} = \sum_{i=1}^{N} i(1 - PER)PER^{i-1} = \frac{1}{1 - PER} \tag{3}$$

decreases too allowing timely delivery, but also reduced traffic.

Furthermore, by comparing the benefits of UDP and MAC partial checksum, we should notice that wired to wireless voice communications are a very attractive scenario for the latter solution.

Figure 1 presents a typical VoIP transmission between a mobile node and a wired terminal. In that scenario the time a packet spend to reach its destination could be reasonably assumed to be higher than 75 ms. Considering that the end-to-end delay tolerated by an interactive speech communication is at most 150 ms, two transmission can hardly be allowed. So, if the packet integrity is only checked at the receiver the possible choices for an erroneous packet are only two: discard the packet or forward it to the application. Instead, if the wireless link layer can verify a partial checksum on the data payload, then forward and backward packets can be quickly retransmitted over the error prone radio link to reduce the PLR without exceeding the allowed total delay.

Performance evaluation of the proposed technique should not only be limited to considering the Packet Loss Rate. Partial checksum, in fact, allows partially corrupted packets and their effect on the decoding process has to be taken into account. Since the checksum covers the most important bits of a speech frame, errors can only occur where they are know to introduce limited distortion. When retransmissions are not allowed this guarantees a better perceived quality compared to the case where that packet is considered lost and concealed. Retransmissions can instead play a fundamental role in favoring total or partial checksum techniques. The latter solution guarantees a lower PLR often accepting lightly corrupted packets, while the former presents more packet losses, but it ensures the integrity of the received ones. The next section will present simulation results used to evaluate the decoded speech quality.

4. SIMULATIONS

Speech transmission with partial checksum over a wireless channel has been simulated using the Wideband GSM Adaptive Multi-Rate coder. Different BER probabilities and allowed number of retransmissions are considered, objective speech quality measures and *informal listening tests* are then employed to compare the perceived speech quality of the decoded streams.

At the application level speech is encoded by the 23.85 kb/s GSM AMR-WB coder in frames of 477 bits: the first 72 bits belong to Class A and should be protected by the checksum. The checksum should also be computed over the protocol headers used to send the multimedia data over the network. For this purpose we employ the Real-time Transport Protocol (RTP) as defined in the recent RFC 3267 [8] that specifies the payload format to be used for AMR encoded speech signals. Ten control bits are present at the beginning of the payload and additional bits are added to the end as padding to make the payload byte aligned for a total of 61 bytes. For the RTP and the additional UDP and IP headers, a compression scheme has been assumed that allows the 40-byte header to be compressed in 2 bytes as defined for Robust Header Compression [9]. Finally each data-type MPDU (MAC Protocol Data Unit) has a 24-byte header plus a 4 byte checksum.

To support partial checksum at the MAC level an additional field should be introduced to specify the number of

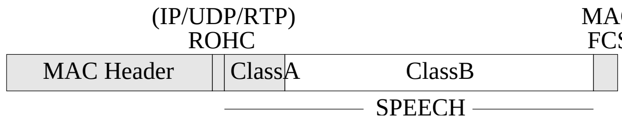


Fig. 2. Unequal Error Detection coverage (gray area).

covered bit, starting from the beginning of the MAC data unit. Since the maximum allowed size is 2346 bytes a two byte field is required. The total packet length at the MAC layer is then 744 bits, but only 338 will be protected by the partial checksum, as in Figure 2.

The time-varying error characteristic of the wireless channel has been simulated with a two state Gilbert-Elliot model where each state represents a Binary Symmetric Channel [10][11]. Each state is assigned a specific constant BER: in the "good" state (G) errors occur with low probability $p_G = 10^{-5}$, while in the "bad" state (B) they happen with high probability $p_B = 10^{-3}$. Within one state errors are assumed to occur independently from each other. Varying the probability to switch from the good state to the bad state (p_{GB}) and vice versa (p_{BG}) different average bit error rates can be achieved as given by the following equation

$$\overline{BER} = \pi_G p_G + \pi_B p_B \tag{4}$$

where π_G and π_B are defined as

$$\pi_G = \frac{p_{BG}}{p_{BG} + p_{GB}}; \pi_B = \frac{p_{GB}}{p_{BG} + p_{GB}}. \tag{5}$$

In the following simulations p_{BG} is kept constant to $0.\overline{6}$ to have the same average burst error length $(1/p_{BG})$.

For simplicity we assume that state transitions occur only at multiples of 20 ms and that a packet is entirely sent in one of the two states (no state transitions occur in the middle of the packet). In addition all the packet transmissions are supposed to happen in the same channel state. The packet error rate can then be expressed as

$$PLR = \pi_G PLR_G + \pi_B PLR_B, \tag{6}$$

where PLR_G and PLR_B are defined as

$$PLR_G = [1 - (1 - p_G)^n]^N; PLR_B = [1 - (1 - p_B)^n]^N, \tag{7}$$

while n is the number of bit covered by the checksum and N is maximum number of allowed transmissions.

