OPTIMIZED UNEQUAL ERROR PROTECTION FOR VOICE OVER IP

Mingyu Chen and Manohar N. Murthi

Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33124-0640 USA
mchen@umsis.miami.edu, mmurthi@miami.edu

ABSTRACT

In Voice over IP, typical Forward Error Correction (FEC) schemes to combat packet loss allocate an equal amount of error-control resources to each voice packet, regardless of the perceptual importance of a packet. Recognizing the unequal perceptual importance of voice packets, we propose signal-adaptive unequal error protection methods in which certain packets are allocated more error-control resources than others. In particular, the amount of error protection provided to a packet is determined through an analysis by expected decoder synthesis paradigm ensconced within a rate-distortion Lagrangian optimization framework. Therefore, the sender evaluates various protection policies by anticipating the behavior of the decoder's Packet Loss Concealment (PLC) algorithm for various loss event probabilities. In this manner, perceptually critical voice packets that cannot be easily replaced by a PLC are provided with greater error protection. For a given average bit-rate, a simple unequal error protection scheme provides a 0.2 to 0.3 advantage in PESQ-MOS over the conventional equal error control schemes.

1. INTRODUCTION

In order to combat packet loss in Voice over IP (VoIP), a combination of an active error control method such as Reed-Solomon (RS) Forward Error Correction (FEC) coding with a Packet Loss Concealment (PLC) method is typically used [1][2][4].

However, conventional schemes for packet-level FEC in VoIP are based on equal error protection in that each voice packet is allocated an equal amount of resources for protection against packet loss. Therefore, conventional VoIP FEC schemes do not take into account the relative perceptual importance of packets and the effect of the PLC. For example, a sequence of packets representing a steady-state voiced speech signal is allocated the same amount of FEC resources as a packet containing a transition between unvoiced and voiced speech, ignoring the perceptual importance of the packets. In addition, the fact that a PLC can conceal packet loss within a steady-state voiced segment more easily than packet loss during a transition is not utilized within the FEC allocation framework. In previous work [6], the unequal perceptual importance of packets was recognized. However the solution offered only works on a network with differentiated services (e.g., Diffserv) unlike the current best-effort networks.

In this paper we propose an optimized unequal error control scheme based on a Rate-Distortion framework [3]. In this scheme, the active error control resources are allocated in an unequal manner, meaning that certain voice packets are allocated more FEC protection than other packets. In particular, our basic unequal protection scenarios include varying the number of copies of a packet that are piggybacked onto subsequent packets, and an adaptive RS coding scheme in which only certain packets are provided with RS FEC. In determining the amount of error control resources to be allocated to a packet, the sender anticipates what the PLC will do in the case of packet loss, and calculates the expected distortion for various protection scenarios. In this manner, the sender computes various operating points on a Rate-Distortion plane, and subsequently uses Lagrangian optimization to select an optimal protection policy.

With a more efficient use of bandwidth, the optimized unequal scheme outperforms the conventional equal error protection schemes, and provides better perceptual quality. Compared with previous equal protection schemes, the gain by the unequal scheme is about 0.2 to 0.3 in Perceptual Evaluation of Speech Quality (PESQ) [9] value for a fixed average bandwidth in simulations with the fixed-rate Internet Low Bit-rate Codec (iLBC) [7] codec. The remainder of this paper is organized as follows. Section 2 points out the limitations of the current equal error protection methods. Section 3 proposes an optimized unequal error control scheme. Section 4 shows our simulation results. Section 5 concludes the paper and lists future work.

2. LIMITATIONS OF CURRENT FEC SCHEMES

In current VoIP practice, the most common FEC schemes consist of simple piggybacking, Low Bit-rate Redundancy

(LBR) or packet-level Reed-Solomon (RS) codes [1]. The details of the conventional packet-loss FEC can be found elsewhere [1][4]. In general, all the schemes rely upon methods in which information about packet n is sent along with subsequent packets.

In these traditional FEC schemes, all speech frames are treated equally and given the same amount of error-control resources. However, some packets are more important to perceptual quality while others can be easily concealed by the PLC algorithm. For example, packets containing transitions between speech signal types are perceptually more important than a single packet within a sequence of packets representing a stationary voiced segment. In fact, the PLC will be able to more easily conceal the loss of a packet within a stationary segment than the loss of a packet containing a transition segment.

Further compounding the problem is that the ITU-T E-Model [1][4][8] commonly used to estimate the conversational voice quality in VoIP cannot reflect the perceptual significance of certain voice packets because its method for assessing performance is based on static curves from MOS tests that are both signal and time-invariant. Therefore, the many methods that use the E-model for optimizing FEC cannot provide a framework for assessing the quality and resource allocation tradeoffs inherent in VoIP.

