An OPNET-based Simulation Approach for Deploying VoIP

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

Abstract. These days a massive deployment of VoIP is taking place over IP networks. VoIP deployment is a challenging task for network researchers and engineers. This paper presents a detailed simulation approach for deploying VoIP successfully. The simulation uses the OPNET network simulator. Recently OPNET has gained a considerable popularity in both academia and industry, but there is no formal or known approach or methodology in how OPNET can be used to assess the support and readiness of an existing network in deploying VoIP. Our approach and work presented in this paper predict, prior to the purchase and deployment of VoIP equipment, 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 simulation approach on a typical network of a small enterprise. The paper presents a detailed description of simulation models for network topology and elements using OPNET. The paper describes modeling and representation of background and VoIP traffic, as well as various simulation configurations. Moreover, the paper discusses 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.

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

1   Introduction

IP telephony or VoIP deployment over data networks is gaining popularity these days. 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, 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?

Some commercial tools have been developed in order to address some of these challenging questions. Most existing commercial tools are focused on testing the performance of multimedia applications in data networks.

[**] Correspondence to: Prof. K. Salah, PO Box 5066, ICS Department, KFUPM, Dhahran 31261, Saudi Arabia, phone: +96638604493 fax: +96638602174

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.

None of the commercial tools offer a comprehensive approach or methodology 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 presents an approach based on simulation. In particular, the approach is based on using OPNET network simulator. OPNET contains a vast amount of models of commercially available network elements, and has various real-life network configuration capabilities. This makes the simulation of real-life network environment close to reality. Other features of OPNET include GUI interface, comprehensive library of network protocols and models, source code for all models, graphical results and statistics, etc. More importantly, OPNET has gained considerable popularity in academia as it is being offered free of charge to academic institutions. That has given OPNET an edge over DES NS2 in both market place and academia.

Despite OPNET popularity, there is no formal or known approach or methodology in how OPNET can be used to assess the support and readiness of an existing network in deploying VoIP. Our approach and methodology presented in this paper predict, prior to the purchase and deployment of VoIP equipment, 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. Such an approach can be applicable to the 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 contains many useful engineering and design guidelines. 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 describes practical steps to be initially performed prior to simulation. Section 4 presents the OPNET-based simulation approach for deploying successfully VoIP. The section details and summarizes the OPNET model and results. Section 5 describes important design and engineering decisions to be made based on the simulation study. Section 6 concludes the study and identifies future work.

2   A Typical Network Topolgy

A typical network topology of a small enterprise residing in a high-rise building is shown in Figure 1. 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. The network is Ethernet-based and has two Layer-2 Ethernet switches connected by a router. 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. 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.
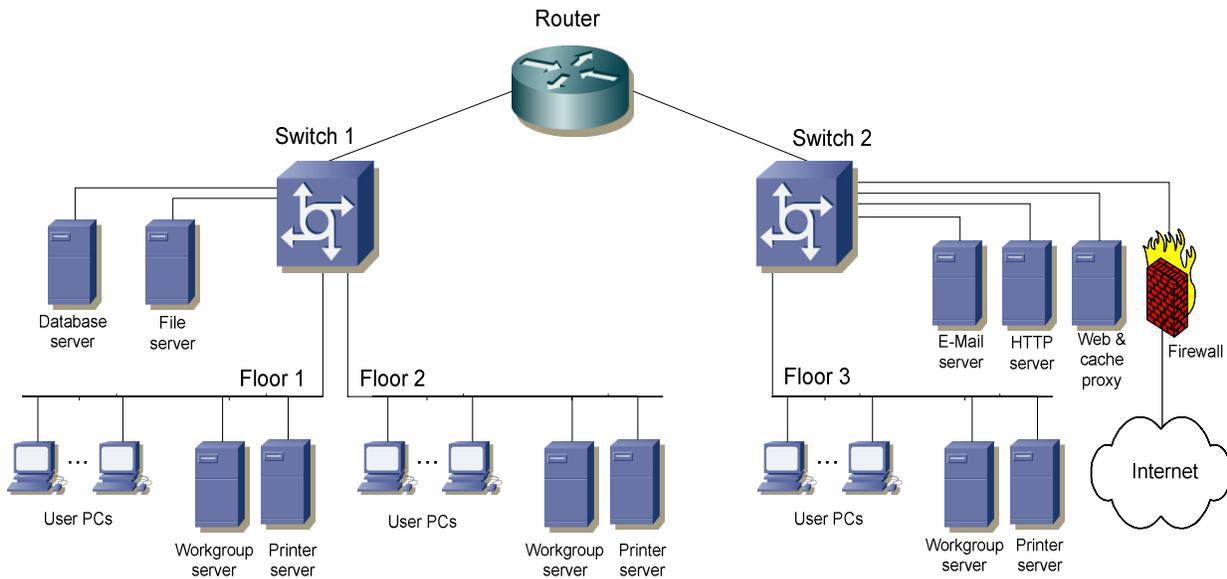

Figure 1. Logical diagram of a small enterprise

3   Initial VoIP Study and Design Phase

At this initial phase, a number of issues pertaining to the deployment of VoIP needs to be investigated and tackled. These issues feed into simulation for determining the actual number of VoIP calls that can be supported by the network.

3.1   Network measurements

Measurements have to be performed to determine the existing traffic load, utilization, and flow. This is a crucial step as it can potentially affect results to be used in our 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 [3]. 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. At 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. 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 1 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 simulation study.

Table 1. Peak-rate traffic 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 %

It is important to recognize that the network traffic measured shown in Table 1 is symmetric, i.e., the data rate for the egress and ingress directions are the same. In reality, this may not be true all the time, but for the network shown in Figure 1, it is assumed to be the same. The background traffic of egress and ingress are key factors that affect the end-to-end delay of VoIP. While the egress and ingress data rate for a singe VoIP call is symmetric, the end-to-end delay may not be symmetric and it all depends on the path that each call flow is taking and more importantly on the egress and ingress traffic presented at network elements of links, routes, and switches. In the simulation study of Section 4.2.1, we will address how symmetric and asymmetric background traffic can be generated in OPNET.

3.2 VoIP Traffic Flow and Call Distribution

Collecting statistics about of the present call volume and profiles is essential for deploying VoIP successfully. 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 2 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.
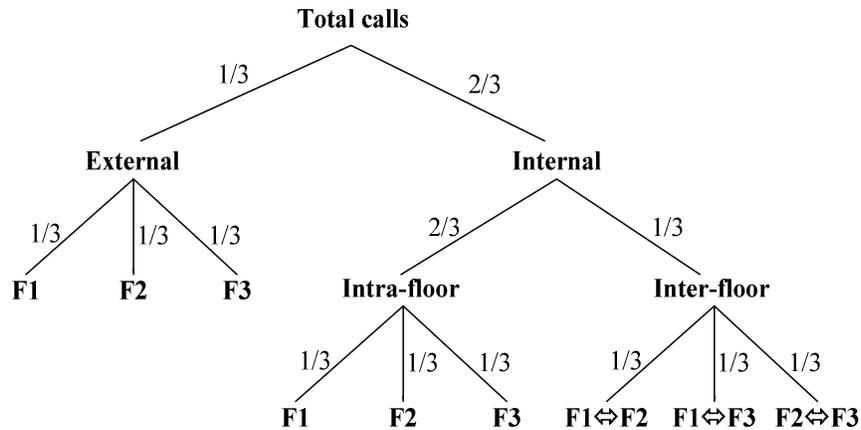

Figure 2. Probability tree describing the VoIP call distribution

3.3 VoIP Traffic Characteristics, Requirements, and Assumptions

When 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 [4,5,6]. 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 Section 3.5. 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 [7-9].

G.714 imposes a maximum total one-way packet delay of 150ms end-to-end for VoIP applications [10]. In [11] a delay of up to 200ms was considered to be acceptable. We can break this delay down into at least three different contributing components [12], 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. Encoding refers to the conversion of A/D signal into samples. Packetization refers to encapsulation of samples into IP packet. In G.711, the delay introduced by (i) for encoding and packetization are 1 ms and 20 ms, respectively. Hence with compression, it is appropriate to have a fixed delay of 25 ms being introduced at the source. The delay

5

incurred by (iii) at the receiver includes jitter buffer delay. The jitter delay is at most 2 packets, i.e. 40ms. When the receiver's processing delay of depacketization and decoding is added, we obtain a total fixed delay of 45 ms at the receiver. Hence the delay for (ii) to be introduced by the network should not exceed (150 – 25 – 45) or 80 ms.

