Proc. IEEE Workshop on Multimedia Signal Processing (Victoria, BC), pp. 155-160, Oct. 2006

Optimizing Voice-over-IP Speech Quality Using Path Diversity

Mohamed Ghanassi and Peter Kabal
Electrical & Computer Engineering
McGill University
Montreal, Quebec H3A 2A7
Email: {mohamed.ghanassi, peter.kabal}@mcgill.ca

***Abstract*— In last few years, voice over Internet protocol (VoIP) has been gaining popularity as an alternative to traditional telephone by transmitting voice signals as packets over the Internet and private IP-based networks. However, voice packets experience loss, delay, and delay variation, which requires buffering, playout scheduling and loss concealment at the receiver. In this paper, we give an overview of a VoIP application and show how playout scheduling and loss concealment are jointly used to optimize perceived speech quality. We use this optimization criterion to design a histogram-based playout scheduling algorithm. Then, we identify the limitations of the VoIP application for this scheme and propose improvement using path diversity approach that can be implemented via a Service Overlay Network (SON). We present simulations results that show significant improvement of VoIP quality by using this approach.**

I. Introduction

VoIP is a technology for transmitting voice over a data network, such as the Internet, using the Internet protocol (IP). Voice information is sent in digital form in discrete packets rather than in the traditional circuit-committed public switched telephone network (PSTN). Packets are moved along the most efficient path to their destination, where they get reassembled and delivered as voice. A major advantage of VoIP is cost saving by integrating data and real-time voice traffic on the same packet network.

However, packet switched networks were not originally designed for real-time transmission and VoIP technology does not always provide high-quality speech. This is mainly due to the fact that Internet protocol is a best-effort delivery service and Quality of Service (QoS) is not guaranteed. Voice packets experience varying delay (known as jitter) and in order to reconstruct a continuous signal, the received packets have to be buffered for sufficient time so that most of the packets are received before their scheduled playout time. With buffering and playout scheduling, there is a trade-off between delay and packet loss. If the playout delay is set long enough so that most of the packets will arrive before the scheduled playout time, the *packet loss ratio (PLR)* is very low. However, in this case, the *end-to-end delay* between the generation of speech and its playout at the receiver is large, which is not well tolerated in voice conversation. If the playout delay is set shorter, the PLR increases since some packets arrive after the scheduled playout time and are effectively lost.

Playout scheduling methods allow a certain amount of late packets. These packets, which are declared lost, as well as the packets dropped in the network, have to be replaced using *Packet Loss Concealment (PLC)* methods. PLC techniques that have been developed [1] are effective for loss of a single or a few packets. However, when packets are lost successively in a long burst, which occurs when the network delay suddenly increases for a short period of time, concealment of the lost packets is not very effective. Situations of burst loss have to be detected by the playout scheduling algorithm in order to adapt the playout delay accordingly.

Playout scheduling of packets is performed by searching for the best trade-off between delay and loss, corresponding to the optimum speech quality as perceived by service users. However, with severe network impairments, even the best trade-off achieved may translate to an unsatisfactory service level. One approach we can use to improve the VoIP service quality is to add redundancy to the voice signal. The information sent as redundant will be used by the receiver in order to recover the signal in case of packet loss. A commonly used redundancy scheme, known as Forward Error Correction (FEC), transmits information from a packet along with the original information in subsequent packets [2]. FEC methods are effective for isolated loss but in order to combat burst losses, they have to introduce additional delay in the speech signal. As an alternative to this technique, a path diversity based redundancy scheme was used to send redundant information on a second path and take advantage of uncorrelated loss and delay characteristics of the two paths [3].

