On the Deployment of VoIP in Ethernet Networks: Methodology and Case Study

Khaled Salah[**]
Department of Information and Computer Science
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Email: salah@kfupm.edu.sa

Abstract

Deploying IP telephony or voice over IP (VoIP) is a major and challenging task for data network researchers and designers. This paper outlines guidelines and a step-by-step methodology on how VoIP can be deployed successfully. The methodology can be used to assess the support and readiness of an existing network. Prior to the purchase and deployment of VoIP equipment, the methodology predicts the number of VoIP calls that can be sustained by an existing network while satisfying QoS requirements of all network services and leaving adequate capacity for future growth. As a case study, we apply the methodology steps on a typical network of a small enterprise. We utilize both analysis and simulation to investigate throughput and delay bounds. Our analysis is based on queueing theory, and OPNET is used for simulation. Results obtained from analysis and simulation are in line and give a close match. In addition, the paper discusses many design and engineering issues. These issues include characteristics of VoIP traffic and QoS requirements, VoIP flow and call distribution, defining future growth capacity, and measurement and impact of background traffic.

Keywords: Network Design, Network Management, VoIP, Performance Evaluation, Analysis, Simulation, OPNET

1 Introduction

These days a massive deployment of VoIP is taking place over data networks. Most of these networks are Ethernet-based and running IP protocol[1]. Many network managers are finding it very attractive and cost effective to merge and unify voice and data networks into one. It is easier to run, manage, and maintain. However, one has to keep in mind that IP networks are best-effort networks that were designed for non-real time applications. On the other hand, VoIP requires timely packet delivery with low latency, jitter, packet loss, and sufficient bandwidth. To achieve this goal, an efficient deployment of VoIP must ensure these real-time traffic requirements can be guaranteed over new or existing IP networks.

** Corresponding Author: Prof. K. Salah, PO Box 5066, ICS Department, KFUPM, Dhahran 31261, Saudi Arabia

When deploying a new network service such as VoIP over existing network, many network architects, managers, planners, designers, and engineers are faced with common strategic, and sometimes challenging, questions. What are the QoS requirements for VoIP? How will the new VoIP load impact the QoS for currently running network services and applications? Will my existing network support VoIP and satisfy the standardized QoS requirements? If so, how many VoIP calls can the network support before upgrading prematurely any part of the existing network hardware?

These challenging questions have led to the development of some commercial tools for testing the performance of multimedia applications in data networks. A list of the available commercial tools that support VoIP is listed in. For the most part, these tools use two common approaches in assessing the deployment of VoIP into the existing network. One approach is based on first performing network measurements and then predicting the network readiness for supporting VoIP. The prediction of the network readiness is based on assessing the health of network elements. The second approach is based on injecting real VoIP traffic into existing network and measuring the resulting delay, jitter, and loss.

Other than the cost associated with the commercial tools, none of the commercial tools offer a comprehensive approach for successful VoIP deployment. In particular, none gives any prediction for the total number of calls that can be supported by the network taking into account important design and engineering factors. These factors include VoIP flow and call distribution, future growth capacity, performance thresholds, impact of VoIP on existing network services and applications, and impact background traffic on VoIP. This paper attempts to address those important factors and layout a comprehensive methodology for a successful deployment of any multimedia application such as VoIP and videoconferencing. However, the paper focuses on VoIP as the new service of interest to be deployed. The paper also contains many useful engineering and design guidelines, and discusses many practical issues pertaining to the deployment of VoIP. These issues include characteristics of VoIP traffic and QoS requirements, VoIP flow and call distribution, defining future growth capacity, and measurement and impact of background traffic. As a case study, we illustrate how our approach and guidelines can be applied to a typical network of a small enterprise.

The rest of the paper is organized as follows. Section 2 presents a typical network topology of a small enterprise to be used as a case study for deploying VoIP. Section 3 outlines practical eight-step methodology to deploy successfully VoIP in data networks. Each step is described in considerable detail. Section 4 describes important design and engineering decisions to be made based on the analytic and simulation studies. Section 5 concludes the study and identifies future work.

