Investigating the Effects of Encoder Schemes, WFQ & SAD on VoIP QoS

Ajay Shrestha, Khaled M. Elleithy, Syed S. Rizvi
Computer Science and Engineering Department
University of Bridgeport, Bridgeport, CT 06601, U.S.A
{shrestha, elleithy, srizvi}@bridgeport.edu

***Abstract-* Voice Encoder Schemes, Weighted Fair Queuing (WFQ) and Speech Activity Detection (SAD) techniques affect the overall Quality of Service (QoS) of Voice over Internet Protocol (VoIP) services. VoIP is one of the most discussed and rapidly emerging technologies in telecommunication. We are slowly witnessing a change in telephony from Public Switched Telephone Network (PSTN) to IP based VoIP Network. Despite the benefits being enormous, the switch to VoIP hasn't been swift, primarily due to various performance (delay, jitter, packet loss, echo etc.) and security issues plaguing the VoIP telephony network. To achieve minimal QoS for telephony, the voice (packets) must be delivered within 150 ms to 200 ms. This paper presents a performance model to quantifying the influence of VoIP which gives an in-depth understanding of how Voice Encoder Schemes, WFQ and SAD influence VoIP QoS from a theoretical and implementation point of views.**

***Keywords-* VoIP, voice encoder schemes, voice activity detection, speech activity detection**

I. INTRODUCTION

VoIP is an upcoming technology, which will and already has revolutionized the way we communicate through telephony. VoIP is a vast subject and trying to touch all aspects of it is beyond the scope of this paper. Rather, I have focused on particular section of it. The paper is laid out in two parts:

- Literature Research
- Implementation: Encoders Schemes/WFQ Techniques/Speech Activity Detection

II. RELATED WORK

In VoIP Network, an IP device (PC or IP enabled Phone) can make calls through the Internet. Here the call/voice bypasses the traditional switched network PSTN and travels as broken-down fixed-size independent IP packets through the Internet. Unlike in the switched network of PSTN, here each packet finds its own route and are reassembled in the correct order at the destination. The commercial feasibility and benefits from VoIP has fuelled its tremendous growth since the mid 1990s [5] and just the residential VoIP services is estimated to generate $4.1billion in 2010 [6].

VoIP Basics

We know that VoIP stands for Voice over IP (Internet Protocol), and IP indeed, is the most important aspect of VoIP. In fact the voice rides over IP, and this is how VoIP comes into existence. IP belongs to the Internet Protocol suite TCP/IP, which is the de-facto communication protocol for the Internet. In reference to the OSI model, IP equates to the Network Layer. As we know the network layer of OSI reference model is responsible for addressing, address resolution, routing, creating and maintaining routing tables and packet formatting. VoIP uses packet switching to transfer voice over the Internet, just as data is travels over the Internet. There are several technical details that need to be understood, to fully know how voice is carried though the data packet-switched network. The following background will give a good start to understanding it.

Pulse Code Modulation (PCM)

PCM is the way in which our analog voice is converted to a digital format, in the PSTN (Public Switched Telephone Network). Below are the steps that take place in PCM.

- First, the analog waveforms are filtered to remove anything greater than 4000 Hz, to remove any crosstalk from the voice signal. 0-4kHz is considered to be voice band.
- Then the filtered signal is sampled at 8000 times per second. The amplitude of the signal at the time of sampling is a 8-bit code.
- Since we are sampling at 8000 times a second. We have in hand 64000 bps. This is exactly how much the PSTN telephone infrastructure uses: 64kbps [1].

III. PULSE CODE MODULATION (PCM) IN VOIP

PCM is also used in VoIP, but the bit code created is of different length for VoIP due to different voice compression methods (Voice Encoder Schemes) that are used. E.g. in G.729 voice compression technique, samples are taken at 8kbps and at that rate, it creates 10ms voice samples. By calculation each such sample works out to 10Byte (80 bits). Cisco IOS groups together two such samples in one packet. Also a header is attached to every packet [1]. Below is the calculation of total bandwidth required for such operation.

