2005-07-01

An Assessment of the Audio Codec Performance in Voice over WLAN (VoWLAN) Systems

Miroslaw Narbutt
Technological University Dublin, miroslaw.narbutt@tudublin.ie

Mark Davis
Technological University Dublin, mark.davis@tudublin.ie

An Assessment of the Audio Codec Performance in Voice over WLAN (VoWLAN) Systems

Miroslaw Narbutt, Mark Davis
Communications Network Research Institute,
Dublin Institute of Technology,
Dublin 8, IRELAND
narbutt@cnri.dit.ie, mark.davis@dit.ie

Abstract

In this paper we present results of experimental investigation into the performance of three audio codecs (ITU-T G.711, G.723.1, and G.729A) under varying load conditions on a Voice over WLAN system utilizing the IEEE 802.11b wireless LAN standard. The analysis is based upon a new technique for estimating user satisfaction of speech quality calculated from packet delay and packet loss/late measurements. We also demonstrate the importance of the de-jitter buffer playout scheme for insuring speech quality. From our results we conclude that the use of the G.711 audio codec in conjunction with the new adaptive playout scheme gives the highest user satisfaction of the Voice over WLAN schemes considered.

1. Introduction

As VoIP spreads from the wireline to the wireless world, performance issues arise because the characteristics of wireline and wireless networks differ. Delay, jitter and packet loss, the key factors that impact packet voice quality in the fixed Internet, are further magnified in a WLAN environment. Due to access point congestion and poor link quality high delay variation is not unusual in an 802.11b network. Such a high jitter complicates proper reconstruction of the speech signal at the receiver and packet voice quality in WLAN environment can be severely degraded.

To compensate for jitter a typical voice over IP application buffers incoming packets in the jitter buffer before playing them out. This allows slower packets to arrive on time to be played out. The buffering delay cannot be too long or too short. If the buffering delay is too short, "slower" packets will not arrive before their designated playout time and voice quality suffers. If the buffering delay is too long, it noticeably disrupts interactive communications. It is not possible to find an optimum fixed buffer size when network conditions vary in time. Playout buffers with dynamic size allocation, so called adaptive playout buffers, are becoming more and more popular. A good playout algorithm should be able to keep the buffering delay as short as possible while minimizing the number of packets that arrive too late to be played out.

The two conflicting goals of minimizing buffering delay and minimizing late packet loss have led to various adaptive playout algorithms:

- Histogram-based algorithms as “Concord” [1] or Moon’s [2] are not capable of very rapidly increasing the buffering delay during congestion and quickly reducing it when congestion has passed.

- Reactive algorithms as Ramjee’s [3] or Bolot’s [4], that rely on estimates of network delays, either react too quickly to transient noise conditions (when the estimator gain is small) or ignore persistent changes in performance (when the estimator gain is high), but cannot do both [5].

A new playout buffer algorithm was proposed in [6][7][8] that extends the reactive approach. In that solution the estimator gain is updated with each incoming packet according to the observed delay variations. When variations in network delays are high (which implies that network conditions are rapidly changing), the value of gain is set low, and vice-versa. With higher-quality estimates of network delays, the new algorithm adapts quickly to changing network conditions, which reduces the frequency of late packets and the amount of buffering delay.

In this paper we compare the performance of reactive and histogram-based algorithms with the proposed solution in an 802.11b WLAN environment. A number of connections through one access point were used to emulate different network conditions (e.g. UDP background traffic). The experimental results show that the new algorithm predicts and follows network delays more efficiently that traditional algorithms. We also compared the performance of three audio codecs (ITU-T G.711, G.723.1, and G.729A) under varying load conditions using the ITU-T E-model methodology. From our results we conclude that the use of the G.711 audio codec in conjunction with the new adaptive playout scheme gives the highest user satisfaction of the Voice over WLAN schemes considered.