Implementation of path diversity by explicitly sending packets over different paths can be achieved via IP source routing or Relay approaches [4]. With IP source routing, we can explicitly specify the intermediate nodes a packet should visit on route to its destination. However, this is currently difficult to implement in the Internet as not all ISPs support source routing. With the second approach, flexible routing is provided using relays placed at a number of nodes in the network. To send a packet from a point to another through an alternate path, the packet is simply sent to a relay where, upon reception, it is forwarded to its destination. Service Overlay Networks (SONs) have been developed as an effective means to provide flexibility in the routing process and have been proposed as a means to address the QoS issue for VoIP [5]. SONs are formed by inserting service gateways, which act as relay nodes and perform service-dependent routing, at access boundaries in the Internet. Currently, SONs, combined with the peer-to-peer network model, are commercially implemented by VoIP service providers (see for example [6]).

In the work of Liang et al. [3], the advantage of using path diversity with multiple description coding (MDC) scheme relative to a single path scheme with FEC and with the same data rate was demonstrated. They used adaptive playout scheduling that sets a playout delay according to a specified loss rate. However, they did not evaluate the benefits of using path diversity when the trade-off between buffering delay and packet loss is optimum. In our work, we show how using a path diversity approach improves the optimum perceived speech quality in a VoIP conversation. Furthermore, we evaluate the improvement for a full redundancy scheme where all packets are sent as redundant and for a partial redundancy scheme where only a fraction of packets, judiciously selected, are sent as redundant information.

In this paper, we explore the approach of path diversity to improve VoIP performance. We design a playout scheduling algorithm based on the optimization of the perceived speech quality and that uses a histogram-based method for delay prediction. In Section 2, we present an overview of the existing playout scheduling algorithms and the design of our algorithm. In Section 3, we present the main ideas of improving VoIP service using path diversity and show the benefits of this approach.

II. Design of playout scheduling algorithm

Fig. 1 shows a network trace of IP packet delays transmitted from UK to China [7][1]. Delay randomly fluctuates above a minimum value (corresponding to the propagation delay). Sometimes, it shows a spike by suddenly increasing and then decreasing to a level similar to that before the spike. A playout scheduling process is used to buffer packets before their playout. In a simple playout scheduling process, a fixed delay can be used but many packets could be lost in the case of a spike. Then, based on the fact that in a typical real-time voice conversation there are talkspurts (when the user is talking), and periods of silence (when the user is listening), per-talkspurt scheduling algorithms have been developed. In a per-talkspurt approach delays are adjusted at the beginning of each talkspurt and the silence periods are compressed or stretched. However with this approach, when a delay spike occurs during a talkspurt many packets can be lost. A recently introduced approach reduces the effect of spikes by adjusting the delay for packets during a talkspurt [8][9]. Packets for which delay has been adapted may have to be time-scale modified.

The starting point for designing an effective playout scheduling algorithm is to estimate the network delay for future packets. Methods commonly used are statistically-based, in the sense they use the statistics (either statistical parameters or distributions) of past delays to estimate the current playout delay. An early algorithm for playout delay was designed based on an auto-regressive estimation of mean and variation of the network delay [10]. Then, another family of scheduling algorithms that exploit delay statistics (such as pdf, histogram) have been developed [11][12][9]. A delay distribution is stored for the last w packets and used to estimate the playout delay for the next packet. With this approach, we can specify the acceptable loss rate and the algorithm automatically adjusts the delay accordingly. Fig. 1 illustrates the concept of this approach. We generate a histogram of delays for the most recent w packets (we used $w = 1000$), and the playout delay for the next packet is set such that the proportion of packets with an end-to-end delay larger than the end-to-end delay d_s for that packet is equal to the specified target PLR.

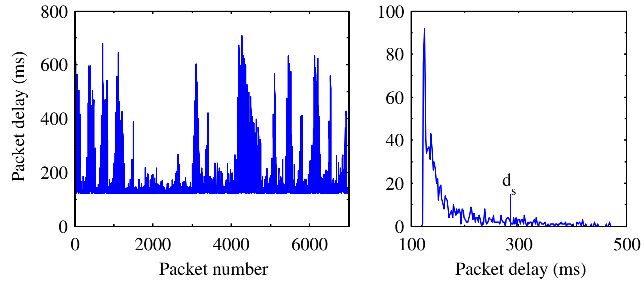

Fig. 1. Network delay trace and the histogram of delays. d_s corresponds to a packet loss ratio of 10%.