2 Existing Network

Figure 1 illustrates a typical network topology for a small enterprise residing in a high-rise building. The network shown is realistic and used as a case study only; however, our work presented in this paper can be adopted *easily* for larger and general networks by following the same principles, guidelines, and concepts laid out in this paper. The network is Ethernet-based and has two Layer-2 Ethernet switches connected by a router[2]. The router is Cisco 2621, and the switches are 3Com Superstack 3300. Switch 1 connects Floor 1 and Floor 2 and two servers; while Switch 2 connects Floor 3 and four servers[3]. Each floor LAN is basically a shared Ethernet connecting employee PCs with workgroup and printer servers. The network makes use of VLANs in order to isolate broadcast and multicast traffic. A total of five LANs exist. All VLANs are port based. Switch 1 is configured such that it has three VLANs. VLAN1 includes the database and file servers. VLAN2 includes Floor 1. VLAN3 includes Floor2. On the other hand, Switch 2 is configured to have two VLANs. VLAN4 includes the servers for E-mail, HTTP, Web & cache proxy, and firewall. VLAN5 includes Floor 3. All the links are switched Ethernet 100Mbps full duplex except for the links for Floor 1, Floor 2, and Floor 3 which are shared Ethernet 100Mbps half duplex.

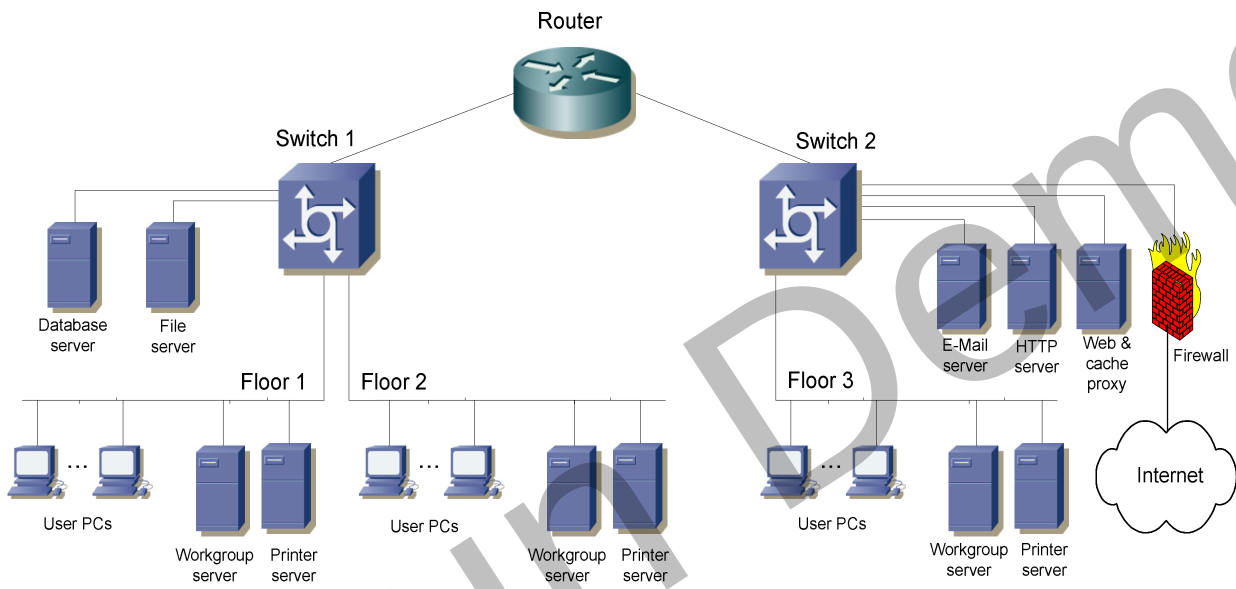

Figure 1. Logical diagram of a small enterprise

3 Step-by-Step Methodology

Figure 2 shows a flowchart of a methodology of eight steps for a successful VoIP deployment. The first four steps are independent and can be performed in parallel. Before embarking on the analysis and simulation study, in Step 6 and Step 7, Step 5 must be carried out which requires any early and necessary redimensioning or modifications to the existing network. As shown, both Step 6 and Step 7 can be done in parallel. The final step is pilot deployment[5].

