ADAPTIVE INTERACTIVE SPEECH TRANSMISSION OVER 802.11 WIRELESS LANS

Antonio Servetti[1], *Juan Carlos De Martin*[2]

[1]Dipartimento di Automatica e Informatica / [2]IEIIT-CNR
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: [servetti|demartin]@polito.it

ABSTRACT

We present an adaptive technique to transmit speech over 802.11 wireless packet networks. According to the proposed scheme, the speech coding rate of a network-driven variable bitrate coder is selected to match the istantaneous wireless channel conditions: higher rates (i.e., larger packets) for low error rates, lower rates (i.e., smaller packets) when the channel is noisy. Packet size is, in fact, directly related to the probability of retransmission, one of the major sources of delay in contention-based medium access control. Network simulation results show that the adaptive approach can address the stringent quality of service requirements for two way interactive speech applications over wireless packet networks, reducing packet loss rates and end-to-end delays.

1. INTRODUCTION

Wireless technology keeps changing the communications scenario. Wireless local area networks (WLANs), in particular, are being enthusiastically adopted by users worldwide, shaping a new world where tetherless access will be possible not only in homes and offices, but also in an increasing number of previously unconnected places, like shopping malls, libraries, trains and other means of mass transportation, even private motor vehicles. As soon as seamless integration with wide-area coverage provided by 2.5G/3G cellular wireless infrastructures is reached, wireless access will likely become the most common form of network access for an increasing number of users.

The IEEE 802.11 WLAN standard, based on the definition of the medium access control (MAC) protocol and the physical layer (PHY) specifications, became available in 1999 [1] and since then has emerged as the most successful and most widely deployed WLAN standard. Figure 1 shows a simple 802.11-based network scenario, with two mobile stations and an access point (AP) connected to a wired LAN.

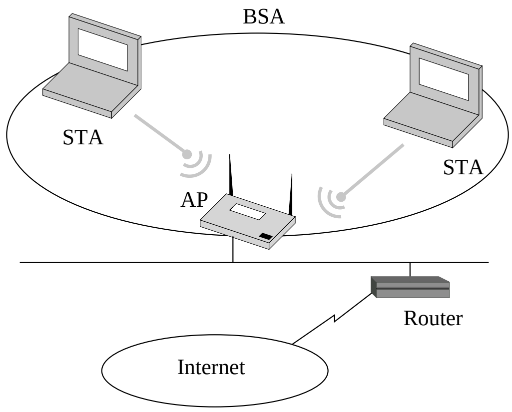

Fig. 1. 802.11-based network communications scenario.

So far, the main usage of Wireless LANs has been limited to Internet based services like Web browsing, e-mail, and file transfers. However, as already happened in the traditional wired LANs, a strong interest is quickly emerging towards multimedia applications over WLANs, and interactive voice communications are appearing more and more as the natural evolution of cordless telephony. Not only such technology would have all the advantages of IP-based communications, including a single infrastructure for both data and voice, but it would also deliver significant cost savings, and possibly better voice quality, with respect to cellular telephony.

Several challenges, however, need to be addressed to make WLAN telephony as successful as cellular and wired telephony. Not only the available bandwidth for WLANs is significantly below that of their wired counterparts, but wireless links are also strongly time-varying and may have high error rates. Other issues are specific of

This work was supported in part by CERCOM, the Center for Wireless Multimedia Communications, Torino, Italy, http://www.cercom.polito.it.

802.11 WLANs, including the MAC layer effects on performance, the consequences of interfering data traffic, and the best configurations for both Access-Point-based and ad-hoc 802.11 networks.

Previous research evaluated the performance of interactive voice traffic over Wireless LANs [2][3], mainly by means of statistical analysis of throughput and packet losses to assess the number of supported voice conversations. Advanced Quality of Service techniques are also yet to be fully investigated.