At the receiver the speech frames are decoded and the new ITU perceptual measurement algorithm, the Perceptual Evaluation of Speech Quality (PESQ) [12], is used to measure their perceived speech quality [13]. The PESQ compares the degraded speech with the reference speech (in uncompressed PCM format) and computes an objective Mean Opinion Score (MOS) value in a 5-point scale. In Figure 3

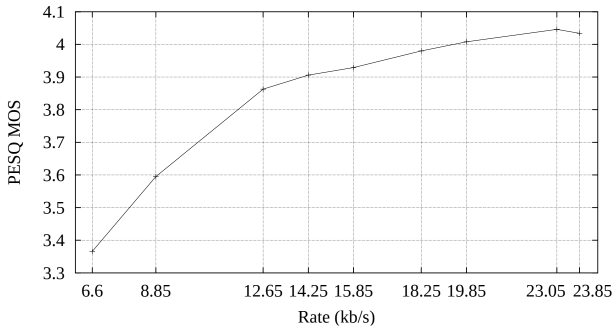
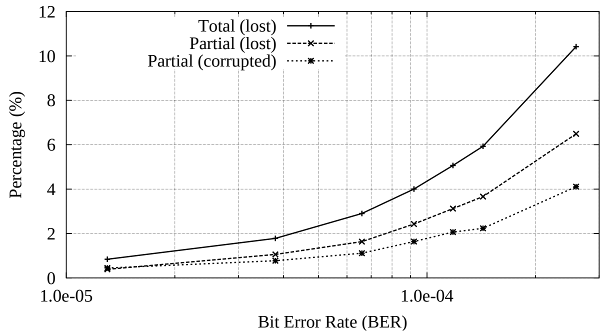
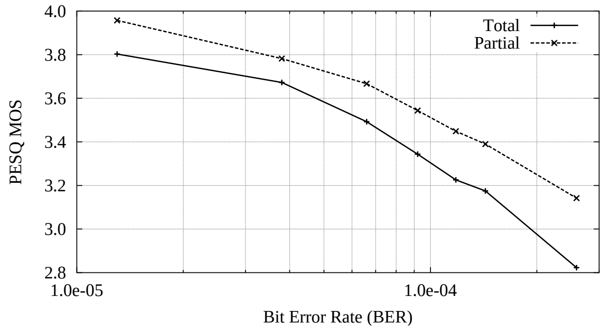


Fig. 3. GSM AMR-WB performance on the test sequence using ITU-T P.862 (PESQ) algorithm at different coding rates.



(a) Packet Loss/Corrupt Rate



(b) PESQ MOS

Fig. 4. Performance in terms of (a) packet loss rate and (b) PESQ score of IEEE 802.11 GSM AMR-WB speech transmission at 23.85 kb/s, with total and partial checksum, for different Bit Error Rates, without retransmissions.

we can initially evaluate the quality of the GSM AMR-WB coder at different rates on the speech stream used in the simulations: 24 different clear sentences spoken by a male and a female speaker for a total of 4800 frames and 96 seconds.

4.1. Results

A single transmission is simulated over a wireless channel at different error rates allowing zero or three retransmissions. Total and partial checksum performance is evaluated for the GSM AMR-WB coding mode at 23.85 kb/s. Corrupted packets are decoded as is without further processing by the AMR decoder, lost frames are concealed as defined in the standard [14]. Perceived quality results are expressed my means of the objective quality measure given by the PESQ MOS.

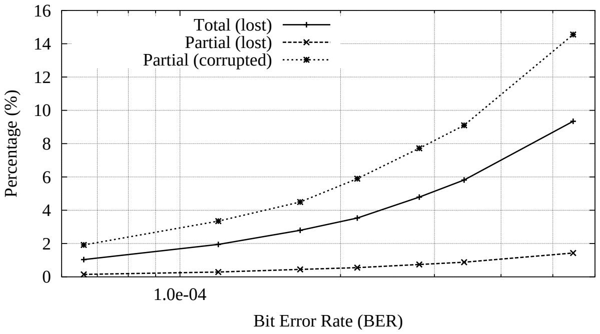
Firstly we tested the proposed solution without any retransmission: erroneous packets are dropped by the standard IEEE 802.11 checksum technique, while partial checksum forwards the packets only if errors occur in the perceptually least important part of the speech payload (406 bits out of 744). Figure 4 presents the simulation results in terms of lost and corrupted packets, and objective speech quality at different bit error rates. With partial checksum the percentage of received frames is clearly higher. The PESQ score of the corresponding decoded speech also confirms that, for the scenarios under consideration, it is better to receive and decode partially corrupted packets than to lose them altogether.