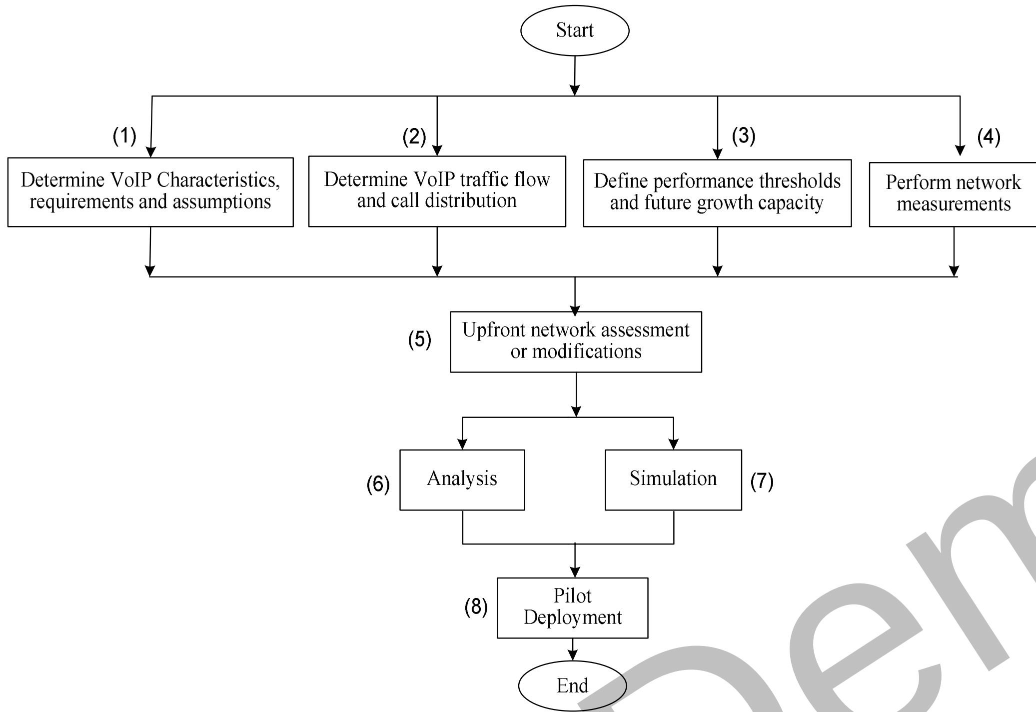

Figure 2. Flowchart illustrating methodology steps

3.1 VoIP Traffic Characteristics, Requirements, and Assumptions

For introducing a new network service such as VoIP, one has to characterize first the nature of its traffic, QoS requirements, and any additional components or devices. For simplicity, we assume a point-to-point conversation for all VoIP calls with no call conferencing. For deploying VoIP, a *gatekeeper* or *CallManager* node has to be added to the network[4] . The *gatekeeper* node handles signaling for establishing, terminating, and authorizing connections of all VoIP calls. Also a VoIP *gateway* is required to handle external calls. A VoIP *gateway* is responsible for converting VoIP calls to/from the Public Switched Telephone Network (PSTN). As an engineering and design issue, the placement of these nodes in the network becomes crucial. We will tackle this issue in design step 5. Other hardware requirements include a VoIP client terminal, which can be a separate VoIP device, i.e., IP phones, or a typical PC or workstation that is VoIP-enabled. A VoIP-enabled workstation runs VoIP software such as IP SoftPhones [6-8].

