Adaptive Rate Control for Real-time Packet Audio Based on Loss Prediction

Lopamudra Roychoudhuri and Ehab S. Al-Shaer
School of Computer Science, Telecommunications and Information Systems
DePaul University
Chicago, Illinois 60604
Email: {lroychou,ehab}@cs.depaul.edu

***Abstract*— This paper presents an *Adaptive Rate Control Framework*, which performs proactive rate control based on *packet loss prediction* and *on-line audio quality assessment* for real-time packet audio. The proposed framework determines the optimal combination of various audio codecs for transmission, utilizing a thorough analysis of audio quality based on individual codec characteristics, as opposed to ad-hoc codec redundancy used by other approaches. As a main component of this framework, this paper shows a preliminary formulation of the *Packet Loss Predictor*, that determines the likelihood of packet loss in terms of available bandwidth, delay variation trend, and network history. We present simulation and experiment results to show the accuracy and efficiency of this technique.**

I. Introduction

Internet measurement experiments continue to show that bandwidth, delay, and packet loss on the Internet vary considerably, posing a challenge for real-time multimedia applications, which require bounded errors in these aspects in order to ensure quality. In the absence of such guarantees of reliability or quality, it is thus necessary to design control mechanisms for multimedia transmission, which will dynamically adapt the behavior of the application to maximize the media quality under network constraints.

Quality of an audio communication is highly sensitive to packet loss [3],[15] caused by congestion in the links. Packet loss for audio is normally rectified by adding redundancy using Forward Error Correction (FEC) [10]. However, unnecessary high degree of FEC can cause excessive traffic and can be detrimental to the ongoing communication. Here the challenge is to ensure a bandwidth-friendly transmission with an effective degree of loss recovery by dynamically changing the degree of FEC. We present a mechanism to predict packet loss in real-time audio streams based on delay variation and trend, that will enable proactive error recovery and rate control. The loss prediction value returned by the Predictor indicates current degree and severity of congestion and lack of available bandwidth created by burstiness of cross traffic, and is fed back from the receiver to the sender. The Rate Control system reacts to changing bandwidth and delay by changing the optimal codec combination, while maintaining the audio quality. This is superior to the static feedback used in current Sender based Error and Rate Control mechanisms, which adapt to variations of available bandwidth by depending on RTCP feedback for packet loss and are limited by reacting to the packet drop information collected at the receiver [1],[11].

Current audio compression techniques have a very diverse range in terms of degree of compression and underlying technologies. Thus they react differently under different network conditions of delay, jitter and degrees of packet loss. We present a real-time audio quality assessment technique to determine the quality of an ongoing audio transmission objectively by analyzing the current loss and delay and their effects on different codecs. The combination of various codecs is thus decided by the Rate Control based on the on-line audio quality assessment, guaranteeing an optimized quality under the current loss and delay conditions. This is superior to ad-hoc mixing of codecs and FEC used in other techniques [1].

Subsequent sections are organized as follows. Section II contains a discussion of the related work. Sections III, IV and V present the Loss Predictor, the Audio Genome (an online audio quality assessment technique), and the Adaptive Rate Control Framework. In Section VI we show simulation and experiment results. Section VII presents the Conclusion and the Future Work.

II. Related Work

Paxson examined the delay-loss correlation issue in his PhD thesis [14]. But he concluded that the linkage between delay variation and loss was weak, though not negligible. In contrast with Paxson's observations, we predict packet loss based on the observed patterns of delay variation, rather than depending on the overall amount of delay variation as indicator of congestion. Moon's PhD thesis [12] explored this issue to a certain degree, but they did not attempt to take a further step of real-time prediction of packet loss from the delay variation data of an ongoing communication. Pathload [9] uses the delay variation principle to measure available bandwidth. The same principle is used in TCP Vegas [2], which exploits RTT variation to measure the difference between the actual and the expected sending rate to provide better congestion detection and control. Packet loss is highly probable when the available bandwidth is low and is consumed by the ongoing cross traffic. Our Loss Predictor method is based on this premise.

