ANALYTICAL METHOD FOR MODELING PBX SYSTEMS FOR SMALL ENTERPRISE

Lawan A. Mohammed and Kashif Munir

University of Hafr A-Batin, 31991, Hafr Al-Batin, Saudi Arabia

ABSTRACT

The current technology advancements have enabled Telephony system manufactures to Integrate IP with existing telephony networks and offer converged solution for more optimum utilization of organization's IT infrastructure as well as internet bandwidth. The paper proposes an analytical method for determining basic traffic characteristics of systems servicing multi-rate overflow traffic streams generated by a finite number of sources. In particular, we describe an optimum solution for the implementation of telephony system. Analytical results of blocking probability calculated using the presented methodology has been compared with the data obtained from the system simulation process.

KEYWORDS

Loss system, blocking probability, grade of service, finite source, finite storage and finite server

1. INTRODUCTION

Internet has become the backbone of day to day business processes for businesses of all sizes. There are various means by what enterprises get connected to their stakeholders over internet. The latest technology has provided diverse media for availing internet access either at home or office and even during transit. Cost of IT infrastructure and quality of internet services, account major share in organizational overheads. In major cases the infrastructure laid down for data access and bandwidth availed for day to day business processes are not optimally utilized, so the IT infrastructure becomes a non-optimized asset.

A PBX is a Private Branch Exchange, a hardware system that handles routing and switching of calls between a business location and the telephone network. Their name originates from the way they interact with the PSTN (public switched telephone network).PBX systems enable employees to call other employees, and businesses can route phone numbers to intercom systems, enabling broader communication. One of the primary benefits of these systems is that employees can share phone lines when placing outgoing phone calls, which allows businesses to reduce the number of lines for which they have to pay.

For these reasons, small to large enterprises, now days all prefer IP telephony enabled PBXs for their telecom infrastructure. Enterprises that have upgraded to pure or hybrid IP-PBXs are looking for functionalities and features to optimize their investment in infrastructure, drive call

cost reduction, reduce overheads and increase employee productivity. Use of IP Telephony offers all of these and is therefore the unrivaled communication technology now and for future.

Some of the common PBX Features as reported in [1] include the following:

- Automatic Route Selection: (Also called Least Cost Routing) This feature allows a system administrator to program a sequential list of outgoing routes usually configured based on the cost incurred by the organization.
- Time of Day Routing (TOD): This feature allows a system administrator to apply time of day restrictions to the route selection process.
- Access Restrictions: This feature allows a system administrator to limit a terminal devices' access to the public network, private network, and certain services and features.
- Authorization Codes: (Also called "auth codes") These codes are used to identify and control users.
- Automatic Call Distribution (ACD): A feature used when a large number of incoming calls are queued and answered by a group of telephones (referred to as agents). The longest waiting call is sent to the agent that has been idle for the longest time.
- Call Forward: This feature allows a user to manually forward incoming calls to an internal extension or a valid external number.
- Hunting: This feature allows incoming calls that encounter a "busy" or "no answer" extension to route automatically to another extension in a predefined path, known as the hunt chain.
- Call Transfer: This feature allows an internal user (Caller A) on a two-way call to put the existing party (Caller B) on hold and place a call to any valid internal extension or external number (Caller C). Once the Caller C answers the call, the internal user (Caller A) can transfer the existing party, Caller B, to Caller C.
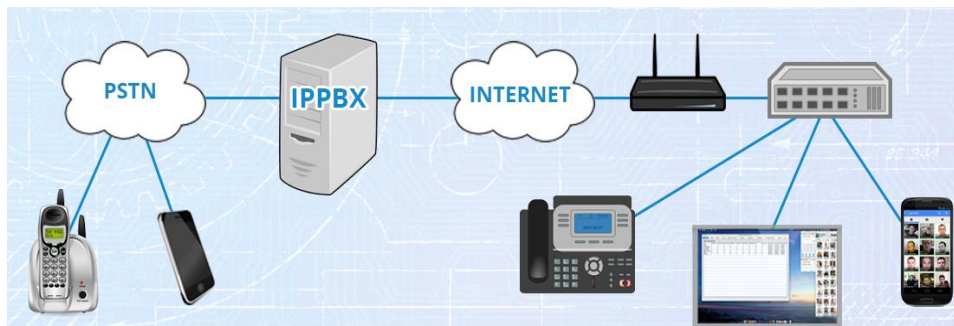

