Performance Evaluation of Wireless IP Telephony (W-IPT) over Wi-Fi Networks

Maan A. Kousa
Electrical Engineering Department,
King Fahd University of Petroleum & Minerals
Dhahran 31261, Saudi Arabia
kousa@kfupm.edu.sa

Abstract -- **As one of the fastest growing voice service technologies, IP Telephony is currently the greatest benefactor of IP Convergence. Apart from the cost and management benefits of a converged network, the exciting array of productivity enabling applications such as unified messaging, collaboration, and presence services within an IP Telephony infrastructure are driving this rapid growth. Wi-Fi, on the other hand, is the widest deployed technology for indoor Internet access. It is therefore the default candidate for enabling wireless IP Telephony (W-IPT). The study aims at assessing the performance of W-IPT over Wi-Fi networks. In particular, the paper describes an experiment that was carried out on a running network at a university campus. The results shed some light on the readiness of Wi-Fi networks to embrace this fast emerging technology.**

Keywords- IP telephony;, VoIP; Wi-Fi.

I. INTRODUCTION

The evolution of Internet Protocol Telephony products has drastically increased in scale and popularity. Vugrinec and Tomazic [1] discussed several issues of IP telephony deployment, what can be expected from such communication methods, what kind of benefits does IP telephony bring, and what drawbacks should users expect in comparison to the Plain Old Telephone System (POTS).

Yet, IP Telephony faces at least two significant challenges. The first challenge is to ensure virtual connection across this connectionless packet IP network using new protocol standards, such as H.323, MGCP, MEGACO/H.248, or SIP. The second is to transport packets over the IP network in a timely manner with high integrity, thereby ensuring acceptable voice quality.

The successful deployment of IP telephony depends on the performance of the underlying data network. Consequently, assessing a network to determine whether it can accommodate the stringent Quality of Service (QoS) requirements of IP telephony is critical.

This work aims at assessing the performance of IPT over the widely-spread Wi-Fi network at King Fahd University of Petroleum and Minerals (KFUPM). Various experiments were run on different parts of the network covering good, average and poor links. Key performance indicators, namely latency, packet loss, jitter and Mean Opinion Score (MOS) were measured and analyzed. The results provided guidelines for the level of voice traffic that can be carried out over Wi-Fi links while maintaining good or acceptable call quality. Furthermore, the paper put forward some recommendations for network upgrade for better call quality or more voice traffic.

The paper is organized as follows. Section II surveys the most relevant work to the problem under investigation. Section III introduces the key performance indicators adopted for the evaluation of IPT. Section IV describes the experiment set up and assessment tool, followed by Section V where results and findings are discussed. The paper concludes by stating some useful lessons learned from this experiment.

II. RELATED WORK

The area has attracted many researchers very early. Hsiaosu, Martin, Denise, and Darren [2], and El-Sherbini, El-Sherif, Kamel, and Fayez [3] have conducted theoretical evaluations as well as computer simulations for IP Telephony assessment. Bearden, Denby, Karacali, Meloche, and Stott [4] have described a technique for evaluating a network for IP telephony readiness. Their technique relies on the data collection and analysis support of their prototype tool, ExamiNet/spl trade/. It automatically discovers the topology of a given network and collects and integrates network device performance and voice quality metrics. They report the results of assessing the IP telephony readiness of a real network of 31 network devices (routers/switches) and 23 hosts via ExamiNet/spl trade/. Their evaluation identified links in the network that were over utilized to the point to which they could not handle IP telephony.

Stefic and Prib [5] presented the results of the subjective testing of user perception of the quality with which IP telephony service is delivered. Both listening and conversational tests were considered. The results were

 ISBN: 978-1-61208-171-7

further used to test the existing commercial QoS mechanisms and their suitability for the immediate service offering. The analysis of tests enables a provider offering IP telephony to not only understand technical features of the service, but also to recognize the users' needs, behavior and their acceptance of the service and its quality. Furthermore, test results may be used as a basis when designing a network supporting the IP telephony service while using existing QoS mechanisms.

Karacali, Denby, and Meloche [6] have described a technique for efficiently assessing network readiness for IPT. Their technique relies on understanding link QoS behavior in a network from an IPT perspective. They used network topology and end-to-end measurements collected from the network in locating the sources of performance problems that may prevent a successful IP telephony deployment. They present an empirical study conducted on a real network spanning three geographically separated sites of an enterprise network.