Figure 3 identifies the end-to-end VoIP components from sender to receiver . The first component is the *encoder* which periodically samples the original voice signal and assigns a fixed number of bits to each sample, creating a constant bit rate stream. The traditional sample-based encoder G.711 uses Pulse Code Modulation (PCM) to generate 8-bit samples every 0.125 ms, leading to a data rate of 64 kbps. The *packetizer* follows the *encoder* and encapsulates a certain number of speech samples into packets and adds the RTP, UDP, IP, and Ethernet headers. The voice packets travel through the data network. An important component at the receiving end, is the *playback buffer* whose purpose is to absorb variations or jitter in delay and provide a smooth playout. Then packets are delivered to the *depacketizer* and eventually to the *decoder* which reconstructs the original voice signal.

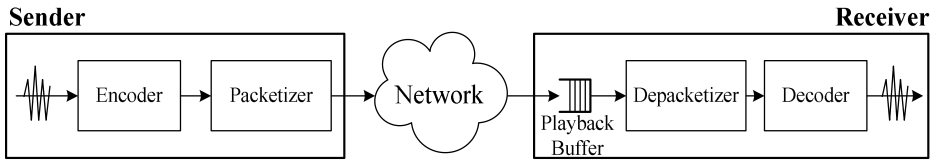

Figure 3. VoIP end-to-end components

We will follow the widely-adopted recommendations of H.323, G.711, and G.714 standards for VoIP QoS requirements. Table 1 compares some commonly-used ITU-T standard codecs and the amount of one-way delay that they impose. To account for upper limits and to meet desirable quality requirement according to ITU recommendation P.800 , we will adopt G.711u codec standards for the required delay and bandwidth. G.711u yields around 4.4 MOS rating. MOS, *Mean Opinion Score*, is a commonly used VoIP performance metric given in a scale of 1 to 5, with 5 is the best. However, with little compromise to quality, it is possible to implement different ITU-T codecs that yield much less required bandwidth per call and relatively a bit higher, but acceptable, end-to-end delay. This can be accomplished by applying compression, silence suppression, packet loss concealment, queue management techniques, and encapsulating more than one voice packet into a single Ethernet frame .

Table 1. Common ITU-T codecs and their defaults

Codec	Data rate (kbps)	Datagram size (ms)	A/D Conversion delay (ms)	Combined bandwidth (bi-directional) (kbps)
G.711u	64.0	20	1.0	180.80
G.711a	64.0	20	1.0	180.80
G.729	8.0	20	25.0	68.80
G.723.1 (MPMLQ)	6.3	30	67.5	47.80
G.723.1 (ACELP)	5.3	30	67.5	45.80

3.1.1 End-to-End Delay for a Single Voice Packet

Figure 3 illustrates the sources of delay for a typical voice packet. The end-to-end delay is sometimes referred to by M2E or Mouth-to-Ear delay. G.714 imposes a maximum total one-way packet delay of 150ms end-to-end for VoIP applications. In a delay of up to 200ms was considered to be acceptable. We can break this delay down into at least three different contributing components, which are as follows (i) encoding, compression, and packetization delay at the sender (ii) propagation, transmission and queuing delay in the network and (iii) buffering, decompression, depacketization, decoding, and playback delay at the receiver.

3.1.2 Bandwidth for a Single Call

The required bandwidth for a single call, one direction, is 64 kbps. G.711 codec samples 20ms of voice per packet. Therefore, 50 such packets need to be transmitted per second. Each packet contains 160 voice samples in order to give 8000 samples per second. Each packet is sent in one Ethernet frame. With every packet of size 160 bytes, headers of additional protocol layers are added. These headers include RTP + UDP + IP + Ethernet with preamble of sizes 12 + 8 + 20 + 26, respectively. Therefore, a total of 226 bytes, or 1808 bits, needs to be transmitted 50 times per second, or 90.4 kbps, in one direction. For both directions, the required bandwidth for a single call is 100 pps or 180.8 kbps assuming a symmetric flow.

3.1.3 Other Assumptions

Throughout our analysis and work, we assume voice calls are symmetric and no voice conferencing is implemented. We also ignore the signaling traffic generated by the *gatekeeper*. We base our analysis and design on the worst-case scenario for VoIP call traffic. The signaling traffic involving the *gatekeeper* is mostly generated prior to the establishment of the voice call and when the call is finished. This traffic is relatively small compared to the actual voice call traffic. In general, the *gatekeeper* generates no or very limited signaling traffic throughout the duration of the VoIP call for an already established on-going call.