Table I
G.729 FRAMES PER PACKET AND BANDWIDTH

PCM Coding	G. 711	G. 726	G. 728	G. 729	G.723.1
Rate (Streams)	64 Kbps	16, 24, 32, 40 K bps	16 K bps	8 K bps	5.3, 6.3 K bps
5 being best and worst quality	4.1	3.85	3.61	3.7 – 3.92	3.65 – 3.9
User/Type	PSTN, PBXs Networks	PBX Networks	Low Delay Networks	Efficient ADPCM	Multi media service

Table II
PCM CODING G SERIES

G.729 Samples per Frame or Packet	IP/RTP/UDP Header	Bandwidth Consumed	Compression or Packet Delay
Default (2 samples per frame)	40 Bytes	24 K bps	25 ms
Satellite (4 samples per frame)	40 Bytes	16 K bps	45 ms
Low Latency (1 sample per frame)	40 Bytes	40 K bps	15 ms

A. *Bandwidth Calculation for G.729 Encoded Packet*

The following are the standard that we follow in our proposed solutions.

- G.729 samples at 8000 times per second creating a 8kbps code stream.
- 8kbps = 1KBps = 1 Bp(ms) i.e. 1 byte per millisecond
  Therefore every G.729 10 ms voice sample results in = 10Byte.
- By default 2 such voice sample are put in a Packet, so its gives 20Byte
- Thus 20 Bytes/frame worked out to 8kpbs.
- Add 40byte header to the packet. By above formula 40byte would require 16kbps. Therefore the total bandwidth required 8kbps G.729 codec = 24kbps
- There is an initial 5ms look-ahead delay (1^{st} frame), thus Latency = 25ms

Using similar logic, Table I shows various results with variation of the G.729 parameters. Thus it can be said that, the lesser the number of samples per frame, the lesser the compression/packetization delay becomes, but it adversely consumes more bandwidth.

B. *Voice Encoder Schemes (Voice Compression)*

In PSTN 64 kbps PCM is used. Voice compression (codecs) uses several methods to compress the code so that less bandwidth is taken up. Codecs exploit repetitive characteristics in the voice wave to generate a compressed version of the waveform. There are several voice compression techniques, e.g. ADPCM (Adaptive Differential Pulse Code Modulation), CELP (Code Excited Linear Prediction Compression) and MP-MLQ PCM (Multi-Purpose Multi-Level Quantization PCM) [2]. Each of these techniques has their use in specific areas and condition. ITU-T (United Nation's governing body for Telecom Networks and Services) has grouped them in a series of recommendations named G-Series. Table II gives the G Series Coding Standards with its PCM streams rates.

Thus, we can see that the lower the Voice Encoder Scheme's Rate (streams), the quality of voice degrades accordingly.

C. *Voice Activity Detection (VAD) / Speech Activity Detection (SAD)*

Voice Activity Detection is an important part of the VoIP network. In a conversation, only one party talk at any given time, but today's network is made of bi-directional 64000 bps channel. Thus more than 50 percent (when accounting for breaks in speech) of the bandwidth is wasted, as voice is being sampled continuously irrespective if someone is speaking or not. SAD if enabled, can detect the magnitude (in decibels - dB) of speech and will stop voice from being framed if it detects no speech activity. Generally, SAD waits for a hangover time of 200ms for which there is no speech amplitude (decibels) before it stops putting the speech frames in packets. One inherent problem with SAD is that it cannot differentiate between noise and voice. The benefit of SAD is obvious, that the wasted (not used when party not speaking) bandwidth is put to use for something else.

D. *Quality of Service (QoS)*

QoS is the probability of meeting a given traffic contract, e.g. bandwidth and latency required for specific application. QoS can be broken down into CoS (Class of Service) and ToS (Type of Service). ToS is a field in the IP header that occupies 3 bits, enabling eight different types of CoS, 0-7. CoS categories packets into groups 0 through 7, depending on their bandwidth and latency requirements [4].

E. *Bandwidth Usage & Delay*

Bandwidth has always been the major concern with telephony, be it PSTN or VoIP. The voice-encoding scheme (codec) used and the number of voice samples per packet determines how much bandwidth is required for the VoIP network. Table III gives the bandwidth usage and delay for G.711 and G.729 encoder schemes depending on the samples per frame used.

