

An Analytical Tool to Assess Readiness of Existing Networks for Deploying IP Telephony

K. Salah M. Almashari

*Department of Information and Computer Science
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Email: {salah, meshal}@ccse.kfupm.edu.sa*

Abstract

Deploying IP telephony or voice over IP (VoIP) is a major and challenging task. This paper describes an analytical approach and tool to assess the readiness of existing IP networks for the deployment of VoIP. The analytical approach utilizes queueing network analysis and investigates two key performance bounds for VoIP: delay and bandwidth. The analytical tool is GUI-based and has an engine that automates the analytical approach. The engine determines the number of VoIP calls that can be sustained by a given generic network while satisfying VoIP QoS requirements and leaving adequate capacity for future growth. As a case study, the paper illustrates how the analytical tool can assess the readiness to deploy VoIP for a typical network of a small enterprise.

1 Introduction

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? 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 [1,2]. 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 gives any prediction for the total number of calls that can be supported by an existing network when considering important design and engineering factors. These factors include VoIP flow and call distribution, future growth capacity, performance thresholds, and impact background traffic on VoIP. This paper attempts to address those important factors utilizing an analytical approach based on queueing networks. The approach utilizes queueing network analysis and investigates two key performance bounds for VoIP: delay and bandwidth. The paper also describes a GUI-based analytical tool that implements the analytical approach to compute the maximum number of VoIP calls that can be sustained by a given generic network while satisfying VoIP QoS requirements and leaving adequate capacity for future growth. As a case study, the paper illustrates how the presented analytical tool can assess the readiness to deploy VoIP for a typical network of a small enterprise.

The rest of the paper is organized as follows. Section 2 presents an analytical approach to assess the readiness of IP telephony. Section 3 describes a GUI-based analytical tool that has an engine that implements our proposed analytical approach. Section 3 also gives an example of a typical network topology of a small enterprise to be used as a case study for deploying VoIP. Finally, Section 4 concludes the study and identifies future work.

2 Analytical Approach

VoIP is bounded by two important metrics. First is the available bandwidth. Second is the end-to-end delay. The actual number of VoIP calls that the network can sustain and support is bounded by those two metrics. Depending on the network under study, either the available bandwidth or delay can be the key dominant factor in determining the number of calls that can be supported.

Our analytical approach considers important factors as background traffic, traffic flow, call distribution and growth factor. For background traffic, network measurements must be performed to determine the traffic rates in bps (bits per second) and pps (packets per second) for links directly connected to the router and switches. Traffic flow has to do with the path that a session travels through. Session distribution has to do with the percentage of sessions to be established within and outside of a floor, building, or department. In order to allow for future growth, we will consider a 25% growth factor for all network elements including router, switches, and links.

2.1 Bandwidth Bottleneck Analysis

The required bandwidth for a voice call, one direction, is 50 pps or 90.4 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.

Bandwidth bottleneck analysis is an important step to identify the network element, whether it is a node or a link, that puts a limit on how many VoIP calls can be supported by the existing network. For any path that has N network nodes and links, the bottleneck network element is the node or link that has the minimum available bandwidth. This minimum available bandwidth can be defined as follows

$$A = \min_{i=1, \dots, N} A_i,$$

and

$$A_i = (1 - u_i)C_i,$$

where C_i is the capacity of network element i and u_i is its current utilization. The capacity C_i is the maximum possible transfer or processing rate. Therefore the theoretical maximum number of calls that can be supported by a network element E_i can be expressed in terms of A_i as

$$MaxCalls_i = \frac{A_i(1 - growth_i)}{CallBW}, \quad (1)$$

where $growth_i$ is the growth factor of network element E_i , and takes a value from 0 to 1. $CallBW$ is the VoIP bandwidth for a single call imposed on E_i . In order to find the bottleneck network element that limits the total number of VoIP calls, one has to compute the maximum number of calls

that can be supported by each network element, as in equation (1), and the percentage of VoIP traffic flow passing by this element. The percentage of VoIP traffic flow for E_i , denoted as $flow_i$, can be found by examining the distribution of the calls. The total number of VoIP calls that can be supported by a network can be expressed as

$$TotalCallsSupported = \min_{i=1, \dots, N} \left(\frac{MaxCalls_i}{flow_i} \right). \quad (2)$$

2.2 Delay Analysis

In order to achieve a natural voice conversation, the end-to-end upper bound delay (sometimes termed latency) for a voice packet should be kept to minimal. Essentially, such a delay can be broken into at least three contributing components, which are as follows (i) voice sampling or frame grabbing, 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. According to recommendations by ITU [3], when delays are less than 150 ms, most interactive applications, both speech and non-speech, will experience essentially transparent interactivity. For voice, the end-to-end delay is sometimes referred to by M2E or Mouth-to-Ear delay.

