International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering Vol. 2, No. 2, 2015, pp. 121-145.
ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Step by Step Methodology for Implementation of the Telephony Services on Data Networks

Esmaeil Nik Maleki[a] , Ghasem Mirjalily[b]

[a] *Faculty of Electrical Engineering Sheikhbahaee University, Isfahan, Iran,*

[b] *Faculty of Electrical and Computer Engineering Yazd University, Yazd, Iran.*

Abstract

Voice-over-IP (VoIP) services offer a cheap alternative to the traditional mobile operators to make voice calls. In this paper, we present a comprehensive method for step by step implementation of VoIP services on data networks. This method includes five steps. After determining the requirements of QoS and call distribution features, we should be analyzing network conditions. In this article, we provide a set of qualitative and quantitative criteria to evaluate the network quality of service. The quantity criterions are packet loss ratio and delay that evaluate with new method. On the other hand, in order to calculate the call capacity of the network based on qualitative analysis , Extended_E-model is used. Extended_E-model estimates the Mean Opinion Score (MOS) rate of the voice quality using the measured network quantities. As a case study, we consider the simulation of an enterprise network connects two branches of the organization. We used the OPNET modeler and MATLAB to simulate the quality and quantity criterions in order to assess the given network.

Keywords: Voice over IP, Quality of Service, The quantitative and qualitative analysis, Packet loss ratio, Delay ,Mean Opinion Score (MOS), Extended_E-model

Introduction:

Using data communication networks for telephony services has evolved the telecommunication industry. Delivering telephony services in packet switching networks is one of the main fields in network domain. Ethernet is a well-known, simple, cheap and reliable technology and also has widespread use in local and metro networks. These benefits make Ethernet the most favorite network technology such that more than 90% of the Internet traffic starts from an Ethernet network and moves towards another Ethernet network [1].

Given the emerging deployment of Ethernet networks, a major aspect of interest is to employ these networks to deliver legacy telephone services. This protects investments in existing infrastructure and creates new revenues for network providers. The basic requirement in this field is achieving confidence about the quality of the telephony service. However, to implement this technology, these networks have to be configured in a way that customary QoS requirements for telephony, can be met[2]. This is a crucial problem given that the data cross traffic in these networks is bursty. This bursty traffic can cause excessive queuing delay and frame losses due to buffer overflow.

In this paper, we introduce a new approach to readiness assessment of the networks based on Ethernet, IP, UDP and RTP protocols for telephony service. This approach is based on the calculation of call capacity by using quantity and quality criterions. In [1], these measures were introduced. Call capacity is the maximum number of calls we can offer by considering the required QoS. The quantity criterions are packet loss ratio and Delay that constraint due to them evaluate with new algorithm in this paper. Mean Opinion Score (MOS) is the major quality criterion for evaluation of the speech quality defined in ITU-T.P800. MOS is based on the opinions of a large number of listeners about the quality of the speech. Each listener evaluates the quality and assigns a grade from 1 to 5 (1:bad, 2:poor, 3:fair, 4:good, 5: excellent). MOS is the average of listeners' grades.

During the past years, there is considerable interest in technologies, systems, and architectures supporting voice over packet applications. Mirjalily, et. al. [3], studied the performance of connecting telephone exchange centers over a metro Ethernet network using MPLS and DiffServ QoS model. They derived some simulations to evaluate the performance in terms of delay, jitter and loss. In [4], Wright compared the advantages of VoMPLS with VoATM, VoIP, and VoFR, from the viewpoints of bandwidth utilization, implementation issues, and the region of the network in which implementation takes place. Nair, in his thesis [5], proposed some solutions to provide QoS for real-time applications in metro Ethernet networks. This involves traffic engineering with load balancing, uniform resource utilization and reliability. Salah [6], considered the deploying of VoIP in a local area network. He studied the QoS requirements and utilized both analysis and simulation tools to determine the number of VoIP calls that can be supported for a typical Ethernet network. Salah only considered LANs, whereas the basic advantage of packet-based telephony service is reducing the cost of conversations between different branches of an organization. On the other hand, he uses only qualitative analysis.

In this paper, a new approach based on the both quantitative and qualitative analysis is presented in order to evaluate the readiness of an Ethernet network for telephony service. This approach is suitable for Wide Area Networks as well as LANs.

This paper is organized as follows: Section 2 outlines practical five steps methodology to evaluate data networks for telephony service. Each step is described in considerable detail. In section 3, we consider a typical network topology as a case study. Simulation results are presented in Section 4 and finally, Section 5concludes the study.

1- Step by step methodology

Figure 1 shows the flowchart of a methodology of five steps for VoIP deployment. The first two steps are independent and can be performed in parallel. The Steps 3 present analytic approach for network assessment .The simulation study presents in steps 4.The results compare in step 5.

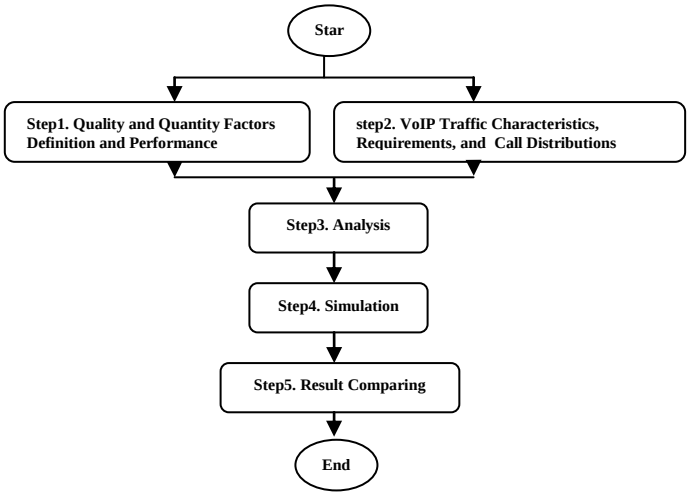


Figure (1). Methodology steps flowchart

2-1- Step 1- Quality and Quantity Factors Definition and Performance Thresholds

2-1-1- Define Quantity Factors

The important quantity factors for QoS evaluation of the telephony service are delay and packet loss. Recommendation G.714 imposes a maximum total one way end-to-end delay of 150 ms for IP telephony applications [2]. We can break this delay down into at least three different contributing components, which are as follows: (i) encoding, compression, and packetizing delay at the sender, (ii) propagation, transmission and queuing delay in the network and (iii) buffering, decompression, depacketizing, decoding, and playback delay at the receiver . 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 . Telephony packet loss should be below 1% [5].

2-1-2- Define Quality Factors

An MOS of 4.0 is considered toll quality by telephone industry. Toll quality is the quality level typically heard on a wired land-line telephone network. The G.711 codec achieves a MOS score between 4.0 and 4.4 on a LAN. On a WAN, the impacts of jitter, latency and packet loss can significantly degrade the speech quality. Speech qualities below 3.0 on the MOS scale are generally unacceptable [7].

Obtaining MOS based on listeners' opinions is time consuming, expensive and do not permit real time measurements. Therefore, using an automatic method for MOS calculation is necessary. The International Telecommunications Union -Telecommunication Standard Sector (ITU-T) has standardized a computational quality measurement mechanism called Emodel [8].

In real networks, the packet losses are often not the same and also delay is variable. So the quality level of speech may change very fast and greatly. These variations must be monitored. Assume that the network is steady and the quality hasn’t considerable variations in short term time. Using E-model we can obtain a sequential of instantaneous values for the quality of speech (Figure 2).

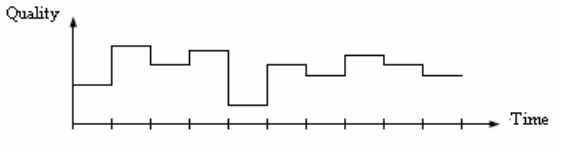


Figure (2). Instantaneous variation of speech quality

2-2- Step2-VoIP Traffic Characteristics, Requirements, and Call Distributions

2-2-1- VoIP Traffic Characteristics

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 node has to be added to the network [5]. The gatekeeper node handles signaling for establishing, terminating, and authorizing connections of all VoIP calls. 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.

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).VoIP-enabled workstation runs VoIP software such as IP Soft Phones .

Figure 3 identifies the end-to-end VoIP components from sender to receiver [5]. 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 [10]. 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.