In this paper, we will implement no QoS mechanisms that can enhance the quality of packet delivery in IP networks. A myriad of QoS standards are available and can be enabled for network elements. QoS standards may include IEEE 802.1p/Q, the IETF's RSVP, and DiffServ. Analysis of implementation cost, complexity, management, and benefit must be weighed carefully before adopting such QoS standards. These standards can be recommended when the cost for upgrading some network elements are high and the network resources are scarce and heavily loaded.

3.2 VoIP Traffic Flow and Call Distribution

Knowing the current telephone call usage or volume of the enterprise is an important step for a successful VoIP deployment. Before embarking on further analysis or planning phases for a VoIP deployment, collecting statistics about of the present call volume and profiles is essential. Sources of such information are organization's PBX, telephone records and bills. Key characteristics of existing calls can include the number of calls, number of concurrent calls, time, duration, etc. It is important to determine the locations of the call endpoints, i.e., the sources and destinations, as well as their corresponding path or flow. This will aid in identifying the call distribution and the calls made internally or externally. Call distribution must include percentage of calls within and outside of a floor, building, department, or organization. As a good capacity planning measure, it is recommended to base the VoIP call distribution on the busy hour traffic of phone calls for the busiest day of a week or a month. This will ensure support of the calls at all times with high QoS for all VoIP calls. When such current statistics are combined with the projected extra calls, we can predict the worst-case VoIP traffic load to be introduced to the existing network. Figure 4 describes the call distribution for the enterprise under study based on the worst busy hour and the projected future growth of VoIP calls. In the figure, the call distribution is described as a probability tree. It is also possible to describe it as a probability matrix.

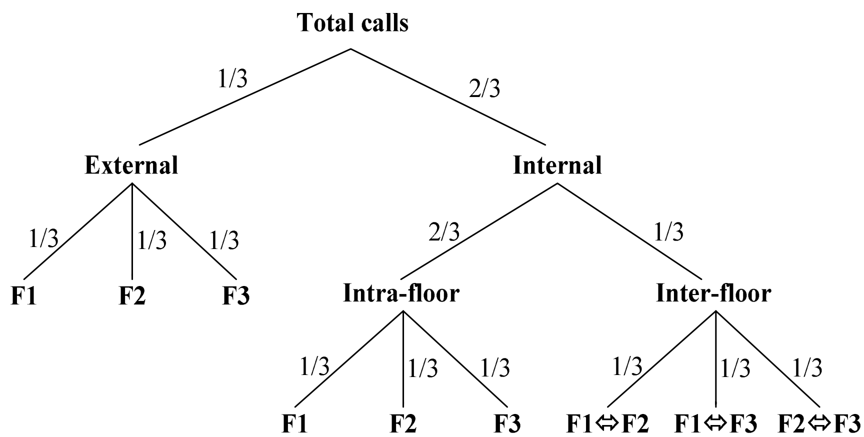

Figure 4. Probability tree describing the VoIP call distribution

Some important observations can be made about the voice traffic flow for inter-floor and external calls. For all these type of calls, the voice traffic has to be always routed through the router. This is so because Switch 1 and Switch 2 are layer 2 switches with VLANs configuration. One can observe that the traffic flow for inter-floor calls between Floor 1 and Floor 2 imposes twice the load on Switch 1, as the traffic has to pass through the switch to the router and back to the switch again. Similarly, Switch 2 experiences twice the load for external calls from/to Floor 3.

3.3 Define Performance Thresholds and Growth Capacity

In this step we define the network performance thresholds or operational points for a number of important key network elements. These thresholds are to be considered when deploying the new service. The benefit is twofold. First, the requirements of the new service to be deployed are satisfied. Second, adding the new service leaves the network healthy and susceptible to future growth.