Figure 1. An Overview of IP- PBX System (source: www.bicomsystems.com)

Private Automatic Branch Exchange (PABX) is another telecommunication technology that is known as private switching system. The main advantage of PABX is that it essentially takes the place of the phone company's Central Office within the company by acting as the exchange point, routing calls. With a PABX in place, each phone only needs an extension, not a phone number, and the PABX handles all calls made from desk-to-desk within the company.

Normally a telephone line is connected to the phone company's local Central Office through the trunk. The Central Office is responsible for routing incoming and outgoing calls. It also provides other services like voice mail, call forwarding, caller ID and so on. For this service the phone

company receives a monthly fee. If a company requires dozens or even hundreds of phone lines, this will quickly incur a very large phone bill.

For this reason, a PABX reduces cost because the company only pays for the number of lines liable to be connected at any given time to the outside. For example, if a company has 200 telephones, it's unlikely that all users will be making an outside call at the same time. If we assume that 10% will require an outside line at any given time, then, the company would only need to lease 10 lines from the telecom company rather than 200.

The loss model can be used to find the optimum solution of the PABX system. Based on the PABX switching techniques and its loss system, a mathematical model is developed to simulate the grade of service (GoS) of the PABX system with respect to the number of cables provided.

In general, there are several advantages businesses can benefit from PBX/PABX System; IVR, Voice Mail, Call Recording, Voice Mail to Email, Mobility Users, Call Queue, Conference Call Bridge, Telephone CRM Integration and Remote office connectivity are some of it.

Waiting lines or Queues, are a very common occurrence both in everyday life and in a variety of business and industrial situations. Queuing theory originated in the research of a Danish engineer, A.K. Erlang in 1913, who studied the fluctuating demands on telephone service [2]. The formation of waiting lines occurs whenever the demand for service from a facility exceeds the capacity of that facility. Waiting line analysis is characterized by the following:

1. Customers, or arrivals, that require service
2. Uncertainty concerning the demand for service, and the timing of the demand for service of the customers.
3. Service facilities, or servers, that perform the service.
4. Uncertainty concerning the time duration of the service operation.
5. Uncertainty concerning the behavior of the customer as they arrive for service and/ or wait in the queue.

Based on these five characters the objective of queuing theory becomes the provision of adequate but not excessive service. Thus, the goal of waiting line modeling is the achievement of an economic balance between the costs associated with the wait required for that service. Three major components are vital to analyze the waiting line as shown in Figure2 below. They are:

1. The arrivals or inputs to the systems;
2. The waiting line, or queue, itself;
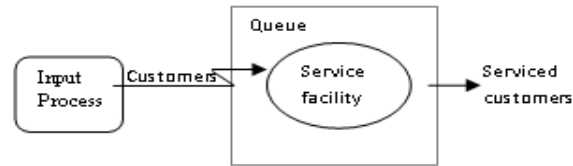3. The service facility;

Figure 2. Components of Waiting Line

Idle time in queueing systems occurs when the server is not rendering service to customers. Starvation and blockage are two forms of idle time in a server ([3], [4], [5]). Starvation occurs when a server must wait for the next customer to arrive from its preceding server. Blockage occurs when a server finishes processing its work, but cannot pass the customer to the next server.