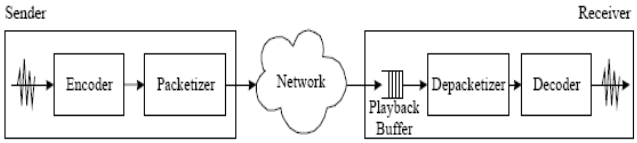


Figure (3). VoIP end_to_end components

2-2-2- Call Distribution

Knowing the current telephone call usage or volume of the enterprise is an important step for a network assessment. Key characteristics of existing calls can include the number of calls and locations of the call endpoints, the sources and destinations. 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. The call distribution is described as a probability tree. It is also possible to describe it as a probability matrix.

2-3- Step3: Analysis Method

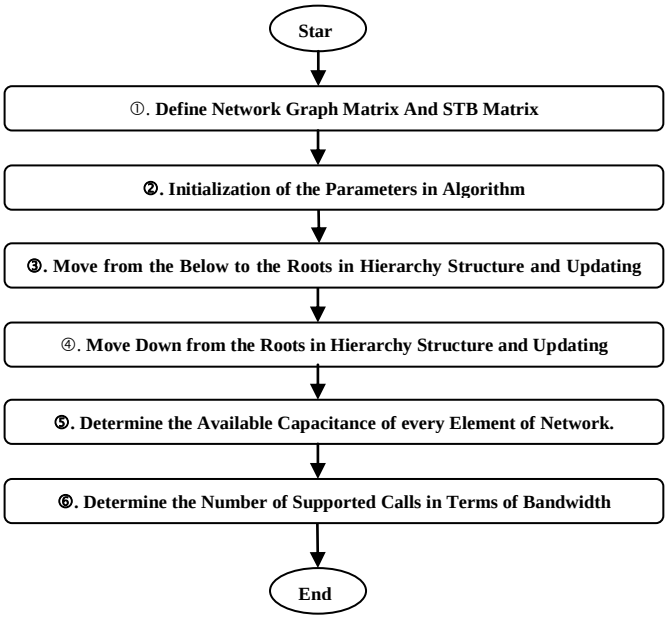
In this section, an operational procedure for evaluating a network based on the quality and quantity criterions is studied.

2-3-1- Quantitative Criterions

- **Bandwidth Analysis**

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.

In this section ,a new approach for determining the available capacitance of a given network is presented and then the number of supported calls with every elements of the this network is acquired. The algorithm is implemented in 6 steps that shown in figure 4.



Figure(4). Bandwidth Analysis Steps

①. Define Network Graph Matrix And STB Matrix.

A two layer Ethernet network can be shown with a graph so that the switches consider as the verteces and the connections between switches as the limb of the graph. In first stage, by using the graph theory and the spanning tree definition, the spanning tree of the network is obtained.

Definition: according to the graph theorem, for a given graph named *G* with *n* nodes, and *A(G)* was the *n*n* matrix of the graph, the $P_{n\times n}$ matrix is the representation of the *A(G)* spanning tree when for every entry one in matrix *P* ,there was an corresponding entry one in *A(G)*.

②. Initialization of the Parameters in Algorithm

In this step , by using the obtained spanning tree from the previous step, the structure of the network is changed to hierarchy state. In the presented algorithm, *N* is the number of switches and routers , *M* is the number of the endpoints(including local networks and servers connected to switches) and *L* is the number of layers in the

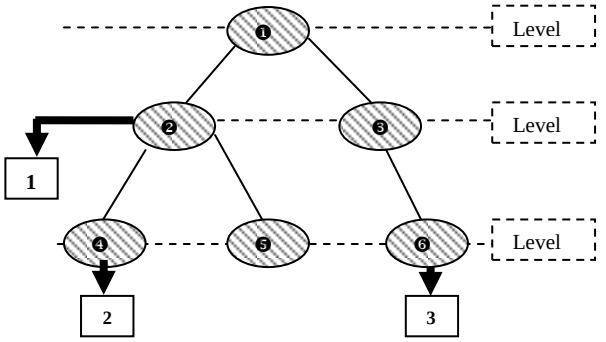


Figure (5). The example of hierarchy structure

hierarchy structure. With respect to figure 5, *L=2,N=6, M=3.*

In the rest of the article, X_{ijk} represent parameter *X* for network *i* in port *j* and switch *k*. I_{ijk} is a three dimentional matrix with logical values of one or zero in every entry. If the endpoint *i* is connected to port *j* from switch *k* with a direct link, then the value of the corresponding entry of the matrix is one. This matrix is obtained from the spanning tree matrix. IT_{ik} is a two dimentional matrix with logical values of one and zero in every entry. If the endpoint *i* is connected to switch *k* with direct link, the value of the corresponding entry in the above mentioned matrix is one. This matrix is obtained by logical OR on the entries of matrix I_{ijk} for all ports of switch *k*.

$$IT_{ik} = \bigcup_{j=1}^{J_k} I_{ijk} \qquad (1)$$

Matrices LS ،C ،u وd ،J state some of the specifications of the switches. The entries of matrix LS_k explained the number of the layer which switch *k* is there. C_k Represented the speed of the switching of k^{th} switch in terms of *PPS*. By having matrices LS_k and I_{ijk} ,the number of the uplink switch for k^{th} switch and the number of the downlink switch connected to port *j* and switch *k* is obtained which inserted in matrices u_k and d_{jk} respectively. J_k Represent the number of ports of switch *k*. The ports in switch *k* are numbered from *1* to J_k which number *1* is belong to uplink port of switch *k*.

By having I , matrix *Im* is defined which is a three dimentional matrix with logical values of one and zero in every entry. If the endpoint *i* was connected to the port *j* from switch *k* with direct link or the switches of the underlying layer, then the value of the corresponding entry in the matrix becomes one and calculated as below:

$$\mathrm{Im}_{ijk} = I_{ijk} + \bigcup_{t=2}^{J_{d_{jk}}} \mathrm{Im}_{itd_{jk}} \qquad (2)$$

In the above equation, entry one in the first term states the direct link from endpoint *i* to the port *j* from switch *k* and the second term represent the logical OR on all endpoints *i* with direct or indirect link to downlink ports of switch d_{jk}. operator "+" represent the logical OR.

By having *Im* matrix *ITm* can be calculated which is a two dimentional matrix with logical values of one or zero in every entry. If the endpoint *i* was connected to switch k with direct or indirect link then the value of the entry in the matrix becomes one. This matrix results from applying logical OR on the entries of matrix *Im* for all ports of switch *k*.

$$ITm_{ik} = \bigcup_{j=1}^{J_k} \mathrm{Im}_{ijk} \qquad (3)$$

The entries of W_{jk} represent the scaling coefficients allocated to the data traffic by using the QoS algorithm on the port j from the switch k. Various methodes of queuing for producing QOS in networks can be used in order to improve the quality of services. Some of the queuing methods are FIFO , PQ and WFQ.

One of the simplest methods of queuing is FIFO structure. In this method, all packets are placed in a common queue and all of them are considered the same. Service of packets is implemented in order the arrival of them to the queue. The objective of PQ queuing is to produce a simple method to distinct between different services. In PQ, first the packets are classified and then are placed in queues with different precedence. The packets in the beginning of every queue is serviced only when all the queues with higher precedence are empty. In the WFQ queuing with allocating scaling to every queue, a percentage of the output bandwidth can be allocated to a specific traffic class.

To apply different methods of queuing, first the traffic must be categorized based on the quality of required services and a *ToS*(Type of Service) must be allocated to every class. In Ethernet networks, ToS often is a number between 0 and 7 and 8 different classes can be defined with it. Class 7 has the highest precedence for providing the quality of required services and class 0 has the lowest precedence[7]. In this article a ToS equal to 6 is allocated to the packets related to telephony conversations and for the rest of the traffics in the network class 0 is allocated and by applying some methods of queuing, the impact of them on the number of the supported calls is studied.

The entry W_{jk} of this matrix for the ports with FIFO queuing is equal to one And is equal to zero For the PQ queuing ,since the telephony traffic is placed in the highest precedence queue and there isn't any services for the rest of the queues until depleting this queue, and finally for the WFQ queuing is calculated in the following form:

$$\alpha = \frac{T_t}{\sum_{m=0}^{7} T_m} \qquad (4)$$

In which T_t is the coefficient related to t^{th} class of traffic.