Therefore, We must always ascertain that the worst-case end-to-end delay for all the calls must be less than 150 ms. It should be kept in mind that our goal is to determine the network capacity for VoIP, i.e. the maximum number of calls that existing network can support while maintaining VoIP QoS. This can be done by adding calls incrementally to the network while monitoring the threshold or bound for VoIP delay. When the end-to-end delay, including network delay, becomes larger than 150 ms, the maximum number of calls can then be known.

There are three sources of delay for a VoIP stream: sender, network, and receiver. An expression is given in [4] to compute the end-to-end delay D for a VoIP flow in one direction from sender to receiver.

$$D = D_{pack} + \sum_{h \in Path} (T_h + Q_h + P_h) + D_{play},$$

where D_{pack} is the delay due to packetization at the source. At the source, there is also D_{enc} and $D_{process}$. D_{enc} is the encoder delay of converting A/D signal into samples. $D_{process}$ is the PC of IP phone processing that includes encapsulation. In G.711, D_{pack} and D_{enc} , are 20 ms and 1ms, respectively. Hence, it is appropriate for our analysis to have a fixed delay of 25 ms being introduced at the

source, assuming worst case situation. D_{play} is the playback delay at the receiver, including jitter buffer delay. The jitter delay is at most 2 packets, i.e. 40ms. If the receiver's delay of $D_{process}$ is added, we obtain a total fixed delay of 45 ms at the receiver. $T_h + Q_h + P_h$ is the sum of delays incurred in the packet network due to transmission, queuing, and propagation going through each hop h in the path from the sender to the receiver. The propagation delay P_h is typically ignored for traffic within a LAN, but not for a WAN. For transmission delay T_h and queuing delay Q_h we apply queueing theory. Hence the delay to be introduced by the network, expressed as $\sum_{h \in Path} (T_h + Q_h)$, should not exceed (150 – 25 – 45) or 80 ms.

We utilize queueing analysis to approximate and determine the maximum number of calls that the existing network can support while maintaining a delay of less than 80ms. In order to find the network delay, we utilize the principles of Jackson theorem for analyzing queueing networks. In particular, we use the approximation method of analyzing queueing networks by decomposition discussed in [5]. In this method, the arrival rate is assumed to be Poisson and the service times of network elements are exponentially distributed. Analysis by decomposition is summarized in first isolating the queueing network into subsystems, e.g., single queueing node. Next, analyzing each subsystem separately, considering its own network surroundings of arrivals and departures. Then, finding the average delay for each individual queueing subsystem. And finally, aggregating all the delays of queueing subsystems to find the average total end-to-end network delay. The queueing models of network elements for link, switch, and router are presented in [6].

In order to determine the maximum number of calls that can be supported by an existing network while maintaining VoIP delay constraint, we developed the following algorithm that basically determines network capacity in terms of VoIP calls. This algorithm is essentially part of the analytical tool engine for calculating number of calls based on delay bound. Calls are added iteratively until the worst-case network delay of 80 ms has reached. The algorithm can be described in the following steps:

- (i) Initially, no calls are introduced and the only traffic in the network is the background traffic.
- (ii) A new call is added, according to the call distribution.
- (iii) For each network element, $\lambda = \lambda_{VoIP} + \lambda_{bg}$ is computed. (λ_{VoIP} is the total added new traffic from a single VoIP in pps, and λ_{bg} is the background traffic in pps). λ_{bg} is known for each element; however, λ_{VoIP} can get affected by

introducing a new call depending on the call traffic flow, i.e. whether or not the new call flow passes through the network element.

- (iv) For each network element, the average delay of a VoIP packet is computed.
- (v) The end-to-end delay is computed by summing up all the delays of step (iv) encountered for each possible VoIP flow. This includes all external and internal flows, with internal flows consisting of intra-floor and inter-floor.
- (vi) The maximum network delay of all possible flows is determined. If the maximum network delay is less than 80 ms, then the maximum number of calls has not been reached. Therefore a new call can be added, and hence go to step (ii).
- (vii) If not, the maximum delay has been reached. Therefore the number of VoIP calls bounded by the delay is one less than the last call addition.

3 Analytical Tool

In order to make use of the analytical approach presented in Section 2 and to ease the implementation of the algorithms of throughput and delay analysis for a given generic network of interest, an analytical tool¹ has been developed in C#. The tool is GUI-based and its engine implements the analytical approach. The tool takes as an input several parameters and configurations. As shown in Figure 1, several tab options are available. The tab of "Nodes & Floors" allows the configuration of nodes (switches, routers, or servers) and their respective capacity and location. "Links" allows for interconnecting network nodes to form the network topology. Each link capacity and background traffic can be configured. "Call Distribution" allows configuration of call distribution per flow in percentage. The call flows between floors can be configured using "Path". Other settings and configurations such as percentage of growth, bandwidth, packet size, host and network latencies can be configured under "VoIP Settings" tab. The results of maximum calls to be supported based on throughput and delay bounds are reported separately under "Analysis" option. The tool has useful pull-down menus to save and restore configurations from a file. Also under the "Tools" option one can view and verify the network topology.

