UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA
SCHOOL OF COMPUTER SCIENCE TECHNOLOGY

REPORT: TELETRAFFIC ANALYSIS IN TELECOMMUNICATION SYSTEMS

# STOCHASTIC PROCESSES AND APPLICATIONS

SUBMITTED TO

**PROF. WU DEAN**

*Kechengzuoye_mi@163.com*

BY

**YEBOAH-DUAKO ELVIS**

*yeboahduako770@gmail.com*

**(202124080120)**

**JANUARY, 2022**

## ABSTRACT

In the Telecommunication Industry, a good network design does not only provide quality network services to subscribers, but also, it saves network operators millions of money. With teletraffic analysis, network operators are able to ascertain the demands of their network and plan accordingly thereby deploying the requisite and cost-effective network resources to serve the needs of their subscribers. In this report, I illustrate how stochastic processes are applied to the two main systems of telecom networks (loss systems and delays systems) to derive their performance metrics. For Loss systems, the probability that an initiated call will be blocked is of interest whiles for delay systems, the probability that a user's call request will experience delay is of interest and more importantly, the probability that the delay experienced will exceed a given time is also of key interest. These two probabilities are mainly dependent on traffic intensities and the channel resources available.

## TABLE OF CONTENTS



## TABLES OF FIGURES

# 1.0 INTRODUCTION

In Telecommunication, *Traffic* refers to the aggregate of all user requests (voice and data requests) being serviced by a given network. As far as the network is concerned, the user requests arrive randomly and usually require unpredictable service times. Unfortunately, the system resources needed to provide the network services are limited and hence the possibility that, not all user requests can be processed within a given time interval. To the Network Service provider, the sum total of all unserviced user requests represents a gross loss financially and it could as well cost them customers as the users tend to get frustrated with time as they are unable to access the network resources and services. Thus, Traffic analysis has been very important as it provides solution to problems concerning planning, performance evaluation, operation and maintenance of telecommunication systems. By Definition, Traffic Theory is the application of probability theory in the design of cost-effective systems that have a predefined grade of service with which estimates of future traffic demands can be made and carefully planned for. The primary objective is to make the traffic measurable in well-defined units through mathematical models and to derive relationships between grade-of-service and system capacity in such a way that the theory becomes a tool by which investments in telecommunication infrastructure can be efficiently developed and deployed [1].

The first step in Traffic Analysis is the characterization of traffic arrivals and the service times in a probabilistic framework. And based on how a given network treats overload traffic, the techniques of traffic analysis can be divided into two general categories: *The Loss Systems* and *The Delay Systems* [ 2]. In a loss system, overload traffic is rejected without being serviced, whereas in a delay system, overload traffic is held in a queue until system resources are released to service the loads. Conventional circuit Switched telecommunication systems are loss systems whereas Packet switched systems are typically delay systems although they can be treated as loss systems when the buffer is full.

## 1.1 Definition of Basic Terms

***Calling rate (*$\lambda$*):*** This is the average number of requests for connection ($n$) that are made per unit time ($T$). It is measured in calls/hour and defined as:

$$\lambda = \frac{n}{T} \qquad (1.1)$$

***Holding time (H):*** is the time it takes to service a request. In simpler terms, it is how long a call or connection lasts.

***Service Rate*** ($\boldsymbol{\mu}$)***:*** is the reciprocal of the average holding time. It describes the rate at which user requests are processed. It is defined as:

$$\mu = \frac{1}{H} \tag{1.2}$$

***Traffic Volume/Intensity (A):*** is the sum of all carried traffic over a period of time. It is also known as the average occupancy and is defined by Equation (1.3). The standard unit is Erlangs (E)

$$A = \frac{nH}{T} = \lambda H = \frac{\lambda}{\mu} \tag{1.3}$$

***Grade of Service (GOS)*****:** Is the measure of system performance which describes the proportion of unsuccessful calls/requests relative to the total number of calls/requests. For loss systems, it is measured as the *probability of call blocking,* where as in delay systems, it is measured as the *probability that delay occurs* and more importantly, the *probability that delay exceeds a specific length of time*.