The required bandwidth for a single VoIP 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 for a symmetric flow.

Throughout this paper, we assume point-to-point voice calls and no voice conferencing is implemented. We also ignore the signaling traffic generated by the *gatekeeper*. We base our study 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 [3].

3.4   Growth Capacity

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 simulation approach, 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.

3.5   Upfront Network Assessment and Modifications

In this design step we assess the existing network and determine, based on the existing traffic load and the requirements of the new service to be deployed, if any immediate modifications are necessary. Immediate modifications to the network may include adding and placing new servers or devices, upgrading PCs, and re-dimensioning heavily utilized links. As a good upgrade rule, topology changes need to be kept to minimum and should not be made unless it is necessary and justifiable. Over-engineering the network and premature upgrades are costly and considered as poor design practices.

Based on the existing traffic load described in Section 3.1, all the links connecting the router and the switches and links connecting the servers and the switches are underutilized. If any of the links were heavily utilized,

e.g. 30-50%, the network engineer should decide to re-dimension the link to 1-Gbps link at this stage. As for shared links of Floor 1, Floor 2, and Floor 3, the replacement or re-dimensioning of these links must be decided on carefully. At first, it looks costly effective not to replace the shared-Ethernet LAN for each floor with a switched LAN. However shared Ethernet scales poorly due to Ethernet-level collisions. More importantly, shared Ethernet offers zero QoS and are not recommended for real-time and delay-sensitive applications as it introduces excessive and variable latency under heavy loads and when subjected to intense bursty traffic [13]. In order to consistently maintain the VoIP QoS, a switched fast full-duplex Ethernet LAN becomes necessary. As a general rule of thumb VoIP should never be installed over networks with shared Ethernet links.

Based on the hardware requirement for deploying VoIP described in Section 3.3, two new nodes have to be added to the existing network: a VoIP *gateway* and a *gatekeeper*. Please refer to Figure 3. As a network design issue, an appropriate node placement is required for these two nodes. Since most of the users reside on Floor 1 and Floor 2 and connected directly to Switch 1, connecting the *gatekeeper* to Switch 1 is practical in order to keep the traffic local. For the VoIP *gateway*, we connect it to Switch 2 in order to balance the projected load on both switches. Also it is more reliable and fault-tolerant not to connect both nodes to the same switch in order to eliminate problems that stem from a single point of failure. For example, if Switch 2 fails, only external calls to/from the network will be affected. It is proper to include the gatekeeper to be a member of VLAN1 of Switch 1 which includes the database and file servers. This isolates the gatekeeper from multicast and broadcast traffic of Floor 1 and Floor 2. In addition, the *gatekeeper* can access locally the database and file servers to record and log phone calls. On the other hand, we create a separate VLAN for the *gateway* in order to isolate the *gateway* from multicast and broadcast traffic of Floor 3 and the servers of switch 2. Therefore, the network has now a total of six VLANs.
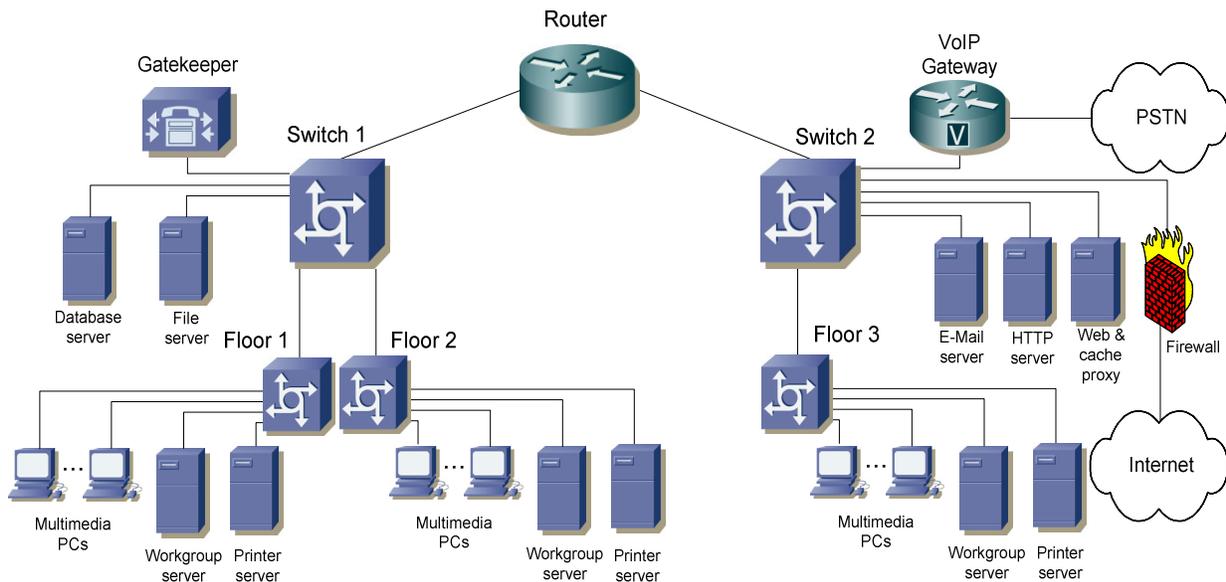

Figure 3. Network topology with VoIP Components

7

4   Simulation Approach

Our simulation approach uses the popular the popular MIL3's OPNET Modeler simulation package[1], Release 8.0.C [14]. OPNET Modeler contains a vast amount of models of commercially available network elements, and has various real-life network configuration capabilities. This makes the simulation of real-life network environment close to reality. Other features of OPNET include GUI interface, comprehensive library of network protocols and models, source code for all models, graphical results and statistics, etc. More importantly, OPNET has gained considerable popularity in academia as it is being offered free of charge to academic institutions. That has given OPNET an edge over DES NS2 in both market place and academia. This section describes in detail the simulation model, traffic model, various simulation configurations, as well as the simulation results.

4.1   Modeling the Network

The simulation model for the existing network under study is illustrated in Figure 4. The simulation model of the organization network, for the most part, is an exact replica of the real network. In OPNET, many vendor-specific models are included in the pre-defined component libraries. However, the specific devices that we need, i.e. Cisco 2621 router and 3Com superstack3 3300 Ethernet switch, were not available. Hence, generic router and Ethernet switch models have been used to represent the router and the switches in our network. VoIP *gateway* is modeled as an Ethernet workstation since we are interested only in collecting statistics inside the enterprise network. The enterprise servers are modeled as Ethernet servers. All network elements have been connected using a 100 Base-T links. Figure 4 shows the described topology. As discussed in Section 3.3, the *gatekeeper* signaling traffic is ignored, and hence modeling such and element and its traffic is not taken into account as we base our study on the worst case situation.

Floor LANs have been modeled as subnets that enclose an Ethernet switch and three designated Ethernet workstations used to model the activities of the LAN users, as shown in Figure 5. One of these workstations generates the background traffic of the floor while the other two act as parties in VoIP sessions. For example, the Ethernet workstations for Floor 1 are labeled as `F1_C2`, `F1_C2`, and `F1_C3`. `F1_C1` is a source for sending VoIP calls. `F1_C2` is a sink for receiving VoIP calls. `F1_C3` is a sink and source of background traffic. Please note that the model of floor LANs does not represent precisely the floor multimedia PCs or IP phones. However, building a model with such exact floor network configurations will make simulation very tedious and almost manual. This is because it requires for each time a new VoIP call (or a group of calls) is added to perform two tasks: first adding individual PCs with individual profiles and settings, and then running the simulation. This has to be repeated manually and results have to be examined after each simulation run. As will be noted later, the simulation run can take up to 15 hours. As will be demonstrated, our simulation approach is an automated one, as the simulation is configured to automatically keep generating three calls every three seconds. Our approach and model has no impact on the performance of internal nodes and links inside the core network. The only impact is only on the utilization and queueing delay at the edge links connecting the floor subnet workstations to their intra-floor switch. The health of these edge links must be checked