Recently, scheduling algorithms have been developed based on the optimization of the perceived speech quality [13][14]. They use a model that translates network impairments (packet loss, packet delay) to perceived speech quality and estimate a playout delay that maximizes the speech quality. In these algorithms, the relation between playout delay and packet loss is based on the modelling of the tail distribution of the delays [13][14]. For designing our playout scheduling algorithm we will use this method of optimization of speech quality, but instead of modelling the packets delay with a specific distribution we use a more general technique that exploits the histogram of delays of received packets.

Spike delays may cause a degradation in speech quality if the playout delay algorithm does not react to their occurrence. A spike detection method has to be integrated in the VoIP application in order to distinguish a normal mode from a spike mode where the playout delay is set differently. A method commonly used sets the playout delay for future packets such that the end-to-end delay is equal to the measured delay for current packet. This technique may improve the performance of VoIP application but delay spikes remain very harmful and the only way to mitigate their effect is to reduce or avoid their occurrence. Using path diversity is a very efficient solution against delay spikes as we will show farther in this work.

Optimum perceived speech quality based algorithm

For designing a playout scheduling algorithm, we use speech quality criteria at the receiver. Fig. 2 shows an overview of a VoIP application. Impairments that affect a VoIP call

[1] The authors thank L. Sun for providing the delay traces.

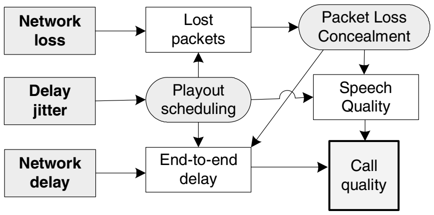

Fig. 2. Overview of a VoIP application.

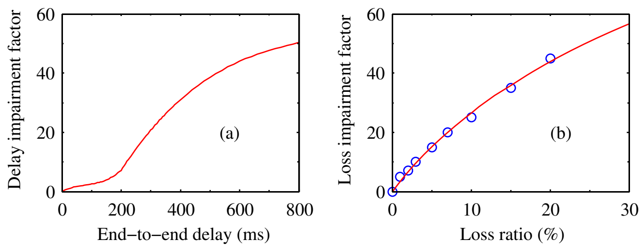

Fig. 3. a. Delay impairment factor I_d calculated from [15]. b. Loss impairment factor I_e for G.711 (from [16]) and a curve fit of the measured data, of the form $7\ln(1+50Loss)$.

quality are network loss, network delay and delay jitter. Delay jitter is mitigated using a well designed playout scheduling algorithm. However, this algorithm induces additional delay and packet loss. It may also affect the speech quality if for example it uses time-scale modification of packets. A PLC technique is used to generate lost packets and the speech quality depends on its performance. This PLC technique may require an additional buffering of received packets before performing signal generation (for example if it uses future packets), which affects the total end-to-end (or mouth-to-ear) delay. The call quality depends on the speech quality achieved and on end-to-end delay.

Depending on the PLC technique used, the playout scheduling algorithm is designed based on the optimization of the call quality (or the perceived speech quality (PSQ)). Since the relation between network performance and PSQ is not straightforward, the E-model was created by the ITU to offer speech assessment based on network performance [15]. This model combines network impairments to provide a scalar quality rating value R, ranging from 0 to 100. An estimated Mean Opinion Score (MOS) for the call quality in the scale 1 to 5 can be obtained from the R-factor [15]. By considering only the impairments relevant to the design of a VoIP playout scheduling algorithm (end-to-end delay and loss), the R-factor can be written as [15]

$$R = 93.2 - I_d - I_e \tag{1}$$

where I_d is the impairment factor caused by packet delay and I_e is the impairment factor caused by packet loss. I_d can be calculated using an analytical expression given in [15] and is shown in Fig. 3a. I_e has been evaluated by subjective listening tests for different coders and for random loss [16]. For G.711, the coder we use in our analysis, results are shown in Fig. 3b.