In the proposed approach the B_{jk} states the bandwidth of the link connected to port j of switch k. The amount of traffic between endpoints (servers and local networks connecter to switches) is represented with matrix D^X. X represents the unit of traffic and it can be "*b"* or "*p"* that represent the traffic in terms of *Mbps* and *Pps* respectively. All the calculations related to input and output traffic of every element in the network, is done in terms of *Mbps* and *Pps* that use in the calculations of the capacitance of the links and switches, respectively. By definition of matrices, initialization in algorithms is done in the following way:

$$TI_{ik}^{X} = \sum_{\substack{h=1 \\ h \neq i}}^{M} \left(D_{hi}^{X}\right) \times IT_{hk} \qquad (5)$$

$$d^{jk} = 0, \quad if \;\; j \;\; is\, access\; port \qquad (6)$$

TI_{ik}^{X} represent the input traffic to switch k, with the destination of endpoint i. In equation (6), the availability ports of switches means the ports connected to endpoints.

③.Move from the Below to the Roots in Hierarchy Structure and Updating Matrices.

In this step, by moving to the root of tree, In order to updating the input and output traffic to every switch ,the following formulas is used:

$$TO_{i1k}^{X} = TI_{ik}^{X} \times \left(\mathrm{Im}\right)'_{ik} \qquad (7)$$

$$TTO_{1k}^{X} = \sum_{i=1}^{M} TO_{i1k}^{X} \quad (8)$$

$$TO_{i1k}^{X} = \begin{cases} TO_{i1k}^{X} & TTO_{1k}^{X=b} \le B_{1k} \times W_{1k} \\ TO_{i1k}^{X} \times W_{1k} \times \alpha_{1k} & otherwise \end{cases}$$

$$\alpha_{1k} = \frac{B_{1k}}{TTO_{1k}^{X=b}} \quad (9)$$

$$TI_{iu_k}^{X} = TI_{iu_k}^{X} + TO_{i1k}^{X} \quad (10)$$

In the formulas of this step, the range of variations of *i* and *k* is as follows:

$$i = 1,...,M \quad ; k = 1,...,N$$

All the above equations implemented to layer *L* to one and in every layer the calculation must be done for all the available switches on the layer. In equation (7) the amount of output traffic from the uplink port of switch *k* to the destination *i* obtain. $(\mathrm{Im})'_{ik}$ in this equation represent the disjoint of destination to this switch (this means the destination is in the other networks and above the switch). Equation (8) represent the overall output traffic from port one of switch *k* and the amount of this traffic is compared to $B_{1k} \times W_{1k}$ in terms of *Mbps* in the equation (9).If the amount of output traffic of this port was more than the allocated capacitance, then the traffic will pass with $W_{1k} \times \alpha_{1k}$ coefficient. In equation (10) the amount of input traffic to uplink switch of switch *k* is updated.

④.Move Down from the Roots in Hierarchy Structure and Updating Matrices.

In order to updating the input and output traffic to every switch ,the following formulas is used:

$$TO_{ijk}^{X} = TI_{ik}^{X} \times \mathrm{Im}_{ijk} \quad (11)$$

$$TTO_{jk}^{X} = \sum_{i=1}^{M} TO_{ijk}^{X} \quad (12)$$

$$TO_{ijk}^{X} = \begin{cases} TO_{ijk}^{X} & TTO_{jk}^{X=b} \le B_{jk} \times W_{jk} \\ TO_{ijk}^{X} \times W_{jk} \times \alpha_{jk} & otherwise \end{cases}$$

$$\alpha_{jk} = \frac{B_{jk}}{TTO_{jk}^{X=b}} \quad (13)$$

$$TI_{id_{jk}}^{X} = TI_{id_{jk}}^{X} + TO_{ijk}^{X} \quad , d_{jk} \neq 0 \quad (14)$$

In the formulas of this step, the range of variations of *i,j* and *k* is as follows:

$$i = 1,...,M \ ; j = 2,...,J_k \ ; k = 1,...,N$$

All the above equations must be implemented for the switches of layer zero to *L* and for all the available switches on that layer. In equation (11) the amount of output traffic from the port *j*(downlink) of switch *k* to the destination *i* obtain. Equation (12) represents the overall output traffic from port *j* of switch *k* and the amount of this traffic is compared to $B_{jk} \times W_{jk}$ in terms of *Mbps* in the equation (13).If the amount of output traffic of this port was more than the allocated capacitance, then the traffic will pass with $W_{jk} \times \alpha_{jk}$ coefficient . Finally in equation (14) the value of input traffic to downlink switch of switch *k* is updated and there is no need to calculate for availability ports.

⑤.Determine the Available Capacitance of every Element of Network.

The available capacitance of switches in terms of *pps* is calculated as below:

$$CR_k = (1 - g_k)C_k - \sum_{j=1}^{J_k}\sum_{i=1}^{M} TO_{ijk}^{X=p} \qquad k = 1,..., N \qquad (15)$$

In which g_k is the future growth coefficient for switch k . $CR_k \le 0$ means occupying the whole capacitance of switch and all the passing calls from this switch is eliminated from then. Matrix $TO_{ijk}^{X=p}$ represent the amount of output traffic from port *j* of switch *k* to destination *i* in terms of *Pps* which calculated in ③ , ④.

The available bandwidth of links connected to switches is In terms of *Mbps* and calculated as follow:

$$BR_{jk} = (1 - g_{jk})B_{jk} - \sum_{i=1}^{M} TO_{ijk}^{X=b} \qquad \begin{matrix} k = 1,..., N \\ j = 1,..., J_k \end{matrix} \qquad (16)$$

In which g_{jk} is the future growth coefficient for port *j* of switch *k* . $BR_{jk} \le 0$ means occupying the whole bandwidth of link and all the passing calls from this link is eliminated from then. Matrix $TO_{ijk}^{X=b}$ represent the amount of output traffic from port *j* of switch *k* to destination *i* in terms of *Mbps* which calculated in ③ , ④.

⑥.Determine the Number of Supported Calls in Terms of Bandwidth Limitation

Now, we calculate the equations related to the capacitance of network for delivering telephony services. The peak number of supported calls through network elements without considering the distribution of calls is calculated as bellow:

$$MaxCall_k = \frac{CR_k}{CallBW(Mbps)} \qquad k = 1,\ldots,N` \qquad (17)$$

$$MaxCall_{jk} = \frac{BR_{jk}}{CallBW(Mbps)}` \qquad \begin{matrix} k = 1,\ldots,N \\ j = 1,\ldots,J_k \end{matrix} \qquad (18)$$

In which the first equation is for the number of supported calls with switch *k* and the second is for the number of supported calls with port *j* of switch *k*. *CallBW (pps)* and *CallBW(Mbps)* represent the required bandwidth for a telephony conversation in terms of *Pps* and *Mbps* respectively and in accordance with section 2-1-1 a bandwidth equal to 100pps or 180.8kbps is needed for a two way connection for G.711 codec .

Knowing the distribution function of the calls in a network is an important step in evaluating the capacitance of network in delivering the telephony services. Considering f_k as a percentage of the whole calls which their pass is from the k^{th} switch of network and f_{jk} as a percentage of all calls pass from port *j* of switch *k*, then

$$MaxCall_S_k = \left(\frac{MaxCall_k}{f_k} \right) \qquad k = 1,\ldots,N` \qquad (19)$$

$$MaxCall_L_{jk} = \left(\frac{MaxCall_{jk}}{f_{jk}} \right)$$
$$k = 1,\ldots,N$$
$$j = 1,\ldots,J_k \qquad (20)$$

The first and second equations, show the peak number of supported calls through all switches and the ports connected to them in the network respectively. Finally the maximum number of supported calls through network is calculated as bellow:

$$MaxCall_BL = \min\left\{MaxCall_S_k, MaxCall_L_{jk}\right\}$$
$$k = 1,\ldots,N \qquad (21)$$
$$j = 1,\ldots,J_k$$

- **Delay Analysis**

As defined before, the maximum tolerable end-to-end delay for an IP telephony packet is 150 ms. The maximum number of calls that the network can sustain is bounded by this delay. As described in Section 2-1-1, there are three sources of delay: sender, network, and receiver. The end-to- end delay D for a telephony service in one direction from sender to receiver can be expressed by[5]:

$$D = D_{enc} + D_{pack} + \sum_{h \in path} \left(T_h + Q_h + PD_h\right) + D_{play} \quad (22)$$

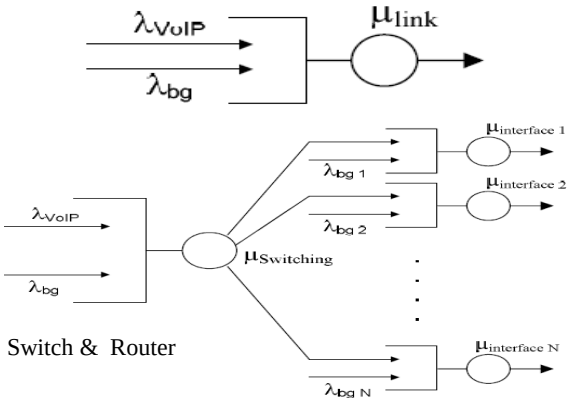
where D_{enc} is the encoder delay of converting A/D signal into samples at the source and D_{pack} is the delay due to packetizing at the source. In G.711, D_{pack} and D_{enc}, are 20 and 1 ms, respectively. In our analysis, we consider a fixed delay of 25 ms at the source, assuming worst case situation. D_{play} is the playback delay at the receiver, including jitter buffer delay. The jitter delay is at most 2 packets (in G.711, 40 ms) [9]. In our analysis, by considering the processing delay, we assume a fixed delay of 45 ms at the receiver. $\sum(T_h + Q_h + PD_h)$ 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. For example using G.711 codec and equation(22) it resulted This delay should not exceed 80 ms. With respect to this threshold, we are capable of estimating the number of calls in a given network.