---

[1] OPNET Modeler was provided under the OPNET University Programs

individually at the end of the simulation run. As will be demonstrated later in simulation results, such an impact was small and negligible.
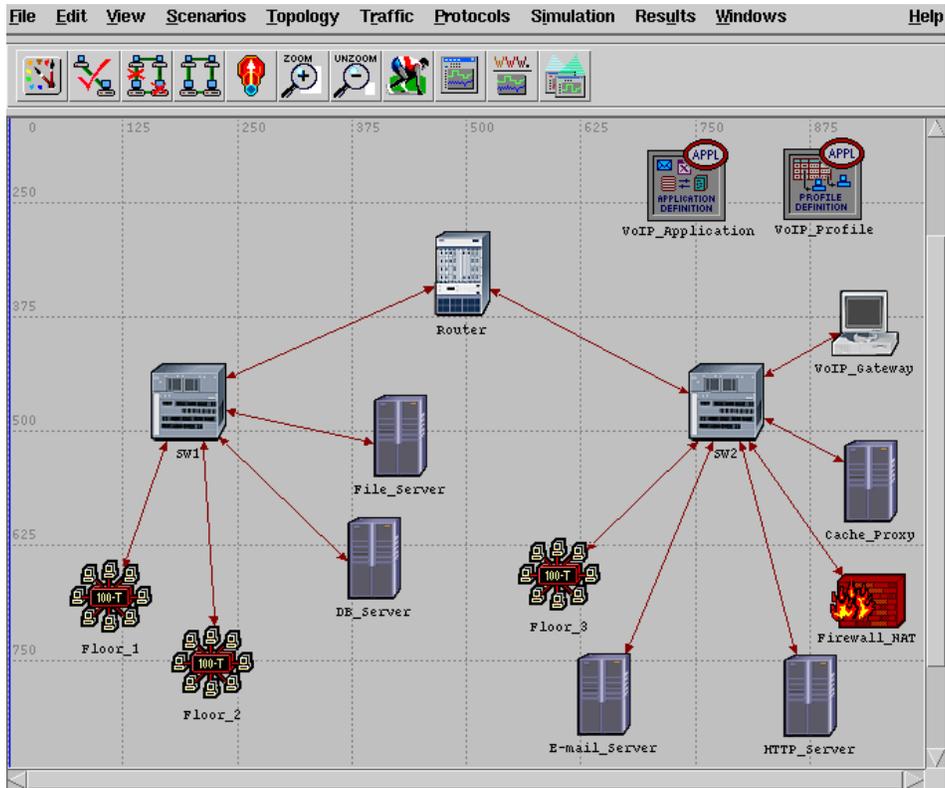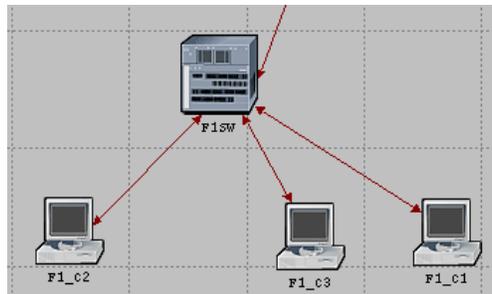

Figure 4. The organization network topology

Figure 5. The floor subnet model

4.1.1   Component Configuration and Growth Factor

In order to develop a realistic model, the default configurations of some network components need to be changed. The following describes the configuration changes applied to each of the components of the network. In all of our configurations, we reserve 25% of the capacity of each network component for future growth.
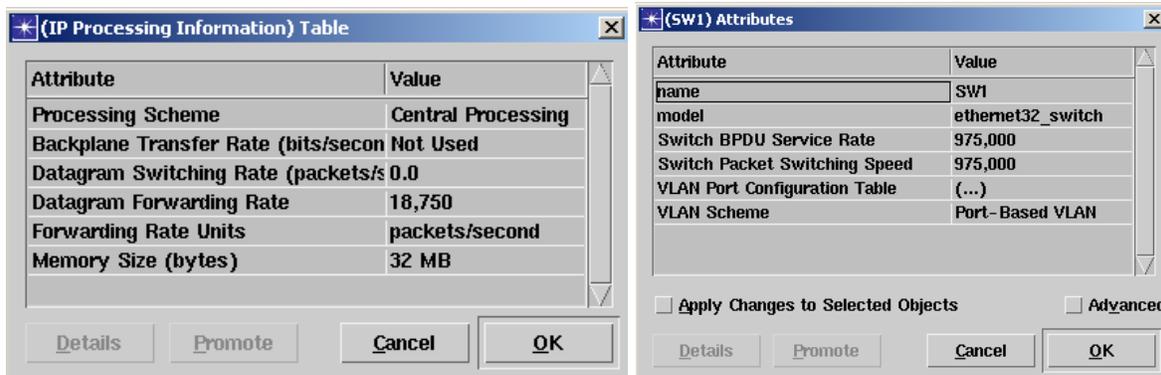
Figure 6. Configuring the attributes of Router (Left) and the Switches (Right)

The Router. The main parameter to be set in the router configuration is the forwarding rate. In our network, Cisco 2621 router has a forwarding rate of 25, 000 pps. However, since we consider a 25% growth factor, the effective forwarding rate of the router is set to 18,750 pps.

The Switches. Similarly, the main parameter to be configured in all the switches is the switching speed, which is 1.3M pps for the 3Com Superstack3 3300 switches. With a growth factor of 25%, the effecting switch speed is set to 975,000 pps.

The Links. For each link in the network model, a background utilization of 25% has been introduced to account for the growth factor. In addition, the ports to which each link have been set in accordance with the network diagram shown in Figure 3.

4.1.2 VLAN Configuration

The modified network topology includes the definition of six VLANs. See Section 3.5. Incorporating these VLANs in the network model is essential to reflect accurately the traffic behavior in the real network. Basically, VLANs can be defined on any switch that supports this property. In OPNET, VLANs are implemented based on ports, i.e. it is required to associate each port with its VLAN. Also, one port can belong to exactly one VLAN. Hence, if a port is required to carry the traffic of multiple VLANs, it must be defined as an `Untagged Trunk Link` in the switch's VLAN port configuration table.
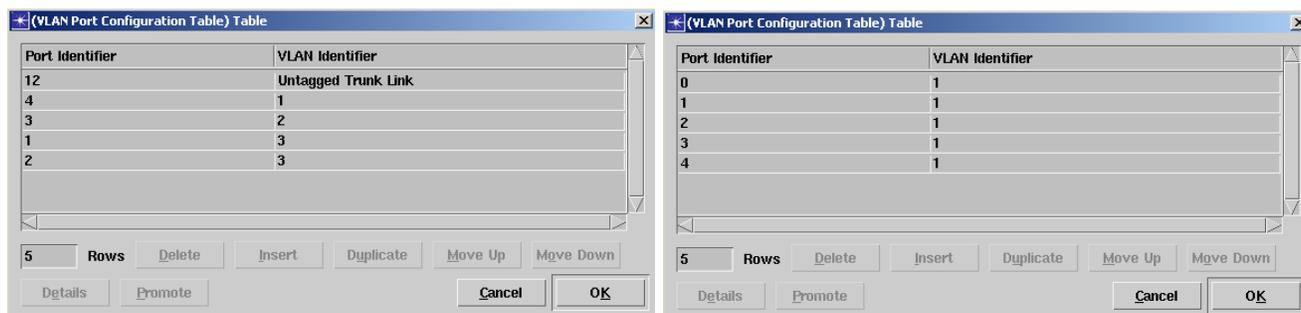

Figure 7. VLAN Configuration for SW1 (Left) and for F1SW (Right)

10

For each switch, the VLAN scheme should be set to port-based VLAN. Then, a VLAN port configuration table should be created and filled with the port numbers associated with the VLAN identifiers. As an example, Figure 7 on the left shows the VLAN port configuration table for the inter-floor switch, SW1. For the intra-floor switches, all the ports are using the same VLAN, which is the one defined for the whole floor. Figure 7 on the right shows the VLAN port configuration table for Floor 1 Switch, F1SW.

It is worthy to note that port 12, which is connected directly to the router, is described as an `Untagged Trunk Link` since it carries the traffic of all VLANs to the router, which routes it to the destination VLAN. This configuration is sufficient for the router to handle the traffic that passes from one port to a different port, e.g. the traffic between Floor 1 and Floor 3. However, this configuration is not sufficient to route the traffic back to the same port, which is the proper flow of traffic between Floor 1 and Floor 2. To enable such a flow, it is required to configure the router as a *one-armed router*. To configure this, OPNET requires the promiscuous mode to be enabled for the specific port(s). Also, the ARP response mode of the port should be changed to *always*, instead of being based on the destination address.