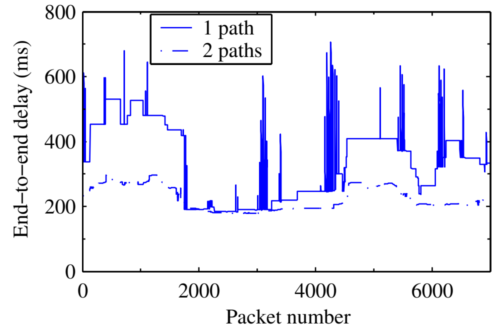

Fig. 4. Results of end-to-end delay calculated by the perceived speech quality optimization based playout scheduling algorithm for one-path and two-paths schemes.

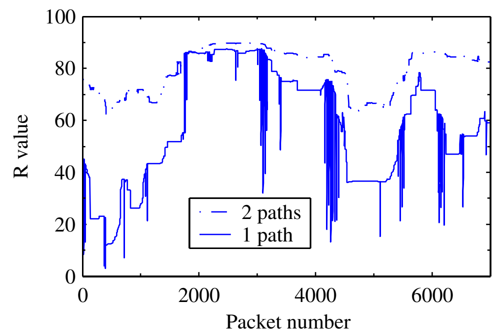

Fig. 5. Results of the R-factor calculated subsequently to using the perceived speech quality optimization based playout scheduling algorithm for one-path and two-paths schemes.

Packet loss includes network loss and scheduling induced loss, which depends on the playout delay. If we can establish a relation $L(d)$ between the total loss and delay, the R-factor can be written in terms of delay as

$$R(d) = 93.2 - I_d(d) - I_e\big(L(d)\big) \tag{2}$$

Then, we can find the value of d that maximizes $R(d)$ in (2) and set the playout delay accordingly. In order to find a relation between loss and delay, recent studies have modelled the delay distributions in the tail region as a Weibull [14] and a Pareto [13] distributions. However, we noticed that the playout delay may be very sensitive to the type of distribution used. In our design we use a more general method that makes no *a priori* assumption on the delay distribution. We simply use the histogram of the most recent w delays as shown in Fig. 1 to calculate the loss as a function of delay.

With our histogram-based method for $w = 1000$ and using measured packet delays of Fig. 1, we simulated the behaviour of our playout scheduling algorithm. In the case of a delay spike, we set the end-to-end delay for the next packet to the measured delay for current packet. Fig. 4 shows the end-to-end delay results. Variations in the end-to-end delay show the ability of the playout scheduling algorithm to track changing conditions in the network.

III. USING PATH DIVERSITY

The playout scheduling algorithm described in the previous section optimizes the PSQ by finding the best compromise between loss and delay in normal mode, and increases the playout delay in spike mode. Then, the PLC method uses the available speech signal to generate the missing packets. Hence, the performance of a VoIP application using this algorithm is limited by the best compromise between loss and delay achieved in normal mode, by the frequency of occurrence of delay spikes, and by the amount of speech available for the PLC method. In order to overcome these limitations and improve the performance of the VoIP application, we can send redundant packets through a different path independent of the main path. If the transmission bandwidth is available, we can improve the trade-off between loss and delay and reduce the frequency of occurrence of spikes simply by duplicating and sending all packets on the second path. If the transmission bandwidth is limited, we can use the second path to transmit only a part of the information. Previous studies showed that in a speech signal, certain packets are more important than others [17] in the sense that they are more difficult to regenerate well in case of loss. Then, by sending these packets as redundant on a second path, we can improve the speech quality at the receiver. We will analyze these path diversity schemes.