## 1.2 Arrival Distributions

In most Teletraffic applications, Call arrivals are modelled to follow the Poisson distribution with a mean equal to the average calling rate, $\lambda$. The fundamental assumption of classical traffic analysis is that call arrivals are independent. That is an arrival from one source is unrelated to an arrival from any other source. Even though this assumption may be invalid in some instances, it has general usefulness for most applications. In the cases where call arrivals tend to be correlated, useful results can still be obtained by modifying a random arrival analysis. Other key assumptions made for the purpose of this analysis as as follows:

- ✓ Only one arrival can occur in any sufficiently small interval.
- ✓ The probability of an arrival in any sufficiently small interval is directly proportional to the length of the interval. (probability of arrival = $\lambda\Delta t$, where $\Delta t$ is the interval length).

- ✓ The probability of an arrival in any particular interval is independent of what occurred in other intervals.

The probability distribution of inter arrival time is $P_0(\lambda t) = e^{-\lambda t}$, which defines the probability that no arrivals occur in a random selected interval, $t$. With the same assumptions, the probability that $\boldsymbol{n}$ arrivals occur in an interval $\boldsymbol{t}$ can be defined as:

$$P_n(\lambda t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{1.4}$$

Usually, there is more interest in determining the probability of $n$ or more arrivals in $t$ interval as shown below:

$$P_{\geq n}(\lambda t) = \sum_{i=n}^{\infty} P_i(\lambda t)$$

$$= 1 - \sum_{i=0}^{n-1} P_i(\lambda t)$$

$$= 1 - P_{<n}(\lambda t) \tag{1.5}$$

## 2.0 APPLICATION OF STOCHASTIC PROCESSES IN LOSS SYSTEMS

In Lost Call Cleared systems, queueing is not provided for call requests. Calls are assumed to arrive with a Poisson distribution, and it is further assumed that there are a nearly infinite number of users. When a user requests service, there is minimal call set-up time and the user is given immediate access to a channel if one is available. If, however, all channels are already in use and no new channels are available, the call is *blocked* without access to the system. The key application of stochastic processes in such systems is the generation of the *Erlang B formula* which describes the Grade of Service (GOS) as the probability that an arbitrary user will experience a blocked call in a Lost system. In network planning, the Erlang B chart attached as '*appendix A'* provides an quick guide on the required system resources (channels) that would be needed to handle a given traffic intensity so as to achieve a satisfactory GOS. This saves time in network planning as admins are able to use the estimated number of subscribers within a geographical area to ascertain how much capital and system investments, they need to provide good services for their customers with minimum to no disruptions in their calls.

### 2.1 The Erlang B Formula

Let $\boldsymbol{C}$ represent the number of trunked channels, and $\boldsymbol{A}$, the total offered load to the trunked system.

$$\Pr[Blocking] = \frac{\left(\frac{A^C}{C!}\right)}{\sum_{n=0}^{C}\left(\frac{A^n}{n!}\right)} \tag{2.1}$$

### 1.2.1 Derivation of Erlang B Formula (Proof)

Consider a system with $\boldsymbol{C}$ channels and $\boldsymbol{U}$ users. Let $\boldsymbol{\lambda}$ be the total mean call arrival rate per unit time and $\boldsymbol{H}$ be the average call holding time (average duration of a call). If $\boldsymbol{A}$ is the offered load for the trunked system and is the average load offered by each user, then $A_U = \lambda_1 H$ where $\lambda_1$ is the average call arrival rate for one user and $A = UA_U = \lambda H$.

The probability that a call requested by a user will be blocked is given by

$$\Pr[Blocking] = \Pr\left[None\ of\ the\ C\ channels\ are\ free\right] \tag{2.2}$$

It is assumed that calls arrive according to the Poisson distribution, such that:

$$Pr = \{a(t+\tau) - a(t) = n\} = \frac{e^{-(\lambda\tau)}}{n!}(\lambda\tau)^n \quad for\ n = 0,1,2, \ldots. \tag{2.3}$$

where $a\ (t)$ is the number of arrivals that have occurred since $t = 0$ and $\boldsymbol{\tau}$ is the call interarrival time (the time between successive call requests).