In this part, the queuing theory is used for evaluating the queuing delay in a network. Using this approach, we can calculate the maximum supported calls by the network through reserving the network delay under permissive threshold (for example 80ms for G.711). To evaluate the network delay, queuing network analysis was used[11]. In this method, the arrival rate is assumed to be Poisson and the service times of network elements are exponentially distributed. this analytic approach is implemented in four following steps:

- Decomposing the network to subsystems (for example a single queuing node)
- Analyzing each subsystem individually considering the rates of receive and servicing
- Calculating the average delay of each subsystem individually
- Adding the delay of all queuing subsystems to find the end to end delay in the network.

Modeling receiving vocal packets with Poisson is a proper approximation, especially when there is a lot of calls. If the rate of receiving packets in the parts of network doesn't have Poisson , approximating delay was so difficult. Figure 6 shows the queuing model for router, switch and connection. The number of output queues for switches and routers are depending to the number of their gates. In this paper, routers and switches are modeled with M/M/1 queue(considering FIFO queuing in output connect) and connections with M/D/1 queue.

In M/M/1, the service time is not deterministic for routers and switches (because ths two parts are based on cpu). In M/D/1 queue , time of servicing to packets is deterministic.



Figure(6). queuing model for router, switch and connection

According to [12] :

$$D_{M/M/1} = \left(\frac{1}{\mu - \lambda} \right) \qquad (23)$$

$$D_{M/D/1} = \left(\frac{1 - \lambda/2\mu}{\mu - \lambda} \right) \qquad (24)$$

while $\lambda = \lambda_{VoIP} + \lambda_{bg}$,where λ_{VoIP} is the total added new traffic from a single VoIP and λ_{bg} is the background traffic. μ is the mean network element service rate network elements. In the presented queuing models, the arrival rate for voice and data traffic is assumed to be Poisson . However in practice, background traffic is bursty in nature. For our analysis and design, using bursty background traffic is not practical. For one thing, under the network of queues being considered an analytical solution becomes intractable when considering non-Poisson arrival. Also, it is important to remember that in order to ensure good QoS at all times, we base our analysis and design on the worst-case scenario of network load or utilization, i.e. the peak of aggregate bursts. And thus in a way our analytical approach takes into account the bursty nature of traffic.

To determine the maximum number of calls that can support through the network with respect to the delay restriction, the following algorithm is presented that is a modified version of [5]. In this modified algorithm the traffic is end to end and other queuing models (besides FIFO which studied in [5]) is evaluated. The telephony calls is added to the network continuously until the network delay reaches to maximum permissive amount (for example 80 ms for G.711). the algorithm is presented in 6 steps:

① **Step 1:** the beginning of algorithm implementation, all the nodes in the network is numbered. At first, there isn't any call in the network and only the background traffic is available. The traffic is usually considered between end points. Using the steps 1-4 of the presented algorithm in section bandwidth analysis, the amount of traffic load on every element of the network can be determined. In the previous section, the amount of output traffic from a switch or a switch port is calculated . In the algorithm of this section with a proper mapping, matrices *TS* and *TL* is derived from matrix *TTO* to express the amount of load on the switches and connections. Matrix TS is used for the amount of load on the network nodes (containing end points, switches and routers). TS_k entry shows the amount of traffic on the k^{th} node In terms of *Pps*. To express the amount of traffic on the connections, matrix TL is used which is a *N*N* matrix. *TLij* is the amount of traffic between *i* and *j* nodes in terms of *Mbps*.

The total capacitance of nodes and connections are determined through matrixes μS and μL. Entries μS_k and μL_{mn} show the switching speed of the k^{th} switch and connection bandwidth between nodes *m* and *n*, respectively. (these matrices are determined with respect to the growth factor from the network elements guideline which introduced in section of bandwidth analysis). For example in a node with switching speed of 25000pps and growth factor of 0.25, is calculated as follows:

$$\mu S_k = (1-0.25)\times 25000 = 18750\, pps$$

And in a connection with the bandwidth of 100Mbps and growth factor of 0.1, μL_{mn} can be calculated as follows:

$$\mu L_{mn} = (1-0.1)\times 100 = 90 Mbps$$

The entries of matrices λS (1*N matrix) and λL (N*N) are initialize in the beginning of operation ,which represent the average arrival rate of packets in switch and bits in connections respectively. Delay in nodes and connections are represented with matrices *DS* and *DL*. The entries DS_k and DL_{mn} show the delay of a packet in the k^{th} switch and the delay of each bit of the traffic in the connection between the nodes *m* and *n* respectively. In the following sections, the letter h in the symbol X^h represent the number of repeat in the algorithm implementation.

$$\lambda S_k^1 = TS_k \qquad k = 1,...,N \tag{25}$$

$$\lambda K_{mn}^1 = TL_{mn} \qquad m = 1,...,N \quad n = 1,...,N \tag{26}$$

$$DS_k^1 = 0 \qquad k = 1,...,N \tag{27}$$

$$DL_{mn}^1 = PD_{mn} \qquad m = 1,...,N \quad n = 1,...,N \tag{28}$$

$$CallNum^1 = 0 \tag{29}$$

CallNum Shows the number of calls in every algorithm implementation. The initial value of delay in the switches is assumed to be zero and in the connections this value is equaled to the propagation delay that stated in the matrix *PD*. For calculating the propagation delay in a connection, the following equation is used:

$$PD_{mn} = \frac{Dis(m,n)}{V(m,n)} \tag{30}$$

The numerator is the distance between the nodes *m* and *n*. *V(m,n)* represent the speed of the light that dependent on the kind of connection.

②**Step 2**: a new call based on call distribution is added

$$CallNum^{h+1} = CallNum^h + 1 \tag{31}$$

which $CallNum^{h+1}$ is the number of calls in the $h+1^{th}$ repeat of algorithm. For adding a call in a specific path, first the the path matrix *P* must be generated, using the spanning tree matrix introduced in section 2-3-1. Every row of this matrix represent the path between the two end point which a call established between them and the entries of every row are the numbers of the available nodes in the path.

③**Step3:** the amount of traffic for each network element in the path which the call is added is updated that stated as follows:

$$\lambda S_k^{h+1} = \lambda S_k^h + \lambda S_{VoIP} \qquad k \in P_i \tag{32}$$

$$\lambda L_{mn}^{h+1} = \lambda L_{mn}^h + \lambda L_{VoIP} \qquad m,n \in P_i \tag{33}$$

Equation (32) calculates the amount of total traffic passing from the k^{th} switch in which λS_{VoIP} is the amount of voice traffic passing from the switch per every call (for example in a call with the G.711 decoder and 20 ms frames, the amount of this traffic for the router and switch is 100pps). For determining the total traffic passing from the connection between nodes *m* and *n* the equation (33) is used. λL_{VoIP} represent the amount of unidirection traffic passing from the connection and is 90.4Kbps for uplink and downlink connections.

④**Step4:** in this step the matrix *DS* is updated using the following equation:

$$DS_k^{h+1} = DS_k^h + \left(\frac{1}{\mu S_k - \lambda S_k^{h+1}} \right) \qquad k \in P_i \tag{34}$$

the entries of the matrix *DL* is also calculated from the equation .

$$DL_{mn}^{h+1} = DL_{mn}^h + \left(\frac{1 - \lambda L_{mn}^{h+1} / 2\mu L_{mn}}{\mu L_{mn} - \lambda L_{mn}^{h+1}} \right) \qquad m, n \in P_i \tag{35}$$

Equation (34) shows the amount of delay considering the FIFO queuing in the output connections of routers and switches. In this section the objective is to study the number of supported calls in the worst case . With this situation and considering the use of PQ queuing , only the voice traffic is passing through the output queue and the average service rate is equal to the whole connection capacity (minus the specified capacitance for the future growth). In the worst case of WFQ queuing when the amount of traffic on output connection is more than the bandwidth of connection, the bandwidth is distributed between traffics based on ToS. So in the case of using this queue for calculating delay in the worst case, the average service rate is determined based on the weighted coefficients designated to the telephony traffic.