2. New playout algorithm

Most of the adaptive playout algorithms described in the literature depend on estimates of network delays to calculate playout deadlines of already received packets. Good network estimators should ignore transient noise conditions, but react quickly to persistent changes in performance. Typically, network estimators in the form of exponentially weighted moving average (EWMA) filters provide one of these properties, but not both [5]. This is because they are constructed with static gain: the smoothing parameter α that determines how aggressively an EWMA filter will track changing network conditions. This gain biases the estimator either towards past history (when α is high) or current observations (when α is low).

The basic adaptive playout algorithm [3] estimates two statistics; the delay itself and its variance:

$$\hat{d}_i = \alpha \cdot \hat{d}_{i-1} + (1-\alpha) \cdot n_i \qquad (1)$$

$$\hat{v}_i = \alpha \cdot \hat{v}_{i-1} + (1-\alpha) \cdot |\hat{d}_i - n_i| \qquad (2)$$

where $\hat{d}_i$ and $\hat{v}_i$ are the i^{th} estimates of delay and its variance respectively, while n_i is the i^{th} packet delay.

According to [3], the weighting parameter α should be fixed at a high value, e.g. α = 0.998002. This was motivated by the work on TCP roundtrip time estimation, and assumed slow changes in roundtrip time.

The idea behind the new algorithm proposed in [6][7][8] is to adaptively adjust the value of α, every time a new packet arrives, depending on the variations in the network delays. When the variation in network delays is high (which implies that network conditions are rapidly changing) the value of α is set low and vice-versa:

$$\alpha_i = f(\hat{v}'_i) \qquad (3)$$

where $\hat{v}'_i$ is a smoothed estimate of the variance of the end-to-end delay and the function $f(\hat{v}'_i)$ was chosen experimentally to maximize the performance of the algorithm over a large set of network traces. This dynamic version of parameter α is used in the estimates of delay and variance:

$$\hat{d}_i = \alpha_i \cdot \hat{d}_{i-1} + (1-\alpha_i) \cdot n_i \qquad (4)$$

$$\hat{v}_i = \alpha_i \cdot \hat{v}_{i-1} + (1-\alpha_i) \cdot |\hat{d}_i - n_i| \qquad (5)$$

The equation for the playout time of the first packet of a talkspurt is the same as in the basic adaptive algorithm:

$$p_i = t_i + \hat{d}_i + \beta \cdot \hat{v}_i \qquad (6)$$

where t_i is the generation time of the i^{th} packet and parameter β controls the delay/packet loss ratio (the larger the value of β, the more packets are played out, at the expense of longer delays). Any subsequent packets of this talkspurt are played out with rate equal to their generation rate at the sender.

We claim that under changing network conditions the accuracy of the estimate (and therefore the resulting VoIP playout quality) can be greatly improved by dynamically choosing the value of α.

3. Experimental measurements

An one-way VoIP session was established between two wireless hosts (VoIP SENDER and VoIP RECEIVER), via the Access Point (AP) in an 802.11b WLAN (Figure 1).

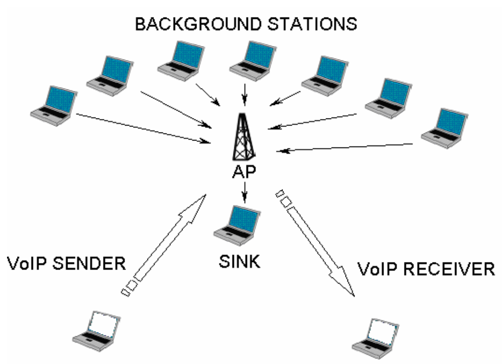

Figure 1: Measurement setup

A number of wireless stations were used to generate background UDP traffic. This was accomplished using the MGEN traffic generator [9]. The stations generated UDP packets of length 1024 bytes at a transmission rate of 50 fps. Voice traffic was generated using RTPtools [10]. The VoIP sender sent voice packets of 80 bytes every 10 ms (i.e. G.711 codec) during voice activity. No packets were generated during silence periods. A sequence of alternating active and passive periods was used following the ITU-T P.59 recommendation [11] with an exponential distribution of talkspurts and gaps (with mean values of 1004ms and 1587ms respectively). The duration of the test was one hour during which time all experimental data (packet arrival times, timestamps, sequence numbers, and marker bits) were collected at the receiving terminal and processed later (off-line) with a program that simulated the behaviour of various playout algorithms. Since the terminal clocks were not synchronized, the clock skew was removed using Paxon's algorithm [12]. The influence of the background traffic on the delay and delay variation is shown in Figure 2.