In this paper we present a new technique for improving the quality of interactive voice communications over 802.11 wireless packet networks. The operating rate of a network-driven variable-bitrate speech coder is chosen on a frame-by-frame basis according to the istantaneous channel conditions: higher rates (i.e., longer packets) when the channel is good, lower rates (i.e., shorter packets) when the channel is poor. Performance is measured in terms of average packet losses and average delay, with and without interfering traffic, using a network simulator. The proposed system consistently outperforms constant-bitrate speech transmission at the same average bitrate.

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

2. VOICE OVER 802.11 WLANS

Voice over IP over wireless packet networks is becoming increasingly attractive. In particular, the widespread adoption of WLAN technology is creating the basis for the introduction of cordless packet telephony in offices, homes, hospitals, etc.

Two-way conversational applications, however, are characterized by stringent requirements on the end-to-end delay. The upper limit for one-way delay is set to only 150 ms, according to the guidelines of ITU-T Recommendation G.114. Moreover, packet losses should be kept below 1% to prevent significant perceptual degradation.

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

Efficient WLAN-based cordless telephony system must thus rely on careful design of advanced speech transmission solutions.

2.1. IEEE 802.11 Wireless LANs

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. 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. All fields except the frame body (28 bytes in total) contribute to the MAC protocol data unit (MPDU) overhead for a data frame. Upon packet transmission the destination station positively acknowledges each successfully received packet 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 hannel to transmit the unacknowledged packet and, in case of further error, retries until a maximum retry limit is reached.

2.2. The GSM AMR Speech Coding Standard

The GSM Adaptive Multi-Rate (AMR) standard [4] is a state-of-the-art network-driven variable-bitrate speech coder. Its operating bitrate can be chosen on a frame-by-frame basis to match the istantaneous channel conditions. In the case of cellular telephony, the objective is to change the ratio between bandwidth devoted to speech and bandwidth devoted to forward error correction. For the proposed technique, the objective is to use the speech rate, i.e. the speech packet size, most suitable for any given 802.11 channel condition.

The GSM-AMR speech coder is a multi-rate ACELP coder with 8 modes operating at bit-rates from 12.2 kbps to 4.75 kbps. The coder modes are integrated in a common structure, where the bit-rate scalability is obtained by adjusting the quantization schemes for the different parameters. The frame size is 20 ms, consisting of 4 subframes of 5 ms each.

3. ADAPTIVE SPEECH TRANSMISSION

As the purpose of AMR codec mode adaptation is to select the codec mode which provides the user with maximum speech quality at a given channel condition, link quality measures are needed.

The GSM AMR standard leaves link quality estimation open. However, it provides an example solution, which is based on burst-wise C/I estimates [5]. A carrier signal estimate is calculated by convoluting a training sequence with the channel impulse response. Other methods have also been proposed [6][7], i.e. based on the acknowledgment history of the most recently transmitted packets.

For codec mode adaptation, the measure of the istantaneous channel quality has to be mapped to codec modes. This is in principle done by quantizing the measurement where the levels of the quantizer used represent the different codec modes.

3.1. Adaptive Rate Selection Algorithm

We propose to rely on a channel estimation algorithm to select the optimal output rate of the speech coder for an interactive speech transmission. Channel quality measurements are roughly quantized in two states that represent good and bad channel conditions. In the bad state, large packets have a higher probability to be in error and, therefore, to be retransmitted, while small packets are more easily received without errors.

When a variable-rate speech codec is available and the wireless link shows a relatively high error-rate, the size of the compressed speech frames can be reduced using a lower rate. The proposed solution should reduce the number of transmissions needed to successfully send a packet when the channel becomes noisy. The end-to-end delays will also be reduced because the sender will contend for the channel less frequently.

4. SIMULATIONS AND RESULTS

Several network conditions, for various kinds of interfering traffic, were simulated using the NS [8] network simulator, modified to include a channel error model.

