VOICE TRANSMISSION OVER 802.11 WIRELESS NETWORKS USING ANALYSIS-BY-SYNTHESIS PACKET CLASSIFICATION

Matteo Petracca, 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: matteop3@inwind.it, [servetti|demartin]@polito.it

ABSTRACT

A new form of telephony is being made possible by the nearly-ubiquitous presence of 802.11 wireless local networks (WLAN's). This paper presents a technique to improve speech quality for WLAN telephony. Speech compressed by the 3GPP GSM AMR coding standard at 12.2 kb/s is packetized and classified according to an analysis-by-synthesis estimate of the perceptual importance of each individual packet. Perceptually important packets are protected against noise and channel collisions by a simple form of forward error correction, i.e., packet repetition. Network simulations and perceptual quality measures have been used to evaluate the performance of the proposed technique. Preliminary results show that protecting the perceptually most important 10% of all speech packets provides the same performance delivered by randomly protecting twice as many packets.

1. INTRODUCTION

Wireless local area networking (WLAN) technology is being enthusiastically adopted by users worldwide. An increasing number of places has now wireless infrastructures and several devices (laptops, cameras, phones, etc.) include wireless local interfaces. As soon as integration with wide-area networks is reached, wireless access will likely become the most common form of network access for an increasing number of users. This emerging scenario will soon be exploited by a new set of Internet applications specifically designed for the needs of mobile users and among them WLAN-based telephony looks particularly appealing.

However, several challenges need to be addressed to provide successful interactive multimedia applications over a network originally designed for generic data traffic and with high error rates. Multimedia communications, in fact, have strictly bounded quality of service requirements in terms of packet losses, end-to-end delays and jitter. The Differentiated Services (DiffServ) architecture [1] is one of the most promising proposals that have been made to introduce Quality-of-Service guarantees in IP networks. In wireless environments, where bandwidth is scarce and channel conditions are variable, differentiated services may be supported by modifications to lower transmission layers [2].

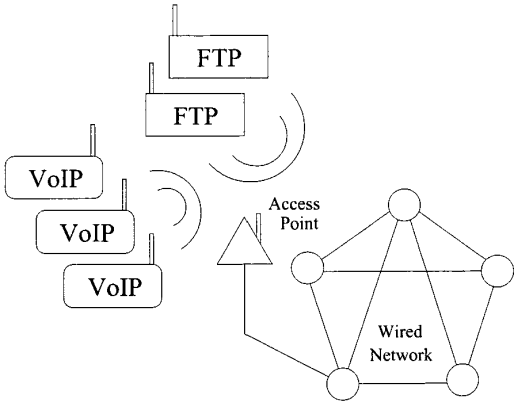

Fig. 1. IEEE 802.11-based telephony network scenario.

According to the DiffServ model, packets are classified and marked to receive a specific forwarding behavior on nodes along their path. In the simple case of just two classes, delay- and losses-sensitive data, such as interactive speech, may be transmitted as *premium* bandwidth (almost no losses and low delay); less critical data as *regular* best-effort. If speech traffic, however, was marked and transmitted as *premium* in its entirety —as it is currently the case— the growth of this kind of service over data networks would soon threaten to saturate the available bandwidth. Techniques to reduce the load on the premium bandwidth have, therefore, been recently proposed; they achieve the desired levels of perceptual end-user quality and usage of network resources by marking as premium only the most perceptu-

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

0-7803-8379-6/04/$20.00 ©2004 IEEE.

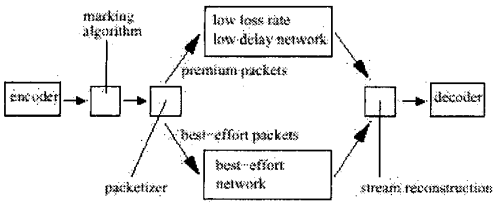

Fig. 2. Block diagram of a 2-class DiffServ network architecture.

ally relevant packets of each multimedia flow [3][4].