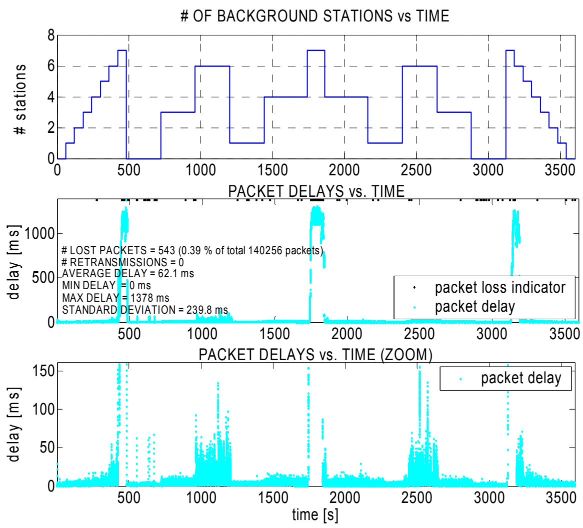

Figure 2: Influence of the background traffic on delay and jitter.

4. Effects of the encoding scheme and playout mechanism on user satisfaction.

To estimate the subjective quality of packet voice the E-Model (ITU-T Recommendation G.107 [13]) was used. The E-Model combines individual impairments (including loss, distortion, echo, delay, and noise) due to both the signal's properties and the network characteristics into a single R-rating that ranges from 0 to 100. The rating factor R is a linear combination of the individual impairments:

$$R = (R_o - I_s) - I_d - I_e + A \qquad (7)$$

From our point of view, the delay impairment I_d and equipment impairment I_e (which captures the effect of information loss due to encoding scheme and packet loss) are relevant. The other impairments: loud connection and quantization impairment I_S, basic signal to noise ratio R_0, and the "advantage factor" A do not depend on the transmission over the network. Therefore, since values of R above 94.15 are unobtainable in narrowband (300 to 3400 Hz) telephony, we can write the R rating for G.711 audio as:

$$R = 94.15 - I_d - I_e \qquad (8)$$

Based on R-rating, we assessed transmission quality and subjective user satisfaction over a one-hour period. First we calculated playout delays and packets loss for a given playout scheme. Then, for each 10 seconds of the session, we calculated delay impairments and equipment impairments according to the ITU-T E-model recommendation. Equipment impairments as a function of information loss due to encoding scheme and packet loss (including loss due to late packet arrival) were calculated for each codec separately based on the ITU-T recommendation G.113 [14]. For calculating delay impairments we assumed echo loss TELR = 65dB. After calculating delay impairments and equipment impairments we finally obtained the time varying quality of the call.

Figures below show average playout delays (using logarithmic scale for the Y axis), average packet loss and corresponding rating factor R for different playout buffer algorithms calculated for the G.711 encoding scheme.

We also took into account user satisfaction in terms of ranges of R [13] that was derivated from delay/loss distribution on the user perception quality plane. The quality plane shows how an average user rates the quality of a call, depending on packet loss and one-way end-to-end delays for a given encoder and a given echo cancellation level (each dot corresponds to average playout delay and average late packet loss for 10 seconds of the transmission) [15].