Cole & Rosenblath [3] described a method for monitoring VoIP applications based upon E-model [5], where they used curve fitting of ITU-published I_e values for selected codecs

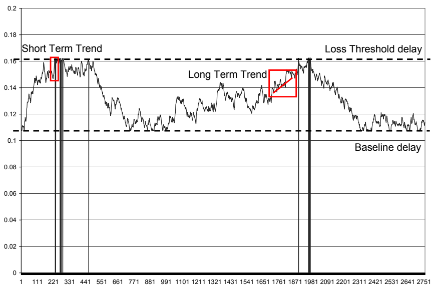

Fig. 1. One Way Delay - Predictor Components

for various loss percentages. VQMon [4] is a non-intrusive passive monitoring system for VoIP using an extended E-model incorporating packet loss and recency effect. We extend these methods to provide real-time audio quality assessment.

Sender based Error and Rate Control mechanisms for audio, such as by Bolot & Garcia [1], Mohamed et al. [11] adapt to packet loss using RTCP feedback from the receiver, thus they react to packet loss. In contrast, we predict loss and take rate control actions based on the current loss likelihood.

III. Packet Loss Prediction

In our approach, we designate the minimum delay of a path as the baseline delay, signifying the delay under no congestion. We also identify the delay at the capacity saturation point of a path as the loss threshold delay, after which packet loss is more likely. We track the increase patterns or trends of the delay as an indication of congestion causing packet loss. We have seen in our experiments that each site shows a consistent minimum baseline delay. We also observe a range of loss threshold delay values after which loss is observed more often (Fig. 1- vertical lines denoting packet loss). The loss threshold delay shows a variety of ranges and behaviors due the unpredictable nature of the cross traffic in the network at that time. To measure these path delay characteristics, we propose certain measurement metrics classified in three categories - Delay Distance ($DelayDist$), Short Term Trend ($STTrend$) and Long Term Trend ($LTTrend$). Delay Distance gives an absolute ratio of the delay value in relation to the baseline and loss thresholds. The Short-term Trend and the Long-term Trend metrics indicate sharpness and consistency of upward and downward delay trends in short and long term window of past packets (Fig.1). We determine these metrics from the ongoing traffic and combine them with different weights based on their importance in order to estimate of the packet loss likelihood. The Loss Predictor is expressed as the following:

$$0 \leq f(DelayDist, STTrend, LTTrend)$$
$$= w_1 * DelayDist + w_2 * STTrend + w_3 * LTTrend \quad \leq 1$$
and $w_1 + w_2 + w_3 = 1$

The Predictor uses dynamic weights w_1, w_2 and w_3 that depend on the current delay situation and congestion level. Details of this work can be found in [16].

1) Delay Distance: This metric can be expressed as:

$$DelayDist = min\left(1, \frac{D^k - base}{thr - base}\right)$$

where, $base$ = the minimum observed delay so far, considered to be the baseline delay, D^k = the delay value of the k-th packet, thr = the threshold delay at which a loss is observed.

The range of this metric is [0,1]. This metric is computationally simple and a good indicator of the absolute delay increase or decrease between the baseline and the loss threshold, and hence is an important component of loss prediction.

2) Long Term Trend Metrics: These metrics indicate if an over all increasing trend is evident for a large number of preceding packets. The length of this packet window is adjusted dynamically based on the delay mean deviation observed so far, and is typically 20 to 50 packets.

Long Term Consistency Measures (Spct/Spdt) We use variations of *PCT* (Pairwise Comparison Test) and *PDT* (Pairwise Difference Test) presented in [9]. Both of them are indicators of consistent increasing trends in a packet train length of Γ. The range of *Spct* is [0,1] and of *Spdt* is [-1,1], scaled to [0,1].

$$Spct = \frac{\sum_{k=2}^{\Gamma} I(D^k > D^{k-1})}{\Gamma - 1}, \; I(X) = 1 \text{ if X, else } 0$$
$$Spdt = \frac{D^{\Gamma} - D^1}{\sum_{k=2}^{\Gamma} |D^k - D^{k-1}|}$$

3) Short Term Trend Metrics: These metrics signify how fast the delay is increasing over last small window of packets. The length of this packet window is adjusted dynamically based on the mean delay deviation observed so far, and is typically 5 to 10 packets. We use *SI* as an indicator of sharpness of increase (the 'slope' of the increase), and *SpctST/SpdtST*, short-term versions of *Spct* and *Spdt*, as indicators of the consistency of increase.