3. OPTIMIZED UNEQUAL ERROR PROTECTION

In our unequal error control scheme, speech packets are treated unequally and more protection is allocated to those packets whose loss cannot be easily concealed by a PLC. For those packets which can be easily concealed, less or even no extra protection will be assigned. In particular, our basic unequal protection scenarios include adaptive piggybacking and adaptive RS coding.

In the adaptive piggybacking method, differing numbers of $K-1$ copies of a specific packet are piggybacked onto subsequent packets. If $K=1$, then only the original primary packet is sent and there is no redundant copy. Assume that packets are generated at intervals of $\Delta = 20ms$. The K'th copy is sent $(K-1) * \Delta$ms later than the first one. To keep the overall delay low, we restricted K to a maximum of $K=3$, which entails a receiver delay similar to the FEC decoding delay imposed by a RS (3,2) code [1].

In the adaptive RS coding scheme, only certain packets are provided with RS FEC. In particular, for every two packets, a decision is made to provide protection through a RS (3,2) code or not. To keep the delay low, we restricted ourselves to the use of RS (3,2) codes.

To determine the particular choice of error protection within an adaptive piggybacking or adaptive RS coding, the sender first calculates the expected distortion at the receiver for a given policy choice. For example, if adaptive piggybacking is used, then there are three possible piggybacking policies for each packet, corresponding to $K=1$, $K=2$, or 3. These three policies correspond to three different rates, and the sender obtains three possible operating points in the Rate-Expected Distortion plane. In the adaptive RS (3,2) scheme, for every two packets, there are two possible policies: adding protection or not, and thus the sender obtains two operating points in the Rate-Distortion plane. The method of calculating Expected Distortion for a given policy is deferred until subsection 3.2. Thus, we get the set of achievable R-D pairs. In Figure 1, each dot is the (R,D) performance of some policy.

After geting the set of possible (R,D) operating points, the sender must choose a particular policy. Towards this end, we employ the commonly used Lagrangian optimization methods primarily used in video communication [3][5]. The Lagrangian optimization method can provide solutions amenable to TCP-Friendly rate control constraints [4], and can be performed independently for each frame [3]. In particular, we introduce a Lagrange multiplier $\lambda \geq 0$, and consider the Lagrangian cost $O_{ij}(\lambda) = D_{ij} + \lambda R_{ij}$, in which i represents frame i, j represents policy j. We can find a policy minimizing O_{ij}, as illustrated by Figure 1.

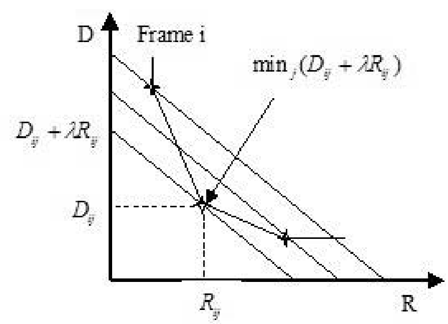

Fig. 1. Rate-Distortion plane and Lagrangian optimization

The Lagrangian multiplier λ allows us to make a tradeoff between rate and distortion. As λ increases, the rate of the optimized policy decreases and the expected distortion increases. Thus, we can meet the rate constraint through controlling the value of λ, [3][5].

3.1. Expected Rate-Distortion

To determine an operating point within the (R,D) plane for a given error protection policy, the sender calculates the expected distortion at the receiver. In particular, the primary packet i under consideration (or a copy of it in a piggybacking policy) successfully arrives at the receiver with a certain probability $P_{playout}(i)$. If packet i arrives at the receiver, then the only distortion incurred is the encoding distortion $D_c(i)$ for frame i. However, if the packet i does not arrive, an event which occurs with probability $(1 - P_{playout}(i))$, then the PLC at the receiver will produce a substitute voice signal i_{PLC} for playout.

The particular signal i_{PLC} produced depends on the state of the receiver's PLC which is determined by the particular packet loss pattern of the previously sent packets. Therefore the sender anticipates that the PLC can produce a number of possible i_{PLC} signals each with a particular probability. Consequently the expected distortion $\hat{D}(i, i_{PLC})$ in the case of PLC can be calculated. So for a given policy, the expected distortion $D_e(i)$ is calculated by

$$D_e(i) = P_{playout}(i) * D_c(i) + (1 - P_{playout}(i)) * (\hat{D}(i, i_{PLC})).$$

Therefore, we are choosing the particular error protection policy through an analysis by expected decoder synthesis paradigm within a Lagrangian R-D optimization framework.