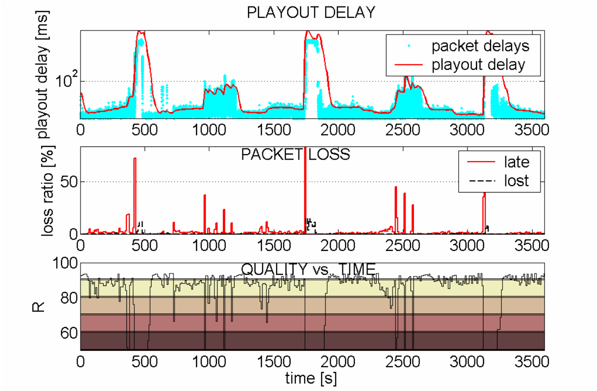

Figure 3: Time varying playout delay, packet loss and quality of the call with the Ramjee's playout alg. (α=0.998002)

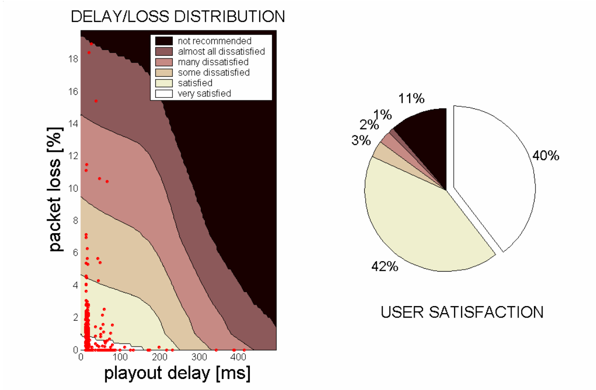

Figure 4: Distribution of playout delays and packet loss on the quality plane with the Ramjee's playout alg. (α=0.998002) and resulting user satisfaction.

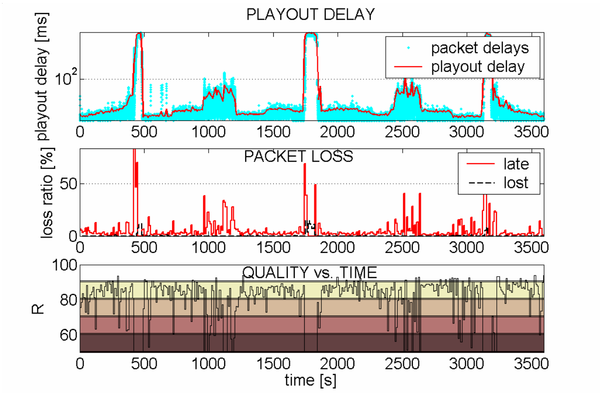

Figure 5: Time varying playout delay, packet loss and quality of the call with theRamjee's playout alg. (α=0.9)

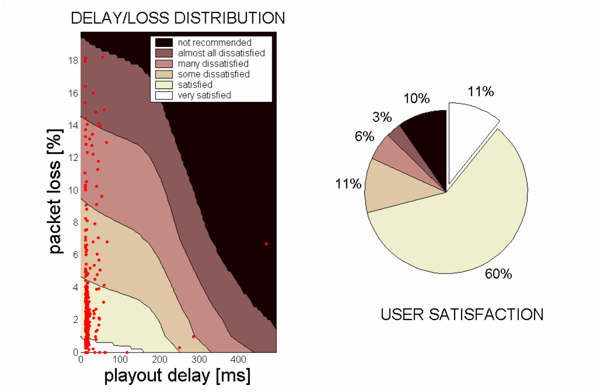

Figure 6: Distribution of playout delays and packet loss on the quality plane with the Ramjee's playout alg. (α=0.9) and resulting user satisfaction.

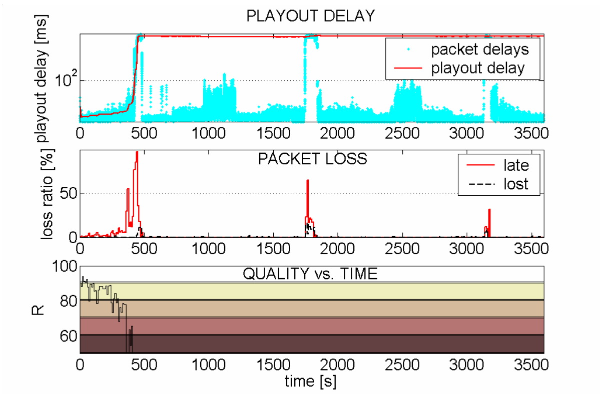

Figure 7: Time varying playout delay, packet loss and quality of the call with the “Concord” alg. (desired loss 1%)