A. Full redundancy

If we are willing to significantly increase the bit-rate to get a better performance, we can duplicate 100% of packets and transmit them on a second path. The receiver selects the first arriving packet to reconstruct speech signal. However, in order to improve the VoIP performance by using this path diversity scheme, the second path has to be independent from the main path. Indeed, if the performances of the two paths are highly correlated, there is no significant gain in using the second path. Furthermore, the performances of the two paths in terms of minimum delay and jitter must not be very different. If for example packets from the second path arrive much later than those from the first path, they only can be used if the buffering delay is dramatically increased, which may compromise performance.

In addition to improving the best compromise loss-delay in normal mode, using this path diversity scheme reduces the effect of delay spikes on the speech quality. Indeed, when the two paths are independent, the probability of having both delays of the first and second paths in spike mode is small. And even if the delays of both paths are in spike mode, the scheduling algorithm sets the playout delay for the next packet according to the minimum of the current delays for the two paths.

The playout delay algorithm in the two-paths scheme is designed using the PSQ optimization method. The R-factor from (2) is used with a function $L(d)$ measured for the two-paths scheme. Assuming the two paths are independent, $L(d)$ is calculated by multiplying the loss functions corresponding to each path.

We simulated the behaviour of our playout scheduling algorithm for a two-paths scenario. Since we have no simultaneous delay measurements for two different paths with the same source and destination, we used for the first path, the measured delay trace of Fig. 1, and for the second path another portion of this trace. Results of end-to-end delay are shown in Fig. 4. Compared to the simulations results for a single path scheme, the end-to-end delay is significantly reduced in the periods where it was large for a single path scheme, and shows a lower variation with time. Furthermore, using path diversity, the effect of delay spikes is completely eliminated, which would highly improve the perceived speech quality.

Using the simulation results for single and two-paths cases and data of Fig. 3, we calculated the R-factor from (1). For each packet, the delay impairment factor is calculated for the simulated end-to-end delay and the loss impairment factor is calculated for a loss ratio measured for the most recent 1000 packets. Fig. 5 shows the improvement to the R-factor by using a two-paths scheme. The R-value shows a minimum of around 65 for the whole simulation period.

B. Partial redundancy

Duplicating 100% of packets in the second path is not efficient in terms of the bit-rate used. If we aim to improve the performance of VoIP application without significantly increasing the bit-rate, we can improve the speech quality at the receiver by sending only the most important packets (or the packets that are difficult to generate in case of loss) through a second path.

The importance of voiced frames preceded with unvoiced frames and transitional (unvoiced to voiced) frames for packet loss concealment and for speech quality has been confirmed in recent studies [17][18]. In a study of redundancy-based scheme for PLC of G.723.1 codec, it was shown that speech quality is significantly improved by using redundant information about important packets in the reconstruction process. With a simple method used for packet classification, it was determined that only 11% of packets are important [17]. Hence, by sending these packets as redundant information through a second path, the bit-rate is only slightly increased.

Design of a playout scheduling algorithm based on the optimization of PSQ requires a real-time identification of important packets and a characterization of the combined effect of packet loss for these packets on the second path and for all packets on the main path. Then, using the loss functions for the two paths $L_1(d)$ and $L_2(d)$, we can write the R-factor in terms of the end-to-end delay d as

$$R(d) = 93.2 - I_d(d) - I_e\big(L_1(d), L_2(d)\big) \qquad (3)$$

where I_e is the loss impairment factor caused by the combined packet loss. The playout delay is set according to the value of d that maximizes the R-factor.

In practice, a real-time selection process for important packets can be for example based on packet classification according to voiced/unvoiced criteria [17]. In our study, and for the purpose of showing the benefit and efficiency of this redundancy scheme, we apply a non real-time method on a recorded speech signal. We identify the important packets by dropping from the speech signal one packet at a time and estimating the MOS after applying the built-in G.711 PLC method [19]. MOS estimation is performed using an objective method of Perceptual Evaluation of Speech Quality (PESQ) [20]. We select important packets as those corresponding to a MOS value less than $MOS_o - 2\sigma$, where MOS_o is the MOS measured in a situation where no packet is dropped, and σ is the variance of MOS for all packets. With this selection criteria, the proportion of important packets in the speech signal is 21%.