Two important performance criteria are to be taken into account. First is the maximum tolerable end-to-end delay; and second is the utilization bounds or thresholds of network resources. The maximum tolerable end-to-end delay is determined by the most sensitive application to run on the network. In our case, it is 150ms end-to-end for VoIP[6]. It is imperative to note that if the network has certain delay-sensitive applications, the delay for these applications should be monitored, when introducing VoIP traffic, such that they do not exceed their required maximum values. As for the utilization bounds for network resources, such bounds or thresholds are determined by factors such as current utilization, future plans, and foreseen growth of the network. Proper resource and capacity planning is crucial. Savvy network engineers must deploy new services with scalability in mind, and ascertain that the network will yield acceptable performance under heavy and peak loads, with no packet loss[7]. VoIP requires almost no packet loss. In literature 0.1% to 5% packet loss was generally asserted . However, in the required VoIP packet loss was conservatively suggested to be less than 10^{-5}. A more practical packet loss, based on experimentation, of below 1% was required in[8]. Hence, it is extremely important not to utilize fully the network resources. As rule-of-thumb guideline for switched fast full-duplex Ethernet, the average utilization limit of links should be 190%, and for switched shared fast Ethernet, the average limit of links should be 85% .

The projected growth in users, network services, business, etc. must be all taken into consideration to extrapolate the required growth capacity or the future growth factor. In our study we will ascertain that 25% of the available network capacity is reserved for future growth and expansion. For simplicity, we will apply this evenly to all network resources of the router, switches, and switched-Ethernet links. However, keep in mind this percentage in practice can be variable for each network resource and may depend on the current utilization and the required growth capacity. In our methodology, the reservation of this utilization of network resources is done upfront, before deploying the new service, and only the left-over capacity is used for investigating the network support of the new service to be deployed.

3.4 Perform Network measurements

In order to characterize the existing network traffic load, utilization, and flow, network measurements have to be performed[9]. This is a crucial step as it can potentially affect results to be used in analytical study and simulation. There are a number of tools available commercially and non-commercially to perform network measurements. Popular open-source measurement tools include MRTG, STG, SNMPUtil, and GetIF . A few examples of popular commercially measurement tools include HP OpenView, Cisco Netflow, Lucent VitalSuite, Patrol DashBoard, Omegon NetAlly, Avaya ExamiNet, NetIQ Vivinet Assessor, etc.

Network measurements must be performed for network elements such as routers, switches, and links. Numerous types of measurements and statistics can be obtained using measurement tools. As a minimum, traffic rates in bps (bits per second) and pps (packets per second) must be measured for links directly connected to routers and switches. To get adequate assessment, network measurements have to be taken over a long period of time, at least 24-hour period. Sometimes it is desirable to take measurements over several days or a week.

Table 2. Worst-case network measurements

Link	Bit rate (Mbps)	Packet rate (pps)	Utilization
Router ⇔ Switch 1	9.44	812	9.44 %
Router ⇔ Switch 2	9.99	869	9.99 %
Switch 1 ⇔ Floor 1	3.05	283	6.1 %
Switch 1 ⇔ Floor 2	3.19	268	6.38 %
Switch 1 ⇔ File Server	1.89	153	1.89 %
Switch 1 ⇔ DB Server	2.19	172	2.19 %
Switch 2 ⇔ Floor 3	3.73	312	7.46 %
Switch 2 ⇔ Email Server	2.12	191	2.12 %
Switch 2 ⇔ HTTP Server	1.86	161	1.86 %
Switch 2 ⇔ Firewall	2.11	180	2.11 %
Switch 2 ⇔ Proxy	1.97	176	1.97 %

One has to consider the worst-case scenario for network load or utilization in order to ensure good QoS at all times including peak hours. The peak hour is different from one network to another and it depends totally on the nature of business and the services provided by the network. Table 2 shows a summary of peak-hour utilization for traffic of links in both directions connected to the router and the two switches of the network topology of Figure 1. These measured results will be used in our analysis and simulation study.