This paper summarizes the results of an experiment conducted to assess the performance of IPT over Wi-Fi networks. This work differs from other works cited above in applying it to a vey large scale network running in real time.

III. KEY PERFORMANCE INDICATORS

The quality of IPT is inferred from a set of indicators. The first indicator is the *Delay*. Delay (or Latency) is the time it takes a packet to make its way through a network end-to-end. It is the sum of packetization delay, propagation delay, transport delay and jitter buffer delay. Generally, it is accepted that the end-to-end delay should be less than 150 ms for toll quality voice calls.

The second indicator is *Packet Loss*. During network congestion, the queue buffers of some routers and switches can overflow. Packet loss for non-real-time applications, such as Web browsers and file transfers, is undesirable but not critical. The protocols used by non-real-time applications, usually TCP, are tolerant to some amount of packet loss because of their retransmission capabilities. However, real-time applications based on UDP are significantly less tolerant to packet loss. In an RTP session, by the time a media gateway could receive a retransmission, it would no longer be relative to the reconstructed voice waveform; that part of the waveform in the retransmitted packet would arrive too late.

The third indicator is *Jitter*. Jitter is the measure of the variation of packet arrival time. Jitter can be positive, where some packets arrive late, or negative where some packets arrive early. Keeping jitter under control is of particular interest to IPT networks in order to prevent calls from developing glitches or sounding "choppy".

The fourth indicator is the *Mean Opinion Score* (MOS). Described in ITU-T P.800, MOS is the most well-known measure of voice quality. It is a subjective method of quality assessment based on users opinion of the perceived quality of a voice transmission. A MOS of 5 is excellent; a MOS of 1 is unacceptably poor. The E-Model, ITU Standard G.107 quantifies MOS by determining which impairment factors produced the strongest user perceptions of lower quality. The E-Model thus includes factors for equipment and impairments and takes into account typical users' perceptions of voice transmissions affected by jitter, lost data, and delay.

IV. EXPERIMENT SETUP

Several tools for assessment were examined, namely ResponseWatch, Vista Insight, Expernet and Vivinet. Vivinet Assessor [7] was found to be a very flexible and feature-rich software, and therefore was selected for our experimental assessment.

The test was run over KFUPM wireless LAN. KFUPM has a well developed wireless LAN based on IEEE 802.11g standard which supports up to 54 Mbps. The wireless access points are back connected to the layer-2 switches within a building, while layer-2 switches are connected to the only layer-3 switch of the building, which then forwards the data over the fiber-optic link to the university core network.

For assessment tests on WLAN environment, six locations were selected, which are distributed on three floors in Building 59 (the largest academic building) based on different criteria as given below:

Room 0032 (Ground Floor, very good signal coverage)

Room 0072 – PC1 (Ground Floor, good signal coverage)

Room 0072 – PC2 (linked with same access point of PC1)

Room 0081 (Ground Floor, far from the Access Point thus poor signal coverage)

Room 1079 (First Floor, far from the Access Point thus poor signal coverage)

Room 2078 (Second Floor, excellent signal coverage)

One IPT probe was installed in each of these locations and simulated IP calls were made between these rooms in a full mesh connection, i.e. every location calling every other location. We have considered three levels of traffic intensity: low, where one call is initiated between every pair of nodes, medium where 2 simultaneous calls are initiated between every pair, and high where 4 simultaneous calls are initiated between every pair. The logical network diagram is shown in Figure 1.

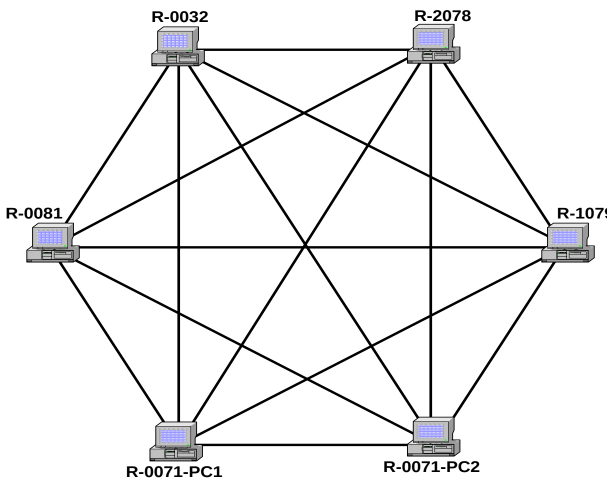

Figure 1. Logical connectivity of the assessment network

All calls are 3-minute long separated by a 5-minute silence. Each configuration (traffic intensity) was run for 24 hours. G711 Codec was used throughout the test. The specifications of G711 are shown in Table 1.