Before proceeding further, it is essential to test the VLAN configuration in the model. In particular, it is required to test the router's response to the traffic flowing between VLANs. One testing scheme is to define some conversation pairs between different VLANs and observe the router's behavior. In OPNET, a `conversation pair` is a flow of traffic between two different VLANs. `Conversation pairs` can be simply specified in a spreadsheet and then imported by OPNET by choosing `Import Conversation Pairs` in the `Traffic` menu. When simulated and if properly configured, flow and amount of traffic routed on each link between the different VLANs should match the specified flow in the spreadsheet.

4.2   Modeling Traffic

Having the network model configured correctly, we turn to the traffic flowing in the network. We first model the exiting current background traffic, and then model the new traffic of deploying VoIP.

4.2.1   Background Traffic

To model the traffic that exists on the network prior to VoIP introduction, the traffic measurements that were obtained are represented in a spreadsheet format so that it can be imported to OPNET. It is important to mention here that all the traffic measurements are between connected pair of nodes, as shown in Table 1. Ideally, we must define a conversation pair between any two connected nodes. However, this is not possible in all cases since switches are defined in OPNET as *passive* devices, i.e. cannot be used as a source or a sink of traffic. To overcome this problem, the router is used as a sink or a source in all conversation pairs that pass through the switches.

Since only the peak values of the background traffic are considered when analyzing the network capacity, a spreadsheet of the background traffic which only includes the peak hours is created. The background traffic information is based on the traffic measurements of the existing network shown in Table 1. The spreadsheet to

be imported by OPNET is shown in Figure 8. For OPNET, the format of the spreadsheet of the background traffic has to be changed to units in Mbits/s and average packet size. The background traffic duration is set to two hours. The simulation run must be configured such its duration is less than two hours. In all of our simulation runs, as will be demonstrated in Section 4.3, the duration was not more than 8 minutes. As discussed in Section 4.1, a designated workstation of each floor subnet, e.g. `F1_C3`, is used to generate and receive background traffic.

Duration: 2.00				
Duration Units: hours				
Step Size: 2.00				
Step Units: hours				
Number of Steps: 1				
Traffic Units: Mbits/s				
#Source Name	Destination Name	Avg Pkt Size	0 - 2 hours	
Enterprise Network.Router	Enterprise Network.Floor_1.F1_C3	0.010776	3.05	End Data
Enterprise Network.Router	Enterprise Network.Floor_2.F2_C3	0.011904	3.19	End Data
Enterprise Network.Router	Enterprise Network.Floor_3.F3_C3	0.011952	3.73	End Data
Enterprise Network.Floor_1.F1_C3	Enterprise Network.Router	0.010776	3.05	End Data
Enterprise Network.Floor_2.F2_C3	Enterprise Network.Router	0.011904	3.19	End Data
Enterprise Network.Floor_3.F3_C3	Enterprise Network.Router	0.011952	3.73	End Data
Enterprise Network.Router	Enterprise Network.DB_Server	0.012736	2.79	End Data
Enterprise Network.Router	Enterprise Network.File_Server	0.012352	1.89	End Data
Enterprise Network.DB_Server	Enterprise Network.Router	0.012736	2.79	End Data
Enterprise Network.File_Server	Enterprise Network.Router	0.012352	1.89	End Data
Enterprise Network.Cache_Proxy	Enterprise Network.Router	0.011192	1.97	End Data
Enterprise Network.E-mail_Server	Enterprise Network.Router	0.011096	2.12	End Data
Enterprise Network.Firewall_NAT	Enterprise Network.Router	0.01172	2.11	End Data
Enterprise Network.HTTP_Server	Enterprise Network.Router	0.011552	1.86	End Data
Enterprise Network.Router	Enterprise Network.Cache_Proxy	0.011192	1.97	End Data
Enterprise Network.Router	Enterprise Network.E-mail_Server	0.011096	2.12	End Data
Enterprise Network.Router	Enterprise Network.Firewall_NAT	0.01172	2.11	End Data
Enterprise Network.Router	Enterprise Network.HTTP_Server	0.011552	1.86	End Data

Figure 8. Spreadsheet representation of peak-period background traffic

There is one important observation to make about the generation of egress and ingress background traffic. Since we considered symmetric background traffic for egress and ingress, the entries for the traffic values reflected in the spreadsheet are duplicates. For example, the entries of traffic values going from the HTTP Server to the Router Server are the same traffic values from the Router to the HTTP Server. Compare the last row of spreadsheet to the fifth row from the bottom. If the traffic is not symmetric these values have to be changed accordingly.

4.2.2   VoIP Traffic

VoIP Application and Profile Settings. One way to model the VoIP traffic in OPNET is to use the predefined voice application. Basically, an application in OPNET is a collection of tasks of which each task is defined as a set of phases. In turn, each phase takes place between two endpoints and has a configurable traffic behavior. The time on which each task starts and the duration that it takes can be configured when defining the application. Applications can be defined and configured using the `Application Definition` while the `Task Definition` object is used to define and configure tasks. In the case of predefined applications, the underlying tasks are already defined but usually some flexibility is given to the user to configure their attributes. For example, the configurable parameters in the case of the predefined voice application are shown

in Figure 9a. One important parameter is the "*Encoder Scheme*", which is set to G.711. Another important attribute is the `Voice Frames per Packet`. In OPNET terminology, a voice frame is a collection of 32 voice samples of which each sample is 8 bits, i.e. each voice frame is 32 bytes in size. However, in the standard we are adopting, each VoIP packet has a payload of 160 bytes. Hence we set `Voice Frames per Packet` attribute to 5. The third attribute to which attention should be paid is the `Symbolic Destination Name`. As will be discussed shortly, this symbolic name is used in defining the destination nodes for VoIP calls.

After defining and configuring the VoIP application, it is required to configure the way in which workstations will be implementing this application. In general, the behavior of a network workstation is defined through its `Profile`, which is basically a collection of applications that can be configured to control their start and end times, in addition to their repeatability. Profiles can be defined and configured through the `Profile Definition` object. In our case, we need to define and configure only one profile, namely `VoIP_Profile`. Afterwards, the floor subnet workstations, designated for generating VoIP calls, will be configured to support this profile, as specified in Section 4.1. The designated workstations include `F1_C1`, `F2_C1`, and `F3_C1`.
While configuring the profile, 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 thresholds or bounds for VoIP delay and network capacity or bandwidth. When any of these bounds has been reached, the maximum number of calls can then be known. Note that the bound for network bandwidth is reflected in seeing packets being lost or observing a mismatch between VoIP traffic being sent and received.

In OPNET terminology, we need to configure the profile to add calls repeatedly at a fixed rate. To accomplish this, the application profile first must be configured to establish one call at the beginning and then add a call each 2 seconds. To achieve such behavior, it is required to include the VoIP application in the `VoIP_Profile` definition as shown in Figure 9b and Figure 9c. In these figures, we let the `Start Time Offset` of the VoIP traffic to start after 60 seconds and for the first VoIP call to start after 10 seconds. This means that the first VoIP call will produce traffic after 70 seconds from the start of the simulation run. By default in OPNET, the background traffic starts after 40 seconds. These start times are tunable and can be helpful in debugging the simulation by making sure simulated traffic being properly generated at certain times. To keep generating calls, the repeatability of the VoIP application is set to be *Unlimited* with an inter-repetition time of 2 seconds, as shown in Figure 9d.

Generating VoIP Traffic Based on Call Distribution. The floor subnet workstations of `F1_C1`, `F2_C1`, and `F3_C1`, designated for generating VoIP calls, are configured to support the `VoIP_Profile`. This is simply done by adding this profile to each workstation's lists of supported profiles. In addition, it is essential to set the destination preferences of each workstation so that the call distribution ratios can be implemented. For a workstation to qualify as a valid destination, its client address attribute should be set to a unique value, which can be used to reference this workstation. Next, we add the symbolic name `VoIP_Dest` in the `Application: Destination Preferences` table of each sending workstation, as shown in Figure

10a. As explained before, the symbolic name defines the workstation to which the call will be made. To define the actual destinations and the call distribution ratios, a set of actual names is mapped to each symbolic name. For example, Figure 10b shows the list of actual names that map the `VoIP_Dest` symbolic name for `F1_C1`. `F1_C1` is the sending workstation of Floor 1.   Figure 10b also shows the selection weights that are used to control the distribution of calls. In this case, the selection weights are based on the probability distribution tree shown in Figure 2.