The Poisson process implies that the time of the $\boldsymbol{n}$th call arrival and the interarrival times between successive calls are mutually independent. The interarrival times between calls are exponential and mutually independent, and the probability that the interarrival time will be less than some time $\boldsymbol{s}$ is given by:

$$\Pr(\tau_n \leq s) = 1 - e^{-\lambda s}, \qquad s \geq 0$$

Where $\tau_n$ is the interarrival time of the $n$th arrival, and $\tau_n = t_{n+1} - t_n$, where $t_n$ is the time at which the $n$th call arrived. In other words, the probability density function for $\tau_n$ is

$$p(\tau_n) = \lambda e^{-\lambda \tau_n}, \quad \tau_n \geq 0 \tag{2.4}$$

For every $t \geq 0, \delta \geq 0$,

$$\Pr\ \{a(t+\delta) - a(t) = 0\} = 1 - \lambda\delta + \mathrm{O}(\delta) \tag{2.5}$$

$$\Pr\{a(t+\delta) - a(t) = 1\} = \lambda\delta + \mathrm{O}(\delta) \tag{2.6}$$

$$\Pr\{a(t+\delta) - a(t) \geq 2\} = \mathrm{O}(\delta) \tag{2.7}$$

Where $\mathrm{O}(\delta)$ is the probability of more than one call arriving over the time interval $\delta$ and is a function of $\delta$ such that $\lim_{\delta \to 0} \left\{\frac{\mathrm{O}(\delta)}{\delta}\right\} = 0$

The probability of $n$ arrivals in $\delta$ seconds is found in equation (3)

$$Pr = \{a(t+\delta) - a(t) = n\} = \frac{e^{-\lambda\delta}}{n!}(\lambda\delta)^n \tag{2.8}$$

The user service time is the duration of a particular call that has successfully accessed the trunked system. Service times are assumed to be exponential with mean call duration $H$, and $\mu = 1/H$ is the mean service rate (average number of calls per unit time). $H$ is also called the average holding time. The probability that the service time of the $n$th user is less than some call duration s is given by

$$P_r\{s_n < s\} = 1 - e^{-\mu s}, \quad s > 0 \tag{2.9}$$

And the probability density function of the service time is:

$$p(s_n) = \mu e^{-\mu s_n} \tag{2.10}$$

The property of Markov chains can be used to derive the Erlang B formula. Consider a discrete time stochastic process $\{X_n | n = 0,1,2, \dots\}$ that takes values from the set of nonnegative integers, so that the possible states of the process are $i = 0,1, \dots$ . The process is said to be a Markov chain if its transition from the present state $i$ to the next state $i + 1$ depends solely on the state $i$ and not on previous states. Discrete time Markov chains enable the traffic to be observed at discrete observation points for specific traffic conditions. The operation of a practical trunked system is continuous in time, but may be analyzed in small time intervals $\delta$, where $\delta$ is a small positive number. If is the number of calls (occupied channels) in the system at time $k\delta$, then $N_k$ may be represented as:

$$N_k = N(k\delta) \tag{2.11}$$

where $N$ is a discrete random process representing the number of occupied channels at discrete times. $N_k$ is a discrete time Markov chain with steady state occupancy probabilities which are identical to the continuous Markov chain, and can have values ranging on $0,1,2,\dots,C$.

The transition probability $P_{i,j}$ is given by

$$P_{i,j} = \Pr\{N_{k+1} = j | N_k = i\} \tag{2.12}$$

Using equations (2.5) through (2.7), and letting $\delta \to 0$, we obtain:

$$P_{00} = 1 - \lambda\delta + \mathrm{O}(\delta) \tag{2.13}$$

$$P_{ii} = 1 - \lambda\delta - \mu\delta + \mathrm{O}(\delta), \quad i \geq 1 \tag{2.14}$$

$$P_{i,i+1} = \lambda\delta + \mathrm{O}(\delta), \quad i \geq 0 \tag{2.15}$$

$$P_{i,i-1} = \mu\delta + \mathrm{O}(\delta), \quad i \geq 1 \tag{2.16}$$

$$P_{i,j} = \mathrm{O}(\delta), \quad j \neq i, j \neq i+1, \; j \neq i-2 \tag{2.17}$$

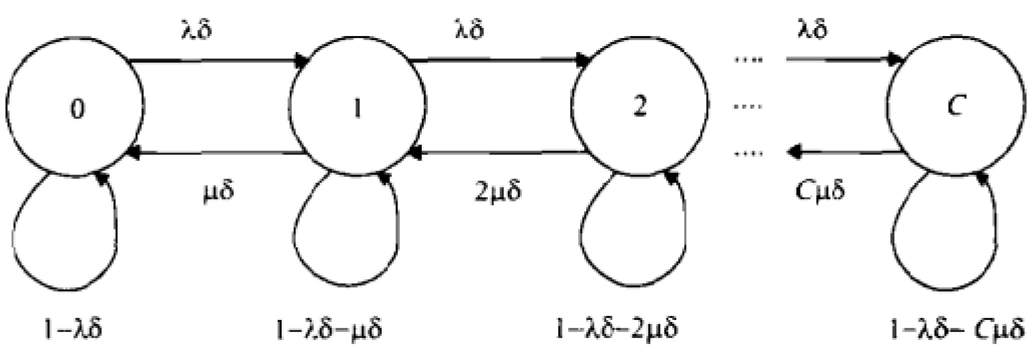


Fig 2.1: The transition probabilities represented as a Markov chain state diagram for Erlang B.