TABLE 1. G711 CODEC SPECIFICATIONS

bit rate	64 kbps	IP frame size	280 bytes
packetization time	30 ms	bandwidth at IP level	74.7 kbps
packet rate	33.3 /sec	bandwidth at ethernet level	84.7 kbps

The routers of KFUPM network were not configured for QoS, therefore all test were run with the absence of QoS protocols.

V. RESULTS AND DISCUSSION

In the study, several configurations were run to cover all scenarios of interest. Due to paper size constraints, only three sample runs are presented. The results of the other runs are supportive of the cases presented here.

Run -1: one call/pair

For a mesh network of 6 nodes, 1 call/pair translates to 5 calls/link The performance was pretty good for more than 99% of the calls. The average delay was always below 45 ms, which is quite acceptable. The lost data was negligible (less than 0.03%). Moreover, all links performed equally well. The variation in their performance was negligible.

Run-2: 2 calls/Pair

The impact of this increase in traffic on call quality was harsh: 33% of the calls were poor (Figure 2). In general, there are three main factors that affect the call quality, namely: delay, jitter, and lost data. The percentage effect of each of these factors for this run is shown in Figure 3. The source of poor quality is mainly delay (51%) and lost data (41%). The average delay varied between 300 ms and 310 ms, and the average packet loss varied between 6.6% and 7.7%; both are on the high side.

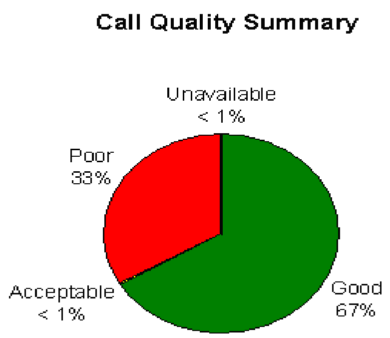

Figure 2. Call quality summary for 2 call/pair traffic.

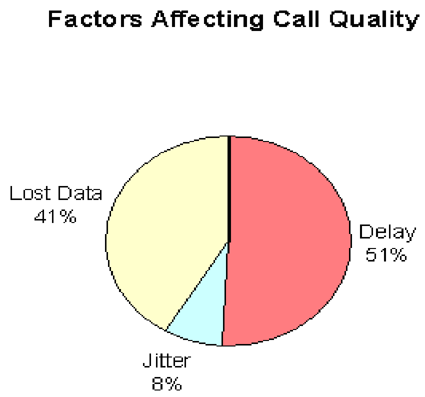

Figure 3. Factors affecting quality, for 2 call/pair traffic

Unlike the 1-call case where all links were comparatively good, there is a high variation in call quality between different pairs of nodes. Figures 4 and 5 show the call quality on the top 5 and bottom 5 links, respectively (referred to as Call "Group" in the figures). The top 5 are always good (MOS = 4.38), while the bottom 5 are always poor (MOS < 1.35). By examining the WLAN links, it can be seen that the link between Room 0081 and its AP is the source of trouble. All communications between Room 0081 and other nodes are poor, and they are the only poor links. The delay on these links exceeds 800 m sec, while the delay on other links was in the range of 45 m sec. On those same poor groups the lost data exceeds 20%.

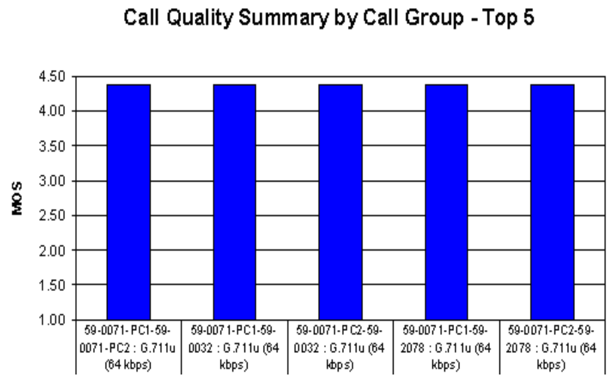

Figure 4. Call quality of the best 5 links for 2 call/pair.