Sharpness Indicator (SI) This metric determines how fast the delay is approaching the loss threshold by measuring the slope of the line joining the delay values of the current packet and the previous packet.

$$SI = max(-1, min(1, (D^k - D^{k-1})/(t^k - t^{k-1})))$$

Under a sudden burst of high congestion, the slope is observed to be steeper. Thus higher degree of slope indicates higher congestion, and hence higher likelihood of packet loss. The range of this metric is truncated from $[-\infty, +\infty]$ to [-1,1] and scaled to [0,1].

IV. Audio Quality Assessment

We had noticed in our Internet experiments that different codecs react differently to various degrees and burstiness of losses [15]. For example, there was a change of behavior of G.728 vs. G.711 in terms of quality when the loss ratio increased. Also, the codecs had very different quality output under various degrees of loss bursts. This motivated us into further investigating the effect of packet loss and delay on various codecs. We thus attempt to establish audio quality as a function of the codec type, loss distribution and delay, store

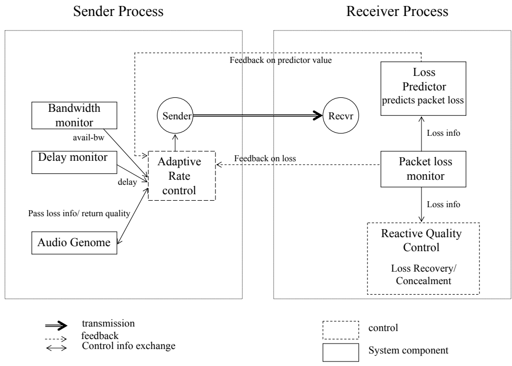

Fig. 2. Adaptive Rate Control Framework

in a system (Audio Genome), and then use it to determine the quality for an ongoing multi-codec communication.

A. Audio Genome

The purpose of this approach is to establish audio quality as a set of functions, each for a specific loss condition, for a number of selected codecs. This repository will be used to derive the audio quality of the ongoing transmission by analyzing the loss data and using the functions for the codec to fit the current loss distribution. We will extend ITU-recommended methods and use E-model (G.107 [5], curve-fitting and interpolation for current loss data to derive I_e (Equipment Impairment factor), hence R-factor [5] and subjective audio quality measure Mean Opinion Score MOS (P.800 [7] and P.830) for an ongoing communication.

1) Limitations of ITU Recommendations: Since we are interested in correlating packet loss for a certain codec and the resulting audio quality, it amounts to measuring the loss effects on I_e. ITU provides I_e values for many codecs under no loss condition [14] and a limited number of codecs under some loss conditions [10]. But ITU does not provide any description of statistical models used for generating random and bursty packet losses. Also, extensive subjective MOS testing with many human listeners is time-consuming, cumbersome, error-prone and non-repeatable.

2) Measuring loss effects on I_e: We address the limitations of ITU recommendations in our method. We propose to use various statistical models approximating the Internet to simulate many degrees of random and bursty packet loss. We use an objective quality assessment technique PESQ (P.862 [8]), for extensive MOS score testing. The proposed procedure is presented as follows: (1) Use statistical models to create random and burst packet loss and inter-loss gaps, (2) Drop packets using a statistical model and compare 'before' and 'after' images and get MOS scores using PESQ; derive I_e values from these MOS scores, converted to R factor, (3) Deduce curve fits for these values, (4) Compare the results of representative files with subjective testing, (5) Use these formulas to interpolate the data/curves for real-time parameters to determine I_e and deduce MOS for an ongoing transmission.

V. Adaptive Rate Control Framework

Here we propose an adaptive Rate Control framework that uses combination of audio codecs of various bitrates, together with Audio Genome, and conserves bandwidth while maintaining an optimized audio quality. In comparison to the sender-based mechanisms reacting to packet loss, we predict packet loss using the Loss Predictor and take rate control actions based on the nature of the prediction. Hence our approach is to detect congestion proactively based on current loss likelihood, and react to changing bandwidth and delay by simultaneously optimizing the rate and audio quality under the current constraints by changing the codec combination.

Fig. 2 depicts the main components of the proposed system. The Sender takes proactive actions in terms of Rate control and Error Control, whereas the Receiver provides various feedbacks to the sender and takes reactive control actions such as error concealment. In this paper we focus mainly on the Rate Control part of this mechanism. The adaptive action of the Rate Control depends on the severity of bandwidth degradation and delay increase, and is activated if the Predictor feedback is above a high threshold value for a considerable amount of time, indicating significant long-term bandwidth degradation.