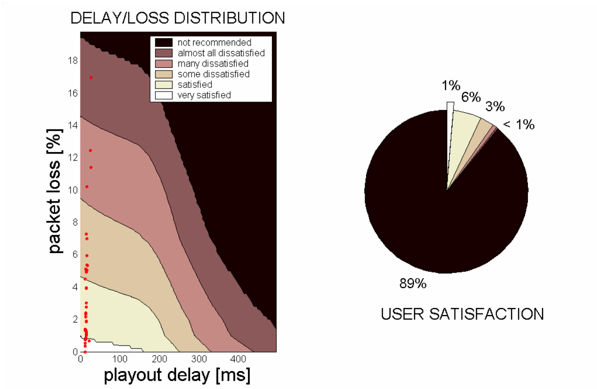

Figure 8: Distribution of playout delays and packet loss on the quality plane with the “Concord” alg. (desired loss 1%) and resulting user satisfaction

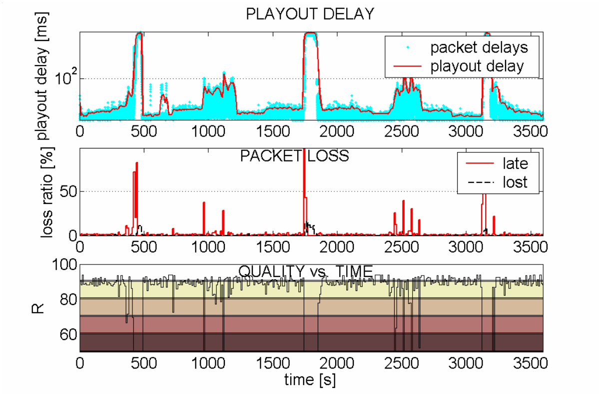

Figure 9: Time varying playout delay, packet loss and quality of the call with the Moon's alg. (desired loss 1%, # samples 400)

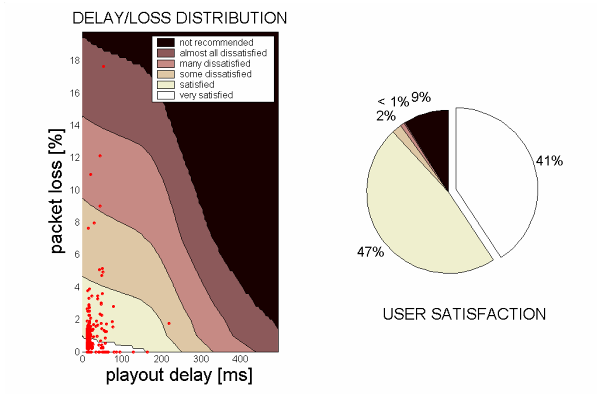

Figure 10: Distribution of playout delays and packet loss on the quality plane with the Moon's alg. (desired loss 1%, # samples 400) and resulting user satisfaction

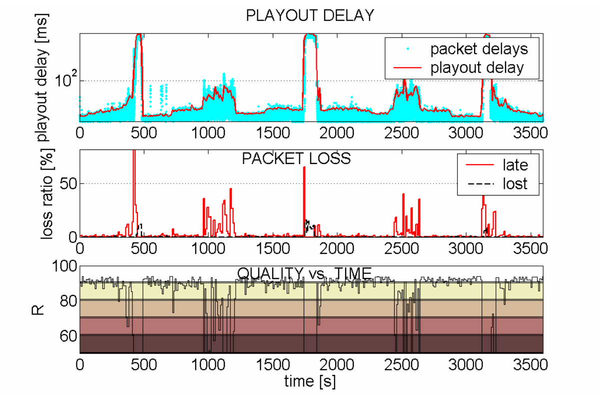

Figure 11: Time varying playout delay, packet loss and quality of the call with the Bolot's alg.