5 Conclusion

The paper outlined a step-by-step methodology on how VoIP can be deployed successfully. The methodology can help network researchers and designers to determine quickly and easily how well VoIP will perform on a network prior to deployment. Prior to the purchase and deployment of VoIP equipment, it is possible to predict the number of VoIP calls that can be sustained by the network while satisfying QoS requirements of all existing and new network services and leaving enough capacity for future growth. In addition, the paper discussed many design and engineering issues pertaining to the deployment of VoIP. These issues include characteristics of VoIP traffic and QoS requirements, VoIP flow and call distribution, defining future growth capacity, and measurement and impact of background traffic.

We considered a case study of deploying VoIP in a small enterprise network. We applied the methodology and guidelines outlined in this paper on such a network. We utilized both analysis and simulation to determine the number of VoIP calls that can be supported for such a network. From results of analysis and simulation, it is apparent that both results are in line and give a close match. Based on the analytic approach, a total of 313 calls can be supported. Based on the simulation approach, a total of 306 calls can be supported There is only a difference of 7 calls. The difference can be contributed to the degree of accuracy between the analytic approach and OPNET simulation. Our analytic approach is an approximation. Also, the difference is linked to the way the OPNET Modeler adds the distribution of the calls. It was found that external and inter-floor calls are added before intra-floor calls. In anyways, to be safe and conservative, one can consider the minimum number of calls of the two approaches.

In this paper, only peer-to-peer voice calls were considered. As a future work, one can consider implementing important VoIP options such as VoIP conferencing and messaging. Also as a future work, one can look into assessing the network support and readiness of deploying other popular real-time network services such multimedia, video, and web conferencing. As a near-term work, we are in the process of developing a GUI-based analytical design tool that automates the analytical approach presented in this paper in order to find the maximum number of VoIP calls that can supported by any given generic network topology.

Acknowledgements

The author acknowledges the support of King Fahd University of Petroleum and Minerals in the development of this work. Special thanks go to previous ICS Department graduate students (Mr. A. Alkhoraidly, Mr. M. Turki, Mr. R. Alghanmi and Mr. A. Alsanad). The technical work of Mr. A. Alkhoraidly has been greatly appreciated. The author also acknowledges the anonymous reviewers for their valuable comments on the earlier versions of this article.

References

[1] M. Bearden, L. Denby, B. Karacali, J. Meloche, and D. T. Stott, “Assessing Network Readiness for IP Telephony,” Proceedings of IEEE International Conference on Communications, ICC02, vol.4, 2002, pp. 2568-2572

[2] B. Karacali, L. Denby, and J. Melche, “Scalable Network Assessment for IP Telephony,” Proceedings of IEEE International Conference on Communications (ICC04), Paris, June 2004, pp. 1505-1511.

[3] Goode B, “Voice over Internet Protocol (VoIP),” Proceedings of IEEE, vol. 90, no. 9, Sept. 2002, pp. 1495-1517.

[4] P. Mehta and S. Udani, “Voice over IP”, *IEEE Potentials Magazine*, vol. 20, no. 4, October 2001, pp. 36-40.

[5] W. Jiang and H. Schulzrinne, “Towards Junking the PBX: Deploying IP Telephony,” Proceedings of ACM 11 International Workshop on Network and Operating System Support for Digital Audio and Video, Port Jefferson, NY, June 2001, pp. 177-185

[6] W. Jiang, K. Koguchi, and H. Schulzrinne, “QoS Evaluation of VoIP End-Points,” Proceedings of IEEE International Conference on Communications, ICC’03, Anchorage, May 2003, pp. 1917-1921

[7] Avaya Inc., “Avaya IP voice quality network requirements,” http://www1.avaya.com/enterprise/whitepapers, 2001.

[8] A. Markopoulou, F. Tobagi, M. Karam, “Assessing the quality of voice communications over internet backbones”, *IEEE/ACM Transaction on Networking*, vol. 11, no. 5, 2003, pp. 747-760

[9] Recommendation G.711, “Pulse Code Modulation (PCM) of Voice Frequencies,” ITU, November 1988