To simplify the computations, the sender approximates the signal i_{PLC} produced by the PLC by a particular signal from a previous frame. This approximation is justified by the fact that most conventional PLC algorithms typically repeat or extend signals from previous packets to approximate a missing packet. Furthermore, we assumed that the PLC only uses the two previous frames $i-1, i-2$ to replace a lost frame in the calculations. Therefore, in considering the event that the current frame and two previous frames are all lost, we assumed that the PLC solely produces background noise. Given an error protection policy for speech frame i, the Rate $R_e(i)$ is the sum of source coding rate and the amount of error protection redundancy, and the Expected Distortion $D_e(i)$ is estimated by

$$\begin{aligned}
&D_e(i) = P_{playout}(i) * D_c(i) + (1 - P_{playout}(i)) * (P_{playout} \\
&(i-1|no\ i) * D(i, i-1) + (1 - P_{playout}(i-1|no\ i)) P_{playout} \\
&(i-2|no\ i, i-1) * D(i, i-2) + (1 - P_{playout}(i-1|no\ i)) \\
&(1 - P_{playout}(i-2|no\ i, i-1)) * D(i, background\ noise)).
\end{aligned}$$

$P_{playout}(i)$ is the probability of playing out packet i, which is computed by the method in subsection 3.2. $D(i,k)$ is the distortion between the original speech frame i and the degraded frame k after coding, which is estimated by the method in subsection 3.3 . $D(i,k)$ corresponds to the case in which the signal i_{PLC} is based on the recevied frame k.

3.2. Probability of Playout

We assume the network loss model to be the Gilbert Model [4], which is a two state Markov chain with parameters p, the transition probability from no loss to loss, and q, the transition probability from loss to no loss, leading to a steady-state loss probability of $p/(p+q)$. As in [5], we model packet delay as having a shifted Gamma distribution with rightward shift κ and parameters n and α. It can be interpreted as a packet going through n routers, each of which requires a constant processing time κ/n plus waiting time in a steady state M/M/1 queue. The parameters can be estimated by a method given in [5]. With loss and delay models, we can calculate the probability of playout $P_{playout}$ of a packet protected by certain policy.

For example, consider a piggybacking scheme. A packet is played out as long as the receiver receives any one of the K copies on time. We assume that packet loss and delay are mutually independent, that the packets are generated at intervals of $\Delta = 20ms$, and that the playout time at the receiver is D. We remove the assumption that the network delivers packets in sequence because the packets may arrive unordered due to delay variation. Let d_k be the delay of the K'th copy. The formulas below give $P_{playout}$ when $K = 1, 2, 3$.

$$\begin{aligned}
&if K = 1, P_{playout} = \frac{q}{p+q} P[d_1 < D]; \\
&if K = 2, P_{playout} = 1 - \frac{p}{p+q}(1 - qP[d_2 + \Delta < D]) \\
&\quad - \frac{q}{p+q} P[d_1 > D](p + (1-p)P[d_2 + \Delta > D]); \\
&if K = 3, P_{playout} = 1 - \frac{p}{p+q}(1 - qP[d_1 < D]) * \\
&\quad (1 - qP[d_3 + 2\Delta < D]) - \frac{q}{p+q}(p + (1-p)P[d_1 > D]) * \\
&\quad P[d_2 + \Delta > D](p + (1-p)P[d_3 + 2\Delta > D]);
\end{aligned}$$

3.3. Distortion between two speech frames

To measure the distortion between two frames i and j of speech, we compute a simple Log Spectral Distortion between the two all-pole spectral envelopes $S_i(e^{j\omega})$ and $S_j(e^{j\omega})$, i.e.

$$D(i,j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |S_i(e^{j\omega}) - S_j(e^{j\omega})|^2 d\omega$$

The distortion captures changes in the model of the vocal-tract transfer function, and is simple to compute.

4. SIMULATION RESULTS

We have done simulations to assess the performance of the proposed optimized unequal error protection scheme. We use the iLBC codec (15.2Kbits/s) which encodes the speech frames independently. To evaluate the voice quality, we use the Perceptual Evaluation of Speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs [9].

First, we compare the optimized adaptive piggybacking with the conventional equal piggybacking method. The parameters of the Gilbert model were set to $p = 0.25$ and $q = 0.7$. In this case, the average loss rate is 26%. In order to compare the two methods at the same average sending rate, if the average number of copies in a conventional piggybacking scheme is not an integer, 1.6 for example, we

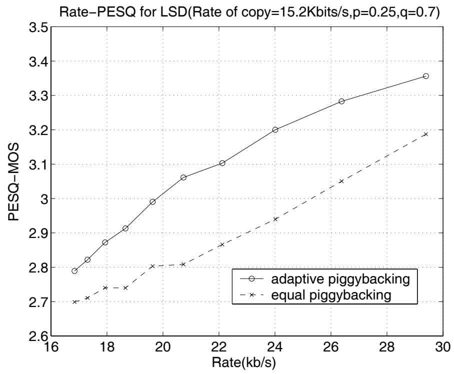

Fig. 2. Adaptive piggybacking vs. equal piggybacking (loss rate 26%). The adaptive one has a gain of about 0.2 to 0.3 in PESQ MOS value.