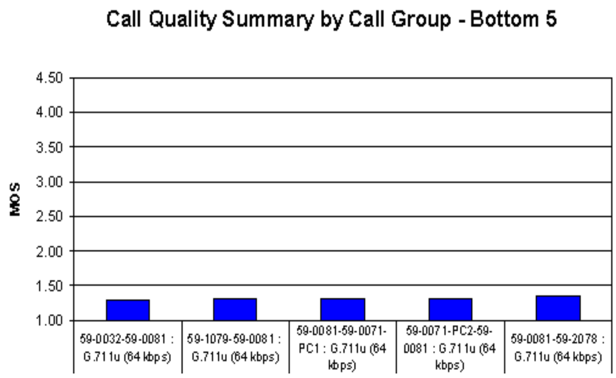

Figure 5. Call quality of the worst 5 links for 2 call/pair

 ISBN: 978-1-61208-171-7

It is to be noted that a poor link is consistently poor at all times. Figure 6 shows the quality on one such poor link (Room 0081- Room 2078), by hour. The MOS is in the range of 1.3-1.4.

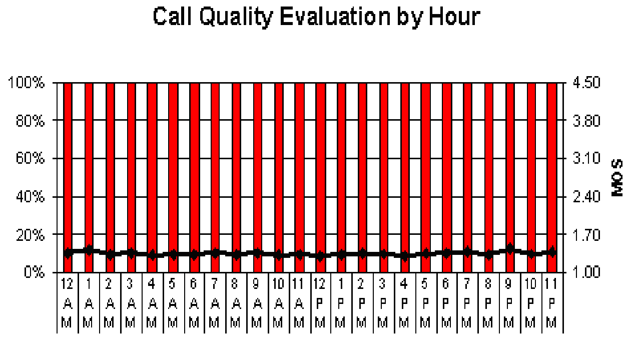

Figure 6. Call quality by the Hour for the worst link for 2 call/pair.

Run-3: 4 calls/Pair

Figure 7 shows the overall statistics of call quality for this level of traffic. The Figure shows that 75% of the calls are poor, 14% are acceptable and only 10% are good.

Figure 7. Call quality summary for 4 call/pair.

The percentage effect of the three performance factors is shown in Figure 8. We can clearly see that the source of poor quality is lost data (52 %) and delay (44%). The effect of jitter is marginal (4%).

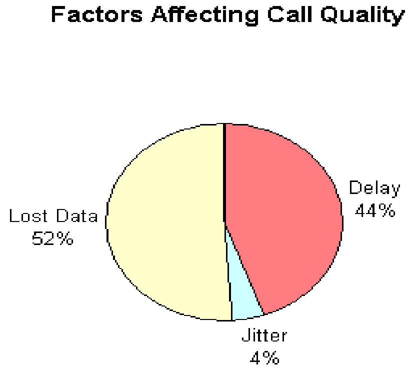

Figure 8. Factors affecting call quality for 4 call/pair.

Figure 9 shows the call quality evaluation by hour. The figure highlights the effect of the data traffic on the quality of IPT traffic (MOS ~ 2.7 in light traffic hours 11 pm – 7 am, MOS ~ 2.1 in Busy Hour (BH) 8am – 9 pm). Measurement showed that the delay has been always excessive (average delay over 600 ms, approaching 740 ms in BH), and the average percentage of lost data has been always above 16%, exceeding 28% during BH.

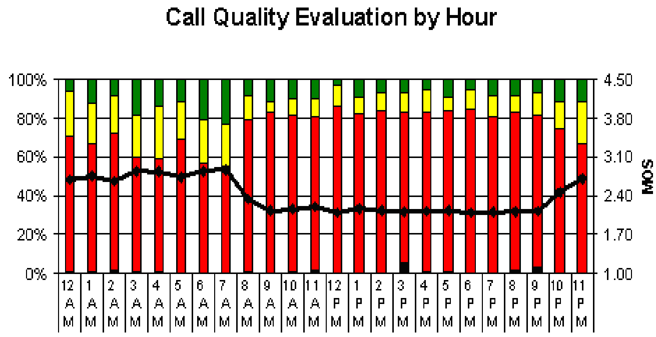

Figure 9. Average call quality by the hour for 4 call/pair.

Similar to the 2-call/pair case, not all calls on WLAN links had the same quality. Figures 10 and 11 show the MOS for the best 5 links and the bottom 5 links, respectively. The Figures show wide variation of call quality between links (from MOS = 4 to 1).