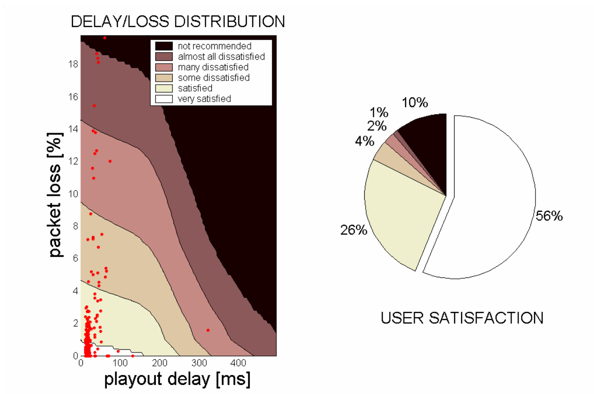

Figure 12: Distribution of playout delays and packet loss on the quality plane with the Bolot's alg. and resulting user satisfaction

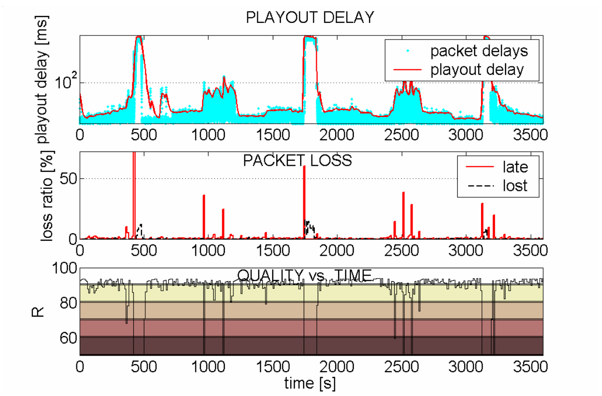

Figure 13: Time varying playout delay, packet loss and quality of the call with the "dynamic α" alg.

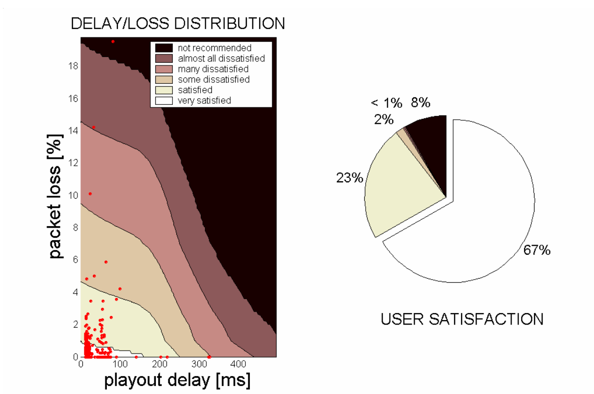

Figure 14: Distribution of playout delays and packet loss on the quality plane with the "dynamic α" alg. and resulting user satisfaction

Results above show that the new adaptive buffering scheme with dynamic α gave very good user satisfaction 67% of the time, compared to the basic algorithm with fixed α at 40% (α=0.998002), Bolot's alg. 58%, Moon's alg. 41% and Concord 1%. This indicates that the dynamic α approach responds well to the fast variations that are expected in a WLAN environment.

In a similar way we also assessed the time varying quality of the call and overall user satisfaction taking into account two other popular audio codecs i.e. G.723.1 and G.729A.

Table 1 shows user satisfaction for each of the buffering mechanisms and encoding scheme examined.

From Table 1 it can be seen that user satisfaction is highly influenced by both encoding scheme and playout buffering mechanism at the receiver. For example, adaptive buffering scheme with G.711 encoding gave very good user satisfaction 67% of the time. The same algorithm with G.723.1 or G.729A encoding couldn't achieve very good user satisfaction.

5 Conclusions

We compared the performance of three audio codecs (ITU-T G.711, G.723.1, and G.729A) in a WLAN environment under varying load conditions using the ITU-T E-model methodology. Results show that the use of the G.711 audio codec in conjunction with the new adaptive playout scheme gives the highest user satisfaction of the Voice over WLAN schemes considered.