At the application level a different codec mode is assigned to each of the two possible states of the wireless channel. In the good state, a 12.2 kbps source is used, while in the bad state the voice encoder produces an output rate of only 4.75 kbps. Sensing of the channel is performed before coding a speech frame in order to adapt the source output to the time-varying channel conditions. Sensing was assumed to be ideal. The payload is then encapsulated by the RTP protocol, UDP is used for multiplexing different flows, and IP takes care of addressing and delivering the packets to their destination. We use a 4-byte compressed header to overcome the RTP/UDP/IP overhead (40 bytes per packet). The MAC and PHY layer headers are then added according to the 802.11 standard.

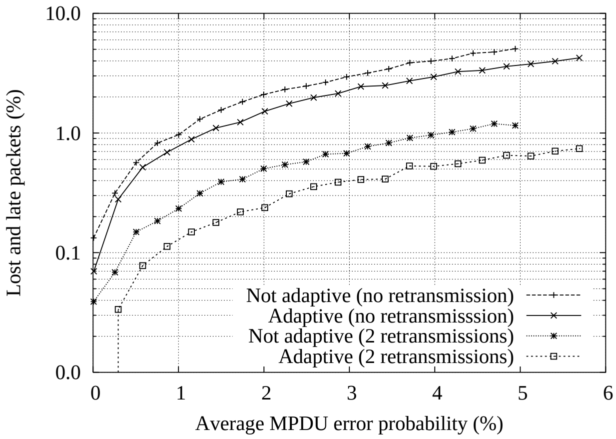

Fig. 2. Lost and late speech packets as a function of the average MPDU error probability; adaptive vs. not-adaptive technique, maximum number of retransmissions set to zero and two.

4.1. Wireless Channel Error Model

The time varying wireless channel is modeled using the Gilbert-Elliot two-state Markov model where each state represents a binary symmetric channel (BSC). In the "good" state (G) losses occur with low probability p_G while in the "bad" state (B) they happen with high probability p_B. p_{GB} and p_{BG} represent the probability to switch from the good state to the bad state and vice versa. The steady state probabilities of being in states G and B are:

$$\pi_G = \frac{p_{BG}}{p_{BG} + p_{GB}}, \pi_B = \frac{p_{GB}}{p_{BG} + p_{GB}} \tag{1}$$

respectively. Hence the average packet loss rate produced by the Gilbert channel is $p = p_G \pi_G + p_B \pi_B$.

We assume that state transitions occur between packets transmissions. In the good state the bit error probability is $p_G = 10^{-6}$, while in the bad state it is $p_B = 10^{-3}$. Different channel error conditions are obtained varying the p_{GB} and p_{BG} transition probabilities.

The packet error probability is a function of the packet size: given an S-byte packet and a bit error probability of p_x, that packet will be considered corrupted and therefore discarded with probability $1 - (1 - p_x)^{8S}$.

4.2. Results

Simulations were performed for an 11-Mbps wireless LAN scenario where two mobile terminals are placed at the same distance from an access point and they are sending packets to a host connected to the wired infrastructure network. Because the wireless path represents only the first transmission hop, we consider 20 ms the maximum acceptable value for the one-way transfer delay over the wireless link. The

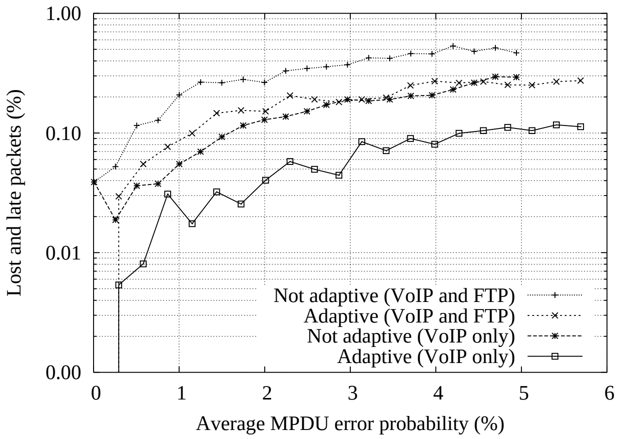

Fig. 3. Lost and late speech packets as a function of the average MPDU error probability; adaptive vs. not-adaptive technique, with and without interfering FTP traffic.