Table III
BANDWIDTH USAGE VS CODECS (AND SAMPLES/FRAME)

Codec & Sampling Rate	G.711 (64Kbps)	G.711 (64Kbps)	G.729 (8Kbps)	G.729 (8Kbps)	G.729 (8Kbps)
Samples per Frame	one 10ms	two 10ms	one 10ms	two 10ms	four 10ms
Bandwidth	112kbps	96kbps	40kbps	24kbps	16kbps
Latency/Delay	10ms	20ms	15ms	25ms	45ms

One noticeable factor in reducing bandwidth usage is the number of samples used per frame, which is inversely proportional to the bandwidth usage. But negatively, the more samples you put in a frame, the more latency becomes. It can also be said from the above table that the bandwidth usage is less when Codec with lower sampling rate (stream) is used. Another conclusion that can be drawn from Table II and III is that, the quality of Voice starts degrading when we move to using lower bit rate stream encoders (codecs), thus the right balance of quality of voice and bandwidth usage needs to be sought, when choosing the Voice Encoder Scheme (codec), by looking at one's particular needs.

F. *Queuing*

As packets approach an interface (Router) for processing, they get queued while the processing is taking place depending on the nature of queuing algorithm used, the are released from the queue. The most simple of the queuing concepts would FIFO (First In First Out), whereby the packet that reaches the interface first gets to go out first. Taking the concept of Queuing to the next level, the packets can be classified into different categories and accordingly sorted into different priority queues. Packets in the higher priority queues pass though the interface faster than packets in the lower priority queues. In general there are three types of queuing method used: FIFO, Priority Queuing and WFQ (Weighted Fair Queuing)[3]. This paper focuses on WFQ.

G. *WFQ (Weighted Fair Queuing)*

WFQ differentiates traffic into several queues to separate flows and assigns equal bandwidth to each flow. This mechanism doesn't let one application (e.g. HTTP) to take over all the bandwidth. WFQ benefits low-volume applications allowing them to transfer faster while high-volume gets proportional amount of bandwidth. A good analogy to understand WFQ would be TDM (time-division multiplexing), whereby bandwidth is equally shared (by time-slots) between several channels or signal-streams. WFQ has an additional dynamic capability to sense absent data streams and then allocates that un-used bandwidth for other flows. In WFQ, streams are prioritized depending on the amount of bandwidth the flow consumes. So, basically the bandwidth is shared fairly by all applications. WFQ analyzes the source/destination address, socket/port number, protocol type and QoS/ToS (Type of Service) to determine flow type to categorize them accordingly. The weighting part of WFQ is determined by the following: IP Precedence, RSVP, IP RTP Priority, IP RTP Reserve, FECN (Frame Relay forward explicit congestion notification) and BECN (backward explicit congestion notification) [1]. FECN and BECN bits signify congestions, so such traffic is transmitted less. Values are assigned to each of the above factors, and the bandwidth allocated according to those values.

IV. PROPOSED IMPLEMENTATION AND SIMULATION RESULTS

Now that the literature research has covered the VoIP topics relevant to my research work, lets go into the implementation part. I've used OPNET IT Guru, the most widely used Network Simulation Tool in the academic arena, to perform my thesis implementation. Even outside the Academia, many small and large corporations use it alike. Even, Department of Defense (DoD) uses it for advanced Network Simulations.

A. *Simulation Tool and Specs*

For the sake of simulation, the OPNET IT Guru Academic Edition (9.1.A, build in 1996) is used. As a system specs, we use Windows XP Home Edition, service pack 2 with typically a small network using a mesh and bus topologies.

B. *Effects of Encoder Schemes and Speech Activity Detection on Load and Throughput*

This is the 1st part of my VoIP implementation. It demonstrates the effects of various voice encoder schemes on the load and throughput. As I mentioned earlier in the report, Encoder Schemes significantly affects the total bandwidth used by the link. Complex codec algorithms are used to reduce the sampling rate streams, which in turn reduce the bandwidth utilized.

Here, say a caller 0 and caller 1 from an office makes call to the another remote office. Caller 0 uses G.711 encoding on outgoing/incoming voice signal. Caller 1 on the other hand uses G.729 encoding on outgoing/incoming voice signal. Fig.1 shows the traffic (bytes/sec) received for both G.711 vs. G.729.