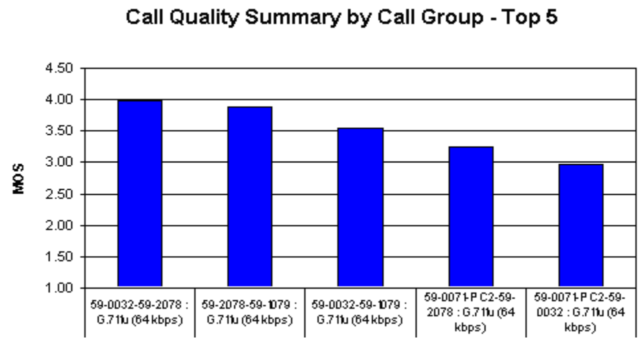

Figure 10. Call quality of the best 5 links for 4 call/pair.

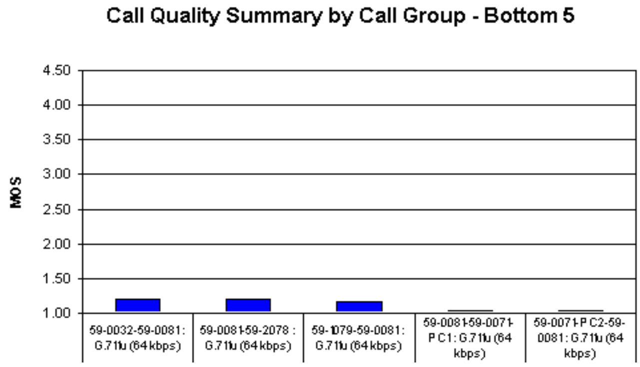

Figure 11. Call quality of the worst 5 links for 4 call/pair.

We examined the performance of the link with highest MOS and that of the lowest MOS. The best link maintained a high MOS in the acceptable and good range (3.7 ~ 4.1). The delay by the hour, Figure 12 fluctuated between 100 – 200 ms, which is within the acceptable / good ranges. However the lost data, Figure 13, is on the high side (4-10%).

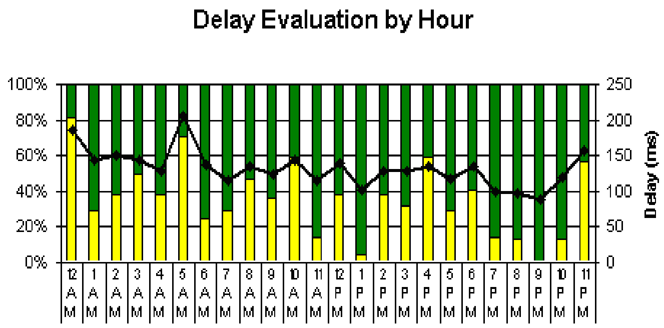

Figure 12. Delay by the hour over the best link for 4 call/pair.

The fact that the delay is in the acceptable range while the lost data is on the high side suggests that the link suffers from intermittent interruptions/disconnections at such high IPT traffic levels. While data traffic may not feel such disconnections, they are noticeable for voice traffic. They are the main reason behind the drop in MOS from the best range of 4.5 to around 4.0.

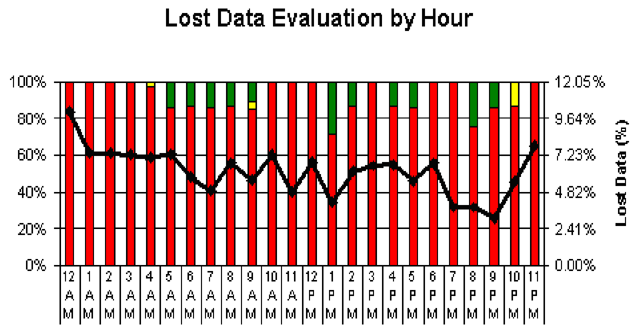

Figure 13. Lost data by the hour over the best link for 4 call/pair.

Figures 14 and 15 show the delay & lost data of the worst link. With delay reaching 1500 ms and packet loss of about 50%, the link is useless and cannot support IPT traffic by any means.

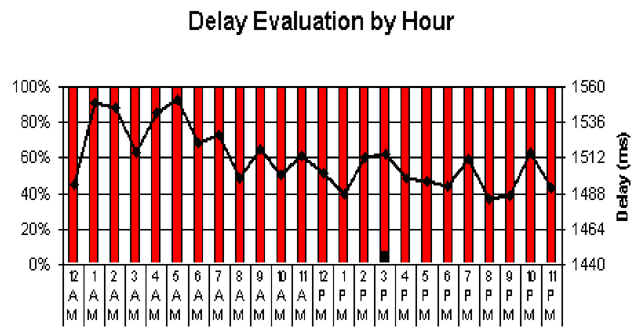

Figure 14. Delay by the hour over the worst link for 4 call/pair.