To characterize the impact of packet loss for the first and second paths on speech quality, we use the same speech signal and simulate different loss scenarios for these packets on the second path as well as for all packets on the first path. For fixed PLRs PLR_1 and PLR_2 for first and second paths respectively, we randomly select from the speech signal a group of packets corresponding to PLR_1 (they correspond to packets lost on path 1), then we select from the speech signal another group of important packets corresponding to PLR_2 (they correspond to important packets lost on path 2). Then, we identify packets that are in both groups, which are considered lost on both paths and are removed from the original speech signal. Then, we apply the PLC method to the resulting speech signal for which a MOS value is estimated using the PESQ tool. For each value of PLR_1 and PLR_2, this experiment is repeated 100 times and the resulting MOS values are averaged. Then, we convert MOS scores to *R*-values using a relation given by the E-model [15], and calculate a raw loss impairment factor as $I_r = 100 - R$. For a single path scheme, we map these I_r values to the ITU's values of Fig. 3b and calculate a mapping function that we use to convert I_r to loss impairment factor I_e for a two-paths scheme.

Fig. 6 shows the loss impairment factor estimated with the method described above, corresponding to different packet loss ratios for the first and second paths. We used the values of $0, 20, 50, 80$, and 100% for PLR_2. For other values, the loss impairment factor can be interpolated. Using these data in (3) and delay measurements of Fig. 1, we calculated the R-factor corresponding to the two-paths scheme. From the histogram of packet delays we calculated the function $L_1(d)$ and we used the same function for $L_2(d)$ (assuming that the two paths have the same performance). Fig. 7 shows the R-factor at different end-to-end delay values for single path and two-paths schemes. The playout scheduling algorithm sets a playout delay that corresponds to the maximum value of the R-factor, which is, according to Fig. 7, improved by 15 in a two-paths scheme.

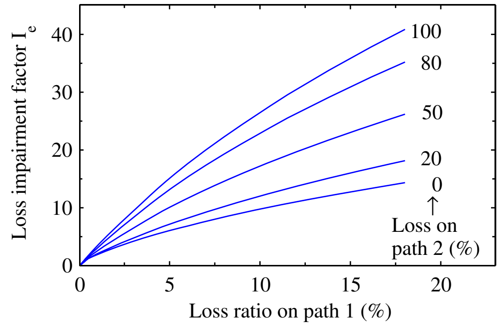

Fig. 6. Loss impairment factor estimated for G.711 for a two-paths scheme at different packet loss ratios for path 1 (all packets) and path 2 (important packets).

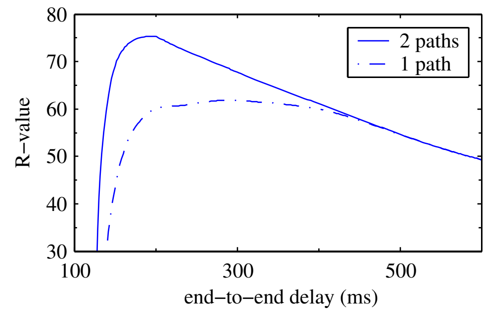

Fig. 7. R-value versus end-to-end delay for one-path and two-paths schemes. In the two-paths scheme, only important packets (21% of the total number of packets) are sent on the second path.

IV. Conclusion

In this paper, we have proposed two different path diversity schemes for improving VoIP performance. After we presented an overview of a VoIP application and showed how its components affect its performance, we first designed a new playout scheduling algorithm based on the optimization of perceived speech quality and using the histogram of packet delays in a single path scheme. Then, we identified the limitations of a VoIP application for this scheme and proposed improvement techniques using full and partial path diversity approaches. We presented simulations results that showed a significant improvement for the perceived speech quality of the VoIP application by using path diversity.