From Fig. 1, we see that traffic received for the Calling Party, is higher when G.711 (64kbps bit stream) encoder is used than when G.729 (8kbps bit stream) encoder is used. Thus we can conclude that when a higher bit encoder scheme

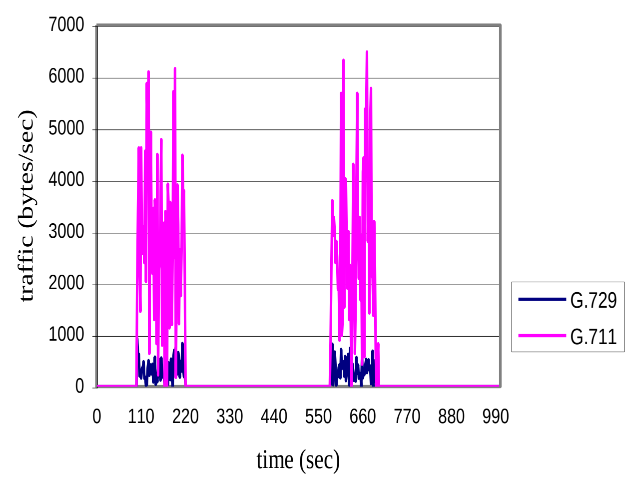


Fig. 1. Traffic Received: G.729 vs G.711

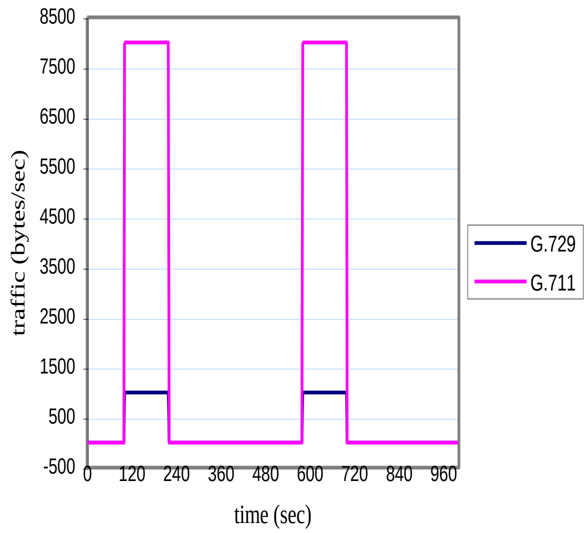


Fig. 2. Traffic Sent: G.729 vs G.711

is used, more voice traffic is generated and more bandwidth is required. Fig. 2 shows traffic sent for the different encoder schemes.

Similarly, from the above graph, we see that traffic sent by the Calling Party, is higher when G.711 (64kbps bit stream) encoder is used than when G.729 (8kbps bit stream) encoder is used. Thus like in the earlier case, we can conclude that when a higher bit encoder scheme is used, more voice traffic is generated and more bandwidth required, and thus adversely affects the VoIP QoS.

C. Effects of Speech Activity Detection on Bandwidth

Using the same setup as earlier implementation, the traffic generated for incoming and outgoing calls by the voice application is configured to be the same. Keeping that intact, now the traffic received is configured to use SAD. Fig.3 shows the traffic variation when SAD is enable and disabled.

In addition, Fig. 3 shows the Traffic Sent (without Speech Activity Detection) is higher than the Traffic Received (with Speech Activity Detection enabled) for the G.729 Application. Thus it can be concluded that enabling Speech Activity Detection lessens the traffic and frees up some bandwidth, and positively enhances VoIP QoS.

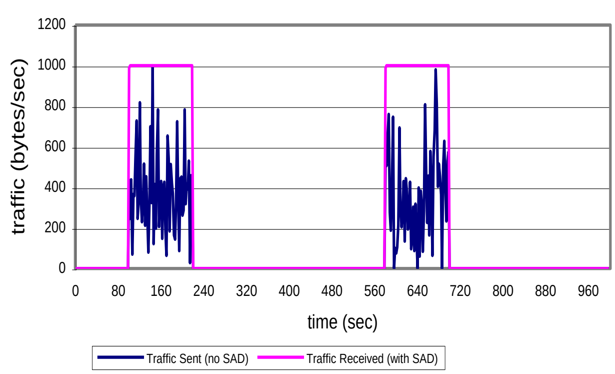


Fig. 3. G.729 Traffic Variation with & without SAD

D. Delay Analysis With WFQ

Now, that I've shown the effects of Voice Encoders and SAD on throughput and load, the next step my thesis implementation is to show the comparison of delay incurred, when WFQ is used with Voice on varying ToS (Type of Service) Applications. Here two nodes compete to send voice traffic through the same link between Router 1 & 2.

In the initial case, for both nodes the ToS (Type of Service) is set "best-effort" i.e. first-come first-serve basis. Fig. 4 displays the delay incurred by voice traffic from both these nodes.