⑤**Step 5:** the amount of network delay in the path *i*, calculated as follows:

$$DP_i^{h+1} = \sum_{k \in P_i} DS_k^{h+1} + \left(\sum_{m,n \in P_i} DL_{mn}^{h+1} \right) \times (Pb)$$
$$i = 1, \ldots, \max_length(P) \tag{36}$$

In which DP_i^{h+1} is the delay of established voice call in the path *i* and $h+1^{th}$ repetition of algorithm. The first and second statements in the equation (36) declare the total delays resulting from the nodes in the path and the total connection delays of the path respectively. Considering the fact that the first and second statements show the amount of delay in terms of a packet and a bit respectively, for correcting the above statement, the units must be equals and this can be done by multiplying the connection delays in terms of bits in the number of bits of a voice packet*(Pb)*. (for example for a 20 ms voice packet with the G.711 decoder, the number of bits is 1808).

$$Max_Delay = Max(DP)_i$$
$$i = 1, \ldots, \max_length(P) \tag{37}$$

⑥**Step 6**: in this step, the maximum delay in the network is calculated with respect to the calculated one in step 5. If it wasn't more than the specified threshold (80 ms for G.711), the algorithm came back to step 2, otherwise the maximum supported calls through the network considering the restriction of delay is calculated as follows:

$$MaxCalls_DL = (CallNum^{h+1} - 1) \tag{38}$$

After calculation the maximum supported calls through the network considering the restriction resulted from the bandwidth and delay in equations (21) and (38) respectively, the capacitance of a given network in delivering telephony services is declared as follows:

$$MaxCalls = \{MaxCalls_BL, MaxCalls_DL\} \quad (39)$$

2-3-2- Qualitative Criterion

In communication networks, E-model is widely used to automatically evaluate the quality of speech. The output of this model is a parameter called *R* that can be mapped to *MOS*.

The simplified version of this model is often used in VoIP. This version of the E-model uses some assumptions. For example, it assumes no echo effect, no environmental noise on a sender and receiver sides, constant and sufficient loudness level, etc. These assumptions are reasonable and suitable for our purposes. The VoIP version of the E-model includes several components:

$$R = R_o - I_d - I_{e-eff} \quad (40)$$

R is the indicator of the speech quality (from 1 to 100), R_0=93.2 is the maximum grade can be obtained by using an encoder according to *MOS=4.41*, I_d is the disturbance factor results from end-to-end delay. E-model uses complicated equations to calculate the impact of delay on speech quality. Cole and Rosenbluth introduced a linear function to describe the delay disturbance [10]:

$$I_d = 0.024D + 0.11(D - 177.3).H(D - 177.3) \quad (41)$$

D is the end-to end delay and *H(x)* is:

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$$

In Equation (40), $I_{e\text{-}eff}$ is defined as:

$$I_{e-eff} = I_e + (95 - I_e)\frac{Ppl}{Ppl + Bpl} \quad (42)$$

where I_e is equipment disturbance factor, *Bpl* is the packet loss robustness and *Ppl* is the packet loss rate in percent. The lost packets can be recovered using PLC algorithms. Different encoders use different PLC techniques. The specification of encoders is described in G.113. For each encoder, ITU-T suggests special values of *Bpl* and I_e (Table1). In this table, the values of equipment disturbance factor, I_e, is defined in respect to the quality of G.711 (worse encoder has higher value of I_e).

Table 1- Equipment disturbance and packet loss robustness factors

codec	PLC type	Ie	Bpl
G.711	non	0	4.3
G.723+vad	native	15	16.1
G.729a+vad	native	11	19.0

E-model proposes a mapping rule between *R* and *MOS*. The range of *MOS* is considered from 1 to 4.5:

$$MOS = \begin{cases} 1 & R < 0 \\ 1 + 0.035R + R(R - 60)(100 - R) \times 7.10^{-6} & 0 \le R < 100 \\ 4.5 & R \ge 100 \end{cases} \quad (43)$$

In real networks, usually we measure *MOS* in short time periods. Rosenbluth proposed a Extended_E-model to calculate total *MOS* in a telephone call by weighted averaging of the *MOSs* measured in short time periods [13]:

$$MOS_I = \frac{\sum_i W_i.MOS_i}{\sum_i W_i} \quad (44)$$

Here, MOS_I is the total call *MOS*, MOS_i is the *MOS* measured in a short time period by using Equations (40)-(43), L_i is the location of the time period (between 0 to 1, 0 indicates beginning of the call and 1 indicates the end of call). The weights can be calculated by:

$$W_i = \max\left[1,1+\left(0.038+1.3L_i^{0.68}\right).\left(4.3-MOS_i\right)^{\left\{0.96+0.61L_i^{1.2}\right\}}\right] \quad (45)$$

2-2-3- Step4: Simulation

In the proposed algorithm, the simulation includes two parts:

- Simulation in order to evaluation of quantity factors.
- Simulation in order to evaluation of quality factors.

In the first phase, simulation is done through MATLAB in order to implement the proposed algorithm for evaluating the bandwidth and delay limits. In this phase , the given network is simulated via OPNET and compares with the previous results to evaluate the correctness of the proposed algorithm. OPNET is a powerful software in network field. It brings the simulation environment to the real network because of it's comprehensive libraries of protocols and commercial models of network elements. In the second phase, the number of supported calls through network is obtained with presented extended_E_ model in section 2-3-2 and compares with the results of the first phase in order to comparing the results of quatity and quality criterions.

2- Case Study

Figure 7 illustrates a typical network topology that connects two local networks together. These networks are Ethernet-based. All of the links within the local networks are assumed 100 Mbps and full duplex. The link between routers is 44.736 Mbps. The routers are assumed Cisco 2621, and the switches are assumed 3Com Superstack 3300. Here, VoIP gateway connects the enterprise network to the PSTN network. Branch 1 has two Ethernet switches connected together by a router and branch 2 has one Ethernet switch. This topology is a extended version of the presented one in the [5].In every floor of the branches, we assumed three workstations connected to an Ethernet switch. One workstation is the source of calls, another is the sink of receiving calls and third one is the source and sink of background traffic. As an example see first floor of branch 1 in Figure 8. F1_C1 workstation is the source of calls, F1_C2 is the sink of calls and F1_C3 is the source and sink of background traffic. The growth factor is 0.1 for all of network elements except Routers that this factor is 0.25 for them.

In this paper, we ignore the signaling traffic generated by the gatekeeper. The signaling traffic mostly presents prior to the establishment of the voice call and when the call is finished and this traffic is small compared to the voice call traffic. For a successful deployment of telephony service, first we must know the call distribution function of the branches and the data traffic volume on the network links. There are a number of tools available to perform network measurements. Figure 9 describes the call distribution for the network under study. Also, Table 2 shows peak-hour utilization of network endpoint in both directions.