At the Sender, the Bandwidth Monitor determines available bandwidth of the connection. The Delay monitor estimates the absolute end to end delay from the Sender to the Receiver. Audio Genome returns the audio quality score for a particular codec when queried for that codec under certain loss condition and delay. At the Receiver, The Loss Predictor keeps track of per-packet one way delays at the receiver side and periodically returns a predictor score denoting packet loss likelihood based on the delay change patterns as observed at the Receiver. The Packet Loss Monitor keeps track of the percentage of loss, loss burst frequency and gap lengths preceding loss. This data is required by the Audio Genome to calculate quality values for the codecs used by the Adaptive Rate Control. Reactive Quality Control provides loss recovery and loss concealment.

The Sender gets feedback on packet loss distribution of the current transmission from Packet Loss Monitor and passes the loss data to Audio Genome to get the current audio quality of the session so far. The Sender also uses feedback on available bandwidth and end-to-end delay from the Available Bandwidth monitor and the Delay monitor. These feedbacks prompt the adaptive mechanism to solve an optimization problem in order to maximize the audio quality of the ongoing connection under the current constraints of available bandwidth, end-to-end delay and packet loss.

A. Rate-Quality Optimization

The objective of the Rate-Quality Optimization problem is to derive the set of codec combination that will maximize the audio quality of the ongoing connection under the current constraints of available bandwidth, end-to-end delay and packet loss. The solution of the optimization problem is a combination of ratios of codecs and/or bitrates that ensures the highest possible audio quality under the current network condition. The various constraints are explained in detail.

1) Constraint of Available Bandwidth: High bitrate encoders can be used when the network is underutilized, whereas low bitrate encoders are more appropriate for a tighter bandwidth constraint. It is possible to mix multiple encoders in a certain ratio for bandwidth optimization, but we need to ensure that the audio quality provided is optimum.

2) Constraint of Delay: The One Way Delay (OWD) in a network path consists of the propagation delay, a service delay and a queuing delay at each hop of the path. The OWD cannot exceed the allowable Mouth-to-ear (M2E) delay of 400ms according to ITU specifications [6]. The difference of the allowable M2E delay and the OWD can be consumed by the delay inherent to the codecs used in the audio transmission.

3) Packet Loss: Different codecs react differently to various degrees and burstiness of losses. The audio quality scores for a particular codec are derived from Audio Genome under a certain packet loss situation, and are used in formulating the objective function of the optimization problem.

Problem Formulation. Maximize the audio quality under the constraint of available bandwidth and link delay.

Maximize $z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$
subject to
$b_1x_1 + b_2x_2 + \ldots + b_nx_n \leq B$
$d_1x_1 + d_2x_2 + \ldots + d_nx_n \leq D$
$c_1x_1 + c_2x_2 + \ldots + c_nx_n \leq 4.3$
$c_1x_1 + c_2x_2 + \ldots + c_nx_n \geq 3.5$
$x_1 + x_2 + \ldots + x_n = 1$
$x_i \geq 0, i = 1 \ldots n$
where
$x_1, x_2, \ldots x_n$ = 'amount'/percentage of each codec (type+bitrate) in the transmission mix
$c_1, c_2, \ldots c_n$ = MOS score for each codec under current loss
$b_1, b_2, \ldots b_n$ = bit rate of each codec
$d_1, d_2, \ldots d_n$ = (packet size in bytes)*(encoding/decoding delay to create/decode 1 byte)
B = available bandwidth, D = 400 - link OWD

The objective function is the audio quality to be maximized, and is expressed in terms of the sum of the product of codec percentage and the codec quality score under current loss condition, as determined by Audio Genome. The rationale behind the constraints is as follows. The total bandwidth consumption by the codecs, expressed as the sum of the products of bitrate and percentage of each codec, should not exceed the available bandwidth. Similarly, the total codec delay, expressed as the sum of products of encode/decode delay and percentage of each codec, should not exceed the difference of the maximum allowable M2E delay (400ms) and the link OWD. The quality sum cannot exceed the maximum quality value 4.3 (the MOS score of G.711 under no loss), and should be greater than or equal to 3.5 (lower bound of acceptable speech quality). An example solution of this optimization problem is presented in the next section.