The VoIP traffic will be generated and received by workstations within the floor. Since we are interested only in determining the network delay and capacity for supporting VoIP calls, we rule out any delays to be introduced by the workstations of subnet floors. Also we rule out having the processing of workstations being the bottleneck for bandwidth. Hence we set the processing rate for each work station to *infinity*. We accomplish this, as shown in Figure 10c, by changing the workstation's attribute of `Datagram Forwarding Rate`. This attribute is one of the attributes for the `IP Processing Information` table of the workstations.
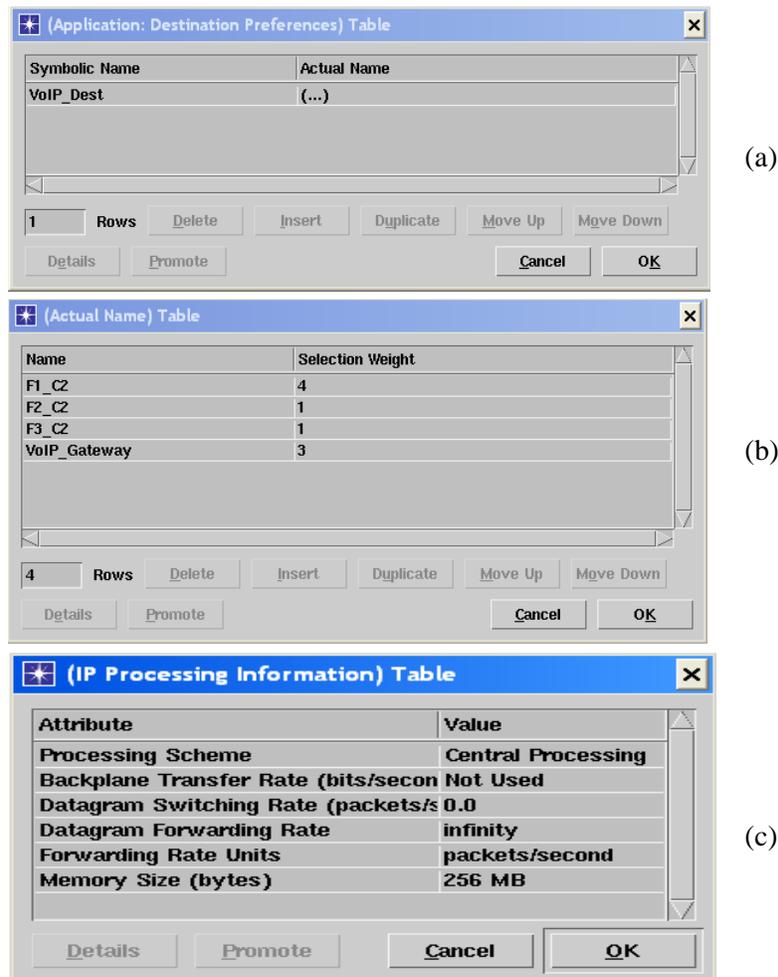

(a)

(b)

(c)

Figure 10. Settings of VoIP workstations

4.3   Simulation Results

Before the simulation started, OPNET has to be configured to obtain graphed results for numerous network components which include VoIP traffic, router, switches, and links. In this section, graphed results for some of the most important components are shown. In particular, we show graphed results for the VoIP traffic, router, and some links. As for the switches and other elements, they for the most part showed healthy behavior and were under utilized.

We configured the duration of the OPNET simulation run for 8 minutes. The generation of background traffic, by default in OPNET, started at 40 seconds from the start time of the simulation run. The VoIP traffic started at 70 seconds at which a total of 3 VoIP bi-directional calls are initially added. Then, every 2 seconds 3 VoIP calls are added. The Simulation stops at 8 minutes in which a total of $3 + \left(\dfrac{7 \times 60 + 58 - 70}{2} \right) \times 3 = 615$ calls got generated. This should translate into a total of 61,500 packets being generated every second. Note that since the simulation stops at 8 minutes, the last 3 calls to be added was at 7 minutes and 58 seconds.

VoIP Traffic. Figure 11 shows the VoIP traffic and the corresponding end-to-end delay as VoIP calls are added every two seconds. Figure 11a shows the total VoIP traffic that was sent, received, and dropped. Figure 11b is a zoom-in version of Figure 11a, focusing on the mismatch region between traffic sent and received. From Figure 11a, it is clear that the total VoIP traffic generated by the end of simulation run is very close 61,500 pps. In fact, simulation results gave 61,429 pps.

One can determine the total number of calls that the network can sustain by examining network bandwidth or delay bounds. We first investigate the bandwidth bound. Figure 11a and Figure 11b show clearly that not all of VoIP packets being sent get received. I.e. there is a mismatch between traffic sent and received. Figure 11b captures clearly the addition of the three calls every 2 seconds; and how this addition is repeated in gradual steps of 300 pps. We can determine the number of calls that can be supported by examining the X and Y axes. Examining the X axis of the simulation run time, it is clear that the last successful addition of three calls was at exactly 4 minutes and 48 seconds, as seen clearly in Figure 11b. The next addition, as shown, was at 4 minutes and 50 seconds which resulted in a mismatch. For the last successful addition of voice calls, which occurred at 4 minutes and 48 seconds, we had a traffic volume (see Y axis) of exactly 33,000 pps or 330 VoIP calls. Also one can arrive at the same number of calls by calculating how many calls have been added until the last successful addition of three calls, i.e. 4 minutes and 48 seconds. This yields to $3 + \left(\dfrac{4 \times 60 + 48 - 70}{2} \right) \times 3 = 330$ calls.

Figure 11c shows the corresponding VoIP end-to-end delay. Remember this delay should not exceed 80 ms, as discussed in Section 3.3. As depicted, the delay stays less than 80 ms until a simulation time of 4 minutes and 54 seconds at which the delay increases sharply. One can then find out the number of VoIP calls that the network can support to satisfy the 80 ms time constraint. The number of calls can be computed as

$$3 + \left(\frac{4 \times 60 + 54 - 70}{2} \right) \times 3 = 339 \text{ calls.}$$ Therefore one can conclude that, based on these simulation results, the number of voice calls to be supported by the network is bounded more by the network bandwidth than the delay. Hence, the number of the VoIP calls that the network can support based on simulation is 330 calls.