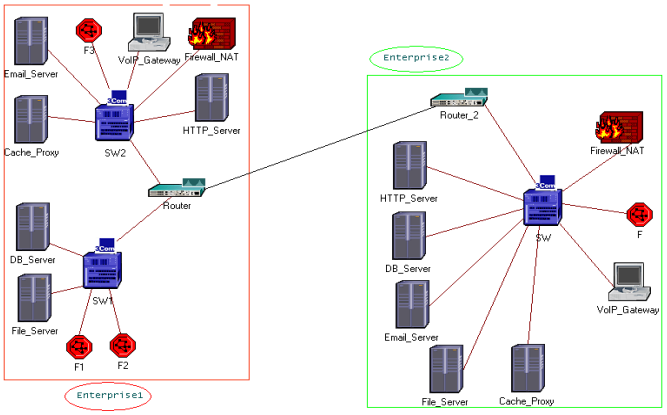


Figure (7). Network Topology

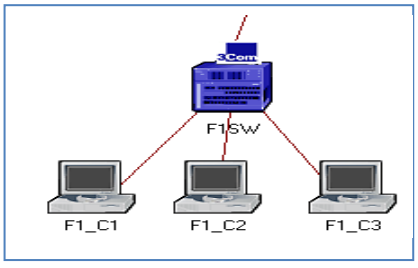


Figure (8). Floor 1 subnet model

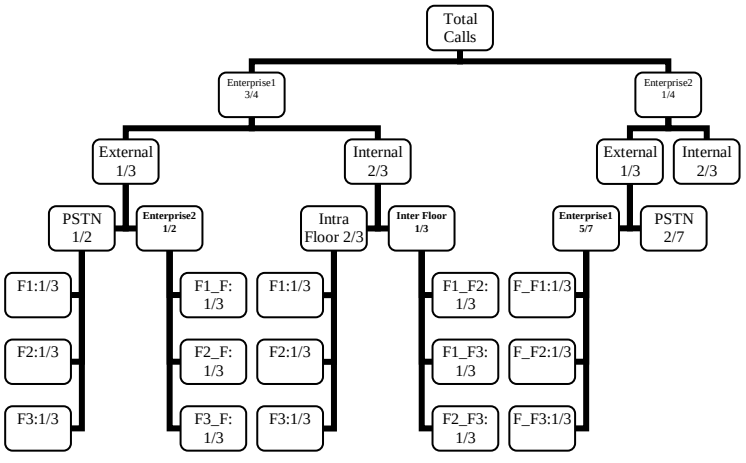


Figure (9). Call Distribution tree

3- Simulation Result

4-1- Simulation for evaluating quantity factors

This section includes evaluating bandwidth and delay factors based on the algorithm of section 2-3-1 using MATLAB and OPNET softwares.

Table 2- End_to_end Traffic

Traffic	Bitrate (Mbps)	Packetrate (pps)	Traffic	Bitrate (Mbps)	Packetrate (pps)
Branch1.Related with Floor1			Branch1.Related with Floor2		
F1↔Cache_Proxy(B1)	0.9	90	F2↔Cache_Proxy(B1)	0.8	80
F1↔DB_server(B1)	1.2	80	F2↔DB_server(B1)	0.75	70
F1↔Email_Server(B1)	1.2	90	F2↔Email_Server(B1)	0.75	60
F1↔File_Server(B1)	0.8	75	F2↔File_Server(B1)	0.7	40
F1↔HTTP_Server(B1)	0.8	95	F2↔HTTP_Server(B1)	0.6	45
F1↔Firewall_NAT(B1)	1	100	F2↔Firewall_NA(B1)T	0.5	30
F1↔F2	0.5	50	F2↔F1	0.5	50
F1↔F3	0.5	50	F2↔F3	1	80
F1↔F	20	1000	F2↔F	10	400
Branch1.Related with Floor3			Branch2.Related with Floor		
F3↔Cache_Proxy(B1)	0.2	16	F↔Cache_Proxy(B2)	1.65	140
F3↔DB_server(B1)	0.2	22	F↔DB_server(B2)	2	138
F3↔Email_Server(B1)	0.1	41	F↔Email_Server(B2)	1.98	165
F3↔File_Server(B1)	0.3	38	F↔File_Server(B2)	1.5	125
F3↔HTTP_Server(B1)	0.4	21	F↔HTTP_Server(B1)	1.5	133
F3↔Firewall_NAT(B1)	0.6	50	F↔Firewall_NAT (B1)	1.86	150
F3↔F1	0.5	50	F↔F1	20	800
F3↔F2	1	80	F↔F2	10	400
F3↔F	8	200	F↔F3	8	200

- **Simulation with MATLAB**

Figure 10 shows the network graph based on the rules defined in section 2-3-1. In this graph , switches , endpoints and ports are numbered. Tables 3 and 4 represent the available bandwidth in every switch and some of their ports and also the number of supported calls in different conditions after accomplishing the algorithm .It can be seen that in the case of FIFO queuing, the limitation is due to the bandwidth of two-branch enterprise connection link and the number of supported calls is 202 calls. By applying PQ and WFQ to the router ports connected to this link,this limitation is ignored and the speed of switching of the router1 is considered as the

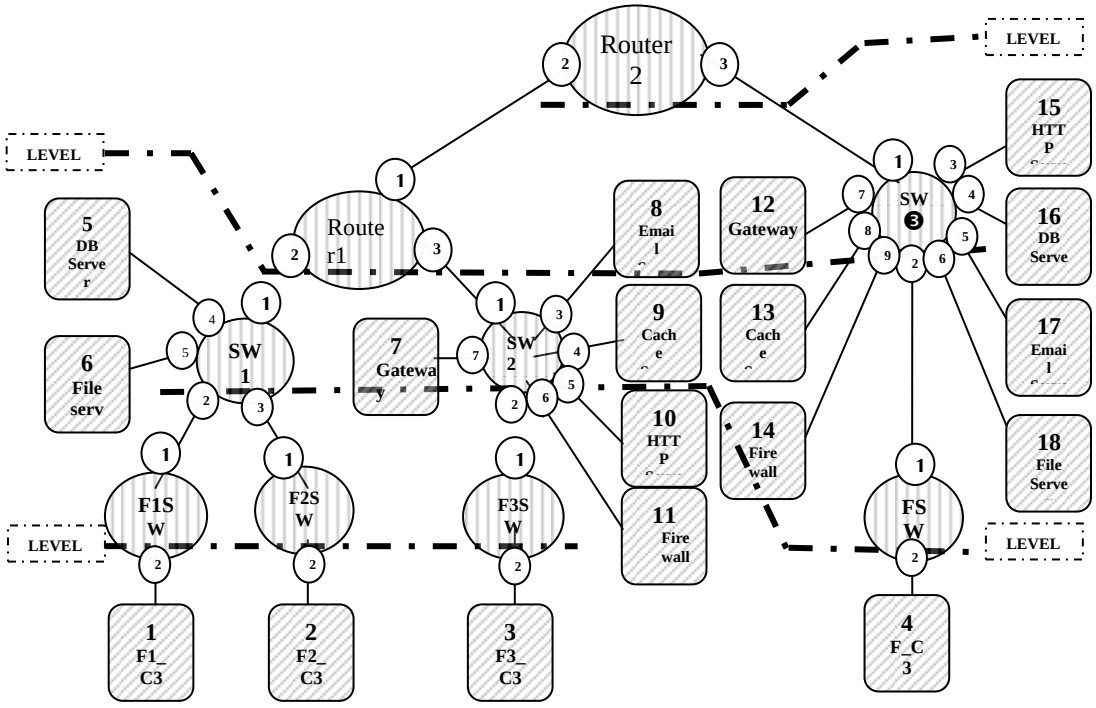


Figure (10). Graph Topology tree

Table 3- Remaining Capacity and Maximum VoIP Calls Supported by Switchs

Switch(k)	fk	CRk (pps) (FIFO)	MAXCALL _Sk (FIFO)	CRk (pps) (WFQ)	MAXCALL _Sk (WFQ)	CRk (pps) (PQ)	MAXCALL _Sk (PQ)
Router2❶	31/168	15550	843	18200	987	18750	1016
Router1❷	101/168	13990	233	16600	277	17190	286
SW ❸	35/168	1165098	55925	1166400	55989	1166698	56002
SW1❹	47/168	1165010	26770	1166200	26797	1166410	26803
SW2❺	143/504	1167784	41158	1168000	41164	1167984	41165
FSW❻	1/4	1165098	46604	1166400	46657	1166698	46668
F1SW❼	82/252	1166740	35856	1167600	35881	1167740	35887
F2SW❽	82/252	1168290	35904	1168600	35914	1168690	35916
F3SW❾	82/252	1168964	35924	1169100	35929	1169164	35930

Table 4- Remaining Capacity and Maximum VoIP Calls Supported by Links that Connected to Switchs

Switch(k)	PORT (j)	fjk	BRjk (Mbps) (FIFO)	MAX CALL_Ljk (FIFO)	BRjk (Mbps) (WFQ)	MAX CALL_Ljk (WFQ)	BRjk (Mbps) (PQ)	MAX CALL_ Ljk (PQ)
Router2❶	2	**5/84**	2.2624	420	33.87	6294	40.2624	7482
	3	**1/8**	52	4601	83.57	7395	90	7964
Router1❷	1	**1/8**	2.2624	202	33.87	2997	40.2624	3563
	2	**19/126**	51.32	3764	76.2448	5593	81.32	5965
	3	**53/252**	73.32	3856	79.96	4206	81.32	4277
SW ❸	1	**5/84**	52	9663	52	9663	52	9663
	2	**87/504**	41.51	2660	73.08	4683	79.51	5095
SW1 ❹	1	**5/28**	51.32	3179	51.32	3179	51.32	3179
	2	**87/504**	63.1	4043	79.71	5108	83.1	5325
	3	**87/504**	74.4	4767	82.71	5300	84.4	5408
SW2 ❺	1	**11/72**	73.32	5308	73.32	53308	73.32	5308
	2	**87/504**	78.36	5021	85	5447	86.36	5534

limiting factor and the number of supported calls is reached to 277 and 286.