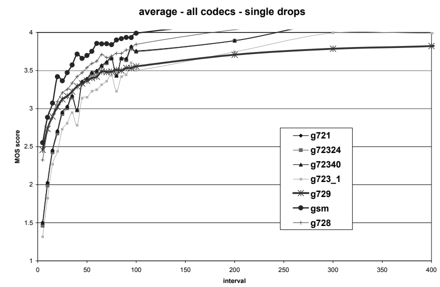

Fig. 3. Comparison of codecs under single packet loss

VI. Simulation and Experiment Results

A. Predictor Evaluation: Simulation Results

Ideally the predictor should behave accurately, that is, the predictor should report high values for the majority of packet loss occurrences. The predictor should also be efficient by not over-estimating when there is no loss. We evaluated the Predictor under many different cross-traffic and intermediate hop scenarios. We created different degrees of congestion at intermediate links, resulting in packet loss, by introducing a varied number of transient cross traffics of different packet sizes, and determined the accuracy and efficiency of the Predictor under these scenarios. We noticed that the loss predictor value is accurate 78% of the time in a path with a two intermediate hops, vs. 59% in a path with five intermediate hops. In the second case we simulated more variable and random degrees of congestion at different parts of the path compared to the first case, resulting in less accuracy for the predictor. This motivates us into refining the predictor for better accuracy under such variable conditions. For efficiency, both the simulation scenarios show very small percentages (5% for the first case, 2% for the second case) of predictor value larger than 0.7 when there is no loss. Thus the Predictor behaves efficiently in the simulation environment. A detailed study of these results can be found in [16].

B. Audio Genome: Simulation Results

As a preliminary simulation, we created a simplified scenario, where we dropped packets (single, bursts of 2, 3, 4 and 5) with a fixed gap of packets (5, 10, 15, etc.) in between for 7 encoders. Fig. 3 shows a representative result with interesting observations. GSM, though lower bitrate than the ADPCM codecs, performs the best in high loss conditions. It starts with the best score (2.5 at 20% loss), and stays consistently better. G.729 starts well in high loss conditions (2.45 at 20% loss), but around gap 50, starts dropping down and performing worse than others. Under low loss, G.729 performs much worse than others.

These preliminary results give us an idea of what to expect in terms of individual codec behaviors under packet loss. We need to extend this to a thorough analysis in our future work.

C. Rate-Quality Optimization: Experiment Results

Table I presents the codecs which were used to test the feasibility of the Rate-Quality optimization problem using

TABLE I

COMPARISON OF AUDIO CODECS

Variable	Codec	Underlying technology	Bitrate (kbps)	Delay for 1 byte (ms)	Packet size (bytes)	MOS Under no loss
x_1	μ-law	Waveform	64	0.5	200	4.3
x_2	G.721	Waveform	32	1	200	4.0
x_3	GSM FR	LTP-RPE	13	2.42	198	3.7
x_4	G.728	LD-CELP	8	2.5	200	4.0
x_5	G.723.1	MP-MLQ	5.6	6.0	210	3.9

TABLE II

RATE-QUALITY OPTIMIZATION EXPERIMENT RESULTS

Scenario	Network condition	Avail-bw kbps	One way delay ms	Feasible solution	Opt transmission Rate (kbps)	Optimum audio quality value
1	High avail-bw/low delay	500	40	<1,0,0,0,0>	64	4.3
2	High avail-bw/high delay	100	150	<1,0,0,0,0>	64	4.3
3	med avail-bw/high delay	50	140	<0.75,0,0,0.25,0>	50	4.23
4	Low avail-bw/low delay	30	40	<0.39,0,0,0.61,0>	29.8	4.11
5	Low avail-bw/high delay	30	150	<0.08,0.72,0,0.19,0>	30	4.03
6	Really low bw/high delay	20	150	Infeasible!!		

Maple V, a mathematical software. The observations from the results achieved as feasible solutions (Table II) match what is expected intuitively, thus seem correct in the practical sense.

Scenarios 1 and 2: Scenario 1 is the Best-case scenario when the path has high available bandwidth and low link delay. As expected, the feasible solution consists of only PCM (x_1), since it is of high bitrate, but of low delay and highest quality. The same is true for Scenario 2.

Scenarios 3 and 4: But as the bandwidth is made tighter, the feasible solution starts tending towards low bandwidth-low delay encoder G.728 (x_4), still maintaining quality.