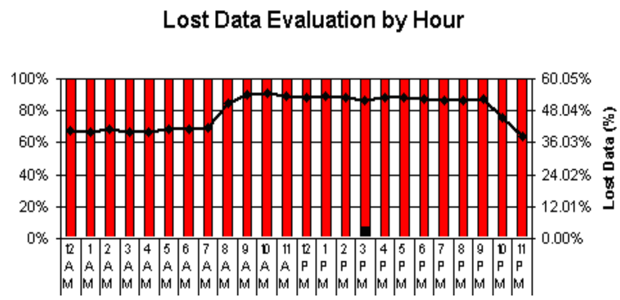

Figure 15. Lost data by the hour over the worst link for 4 call/pair.

V. CONCLUSION

In conventional telephony, the number of simultaneous calls affects the blockage probability but not the quality of the call. This is not the case in IPT. The assessment results of IPT over Wi-Fi network show how the IPT traffic load affects the quality of calls. For the case of one call between any pair of nodes (5 concurrent IPT calls on each link), the performance seems to be pretty good for more than 99% of the calls The average delay was always below 45 ms and the lost data was negligible. Moreover, all links performed equally well.

When the traffic is doubled 33% of the calls became poor. And, unlike the 1-call case where all links were comparatively good, there was a high variation in call quality between different pairs of nodes, some being consistently “good” and others being consistently “poor”. By examining the poor links, we found one node common to all, where the channel between that node and the nearest Access Point is poor.

When the quality is doubled to 4 calls per pair, call quality became poor for 75% of the calls. For this test we started to notice the effect of Busy Hours of data traffic on IPT quality. Being within an academic building, the WLAN network is usually utilized within the hours of 8 am – 9 pm and is hardly utilized after that. The MOS of calls between 9 pm-8 am were found to be 30% above that for calls between 8 am – 9 pm.

The implication of this work is that Wi-Fi networks designed for data traffic can be suitable for reasonable IPT traffic and acceptable performance without modification. However, for heavy traffic or better performance, the routers may have to be configured for QoS enhancement as in IEEE 802.11e. Relying on 802.11e–enabled access points should help diminish the effect of non-IPT traffic and ensure better equity between concurrent IPT calls. For heavier traffic, there may be a need for more AP installations.

In this work we assumed the best quality G711 Codec. The system could be as well configured to G723 (5.3 kbps) which consumes less bandwidth but at the cost of quality (MOS theoretical maximum is 3.69).

 ISBN: 978-1-61208-171-7

ACKNOWLEDGMENT

The author expresses his sincere thanks to Ahmar Shafi and Ajmal Khan for their help in acquiring the tools, preparing the set up and running the experiment. The author would also like to thank the Deanship of Scientific Research at KFUPM for its support. This project was carried out under Fund Programs FT070003 and SB100008.

REFERENCES

[1] A. Vugrinec and S. Tomazic, "IP telephony from a user perspective", 10th Mediterranean Electrotechnical Conference (MELECON) 2000, vol. 1, pp. 344-347.

[2] H. Hsiaosu, F. Martin, M. Denise, and C. Darren, "An Approach to IP Telephony Performance Measurement and Modeling in Government Environments", INET'99 Book, June 1999.

[3] A. El-Sherbini, M. El-Sherif, T. Kamel, and A. Fayez, "A Performance Evaluation of the Integration of Voice and Data in a TCP/IP Local Environment," 36th IEEE Midwest Symposium on Circuits and Systems, Detroit, 1993, pp. 1332-1335.

[4] A. Bearden, L. Denby, B. Karacali, J. Meloche, and D. Stott, "Experiences with evaluating network QoS for IP telephony", 10th IEEE International Workshop on Quality of Service, 2002, pp. 259- 268.

[5] R. Stefic and N. Prib, "Measurement and analysis of users' perception of QoS for IP telephony service", 7th International Conference on Telecommunications (ConTEL) 2003, 11-13 June 2003, vol. 2, pp. 505- 512.

[6] B. Karacali, L. Denby, and J. Meloche, "Scalable network assessment for IP telephony communications", IEEE International Conference on Communications, vol. 3 20-24 June 2004, pp. 1505- 1511.

[7] Performing a VoIP Assessment with Vivinet Assessor, white paper, June 2007. http://download.netiq.com/CMS/WHITEPAPER/NetIQDoingVoIPAssessmentWithVivinetAssessor-June2007.pdf <retrieved: Oct, 2011>.