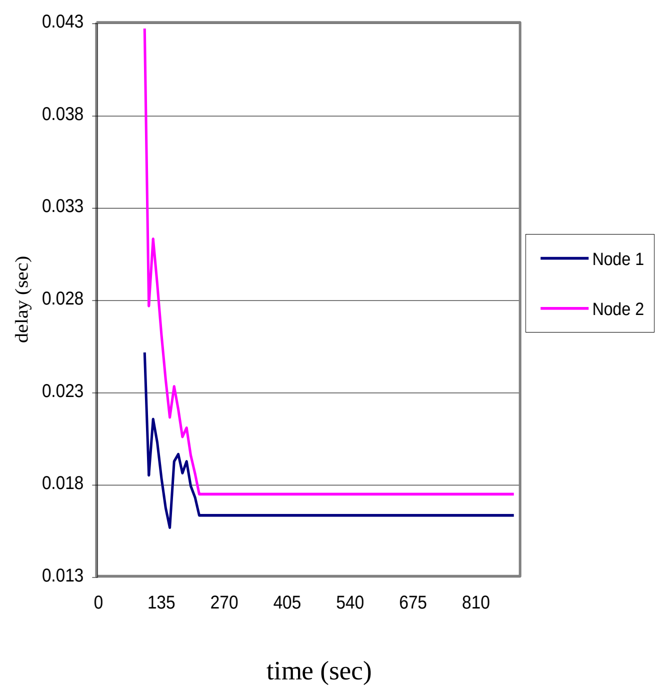
It should be noted in Fig. 4 that both nodes have almost same End-to-End delay when both are using the same ToS. The initial delay is slightly more for Node 1, primarily due to the variation in the traffic sent/received among the nodes.

In the second scenario Node 1 is set to Hi_Priority i.e. the ToS is set to "Interactive Voice" and Node 2 is set to Low_Priority i.e. the ToS is set to 'excellent-effort'. Now the Router 1 and Router 2 will use the WFQ setting configured in the IP QoS attribute to prioritize the traffic from the two nodes.

Fig. 5 demonstrates the results of the proposed implementation. It should be noted in Fig. 5 that different ToS were used for the two traffic. 'Interactive Voice' has higher priority than the "Excellent-Effort" ToS value. Thus, HI_Priority_Traffic experience virtually no delay compared to the delay experience by LOW_Priority_Traffic. Thus we can clearly see that using WFQ significantly reduces the end-to-end delay and enhances VoIP QoS.

E. Speech Activity Detection (SAD) and Bandwidth/Link capacity Utilization

This implementation shows another aspect of how VoIP QoS is influenced by SAD. Speech Activity Detection greatly



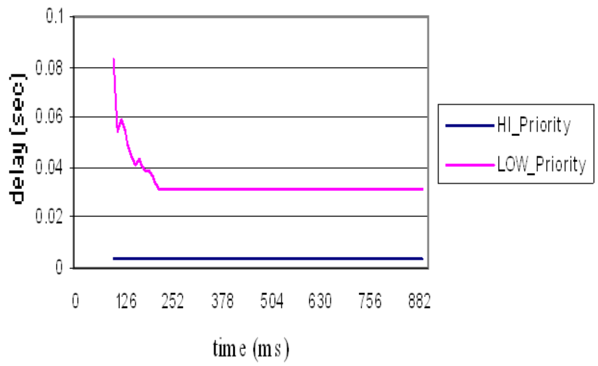


Fig. 5. End-to-End Delay with Different ToS

helps make efficient use of available bandwidth. The 64kbps bi-directional voice channel's bandwidth is wasted more than 50% of the time due to break in conversations. So, SAD senses these breaks in conversations by keeping track of magnitude of speech (decibels) and uses the bandwidth for other traffic during the breaks in conversations.