The simulation's reported delays shown in Figure 11c is the maximum values of a bucket of 100 collected values. The OPNET default reported delay configuration is the sample mean of a bucket of 100 collected values. Figure 11d depicts a different collection mode, in which "all values" are collected and plotted. Figure 11d depicts two types of delays. First, the delays of external and inter-floor VoIP packets passing through the router. These are the bigger delays and they resemble the delays of Figure 11c. Second, the delays of intra-floor VoIP packets that are not passing through the router. These are the smaller delays, in which the majority of these values stay close to 2.5 ms.
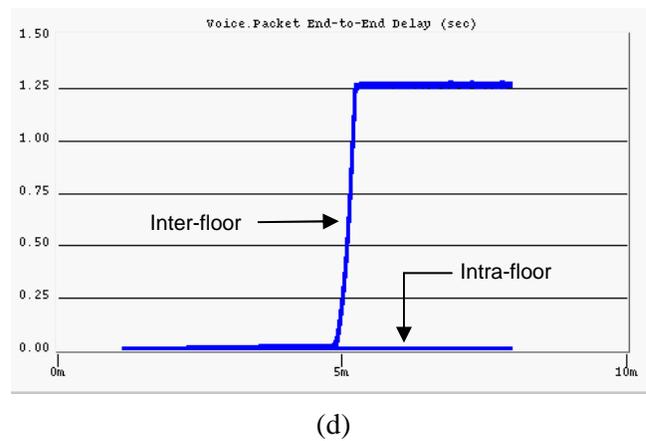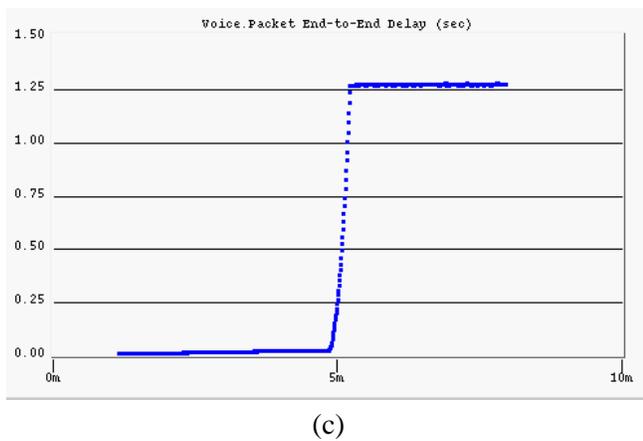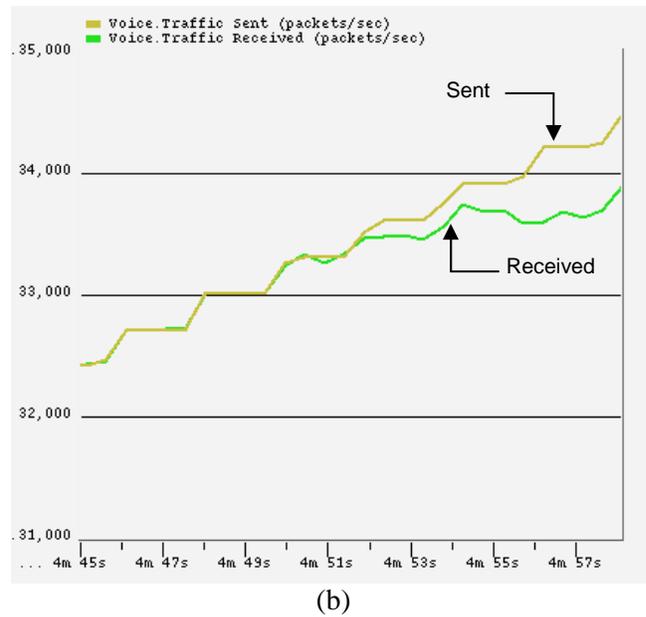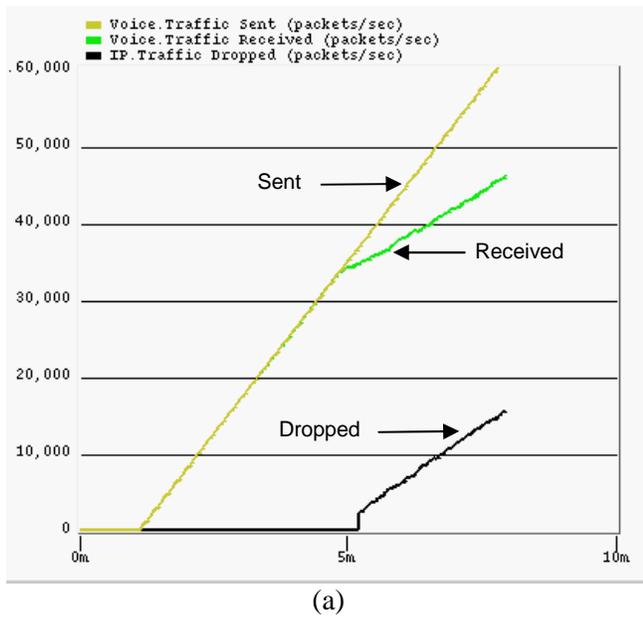

(a)

(b)

(c)

(d)

Figure 11. VoIP traffic and network delay

16

Router. The router is a vital component for the network under study. It is imperative to investigate its health, performance, and behavior. Graphed results of a number of important key health indicators are shown in Figure 12. The figure shows key performance indicators for IP traffic, CPU utilization, processing delay, queue size in packets and bytes, and queueing delay. Figure 12a confirms that the background traffic starts at 40 seconds and stays flat for 30 seconds with a total of 812 and 869 packets coming from SW1 and SW2, respectively. At 70 seconds, the VoIP traffic gets generated with a rate of 3 calls every 2 seconds. Figure 12a shows that the IP traffic being processed by, i.e. sent out from, the router stays flat at 4 minutes and 40 seconds. Figure 12b shows that at this time the corresponding CPU utilization has reached 100%. This is the saturation processing point for the router and at which the mismatch between VoIP traffic sent and received occurs. After this point, the router buffer starts being filled very quickly by new calls, resulting in dropping of packets, as shown in Figure 11a. The dropping of packets occurs after 25 second, precisely at 5 minutes and 15 seconds. At this point the memory of 32 Mbytes of the router's forwarding queue is completely full, as shown in Figure 12c. Also, the corresponding queuing delay, shown in Figure 12d, sharply increases and it stays flat with a delay of approximately 1.25 seconds. This delay can be verified and approximated theoretically, using $M/M/1/B$ queueing model, as B/μ. Where B the number of packets and μ is the routers's forwarding rate of 18,750. The number of packets stored in the memory of the router, B, shown in Figure 12c and Figure 12d, can be calculated by dividing 32M by 1518 bytes. In OPNET, one packet occupies 1518 bytes of memory in routers, regardless of the actual size of the packet.
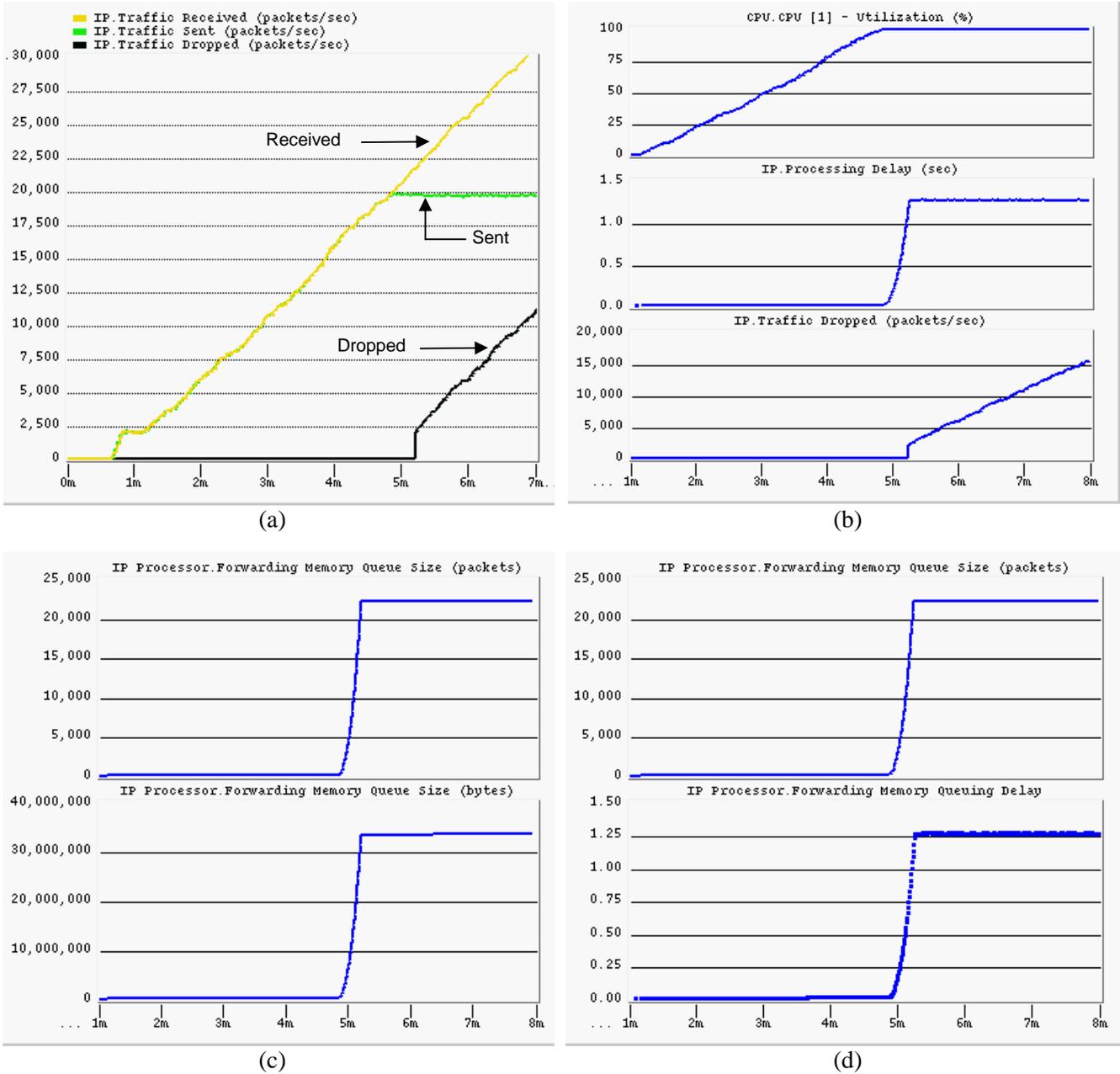
Figure 12. Key health indicators for the Router