Table 1. User satisfaction vs various encoding schemes and playout mechanisms

		USER SATISFACTION CATEGORIES					
CODEC	**PLAYOUT MECHANISM**	not recommended [% time]	almost all dissatisfied [% time]	many dissatisfied [% time]	some dissatisfied [% time]	satisfied [% time]	very satisfied [% time]
G.711	Ramjee's alg. α=0.9980	11	1	2	3	42	40
	Ramjee's alg. α=0.9	10	3	6	11	60	11
	Concord alg.	89	0	1	3	6	1
	Moon's alg.	9	0	1	2	47	41
	Bolot's alg.	10	1	2	4	26	56
	dynamic α alg.	**8**	**0**	**0**	**2**	**23**	**67**
G.723.1	Ramjee's alg. α=0.9980	15	2	11	72	0	0
	Ramjee's alg. α=0.9	18	10	35	37	0	0
	Concord alg.	90	2	2	5	0	0
	Moon's alg.	10	2	8	80	0	0
	Bolot's alg.	14	4	8	80	0	0
	dynamic α	9	2	4	85	0	0
G.729A	Ramjee's alg. α=0.9980	12	3	4	45	36	0
	Ramjee's alg. α=0.9	15	6	19	50	10	0
	Concord alg.	89	1	3	6	1	0
	Moon's alg.	9	1	4	49	37	0
	Bolot's alg.	13	2	4	30	51	0
	dynamic α alg.	9	1	2	25	64	0

Acknowledgements

This work was supported by Science Foundation Ireland grant 03/IN3/1396

References

[1] N. Shivakumar, C. J. Sreenan, B. Narendran, and P. Agrawal, "The Concord algorithm for synchronization of networked multimedia streams", Proc. of the IEEE International Conference on Multimedia Computing and Systems, May 1995

[2] Sue B. Moon, Jim Kurose, and Don Towsley, "Packet audio playout delay adjustment: Performance bounds and algorithms", ACM/Springer Multimedia Systems, January 1998

[3] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks", Proc. of the IEEE Infocom, Toronto, 1994

[4] J. C. Bolot and A. Vega-Garcia, "Control Mechanisms for Packet Audio in the Internet", Proc. of the IEEE INFOCOM, San Francisco, 1996

[5] M. Kim and B. Noble, "Mobile Network Estimation," Proc. of the ACM MOBICOM 2001, Rome, 2001

[6] M. Narbutt, L. Murphy, "VoIP Playout Buffer Adjustment using Adaptive Estimation of Network Delays", Proc. of the 18-th International Teletraffic Congress – ITC-18, Berlin, 2003

[7] M. Narbutt, L. Murphy, "A New VoIP Adaptive Playout Algorithm", Proc. of the IEE QoS 2004 Conf., London, 2004

[8] M. Narbutt, L. Murphy, "Improving Voice over IP Subjective Call Quality", IEEE Communications Letters, Vol. 8, Issue 5, p. 302-304, May 2004

[9] MGEN:
http://mgen.pf.itd.nrl.navy.mil/

[10] RTPtools:
http://www.cs.columbia.edu/IRT/software/rtptools/

[11] ITU-T Recommendation P.59, "Artificial conversational speech", March 1993

[12] Sue. B. Moon, P. Skelly, and D. Towsley, "Estimation and removal of clock skew from network delay measurements," Proc. of the IEEE INFOCOM'99, New York City, NY USA, March 1999

[13] ITU-T Recommendation G.107 "The E-model, A Computational Model for Use in Transmission Planning", 1998

[14] ITU-T Recommendation G.113, "Transmission impairments due to speech processing – Appendix 1", May, 2002

[15] M. Narbutt, Mark Davis, "Assessing the quality of VoIP transmission affected by playout buffer scheme", accepted for publication in Proc. of the Measurement of Speech and Audio Quality in Networks Conference, MESAQIN 2005, Prague, 2005