In this implementation there are two calling nodes (say: Voice_src1 & Voice_src2) and two called nodes (say: Voice_dest1 & Voice_dest2). Voice_src1 and Voice_dest1 is one conversation pair and they use the G.711 voice encoder. Voice_src2 and Voice_dest2 is another conversation pair and they use G.729 voice encoder. The calling nodes are connected to router1 and called nodes are connected to router2, and router1 and router2 are in turn linked together. In the first simulation both conversation pairs use Speech Activity Detection (SAD) (also called silence suppression). In the second simulation, SAD is disabled. This will let us know the effects of SAD on bandwidth utilization. Bandwidth utilization will be shown as a total effect of both conversation

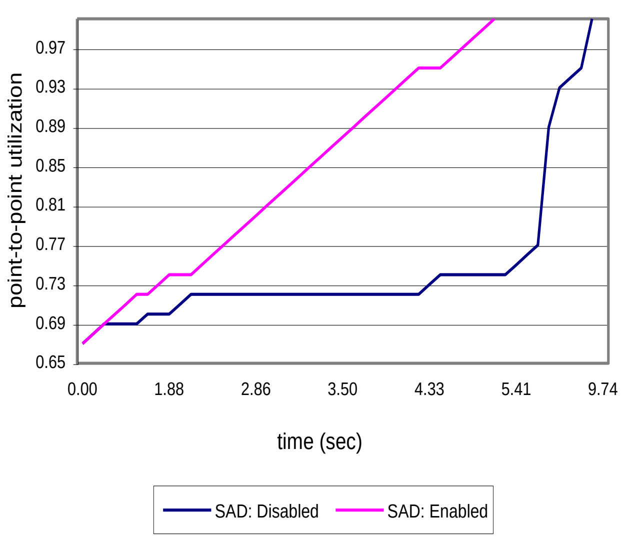


Fig. 6. SAD & Bandwidth Utilization

pairs on the common link between router1 and router2. In addition, Fig. 6 shows how SAD affects the link bandwidth utilization. it can be seen in Fig. 6 that the point-to-point bandwidth utilization was optimally utilized and spanned a shorter period of time when SAD was enabled. So, it can be inferred that more calls can be made more efficiently utilizing the bandwidth, when SAD or silence suppression is used. Enabling Speech Activity Detection (SAD) detects notifies when either caller or called party is not talking (break in conversation for more that given amount of time), then the SAD will free up the bandwidth for other traffic. Thus SAD is an efficient way to utilize bandwidth and enhance VoIP QoS.

V. Conclusion

In this paper, we have explained how Voice Encoder Schemes, WFQ and Speech Activity Detection techniques affects the overall VoIP QoS in terms of Bandwidth and Delay incurred, from both an theoretical and implementation point of view. VoIP offers great benefits over the traditionally PSTN telephony, but it needs to achieve minimal QoS before it can completely replace the existing PSTN telephony. As more research and development work is being done on VoIP, it will only make is more viable for greater use and implementation in both residential and commercial telephony. Voice Encoder Schemes, WFQ and SAD, as described in this paper, are few of the major factors influencing the VoIP QoS, and this paper has in short but successfully show how it affects the VoIP QoS. The OPNET tool has been of tremendous assistance in visualizing the effects of the above factors on VoIP.

References

[1] Voice over IP Fundamentals, Tenth Printing April 2005, ISBN: 1-57870-168-6, Cisco Press, Jonathan Davidson, James Peters.

[2] Availability of Artificial Voice for Measuring Objective QoS of CELP CODECs and Acoustic Echo Cancellers, Paper by: Nobuhiko Kitawaki, Feng Wei, Takeshi Yamada, Futoshi Asano, University of Tsukuba, AIST, Japan, URL: http://wireless.feld.cvut.cz/mesaqin2002/full06.pdf

[3] QoS Queuing Techniques, Microsoft Corporation, Article ID : 233039, Last Review : October 30, 2006, Revision : 3.1, Resource URL: http://support.microsoft.com/kb/233039 as retrieved on Nov 18th, 2006 15:50:53 EST.

[4] Quality of Service Networking, Cisco Systems, Inc: URL: http://www.ciscosystems.com/univercd/cc/td/doc/cisintwk/ito_doc/qos.htm as retrieved on Nov 19th, 2006 13:22:03 EST.

[5] The Business of VoIP, Term Paper for 15.912, Technology Strategy, MIT Sloan School of Management, Jay Liu, Bassam Hajhamad, MBA Class of 2005, May 2005.

[6] World VoIP News, November 2006, Telecom Portal, Webiste, URL: http://www.voipproviderslist.com/world-voip-news/news/residential-voip-to-generate--41-bln-in-2010.html as retrieved on Nov 26th, 2006 17:33:30 EST.

[7] OPNET IT Guru Academic Edition, (9.1.A (Build 1996). URL: http://opnet.com/services/university/itguru_academic_edition.html Lab Manuals, as retrieved on Nov 28th, 2006 23:38:22 EST.