We propose an analysis-by-synthesis (AbS) approach to the marking of packets containing compressed speech. Speech packets are marked depending on the estimated distortion that their loss would introduce at the decoder and the desired level of perceptual quality. High-distortion packets, i.e., perceptually important data, are sent as *premium*, while the remaining part of the stream is sent as *regular* traffic. Performance in wired and wireless environments has been evaluated through network simulations. A statistical model of the decoder state is also introduced.

The paper is organized as follows. In Section 2, we introduce DiffServ IP networks. In Section 3 the proposed technique is described. Simulations and performance results are presented in Section 4. Conclusions follow in Section 5.

2. DIFFSERV IP NETWORKS

A DiffServ network is an IP network in which packets receive a different forwarding behavior on nodes along their path depending on the service class they belong to. This architecture [5] defines different classes each one suitable for specific purposes. One of the simplest cases uses two QoS levels: a *premium* service and a *regular* best-effort service. The first class is meant to transmit delay- and losses-sensitive traffic, such as interactive speech and video, while the second class is for less demanding types of traffic (neither a maximum amount of delay nor the delivery are guaranteed).

Figure 2 shows a DiffServ network architecture. The marking algorithm performs packet classification, packets experience different network behavior depending on their marking, then at the receiver the stream is reconstructed, and —in the case of speech— packet losses are concealed, and speech data decoded.

Algorithms used to manage DiffServ packet forwarding heavily influence the characteristics of QoS classes because packets are treated differently depending on the implementation and the network type. In wired networks, router queue management is the key for assuring differentiated services: packets are placed in different queues depending on the class they belong to. Implementations must attempt to minimize congestion within each class using active queue management algorithms. When packets have to be dropped the DiffServ router selects more frequently the packets belonging to lower QoS classes. In wireless network environments, IP differentiated services require the link layer to compensate for the error-prone time-varying nature of the wireless channel and for the contention-based channel access control mechanism [2][6]. Prioritization can be achieved assuring faster access to the channel (reduced DIFS) for the *premium* service and lower packet loss rate can be guaranteed allowing an higher number of retransmissions in case of successive erroneous packets.

3. ANALYSIS-BY-SYNTHESIS PACKET CLASSIFICATION

Packet classification for speech transmission over DiffServ networks is usually accomplished by marking as premium the entire flow. When no more bandwidth is available, service is denied or degrades without control. Instead of assigning all packets of a given speech flow either to the *premium* class or to the best-effort class, packet classification and marking can be performed on a packet-by-packet basis.

Specifically, *each speech packet can be analyzed and assigned to one class or the other depending on its perceptual importance. To do so, the packet marker needs to analyze the payload and estimate the perceptual impact of the packet at the decoder.* From a complexity point of view, analysis-by-synthesis packet classification is best done at the speech encoder. In that case, in fact, packet classification can be easily generated as a by-product of the encoding operation at little or no extra cost in terms of computation.

3.1. Distortion-Based Analysis

The perceptual importance of a packet can be expressed in terms of the distortion that would be introduced by its loss.

The optimal measure of distortion would be to compare speech decoded using the correct parameters and speech decoded using the parameters estimated by the frame erasure concealment technique. The packet marker needs to:

1. decode the speech parameters, **P**;
2. replicate the behavior of the decoder in presence of a frame erasure and generate estimates of the erased parameters, **P'**;
3. compute distortion measures, **D**, between original parameters and corresponding estimates.

The marking decision depends on this set of parametric distortions. Figure 3 shows the block diagram of the proposed analysis-by-synthesis packet classification.

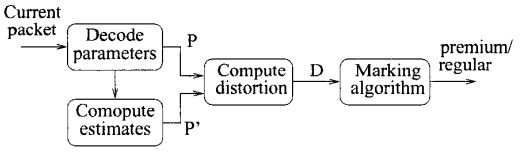

Fig. 3. Block diagram of analysis-by-synthesis packet classification.