Links. Checking the health of key links in the network is essential. The uplinks and downlinks connecting the Inter-floor switches to the router is of great significance. Also, it is imperative to check the health of the uplinks and downlinks connecting the the intra-floor switch to the designated workstations responsible for generating VoIP traffic. Figure 13a and Figure 13b show graphed results of the uplinks going from SW1 and SW2 to the router, respectively. Figure 13c and Figure 13d show graphed results of the downlinks between the router to SW1 and SW2, respectively. It is observed that the links in general are healthy, even with the addition of the growth factor of 25% utilization of each link. As depicted, at the start up of the simulation run, the links are utilized 25% and then shoots higher at 40 seconds to about 10% as the background traffic is

18

generated. As the VoIP calls are generated every 2 seconds, the uplink utilizations, shown in Figure 13a and Figure 13b, are steadily increasing. However, the corresponding downlink utilizations do not steadily increase, but rather flattens off at about 4 minutes and 40 seconds as the router reaches its saturation point.
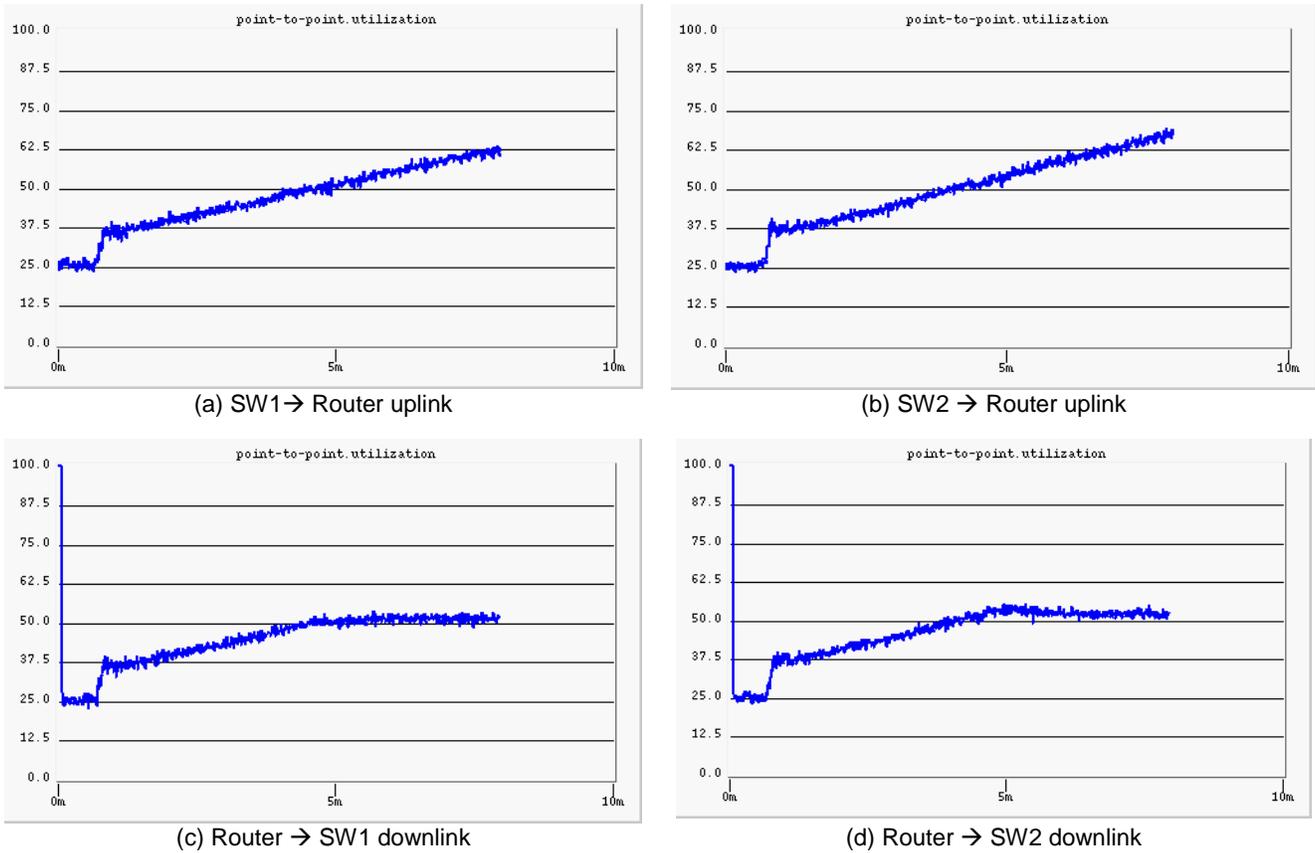

(a) SW1→ Router uplink

(b) SW2 → Router uplink

(c) Router → SW1 downlink

(d) Router → SW2 downlink

Figure 13. Utilization of Router ⇔ Switch links

As discussed earlier, the model of floor LANs does not represent precisely the floor multimedia PCs. In order to automate the simulation run, the floor LANs were approximated by three designated Ethernet workstations to generate traffic, as opposed to separate PCs. We argue that such a model is adequate as long as we keep an eye on the health of the links connecting these designated workstations to the intra-floor switch. Figure 14 shows the utilization and the corresponding queueing delay in seconds for the uplink connecting `F1_C1` to `F1SW` and the downlink connecting `F3SW` to `F3_C2`. This represents edge links of a one-way VoIP call initiated from Floor 1 by `F1_C1` and terminated at Floor 3 by `F3_C2`. The figure shows for the worst case situation, i.e., when generating a total of 615 calls, we obtain an increase only of 15% utilization and a small queueing delay increase of 0.23 ms in the uplink and 0.06 ms in the downlink, or a total delay of 0.29 ms for both. This delay can be considered as the actual edge link queueing delay of one bi-directional VoIP call. From the figure, it is to be noted that for one VoIP call, see the simulation time at 70 seconds, would produce a delay of about 0.02 ms for both links. We demonstrate in Section 4.5 for a stable simulation run of 306 calls, with no packet loss or dropping, the incurred queueing delay and utilization for the edge links is much smaller. In fact, we obtained an extra utilization of 5% and a queueing delay of 0.05 ms for the edge uplinks and