This paper is organized as follows. The next section presents the issue of multiple channels waiting line with respect to Poisson arrival and Exponential service time and indicates some assumptions that are necessary for the solution of the problem at hand. It describes the modeling idea based on M/M/m/K/S multi-server queue but without buffer (i.e. a special finite case where K = m), which is the main backbone of our problem. The concepts of Erlang B formula and traffic intensity were also explained. The issue of dimensioning circuit-switched networks is presented in section 3. Section 4 summarizes the finding and finally section 5 concluded the paper.

2. MODELING

Data networks and server-based services such as Internet bureaux and telemarketing bureaux are examples of queuing with bounded buffer or call-waiting systems. A buffer minimizes, and may even eliminate, any loss of offered traffic. In the case of operator services or call-in telemarketing bureaux, incoming callers wait in a queue listening to music or recorded message until a human operator becomes free.

A. Numerical Example

We have assumed a PBX system for telephone line that has to support 600 users where in average, during office hour, 20 users need to talk to each other with average call holding time of 3 min per user. To develop an optimum model, we need to consider to the following:

- The numbers of needed cables versus the grade of service (GoS) or blocking probability (say from 0.001 to 0.00001).
- The number of needed cables versus load (when the load is variable or not fixed).
- The arrival and service time are assumed to be Poisson and exponential distribution respectively.

In our model no buffering takes place (though this typically occurs in older telephone switches). Whenever an arriving job does not find a free server, it is lost. This typically occurs in telephone switches where the number of lines equals the maximum number of customers that can be coped

with. In case all m lines are busy, no further queuing can occur and the request is not accepted, or queued. We assumed we have a finite population of S customers, each with an arriving parameter λ. In addition, the system has m servers, each with service rate μ. The system has no buffer or storage room, therefore K = m.

The chosen queuing model describe the following properties and is shown in Figure3:
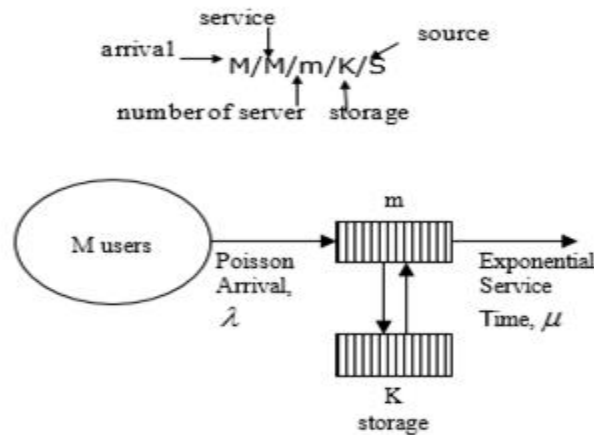

Figure 3: Simulation Model

The state transition diagram for this type of queuing model is illustrated in Figure4. This leads to the following birth-death coefficients. It is a special case of the general birth-death model with the following parameters:

$$\lambda_k = \begin{cases} \lambda(M-k) & 0 \leq k \leq m-1 \\ \\ 0 & otherwise \end{cases}$$

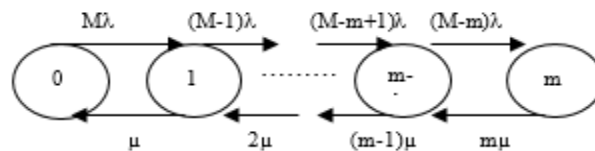$$\mu_k = \begin{cases} k\mu & 0 \leq k \leq m \\ \\ 0 & otherwise \end{cases}$$

Figure 4. State Transition Rate

Theorem 1: PASTA proper – (A well-known and often applied result of queuing theory called Poisson Arrival See Time Averages). The theorem states that, the distribution of jobs in a queuing

station at the moment a new job of a Poisson arrival process arrives is the same as the long-run or steady-state job distribution.