¹ The tool is freeware and can be downloaded from http://www.ccse.kfupm.edu.sa/~salah/VoIP_analytical_tool.zip.

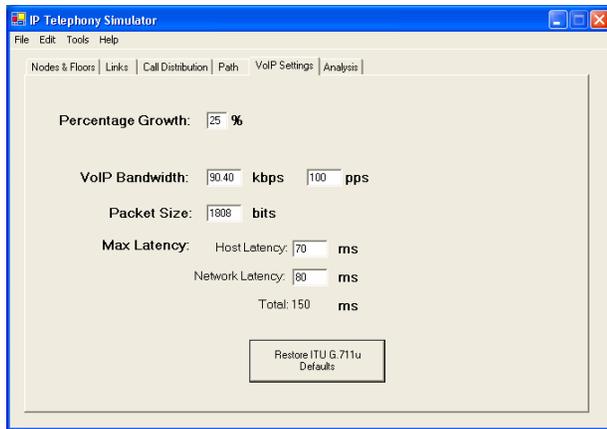


Figure 1. Analytical tool options

In order to illustrate the tool's options further we give an example. Figure 2 illustrates a typical network topology of a small enterprise residing in a high-rise building. This topology was also reported in [6] and used here for comparison purposes. We will also use the same parameter values for background traffic, traffic flow, and call distribution as those reported in [6]. According to [7] and [8], the capacity C_i for the router or the switch, is 25,000pps and 1.3M pps, respectively. All links are switched full-duplex Fast (100Mbps). The network shows the added VoIP nodes of an H.323 *gatekeeper* and *gateway*. The *gatekeeper* node handles signaling for establishing, terminating, and authorizing connections of voice calls. The VoIP *gateway* is responsible for converting VoIP calls to/from the Public Switched Telephone Network (PSTN). 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.

Figure 3 shows the corresponding network diagram produced by the analytical tool. It can be noted that there was no need to have separate nodes for floor's PCs and Servers. However, their aggregate background traffic was taken into account. 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 only generated prior to the establishment of the voice call and when the call is finished. This traffic is relatively limited and small compared to the actual voice call traffic. In general, the *gatekeeper* generates no signaling traffic throughout the duration of the VoIP call for an already established on-going call. Figure 4 (Left) shows the number of calls that can be supported for selected nodes and links. It is shown that the router is the bottleneck with 314 voice calls to support. Figure 4 (Right) shows that network can support 307 voice calls with less than 80 ms (actually shown 74 ms) of network delay. So one can conclude that network delay (for this particular network topology) is the more dominant parameter in determining the

number of voice calls to be supported. Therefore 307 voice calls can be supported.

4 Concluding Remarks

The paper presented an analytical approach and a GUI-based analytical tool to assess network readiness and support for VoIP. The approach 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. The results obtained by the analytical tool are very much in line with reported results of previous work reported in [6]. The results of [6] were based on analytical results obtained using Matlab and OPNET simulation. 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.

Acknowledgements

The author acknowledges the support of King Fahd University of Petroleum and Minerals in the development of this work. Special thanks go to Mr. A. Alkhoraidly and Mr. K. El-Badawi for their valuable discussions, comments.

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] Recommendation G.114, "One-Way Transmission Time," ITU, 1996.
- [4] M. Karam and F. Tobagi, "Analysis of delay and delay jitter of voice traffic in the Internet," *Computer Networks Magazine*, vol. 40, no. 6, December 2002, pp. 711-726 (2002)
- [5] K. M. Chandy and C. H. Sauer, "Approximate methods for analyzing queueing network models of computing systems," *Journal of ACM Computing Surveys*, vol. 10, no. 3, September 1978, pp. 281-317.
- [6] Salah, K., "On the Deployment of VoIP in Ethernet Networks: Methodology and Case Study," *International Journal of Computer Communications*, Elsevier Science, (In Press).
- [7] Cisco Systems Inc., "Cisco 2621 Modular Access Router Security Policy," 2001, http://www.cisco.com/univercd/cc/td/doc/product/access/acs_mod/cis2600/secure/2621rect.pdf
- [8] 3Com, "3Com Networking Product Guide," April 2004, <http://www.3com.co.kr/products/pdf/productguide.pdf>