downlinks. This translates in only 0.10 ms queueing delay for one-way VoIP call, considering one delay for an edge uplink and another for an edge downlink.
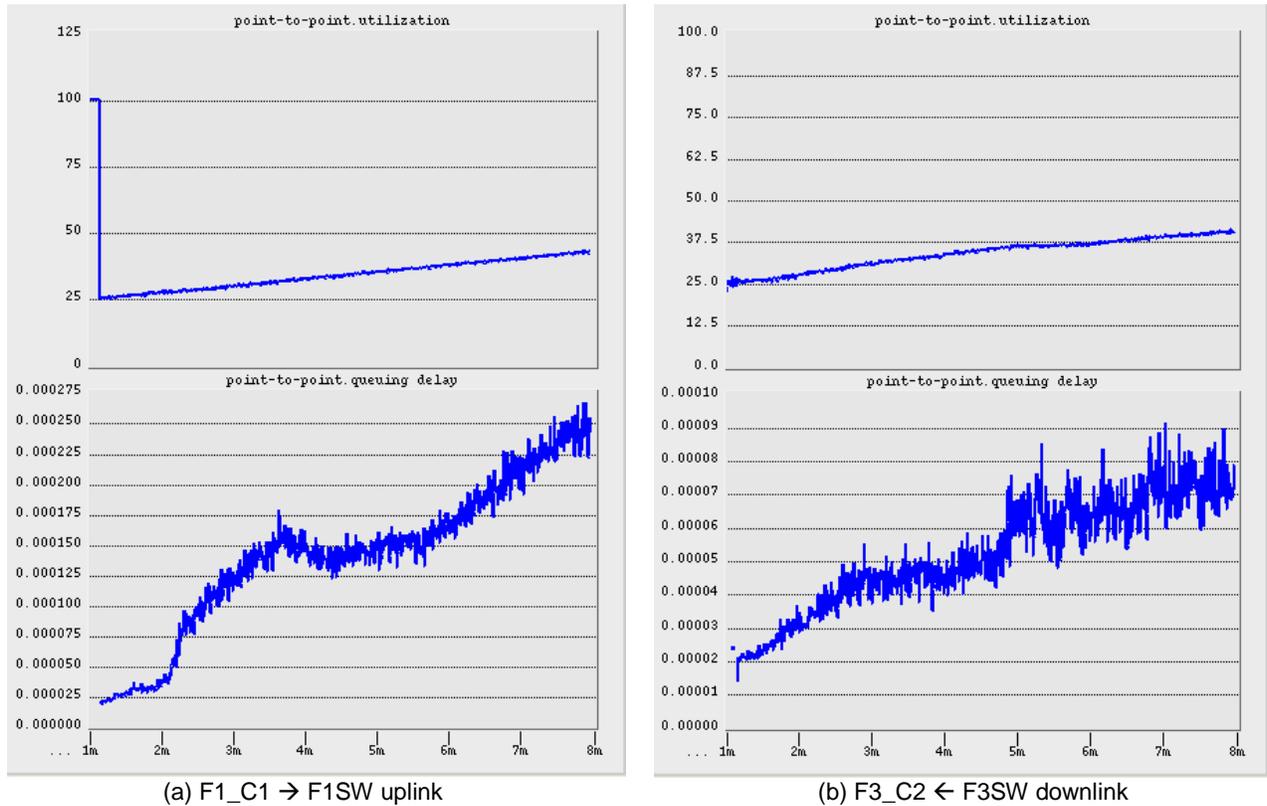

(a) F1_C1 → F1SW uplink (b) F3_C2 ← F3SW downlink

Figure 14. Floor subnet edge links

4.4   Simulation Environment and Accuracy

Our previous OPNET simulation was executed on a Sun E450 server. The Sun E450 server has Solaris 7 OS with a SPARCV9 400 MHz processor and 1G bytes of memory. The elapsed time for the simulation run was approximately 15 hours and produced 722 million events, executed at average speed of 13878 events/second. It is possible to cut down on the simulation run time by configuring the generation of more calls simultaneously at the start of the simulation, say 300 calls, and then add more calls slowly at a rate of 3 calls every 2 seconds. In addition, one can also stop the simulation at earlier time, say at 6 minutes instead of 8 minutes.

In order to gain accuracy (with a narrow confidence interval) of our simulation results, following the popular guidelines presented in [15,16], five simulation replications were run by feeding OPNET with different initial seeds. OPNET's pseudo random number generator is based on BSD's algorithm which permits safely, i.e., with no concern of overlapping of random number streams, any integer value to be an initial seed. Five simulation replications were sufficient.   Each simulation replication produced very similar graphical results, which when interpreted as done in Section 4.3, led to the same total number of calls to be supported.

20

4.5   Final Simulation Run

Based on the simulation results, the existing network can support 330 VoIP calls while satisfying the VoIP QoS of throughput, latency and packet loss. In Section 4.2, the voice calls were added every 2 seconds and the simulation was not allowed to stabilize for a long time. Our attention was focused on finding out the number of voice calls that the network can sustain. As a final check to ensure a healthy network and a normal behavior for all network elements, we perform a final simulation run in which 330 voice calls are added, all at once at the start of the simulation, say after 70 seconds. We let the simulation run execute for a prolong amount of time, say good 5 minutes, to reach a steady state. Then we examine the health of each network element. In our example, this simulation run of supporting 330 calls was not successful. The simulation run showed a mismatch between traffic sent and received and a delay of more than 80 ms. However, a successful simulation run of 306 voice calls showed normal and healthy results with no packet loss, end-to-end delay of 2.25 ms, and adequate utilization of router and switch CPUs and links. To account for the extra queueing delay produced by the edge links due to generating VoIP traffic from the same workstation, as opposed to separate PCs, a delay of 0.10 ms has to be subtracted from 2.25 ms. Hence the effective end-to-end network delay would be 2.15 ms.
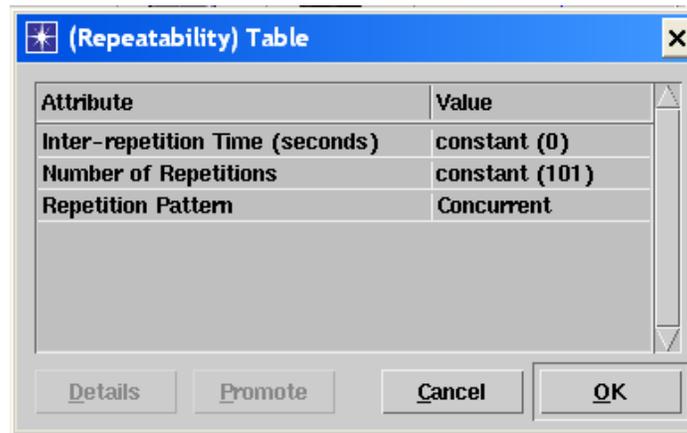

Figure 15. Repeatability table to generate 306 calls

Figure 15 shows how to configure the generation of 303 concurrent calls, after the first initial 3 calls get generated. Each VoIP application profile for the designated sending workstations of the three floors has to be configured as shown in Figure 9, except for Figure 9d where its attributes has to be changed as shown in Figure 15. Figure 15 shows that each workstation will generate 101 calls all at once, and hence the total number of calls that will be generated at once will be $3 + 101 \times 3 = 306$ calls.

5   Design and Engineering Decisions

Based on the simulation results and observations of network under study, the following network design and engineering decisions can be justified:

- The existing network, with a reserved growth factor of 25%, can safely support up to 306 calls while meeting the VoIP QoS requirements and having no negative impact on the performance of existing network services or applications.
- For 306 calls, a network delay of about 2.15 ms is encountered.

- A safety growth factor of 25% is maintained across all network resources.
- The primary bottleneck of the network is the router. If the enterprise under study is expected to grow in the near future, i.e., more calls are required than 306 calls, the router replacement is a must. The router can be replaced with a popular Layer-3 Ethernet switch, and thus relieving the router from routing inter-floor calls from Floor 1 to Floor 2. Before prematurely changing other network components, one has to find out how many VoIP calls can be sustained by replacing the router. To accomplish this, the design steps and guidelines outlined in this paper must be revisited and re-executed.
- The network capacity to support VoIP is bounded more by the network throughput than the delay. This is due to the fact the existing network under study is small and does not have a large number of intermediate nodes. The network delay bound can become dominant if we a have a large-scale LAN or WAN.

6   Conclusion

The paper presented a detailed OPNET-simulation approach to assess network readiness and support for VoIP. The approach can help network researchers and designers to determine how well VoIP will perform on a network prior to deployment. With such an approach, it is possible prior to the deployment of VoIP 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. 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.

Acknowledgement

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] CAIDA, http://www.caida.org/tools/taxonomy, April 2004.

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

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

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

[7] B. Duysburgh, S. Vanhastel, B. DeVreese, C. Petrisor, and P. Demeester, "On the Influence of Best-Effort Network Conditions on the Perceived Speech Quality of VoIP Connections," Proceedings of IEEE 10th International Conference of Computer Communications and Networks, Scottsdale, AZ, October 2001, pp. 334-339.

[8] 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

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

[10] Recommendation H.323, "Packet-based Multimedia Communication Systems," ITU, 1997.

[11] J. H. James, B. Chen, and L. Garrison, "Implementing VoIP: A Voice Transmission Performance Progress Report," *IEEE Communications Magazine*, vol. 42, no. 7, July 2004, pp. 36-41

[12] 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

[13] S. Riley and R. Breyer, "*Switched, Fast, and Gigabit Ethernet*," Macmillan Technical Publishing, 3rd Edition, 2000.

[14] OPNET Technologies, http://www.mil3.com

[15] K. Pawlikowski, H. Jeong and J. Lee, "On Credibility of Simulation Studies of Telecommunication Networks", *IEEE Communications Magazine*, vol. 40, no. 1, January 2002, pp. 132-139

[16] A. Law and W. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, 2nd Edition, 1991.