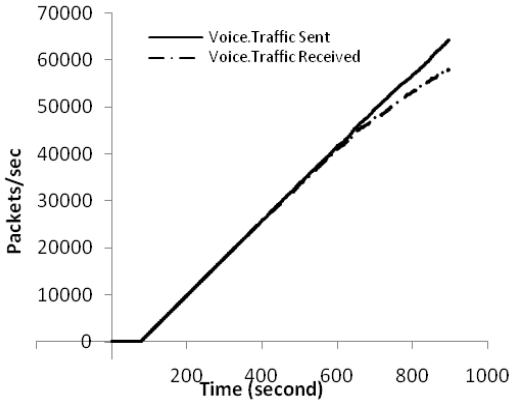
Following the trend of the presented delay analysis in section 2-3-1 ,it can be seen that the maximum supported calls considering the delay restriction are 223, 1650 and 1570 calls for FIFO, PQ and WFQ queuing respectively and after that the delay is increasing dramatically. Comparing the results from analyzing the bandwidth and delay we can see that in the given network ,the restriction resulted from the bandwidth is more than that of delay.

- **Simulation with OPNET**

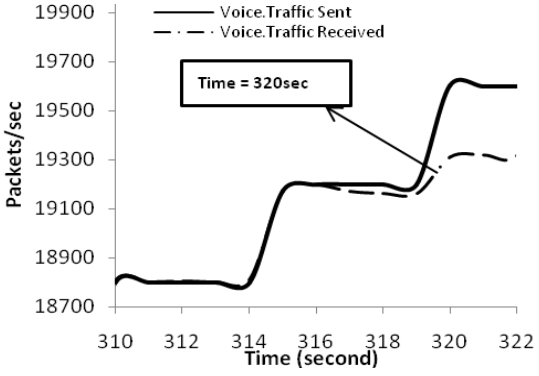
The duration of the OPNET simulation is 10 min. The generation of background traffic started at 40s from

the start time of the simulation. The voice traffic started at 70s at which a total of four bi-directional calls are initially generated. Then, every 5 seconds, four calls are added. The Simulation stops at 10 min in which a total of $4+((10\times60 - 70)/5)\times4=428$ calls are generated. The simulation is done with FIFO queuing, G.711 codec and 20ms frames. First, we consider the packet loss restriction. This parameter must be below %1.Total sent and received telephony traffic is shown in Figure 11a. Figure 11b is a zoom-in version of Figure 11a ,focusing on the mismatch region between traffic sent and received. From Figure 11b it can be seen that the last successful call is at 320 sec (5 min and 20 sec). Therefore, the maximum number of calls considering the packet loss restriction is 204.

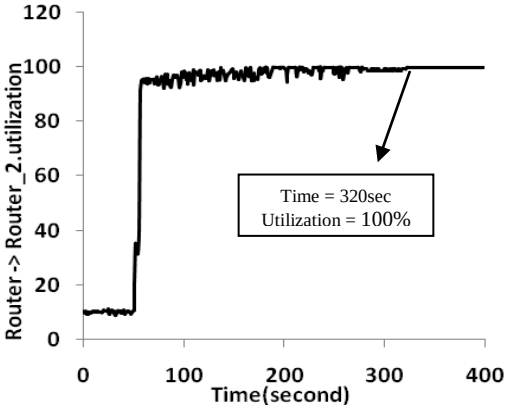
In this network, the bottleneck element is the link between two routers. As you can see from Figure 11c, this link is saturated at 320 sec and after this time, packets are dropped. Now we consider the delay restriction. As we mentioned before, the network delay must be below 80 msec. Figure 10d shows network delay of the voice packets.The maximum number of calls considering this restriction is 212. Therefore, the restriction comes from the packet loss is more than that comes from delay.


Voice.Traffic Sent
Voice.Traffic Received
Packets/sec
70000
60000
50000
40000
30000
20000
10000
0
200
400
600
800
1000
Time (second)

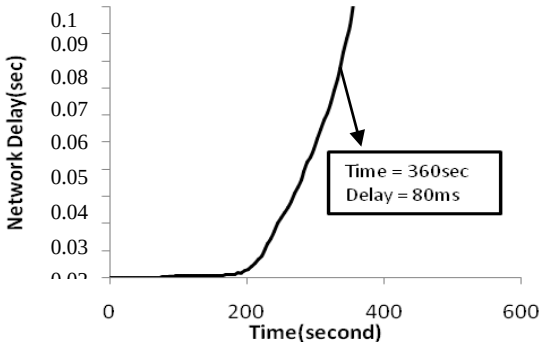
(a)


Voice.Traffic Sent
Voice.Traffic Received
Time = 320sec
Packets/sec
19900
19700
19500
19300
19100
18900
18700
310
312
314
316
318
320
322
Time (second)

(b)


Router -> Router_2.utilization
120
100
80
60
40
20
0
0
100
200
300
400
Time(second)
Time = 320sec
Utilization = 100%

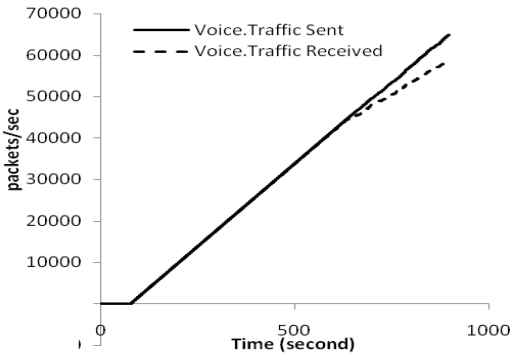
(c)



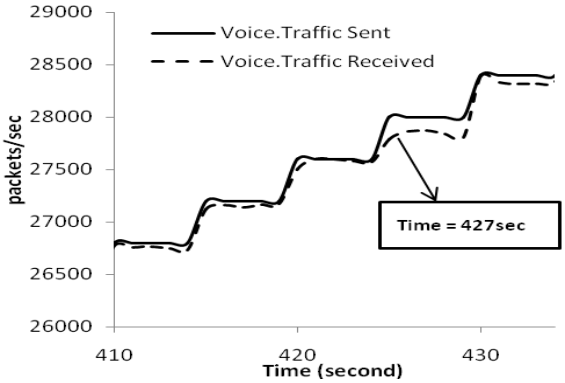
(d)

Figure (11). (a) Sent and received voice traffic with FIFO queue, (b) zoom-in version of Figure 11a, (c)Utilization of the link between routers with FIFO queue , (d) Voice network delay with FIFO queue

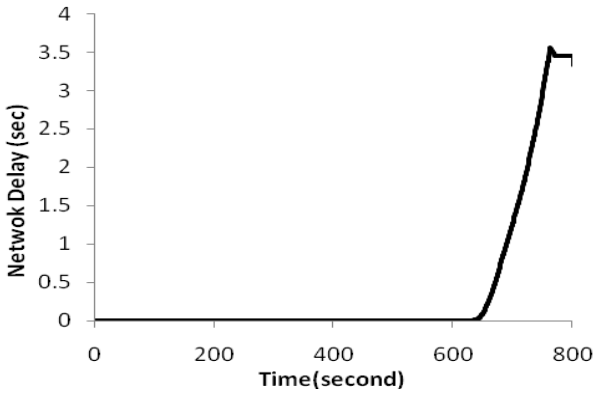
By applying PQ and WFQ methods to Router1-Router2 link and simulation , it can be seen that the number of successfull calls considering the bandwidth limitation will reach to 290 (figure 12b). which is close to result of simulation by MATLAB. In figure 12c the network delay of voice packets is shown . it can be seen that delay of voice packets is very negligible and the main limitation is packet losses ,not the delay.