Acknowledgment

The authors would like to thank NSERC, Canada for supporting this research.

References

[1] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, pp. 40–48, Sept-Oct 1998.

[2] V. Hardman, A. Sasse, and A. Watson, "Reliable audio for use over the Internet," in *Proceedings of INET '95*, June 1995, pp. 171–178.

[3] Y. J. Liang, E. G. Steinbach, and B. Girod, "Real-time Voice Communication over the Internet Using Packet Path Diversity," *Proceedings of the ninth ACM international conference on Multimedia*, vol. 9, pp. 431–440, Sept. 2001.

[4] J. G. Apostolopoulos, "Reliable Video Communication over Lossy Packet Networks using Multiple State Encoding and Path Diversity," *Visual communications and Image Processing (VCIP)*, pp. 392–409, Jan. 2001.

[5] Z. Duan, Z.-L. Zhang, and Y. T. Hou, "Service Overlay Networks: SLAs, QoS, and Bandwidth Provisioning," *IEEE/ACM Transactions on Networking*, vol. 11, no. 6, pp. 870–883, Dec. 2003.

[6] S. Guha, N. Daswani, and R. Jain, "An Experimental Study of the Skype Peer-to-Peer VoIP System," in *Proceedings of The 5th International Workshop on Peer-to-Peer Systems (IPTPS '06)*, Santa Barbara, CA, Feb. 2006.

[7] L. Sun and E. Ifeachor, “Prediction of perceived conversational speech quality and effects of playout buffer algorithms,” in *Proc. IEE ICC*, Anchorage, USA, May 2003, pp. 1–6.

[8] F. Liu, J. Kim, and C.-C. J. Kuo, “Adaptive delay concealment for Internet voice applications with packet-based time-scale modification,” in *Proc. IEEE ICASSP*, vol. 3, May 2001, pp. 1461–1464.

[9] Y. J. Liang, N. Farber, and B. Girod, “Adaptive playout scheduling and loss concealment for voice communications over IP networks,” *IEEE Trans. Multimedia*, vol. 5, no. 4, pp. 532–543, Dec. 2003.

[10] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, “Adaptive playout mechanisms for packetized audio applications in wide-area networks,” in *Proc. IEEE INFOCOM*, vol. 2, June 1994, pp. 680–688.

[11] S. B. Moon, J. Kurose, and D. Towsley, “Packet audio playout delay adjustment: Performance bounds and algorithms,” *Multimedia Systems*, vol. 6, pp. 17–28, Jan. 1998.

[12] K. Fujimoto, S. Ata, and M. Murata, “Statistical analysis of packet delays in the Internet and its application to playout control for streaming applications,” *IEICE Trans. Communications*, vol. E84-B, no. 6, pp. 1504–1512, June 2001.

[13] ——, “Adaptive playout buffer algorithms for enhancing perceived quality of streaming applications,” in *Proc. IEEE GLOBECOM*, vol. 21, no. 1, Nov 2002, pp. 2463–2469.

[14] L. Sun and E. Ifeachor, “New models for perceived voice quality prediction and their applications in playout buffer optimization for VoIP network,” in *Proc. IEEE ICC*, Paris, France, June 2004, pp. 1478–1483.

[15] I. R. G.107, “The E-Model, a computational model for use in transmission planning,” Mar. 2003.

[16] I. R. G.113, “Transmission impairements due to speech processing,” Feb. 2001.

[17] L. Tosun, “Dynamically adding redundancy for improved error concealment in packet voice coding,” MSc thesis, McGill University, Dec. 2004.

[18] L. Sun, G. Wade, B. Lines, and E. Ifeachor, “Impact of packet loss location on perceived speech quality,” in *Proc. of 2nd IP-Telephony Workshop (IPTEL ’01)*, Columbia University, New York, Apr. 2001, pp. 114–122.

[19] ITU-T Recommendation G.711, “Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711,” Sept. 1999.

[20] ITU-T Recommendation 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,” Feb. 2001.