The second simulation scenario allows a maximum of three retransmissions: the two techniques under analysis present different loss behavior, but also different rates given a particular BER. Figure 5 shows that partial checksum on speech data not only guarantees higher perceived quality for the received speech, but also that it is an effective solution to reduce network load and delay in presence of high bit error rates.

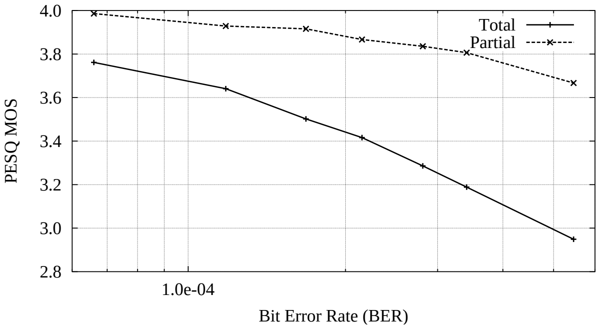
When the number of allowed retransmissions increases, the performance of the standard (i.e., full checksum) approach improves, reaching and eventually exceeding the proposed technique. Such gains, however, come at the cost of increased congestion, larger delays and greater amount of traffic transmitted.

5. CONCLUSIONS

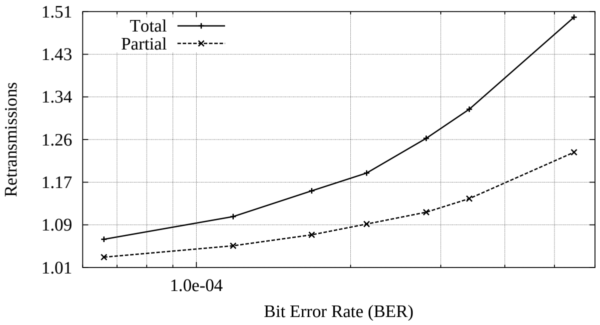
A new technique to enhance the IEEE 802.11 link-layer effectiveness for interactive speech communications has been presented. Since speech decoders can often deal with corrupted frames better than with lost ones, the standard 802.11 MAC-level error detection on the whole packet has been limited to the most perceptually sensitive bits of a speech frame. Retransmissions are then not required when bit errors occur outside the checksum coverage, on the assumption that they will result in only minor distortion. Simulations performed using the GSM AMR-WB speech coding standard show a clear improvement in speech quality: the negative effect of errors in the less perceptually important bits is, in fact, counterbalanced by the lower number of dis-



(a) Packet Loss/Corrupt Rate



(b) PESQ MOS



(c) Average number of retransmissions

Fig. 5. Performance in terms of (a) packet loss rate, (b) PESQ score and (c) average number of retransmissions of IEEE 802.11 GSM AMR-WB speech transmission at 23.85 kb/s, with total and partial checksum, for different Bit Error Rates, with maximum three retransmissions.

carded speech packets. The proposed unequal error detection technique also reduces end-to-end delays and network load, since successful packet delivery requires, on average, less retransmissions.

6. REFERENCES

[1] ITU-T G.114, "One-way transmission time," *ITU-T*, May 2000.

[2] "AMR speech codec, wideband; frame structure," *ETSI TS 126 201*, August 2002.

[3] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *ISO/IEC 8802-11, ANSI/IEEE Std 802.11*, 1999.

[4] L-A. Larzon, M. Degermark, S. Pink, L-E. Jonsson, and G. Fairhurst, "The UDP-lite protocol," *draft-ietf-tsvwg-udp-lite-02.txt*, August 2003.

[5] H. Dong, I.D. Chakares, C.H. Lin, A. Gersho, E. Belding-Royer, U. Madhow, and J.D. Gibson, "Speech coding for mobile ad hoc networks," in *Proc. Asilomar Conference on Signals, Systems and Computers*, November 2003.

[6] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 8, pp. 620–636, November 2002.

[7] L-A. Larzon, M. Degermark, and S. Pink, "UDP lite for real-time multimedia applications," in *Proc. IEEE Intl. Conference of Communications (ICC)*, April 1999.

[8] J. Sjoberg, M. Westerlund, A. Lakaniemi, and Q. Xie, "Real-time transport protocol (RTP) payload format and file storage format for the adaptive multi-rate (AMR) and adaptive multi-rate wideband (AMR-WB) audio codecs," *RFC 3267*, June 2002.

[9] G. Pelletier, "Robust header compression (ROHC): Profiles for UDP-lite," *draft-ietf-rohc-udp-lite-02.txt*, December 2003.

[10] E.O. Elliot, "Estimates on error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, September 1963.

[11] M. Zorzi, R.R. Rao, and L.B. Milstein, "Error statistics in data transmission over fading channels," *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1468–1477, November 1998.

[12] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," .

[13] "Performance characterization of the adaptive multi-rate wideband (AMR-WB) speech codec," *ETSI TR 126.976*, September 2003.

[14] "AMR speech codec, wideband; error concealment of lost frames," *ETSI TS 126 201*, March 2002.