3.2. Decoder State

Regarding step 2, the generation of the estimates, assumptions about the current state of the decoder need to be made. Implementation by De Martin [3] was limited to considering the previous frame as correctly received. Here a more complex model is presented that takes into account the probability that the most recent frame in the past has been lost. In this case the computation of the distortion generates two possible values, one for each possible state of the model: $\mathbf{D_{lost}}$ (previous frame lost) and $\mathbf{D_{recv}}$ (previous frame received). The decision will then be made on the estimated distortion:

$$\overline{\mathbf{D}} = p \cdot \mathbf{D}_{lost} + (1-p) \cdot \mathbf{D}_{recv}, \qquad (1)$$

where p is the frame loss probability.

4. PERFORMANCE ANALYSIS WITH GSM ADAPTIVE MULTI-RATE SPEECH

We tested the analysis-by-synthesis packet classification algorithm using speech compressed with the ETSI/3GPP GSM AMR coder at 12.2 kb/s. Simulations of voice communications over wired and wireless differentiated service networks have been performed by means of the NS-2 Network Simulator [7]. Received speech perceived quality has then been evaluated with the ITU-T PESQ (Perceptual Evaluation of Speech Quality) algorithm [8] over a 5-point MOS (Mean Opinion Score) scale.

4.1. Analysis-by-Synthesis Distortion Evaluation of GSM-AMR Speech Frames

The GSM Adaptive Multi-Rate standard [9] is a state-of-the-art ACELP coder with 8 modes operating at bitrates from 12.2 to 4.75 kb/s.

The Linear Prediction (LP) coefficients of the input signal are first analyzed and then quantized to be used in an LP synthesis filter driven by the output of the excitation codebooks. Encoding is performed in two steps. Long-term prediction coefficients are calculated in the first step; in the second step, a perceptually weighted error between the input signal and the output of the LP synthesis filter is minimized. This minimization is achieved by searching for an appropriate codevector for the excitation codebooks. Quantized LP coefficients, as well as gains and indexes to the codevectors of the excitation codebooks and the long-term prediction coefficients, form the bit-stream.

GSM AMR standard error concealment procedure [10] is meant to mitigate the effect of lost speech frames and it is based on a state machine with seven states. In order to improve the subjective quality, erroneous/lost frames are substituted with either a repetition or an extrapolation of the previous good speech frame(s). This substitution is done so that it will gradually decrease the output level, resulting in silence at the output after more than six consecutive lost frames.

Analysis-by-synthesis distortion is computed for each set of parameter between the original and the extrapolated version:

– *spectral distortion* in dB for the LP coefficients,

– *percentage difference* for the long-term prediction coefficients,

– *difference* in dB for the codebook gains.

The set of distortion values is then passed to the marking algorithm: if the distortion for any of the three sets is above a given threshold (depending on the desired percentage of *premium* frames), the packet is classified as *premium*.

4.2. Network Simulations

Speech transmission with analysis-by-synthesis packet classification has been simulated for wired and wireless networks. In the proposed scenario, three voice streams contend for the available bandwidth with two interfering FTP connections. Every GSM AMR 20-ms speech frame is sent in a different packet and it is assigned to one of the two DiffServ classes (*premium* or *regular*).

In the *wired network* scenario, routers employ the RED In Out (RIO) queue management algorithm. All flows must pass through the same bottleneck and, in case of congestion, routers drop *regular* best-effort packets to preserve the guarantees of the *premium* class.

In the *wireless network* scenario, packets are sent on a noisy channel whose error characteristics are simulated using a Gilbert-Elliot two-state error model. Corrupted packets are discarded and retransmitted for a limited number of times. To satisfy the low loss rate guarantee on the *premium* class, a simple forward error-correction (FEC) technique has been applied, i.e., packet repetition. More complex techniques, such as using a different retransmission limit for each class, will be evaluated in future works.

4.3. Perceived Quality Results

The results of the analysis-by-synthesis packet classification algorithm have been compared to the results of a reference random classification algorithm, which has been pro-

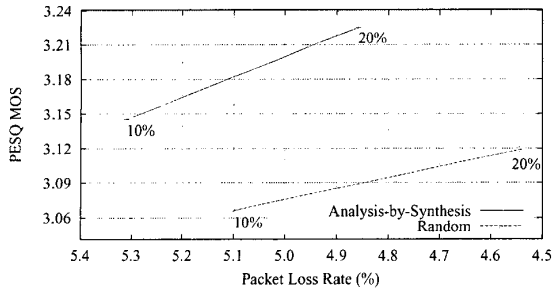

Fig. 4. PESQ-based quality evaluation of AbS packet classification simulated over a wired network, with interfering FTP traffic, as a function of decreasing packet loss rates.