Scenario 5: Worst-case scenario of low available bandwidth and high delay. Here the feasible solution consists of less PCM (about 8%), mostly G.721 (about 72%) and G.728 (19%). GSM (x_3) is not chosen since G.728 (x_4) is of higher quality, lower bitrate and comparable delay. G.723.1 (x_5) is not chosen either. Though it is of lower bitrate, it has much higher delay.

Scenario 6: Finally this is a scenario of extremely low bandwidth and high delay, when the solution is infeasible.

These feasible solutions were tested as codec combination in an audio communication system and the resulting audio quality was tested subjectively by users. There was no noticeable quality degradation due to the codec mixing and the overall audio quality was acceptable.

VII. CONCLUSION AND FUTURE WORK

In this paper we present a rate control system that detects congestion based on the feedback of packet loss prediction (Packet Loss Predictor), and uses an optimal combination of various codecs along with an on-line audio quality assessment (Audio Genome) to perform effective rate control while maintaining optimized audio quality.

The proactive and dynamic nature of various components makes the framework superior to reactive mechanisms with ad-hoc codec mixing and static feedbacks, and a viable technique for majority of network conditions. The results of the Predictor under simulation scenarios show 60%-80% accuracy. The Rate-Quality optimization problem, consisting of 5 variables (a reasonable number for codec combinations), is of low complexity and runs fast to provide feasible solution.

As future work, we need to refine the metrics and the weight factors to improve the accuracy and efficiency of the Predictor. As proposed in Audio Genome, we need to do a thorough analysis of audio quality variations for codecs under different degrees and distributions of packet loss. The design of the Rate Control framework needs to be implemented along with the Predictor and Audio Genome, and evaluated. The framework also needs to be formalized as a TCP-friendly mechanism. We would like to extend and apply the Rate Control mechanism to Video, Overlay and Multicast frameworks.

REFERENCES

[1] J. Bolot and A. Vega-Garcia. "Control Mechanisms for Packet Audio in the Internet." *Proceedings IEEE Infocom*, San Francisco, CA, pp 232-239, April 1996.

[2] L. S. Brakmo, S. W. O'Malley and L. Peterson. "TCP Vegas: New Techniques for Congestion Detection and Avoidance." *ACM SIGCOMM Conference*, pp. 24-35, May 1994.

[3] R.G. Cole and J.H. Rosenbluth. "Voice over IP Performance Monitoring." *ACM SIGCOMM*, 2001.

[4] A.D. Clark. "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality." *IPTel*, April 2001.

[5] ITU-T Recommendation G.107 (12/98). "The E-Model, a computational model for use in transmission planning."

[6] ITU-T Recommendation G.114 (05/93). "One way Transmission Time."

[7] ITU-T Recommendation P.800 (08/96). "Methods for subjective determination of transmission quality."

[8] ITU-T Recommendation P.862 (02/01). "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs."

[9] M. Jain, C. Dovrolis. "End-to-end Available Bandwidth: Measurement Methodology, Dynamics and relation with TCP Throughput." *SIGCOMM*, 2002.

[10] S. Lin, D.J. Costello.*Error Correcting Coding, Fundamentals and Applications*, Englewood Cliffs, NJ, Prentice Hall, 1983.

[11] S. Mohamed, F. Cervantes-Perez and H. Afifi. "Integrating Network Measurements and Speech Quality Subjective Scores for Control Purposes." *IEEE Infocom*, 2001.

[12] S.B. Moon. "Measurement And Analysis Of End-To-End Delay And Loss In The Internet." PhD Thesis, *Department of Computer Science, University of Massachussets*, Amherst, 2000.

[13] The Network Simulator: ns-2: http://www.isi.edu/nsnam/ns/

[14] V. Paxson. "Measurements and Analysis of End-to-End Internet Dynamics." PhD Thesis, *Computer Science Division, University of California*, Berkeley, 1997.

[15] L. Roychoudhuri, E. Al-Shaer, H. Hamed and G.B. Brewster. "Audio Transmission over the Internet: Experiments and Observations." *IEEE International Conference on Communications (ICC)*, Anchorage, Alaska, May 2003.

[16] L. Roychoudhuri and E. Al-Shaer. "Real-Time Analysis of Delay Variation for Packet Loss Prediction." *IEEE Management of Multimedia Networks and Services (MMNS)*, San Diego, California, October 2004.