The Markov chain representation of a trunked system with C channels is illustrated in figure 2.1. To understand the Markov chain state diagram, assume that there are 0 channels being used by the system. Over a small interval of time, the likelihood that the system will continue to use 0 channels is $(1 - \lambda\delta)$. The probability that there will be a change from $0$ channels to $1$ channel in use is given by $\lambda\delta$. On the other hand, if one channel is in use, the probability that the system will transition to $0$ used channels is given by $\mu\delta$. Similarly, the likelihood that the system will continue to use 1 channel is given by $1 - \lambda\delta - \mu\delta$. All of the outgoing probabilities from a certain state sum up to 1. Over a long period of time, the system reaches steady state and has $n$ channels in use. Figure 2.2 represents the steady state response of a Loss system.

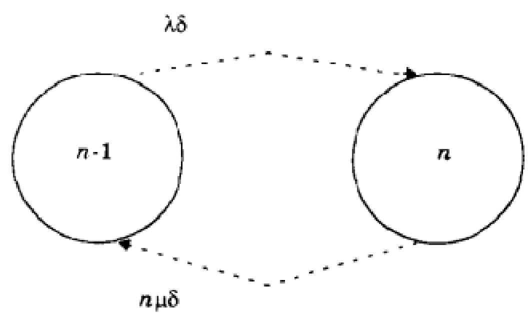


Fig 2.2: A trunked Loss system at steady state with $n$ number of channels in use

At steady state, the probability of having $n$ channels in use is equivalent to the probability of having $(n— I)$ channels in use, times the transition probability $\lambda\delta$. Hence under steady state conditions:

$$\lambda\delta P_{n-1} = n\mu\delta P_n, \quad n \leq C \tag{2.18}$$

Equation (2.18) is known as the Global Balance Equation. Furthermore,

$$\sum_{n=0}^{C} P_n = 1 \tag{2.19}$$

Using the Global Balance equation for different values of $n$, it is seen that

$$\lambda\delta P_{n-1} = P_n n\mu\delta, \quad n = 1,2,3,\ldots\ldots C \tag{2.20}$$

$$\Rightarrow \lambda P_{n-1} = P_n n\mu \tag{2.21}$$

$$P_1 = \frac{\lambda P_0}{\mu} \tag{2.22}$$

Evaluating equation (2.20) for different values of $n$

$$P_n = P_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \tag{2.23}$$

$$P_0 = \left(\frac{\mu}{\lambda}\right)^n P_n n! = 1 - \sum_{i=1}^{C} P_i \tag{2.24}$$

Substituting equation (2.23) into equation (2.24)

$$P_0 = \frac{1}{\sum_{n=0}^{C}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}} \tag{2.25}$$

From (2.23), the probability of blocking for $C$ trunked channels is:

$$P_c = P_0 \left(\frac{\lambda}{\mu}\right)^C \frac{1}{C!} \tag{2.26}$$

Substituting (2.25) into equation (2.26)

$$P_c = \frac{\left(\frac{\lambda}{\mu}\right)^C \frac{1}{C!}}{\sum_{n=0}^{C}\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}} \tag{2.27}$$

The total offered traffic is $A = \lambda H = \lambda/\mu$. Ut = Substituting this in equation (2.27), the probability of blocking is given by:

$$\Pr[Blocking] = \frac{A^C \frac{1}{C!}}{\sum_{n=0}^{C} A^n \frac{1}{n!}} \tag{2.28}$$

Equation (2.28) = Equation (2.1): hence the Erlang B formula has been proved!