A probability P_m signifies the probability that all servers are in use. Due to the PASTA property, this probability equals the long-term probability that an arriving packet is lost. Theformula for P_m was established by Erlang in 1917 and is therefore often referred to as Erlang's loss formula or Erlang's B formula and denoted as $B(m, \lambda/\mu) = B(m, \lambda)$:

$$P_m = B(m, \lambda) = \frac{\lambda^m / m!}{\sum_{k=0}^{m} \frac{\lambda^k}{k!}}$$

or in general form:

$$P_m = \frac{n! \binom{m}{n} \left(\frac{\lambda}{\mu}\right)^n \left(\frac{\mu}{\lambda}\right)^{m-n}}{\sum_{n=0}^{m} n! \binom{m}{n} \left(\frac{\lambda}{\mu}\right)^n}$$

From this description any P_n may be found by the following:

$$P_n = 1/(m-n)! \left(\frac{\mu}{\lambda}\right)^{m-n} P_m$$

Or

$$P_{m-k} = 1/K! \left(\frac{\mu}{\lambda}\right)^k P_m$$

Where m represents the limit of the finite set and k equals the number of machines in operation. From figure 3 and general formula we consider the following range $0 \leq k \leq m-1$. Therefore,

$$P_k = \binom{M}{k} \left(\frac{\lambda}{\mu}\right)^k [P_0]^{-1}$$

$$= \frac{\binom{M}{k} \left(\frac{\lambda}{\mu}\right)^k}{\sum_{k=o}^{m} \frac{\rho^k}{k!} \frac{M!}{(M-k)!}}$$

Since k = m we have:

$$P_k = \frac{\dfrac{\rho^m}{m!}\dfrac{M!}{(M-m)!}}{\sum\limits_{k=o}^{m}\dfrac{\rho^k}{k!}\dfrac{M!}{(M-k)!}}$$

To ensure the stability of the designed system, the utilisation factor, denoted by ρ should be less than 1, i.e. $\rho < 1$, ρ is given by the ratio of $\dfrac{\mu}{\lambda}$. If the number of servers, m is assumed to be finite, then

B. *System Design*

The simulation is designed based on the flow chart described below, MATLAB was used to geenrate the output.
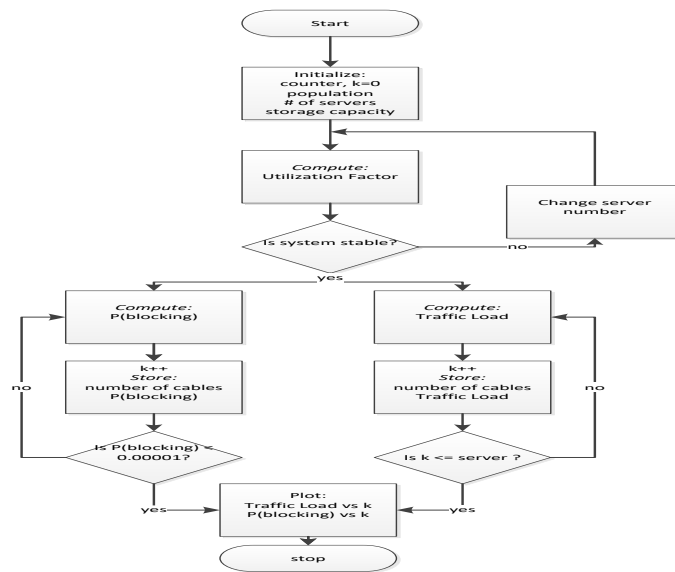