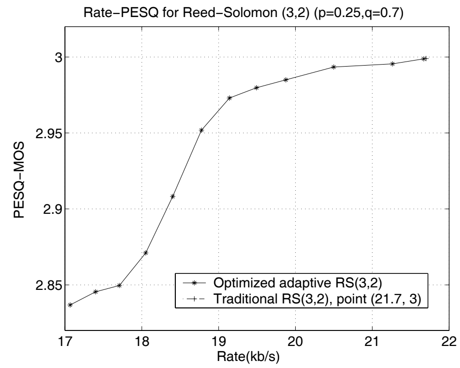

Fig. 4. Adaptive RS(3,2) vs. conventional RS(3,2)(loss rate 26%), To achieve a similar PESQ MOS value, e.g., 2.97, the sending rate can be 2.6Kbits/s lower.

randomly select 60 percent of the packets and send two copies for each of them, while just sending one copy for the remaining 40 percent. From Figure 2, we can see the proposed unequal scheme makes an improvement of about 0.2 to 0.3 in PESQ MOS value, given the same average rate budget. To achieve the same PESQ MOS value, 3.0 for example, the unequal scheme requires the sending rate to be 19.5Kbits/s and saves about 6Kbits/s compared to the equal error protection scheme. For the case of $p = 0.1$, $q = 0.7$ the average loss rate is 12.5% . The simulation result is illustrated by figure 3. The proposed unequal scheme has a 0.2 PESQ-MOS advantage over the conventional scheme. The performance of adaptive piggybacking is 0.1 PESQ-MOS better than the conventional RS(3,2) scheme which has a fixed rate of 21.7Kbits/s.

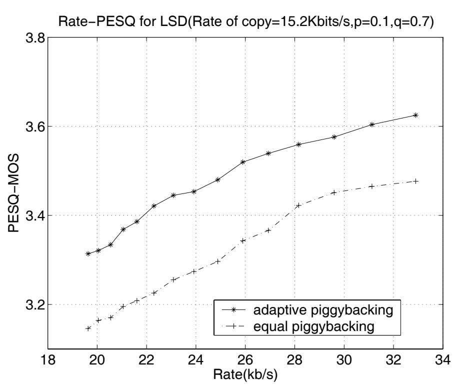

Fig. 3. Adaptive piggybacking vs. equal piggybacking (loss rate 12.5%). The adaptive one has a gain of up to 0.2 in PESQ MOS.

In figure 4, we compare the optimized adaptive RS (3,2) with the traditional RS (3,2), which has a fixed PESQ-value of 3. To achieve a similar PESQ MOS value, say 2.97, the sending rate required by the adaptive RS (3,2) can be 2.6Kbits/s lower.

5. CONCLUSION AND FUTURE WORK

We have proposed an unequal error protection scheme based on a Rate-Distortion framework. The simulation results based on the iLBC codec shows that the optimized unequal scheme outperforms conventional equal error protection scheme such as RS(3,2) and provides better perceptual quality. Based on the current framework, we are going to improve the distortion measure and incorporate an adaptive rate codec into the scheme.

6. REFERENCES

[1] Wenyu Jiang,Henning Schulzrinne, "Comparision and optimization of packet loss repair methods on VoIP perceived quality under Bursty Loss," *NOSSDAV'02,* 2002.

[2] Jean Bolot, Sacha Fosse-Parisis, Don Towsley, " Adaptive FEC-Based error control for interactive audio in the internet" *Proc. Infocom'99,* 1999.

[3] Antonio Ortega, Kannan Ramchandran, " Rate-distortion methods for image and video compression," *IEEE SIGNAL PROCESSING MAGAZINE,* November,1998.

[4] C. Boutremans, J. Y. Le Boudec, "Adaptive Joint Playout Buffer and FEC Adjustment for Internet Telephony," *Proc. IEEE INFOCOM'2003, San-Francisco, CA* April 2003.

[5] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *Microsoft Research Technical Report MSR-TR-2001-35,* February 2001.

[6] Juan Carlos De Martin, "Source-driven packet marking for speech transmission over differentiated-services networks," *Icassp'01,* 2001.

[7] Andersen,S.V.,W.B. Kleijn,R. Hagen, J. Linden,M.N. Murthi,and J.Skoglund, "iLBC–A Linear predictive coder with robustness to packet loss," *In 2002 IEEE Speech Coding Worksshop Proceedings,pp.23-25,* 2002.

[8] R. Cole and J. Rosenbluth, "Voice over IP Performance Monitoring," *Computer Communication Review, vol. 31, no. 2, pp.9C24,* April 2001.

[9] ITU-T P.862, "Perceptual Evaluation of Speech Quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *International Telecommunication Union,* 2001.