**2.3 Simulation and Analysis**

In practice, a GOS of $2\%$ or less is desirable. That is, a system where in every hundred call attempts, a maximum of $2$ experience blocking is considered to be good network design. So, for this simulation, First, the probability of blocking was set at $2\%$ while the number of traffic channels were varied to obtain the corresponding network traffic (and users) that can be supported on the network. It can be seen from Fig. 2.3(a) that, for a constant GOS of $2\%$, traffic intensity is directly proportional to the number of available channels. Hence, for $8$ operational trunking channels, a maximum of approximately $3.627E$ of traffic would be supported by the system. And likewise, $10$ trunking channels would accommodate $5.084E$.

Conversely, given the traffic intensity, the number of channels needed can be estimated. This is mostly desired by Network admins as they desire to know how much infrastructural capital is needed for the given GOS. Still on Fig. 2.3(a), the dash lines in red illustrate this reverse action. For instance, say there are $400$ subscribers / Users (U) of a particular network and the average call traffic intensity per user, $A_u$ is estimated to be $0.02E$. By means of equation (2.29) below, the total traffic intensity can be calculated to be $8E$.

$$A = UA_U = \lambda H \tag{2.29}$$

From the graph, Fig 2.3(a), marked in red dashes, it can be seen that $8E$ corresponds to 13.8 Channels. This approximates to 14 channels. And hence, as a network Engineer, given a target market of 400 subscribers, 14 traffic channels must be deployed in order to achieve the standard GOS of 2% (for an estimated expected Traffic Intensity of $8E$.

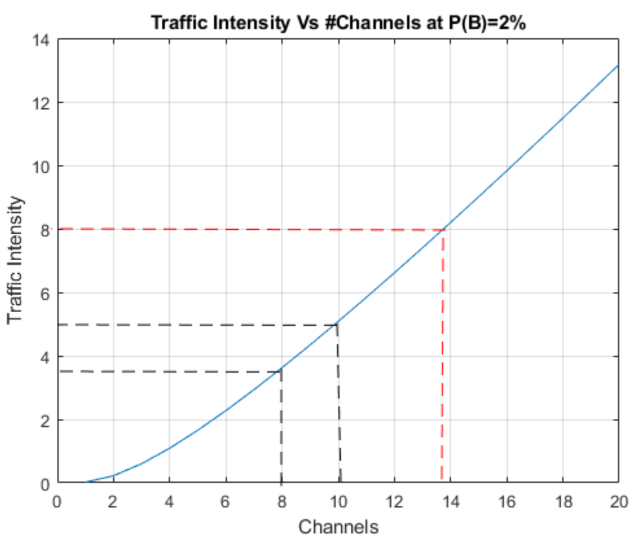


Fig 2.3 (a): Traffic Intensities at GOS of 2%

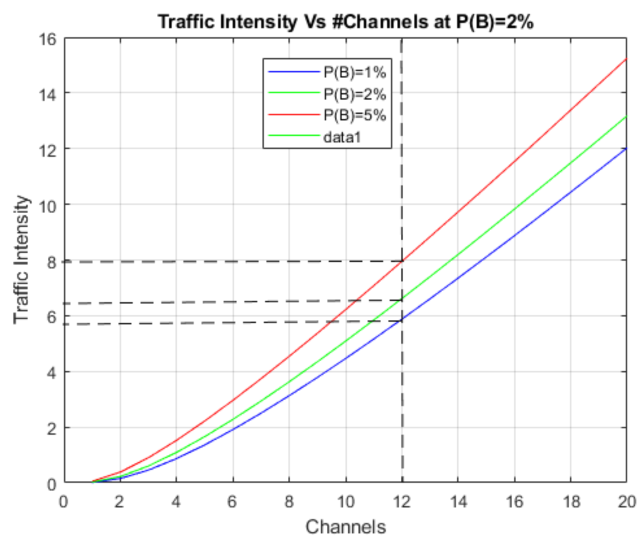


Fig2.3(b): Traffic Intensities for GOS of 1%, 2% and 5%

Moreover, given a limited number of network channel of say $12\ channels$ as marked in fig 2.3(b), it can be seen that, based on the network design goal or GOS, an estimate of the number of users can be drawn from the Total traffic intensity given by referring to equation (2.29). Assuming an average traffic intensity per user, $A_u$ of $0.02E$, it can be seen that for a GOS of 1%, a maximum of $5.876E$ of traffic can be supported by the network. And by using equation (2.29), approximately $294\ users$ or subscribers are expected to enjoy the services of the system. Similarly, for a GOS of 2%, traffic of $6.615E$ can be serviced, and hence approximately $331\ users$ are targeted. And finally, for a GOS of 5%, traffic of $7.960E$ is expected, hence approximately $398\ users$ are targeted to be serviced.