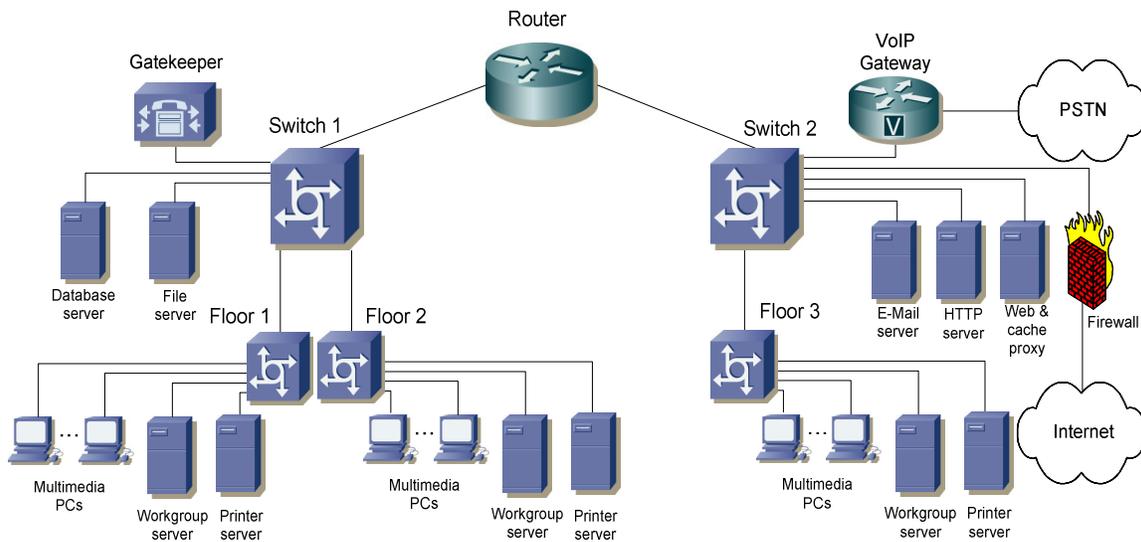


Figure 2. Network topology with added VoIP Components

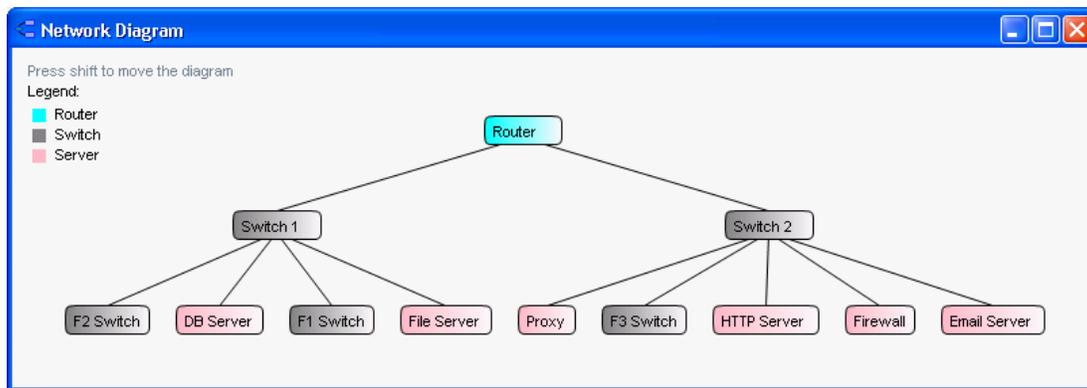


Figure 3. Corresponding network diagram generated by analytical tool

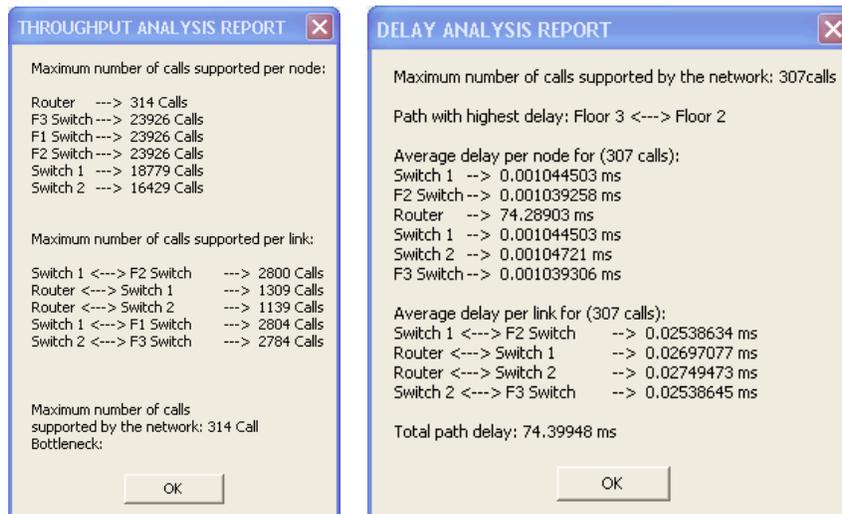


Figure 4. Throughput and delay analysis report