Figure 5: Flowchart of the simulation

The number of server chosen will influence the stability of the system. The traffic load of the system can be computed using the utilisation factor. Since this is a finite server case,
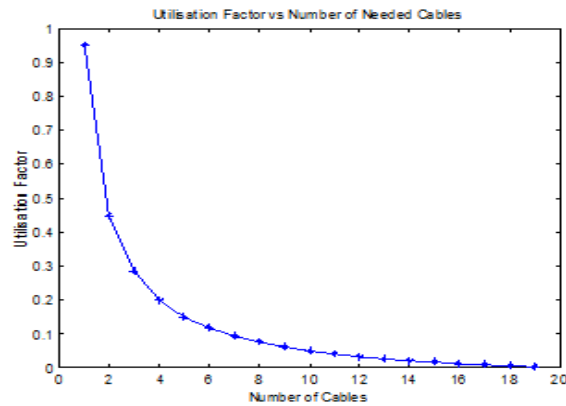
Figure 6: The Utilization Factor vs. Number of Cables

the traffic load, A is given by:

$$Traffic\ Load,\ A\ =\ m\rho = \frac{\lambda_k}{\mu_k}$$

where m is the number of server and k is the number of needed cables.

C. The system stability

Table 1: Utilisation Factor vs. Number of Cables

Number of Cables, k	Utilisation factor, ρ
1	Inf
2	0.9500
3	0.4500
4	0.2833
5	0.2000
6	0.1500
7	0.1167
8	0.0929
9	0.0750
10	0.0611
11	0.0500
12	0.0409
13	0.0333
14	0.0269
15	0.0214
16	0.0167
17	0.0125
18	0.0088
19	0.0056
20	0.0026

For example, the number of users is set at 600, if the number of servers and the storage capacity is equal to 20, then the obtained results are shown in Table 1 above.

D. Traffic Intensity

Definition 1: Traffic intensity of a circuit-switched network (such as telephone) is defined to be the average number of calls simultaneously in progress during a particular period of time. It is measured in Erlangs. Thus an average of one call in progress during a particular period would represent a traffic intensity of one Erlangs [6].

Definition 2: Offered traffic – measure of the unsuppressed traffic intensity that will be transported on a particular route if all customers' calls were connected without congestion (it is simply the demand).

Definition 3: Carried traffic – resultant from the carried calls, it is the value of traffic intensity actually measured.

For a network without congestion the carried traffic is equal to the offered traffic. However, if there is congestion in the network, then the offered traffic will be higher than the carried, the difference being the calls which cannot be connected. Traffic intensity can be expressed as:

$$Traffic\ intensity\ = \frac{\text{sum of circuit holding time}}{\text{duration of the monitoring period}}$$

Now let

A = the traffic intensity in Erlangs
T = the duration of the monitoring period
h_i = the holding time of the it individual call
c = the total number of calls in the period of mathematical summation

Then, from above

$$A = \frac{\sum_i^c h_i}{T}$$

Now, because the sum of the holding times is equal to the number of calls multiplied by the average holding time, then

$$\sum_i^c h_i = ch$$

Where h = average call holding time, and therefore

$$A = ch/T$$

It is interesting to calculate the call arrival rate, in particular the number of calls expected to arrive during the average holding time. Let N be this number of calls, then

N = no of calls arrivals during a period of equal tothe average holding time

  = $h \times$ call arrival rate per unit of time

  = $h \times c/T$

  = $ch/T = A$

In other words, the number of calls expected to be generated during the average holding time of a call is equal to the traffic intensity A.

E. Stationary State Probability

The stationary state probability versus the number of needed cables. For this simulation , the variables used are as follows:

- population = 100
- server = 20
- storage = 20
- load = 1

Table 2 – Stationary State Prob. (Pk) vs. No. of Cables

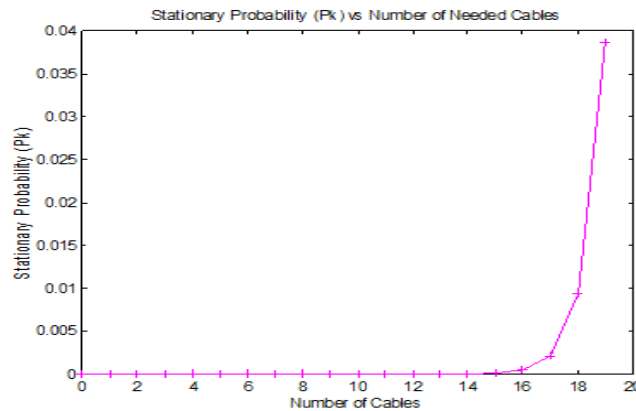| No. Cables, k	Stationary State Prob. (P_k)
1	5.8374e-021
2	2.9187e-019
3	9.6318e-018
4	2.3598e-016
5	4.578e-015
6	7.3248e-014
7	9.9408e-013
8	1.168e-011
9	1.207e-010
10	1.1104e-009
11	9.1862e-009
12	6.8896e-008
13	4.7167e-007
14	2.9648e-006
15	1.7196e-005
16	9.2428e-005
17	0.00046214
18	0.0021567
19	0.0094212
20	0.038627

Figure7: Stationary State Prob. vs No. of Cables

3. DIMENSIONING CIRCUIT-SWITCHED NETWORK

The circuit requirement for route of a circuit-switched network (such as telephone, telex etc) can be determine from the Erlang lost call formula. This can be done by substituting the predicted offered traffic intensity A, and using trial and error values of N to determine the value which gives a slightly better performance than the target blocking or *grade of service* (GoS) B.

It is not an easy task by direct calculation to determine the value of N (circuits required), and for this reason we use *Mathematical* [7]. The formulation is based on Erlang lost-call formula which can simply be written as

$$B(N, A) = \frac{A^N/N!}{\left\{1 + A + A^2/2! + ... + A^N/N!\right\}}$$

Where

B(N, A) = proportion of lost calls, and probability of blocking
A = offered traffic intensity
N = available number of circuits
N! = factorial of N

4. RESULTS AND DISCUSSIONS

The result of our problem is tabulated above (Table 2), the table illustrates a traffic intensity based on some GoS (0.001 – 0.00001). Since in our assumptions A < 1, we consider the first values 1.67 to be immaterial.Down the left hand column of the table the number of needed cables are listed. Across the top of the table various different grades of service (GoS) are shown. In the middle of the table, the values represent the maximum offered Erlang capacity corresponding to the route size and grade of service chosen. A graphical representation of the result is also shown in Figure 8. The optimum solution is to provide six cables, with an offered traffic load of 0.34 Erlang, and a blocking probability of 0.001. Similarly, an increase in one cable is only acceptable

when we simulate base on the next subsequent GoS as shown in Table 3. Therefore, the optimum solution depends on the GoS assumed (in our case we considered 0.001 GoS). It is clear from Figure. 9 that a route of six cables working to a design grade of service of 0.001 has a maximum offered traffic capacity of 0.034 Erlangs. The problem with traffic routes of only a few cables is that only a small increase in traffic is needed to cause congestion. It is therefore good practice to ensure that a minimum number of cables are provided. It should be noted that we tabulate the real values of the output as shown in table 3 (Excluding the complex and negative output values)