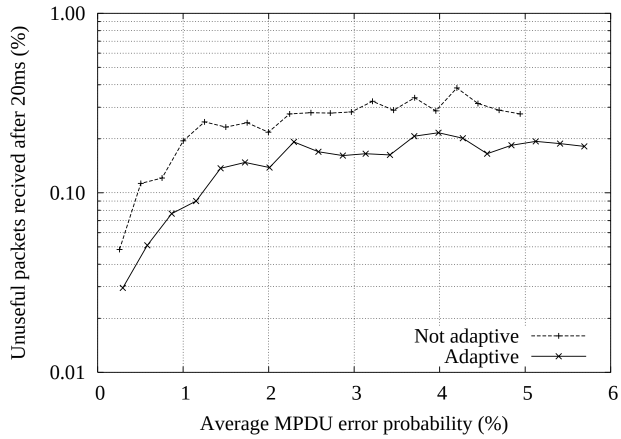

Fig. 4. Packet discarded at the receiver because of their late arrival (delay larger than 20 ms) as a function of the average MPDU error probability. Adaptive vs. not-adaptive technique.

percentage of voice packets lost or received with a delay greater than 20 ms is monitored on the access point.

First we tested the proposed adaptive solution against plain transmission of a single VoIP source without interfering traffic. A wireless node sends a speech frame every 20 ms adapting the payload size (224 or 95 bits) to the channel state. Figure 2 compares adaptive and fixed-rate transmission at the same average bit-rate for the cases of maximum numer of retransmission zero and two. With an estimated packet loss rate at the MAC level of 2% and two retransmissions, the adaptive solution halves the number of lost and late packets.

The adaptation of the transmission rate to channel conditions is effective also when interfering sources are active. In Figure 3 the case of a concurrent FTP source is depicted. The adaptive technique performs almost as well as the non-adaptive solution without interfering traffic.

Regarding end-to-end delay, the proposed adaptive transmission scheme reduces the average delay because less retransmissions are needed. Figure 4 shows the percentage of packets discarded at the receiver due to their late arrival. Adapting the speech frame dimension to the channel conditions, packets tend to arrive on time for successful playback, leading to higher perceptual quality.

5. CONCLUSIONS

An adaptive technique to transmit speech over 802.11 wireless packet networks was presented. According to the proposed scheme, the speech coding rate of a network-driven variable bitrate coder, the GSM AMR, is selected to match the istantaneous wireless channel conditions: higher rates for low error rates, lower rates when the channel is noisy. Network simulation showed that adaptively selecting the speech packet size consistently outperforms the constant bitrate approach in terms of packet loss rates and end-to-end delays.

6. REFERENCES

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

[2] B.P. Crow, I. Widjaja, J.G. Kim, and P.T. Sakai, "IEEE 802.11 Wireless Local Area Netoworks," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 116–126, Sep 1997.

[3] M. Veeraraghavan, N. Cocker, and T. Moors, "Support of voice services in IEEE 802.11 wireless LANs," in *Proceedings of INFOCOM*, April 2001, pp. 488–496.

[4] "GSM 06.90 Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR) speech transcoding," *Draft ETSI EN 301 704*, 1999.

[5] 3GPP TS 05.09, "Digital cellular telecommunications system (Phase 2+); Link Adaptation," *ETSI TS 101 709*, version 8.5.0 Release 1999.

[6] S. Bruhn, P. Blocher, K. Hellwig, and J. Sjoberg, "Concepts and solutions for link adaptation and inband signaling for the GSM AMR speech coding standard," in *Proc. IEEE 49th Vehicular Technology Conference*, July 1999, vol. 3, pp. 2451–2455.

[7] E. Modiano, "An adaptive algorithm for optimizing the packet size used in wireless ARQ protocols," *Wireless Networks*, vol. 5, no. 4, pp. 279–286, July 1999.

[8] UCB/LBNL/VINT, "Network Simulator – ns – version 2," *URL: http://www.isi.edu/nsnam/ns*, 1997.