(a)



(b)



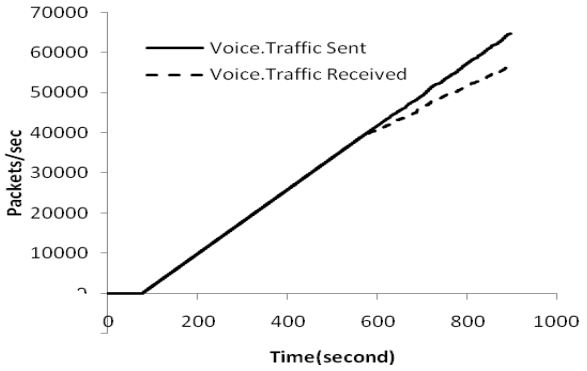
(c)

Figure (12). (a) Sent and received voice traffic with PQ, (b) zoom-in version of Figure 12a

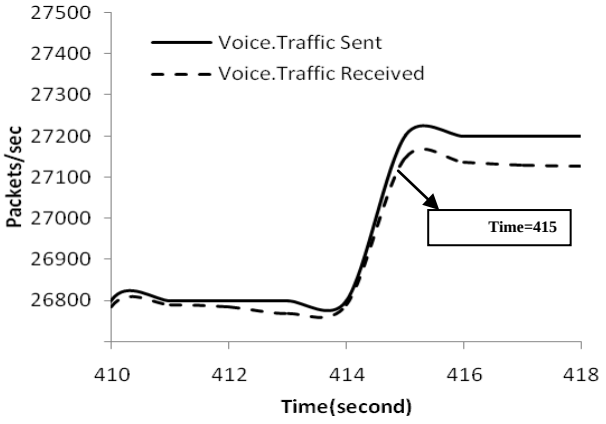
(d) Voice network delay with PQ

It can be seen by reaching the delay to nearly 3.5 sec, the amount of delay will be fixed. After the saturation of the router one and not servicing to the traffic, the network is reached to the maximum amount of delay and it will be fixed in this stage.

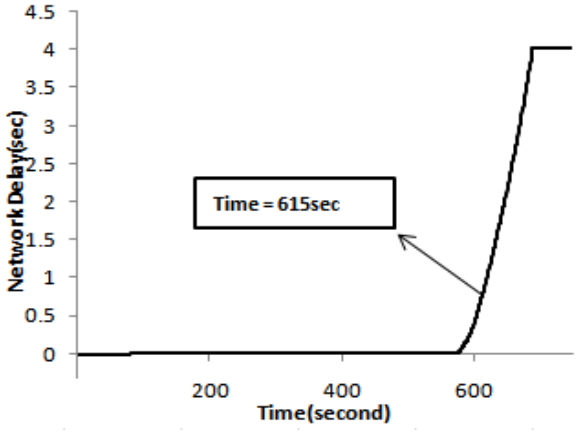
As we can see in figure 13b by applying the WFQ queuing to the router, the last successful added call is in 415 sec. so considering the restriction of the bandwidth and using the WFQ queuing, the number of calls will be reached to 280. In figure 13c the end to end delay of the voice packets is shown.



(a)



(b)



(c)

Figure (13). (a) Sent and received voice traffic with WFQ, (b) zoom-in version of Figure 13a, (d) Voice network delay with WFQ

Like the PQ queuing, because of the kind of queuing, the delay of the voice packets is so neglible and the main restriction is the packet loss not delay. Comparing the results from the PQ and WFQ queuing, it can be seen that the resulted numbers are nearly the same ,while in the PQ queuing, because the entire link bandwidth between the 2 branch is allocated to the voice traffic, the number of calls must be more than the WFQ queuing. The neglible difference between the results is due to the fact that in both the PQ and WFQ queuing , the main restriction is resulted from the switching speed of the Router1 and not the link between the 2 routers. By saturation of the Router1, independent of the queuing, no more call can be add to the network successfully.

Comparing the results from the simulation with OPNET and the presented algorithm it can be seen that the results in the bandwidth domain is nearly the same. While in the delay, in the case when the FIFO queuing is used, the results is also the same and otherwise results from the analytic approach and simulation have a considerable difference. In simulation with OPNET ,the time of reaching the voice packets are modeled with possion .so by using the PQ and WFQ queuing in the connection of the 2 routers and when the routers doesn't receive a new voice packet, servicing to the data traffic is done. While in the analytic approach presented in this article, the rate of voice traffic is considered determinestic and so by considering the worst case in the PQ queuing, it is assumed that no data traffic is passed through the network. Also In WFQ queuing (assuming the worst case) the amount of passing the data traffic is always a percent of the link bandwidth. So considering the delay restriction, results from the simulation, show the capacitance of the network using the PQ and WFQ queuing less than the amounts calculated with the analytic approach.

4-2- E_model Analysis

Figure 14 shows the MOS variant in terms of the number of added calls to the network for different queuing. Maximum supported calls through network with considering MOS=4.2 as the acceptable quality is represented in table 5.The difference in results comes from the fact that quantity criterions only consider the packet loss or delay separately; however with considering both parameters in E-model, the result will be more accurate.

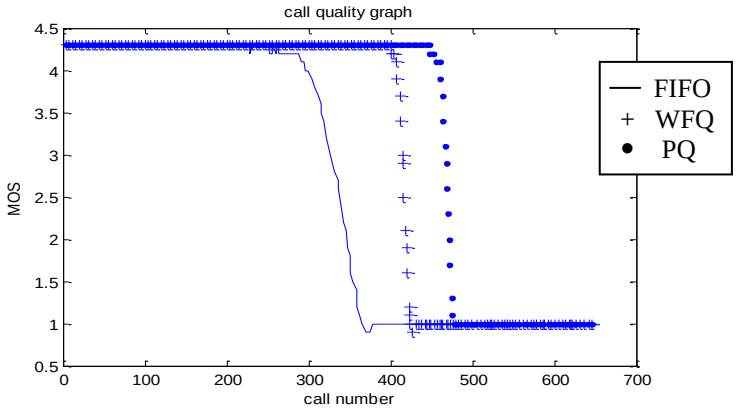


Figure (14). MOS graph in terms of the number of calls for different queuing

Table 5- Extended_E_model results

Switch(k)	FIFO	WFQ	PQ
MAX CALLS (MOS=4.2)	228	400	448

4- Conclusion

In this paper, a new approach based on the quantitative and qualitative analysis is presented in order to evaluate the readiness of an Ethernet network for telephony service. The quantity criterions are bandwidth and delay that evaluate with new method. On the other hand In order to calculate the call capacity of the network base on qualitative analysis , Extended_E-model is used. As a case study, we considered an interconnection of two local networks. By applying quality and quantity criterions to the network, it can be seen that the results have a little difference with each other. The difference in results comes from the fact that quantity criterions only consider the bandwidth or delay limits separately; however with considering both parameters in E-model, the result will be more accurate.

References:

D. Wright, “Voice over MPLS compared to voice over other packet transport technologies”, IEEE Communications Magazine, 2002, vol. 40, no. 11, pp. 124-132.

E. Myakotnykh, Adaptive Speech Quality in Voice-over-IP Communications, Ph.D. Thesis, University of Pittsburgh, 2008.

E. Nik Maleki, G.Mirjalily, “Comprehensive Evaluation of Ethernet Networks for Telephony Service”, Majlesi Conference on Electrical Engineering, 2014.

G. Mirjalily, P. Bastani, and M.T. Almodarresi, "Connecting telephone exchange centers over metro Ethernet networks by using Diffserv-MPLS", Int. Conf. on Computer and Information Technology, 2008, pp. 708-713.

J. H. Rosenbluth, “Testing the Quality of Connections having Time Varying Impairment”, Committee contribution T1A1.7/98-031, 1998.

K. Salah, “On the deployment of VoIP in Ethernet networks: methodology and case study”, Computer Communications, 2006, Vol.29, pp. 1039–105.

K.M. Chandy, C.H. Sauer, “Approximate methods for analyzing queueing network models of computing systems”, Journal of ACM Computing Surveys , Vol. 10 , Issue 3, pp. 281–317, 1978.

L. Kleinrock, “Queueing Systems: Theory”, vol. 1 ,Wiley, New York,1975.

M. Karam, F. Tobagi, "Analysis of delay and delay jitter of voice traffic in the Internet" , Computer Networks Magazine, 2002, Vol.40, pp. 711–726.

Nagano, Takeshi, and Akinori Ito. "Packet loss concealment of voice-over IP packet using redundant parameter transmission under severe loss conditions."Journal of Information Hiding and Multimedia Signal Processing , 2014, vol. , no. 2.

P. M. V. Nair, Quality of Service in Metro Ethernet, Ph.D. Thesis, Southern Methodist University, 2006.

R. Cole and J. Rosenbluth, “Voice over IP performance monitoring”, Journal on Computer Communications Review, 2001.

Recommendation G.107, “The E-model, a computational model for use in transmission planning”, ITU-T, 2003.