Table 3: Traffic intensity

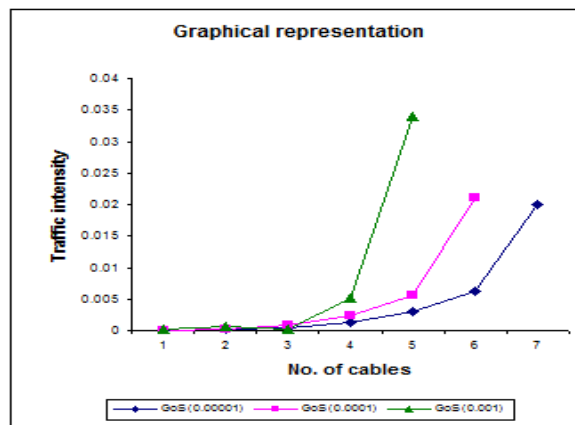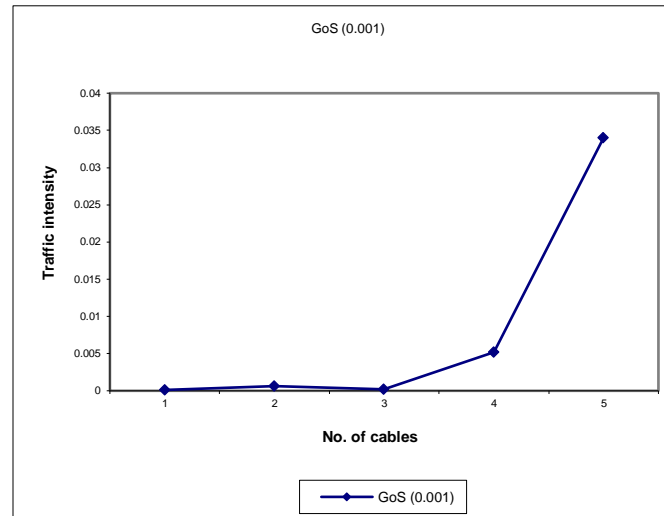| Grade of service (B[N, A])			
Number of cables	1 lost call in 1000 (0.001)	1 lost call in 10000 (0.0001)	1 lost call in 100000 (0.00001)
	Erlangs	Erlangs	Erlangs
1	1.67	1.67	1.67
2	0.00011	0.000033	0.000011
3	0.00062	0.0003	0.00012
4	0.00019	0.00094	0.0005
5	0.0052	0.0024	0.0013
6	0.034 (max)	0.0056	0.003
7		0.021 (max)	0.0063
8			0.02 (max)

Figure 8. Graphical representation

Figure 9. No. of cable vs load (base on GoS 0.001)

It is not easy to simulate for higher traffic intensity (say 0.9, 0.8, 0.7..), the highest value we can get is 0.02 (approximately). This is due to the limitation of the software as it provides all the possible values of A including negative and complex values. However, from the result obtained we can design a method of obtaining these values, as all the values tend to be linear after some few steps.

5. CONCLUSIONS

The optimum number of cables depends on the grade of service (GoS). However, as discussed in the previous section, a small number of cables are preferable to reduce the level of congestion. Therefore, to achieve the maximum traffic intensity, which is 0.034 Erlang, the total of six cables are needed with the GoS of 0.001. This system helps improving the faculty system since it finds the agreeable GoS to optimize the number of cables, in order to obtain the maximum traffic intensity.

REFERENCES

[1]   Brian L. Waldrop GSEC Practical v1.2 Securing the Other System: Basic PBX Functionality and Vulnerabilities, April 24, 2001

[2]   Roger L. Freeman, Reference Manual for Telecommunications Engineering, John Wiley &Sons, 1988.

[3]   Chase J. Operations management. Tata: McGraw-Hill. 2012

[4]   Heizer J. Principles of Operations Management: Student Value Edition. Prentice Hall, 2013.

[5]   Slack N. Chambers S. Johnston R. Operations management. Pearson Education, 2010

[6]   Roger L. Freeman, Telecommunication System Engineering, 4th Edition, John Wiley & Sons, 2004.

[7]   Stephen Wolfram, The Mathematica Book,Wolfram Media Inc, 5th Edition, 2003.

[8]   David Talley, Basic Electronic Switching for Telephone Systems, Hayden Book Company Inc., 1982.

[9]   Bruce E. Brieley, Introduction to Telephone Switching, Addison-Wesley, 1983.

[10] Martin P. Clark, Networks and Telecommunications: Design and Operation, John Wiley & Sons, 2ndEdition, 2001.

[11] Lawan. A. Mohammed,A Simple Heuristics Model for the Design of Telephony Systems, Proceedings of the IEEE International Conference on IT Convergence and Security, (WCMCS'2013), 2013 Tunisia.