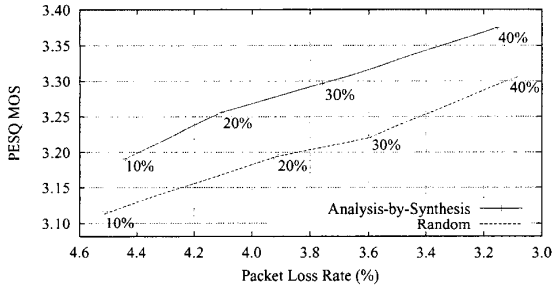

Fig. 5. PESQ-based quality evaluation of AbS packet classification simulated over a wireless network for several percentages of *premium* packets, with interfering FTP traffic, as a function of decreasing packet loss rates.

posed to achieve statistical QoS guarantees for generic data traffic. Simulated loss patterns are applied to speech material taken from the NTT Multi-lingual Speech Database, and lost frames are concealed as defined in the GSM AMR standard. Perceived quality results are expressed by means of the objective quality measure given by the PESQ MOS. Different classification threshold have been used, leading to different percentages of *premium* packets in the speech flow.

In the simulated wired scenario, router queues do not drop any *premium* packet. FTP data is sent as *regular* traffic as well as the perceptively less important speech packets. FTP connections adapt their throughput to match the 'estimated' channel bandwidth: it decreases when packets are lost and it increases again if packets are correctly received. Figure 4 shows that analysis-by-synthesis packet classification provides speech of better quality for all considered percentages of *premium* packets.

In the wireless scenario, there is no guarantee that *premium* packets are all received, but —due to the packet repetition scheme— they suffer from markedly lower error rates than the *regular* traffic. As illustrated in Figure 5, the more the percentage of marked packets increases, the more the packet loss rate decreases. The plot highlights that protecting the perceptually most important 10% of all speech packets provides the same performance of randomly protecting twice as many packets.

5. CONCLUSIONS

A technique to improve speech quality for 802.11-based telephony was presented. Speech compressed by the 3GPP GSM AMR coding standard at 12.2 kb/s is packetized and classified according to an analysis-by-synthesis estimate of the perceptual importance of each individual packet. Perceptually important packets are protected against noise and channel collisions by a simple form of forward error correction, i.e., packet repetition. Network simulations and perceptual quality measures have been used to evaluate the performance of the proposed technique. Preliminary results show that protecting the perceptually most important 10% of all speech packets provides the same performance delivered by randomly protecting twice as many packets.

6. REFERENCES

[1] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," *RFC 2475*, December 1998.

[2] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *Proc. of IEEE INFOCOM*, Anchorage, Alaska, April 2001, vol. 1, pp. 209–218.

[3] J.C. De Martin, "Source-driven packet marking for speech transmission over differentiated-services networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Salt Lake City, Utah, May 2001, vol. 2, pp. 753–756.

[4] C. Hoene, B. Rathke, and A. Wolisz, "On the importance of a VoIP packet," in *Proc. of ISCA Tutorial and Research Workshop on the Auditory Quality of Systems*, Mont-Cenis, Germany, April 2003.

[5] B.E. Carpenter and K. Nichols, "Differentiated services in the internet," *Proceedings of the IEEE*, vol. 90, no. 9, pp. 1479–1494, September 2002.

[6] H. Sanneck, N.T.L. Le, M. Haardt, and W. Mohr, "Selective packet prioritization for wireless voice over IP," in *Proc. Fourth International Symposium on Wireless Personal Multimedia Communication*, Aalborg, Denmark, September 2001.

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

[8] 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," February 2001.

[9] ETSI, "AMR speech codec; general description," *TS 126 071 v5.0.0*, June 2002.

[10] ETSI, "AMR speech codec; error concealment of lost frames," *TS 126 091 v5.0.0*, June 2002.