### 2.4 Conclusion

From the discussions above and the summarized graph shown in fig 2.4 below, It can be seen that, systems designed with higher values of GOS accommodate more traffic and consequently more users on the network. Nevertheless, the performance of the system degrades since the likelihood of a user's call succeeding is inversely proportional to the GOS. Also, for a fixed GOS, number of users directly relates with the available number of channels. That is, the more the available channels, the more users the system can support without any degradation in the quality of service.

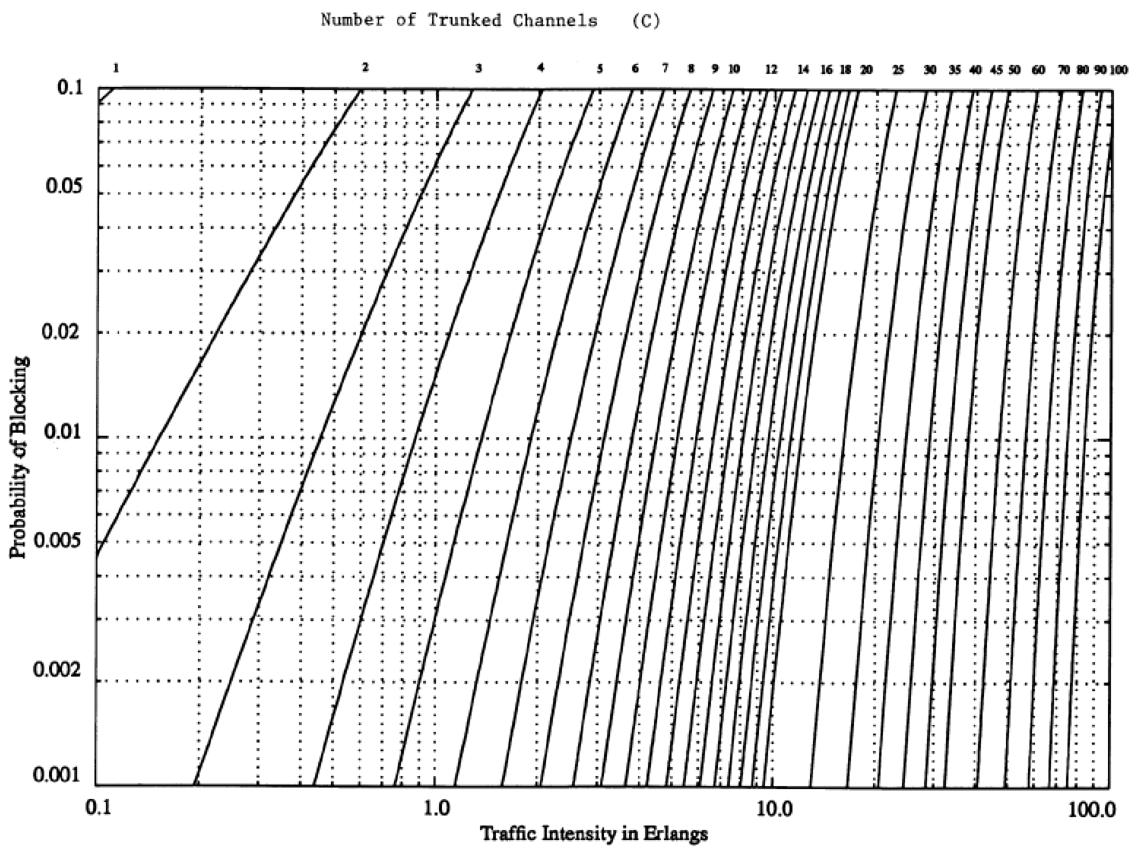


Fig 2.4: Traffic Intensity simulated over 100 channels for different GOS

***Note for Professor:***

A similar analysis is done for delay systems, but due to the page limits imposed for this report, I have attached them to this report as a separate file named as '*appendix B*' for your reference. Thank you.

## REFERENCES

[1] Multimedia University of Kenya, ETI 2506 - Telecommunication Systems, Teletraffic Engineering, **http://eti2506.elimu.net/Teletraffic_Eng/Teletraffic_Default.html**

[2] Agyemang Prempeh Agyekum, Lecture Notes – Introduction to Teletraffic Analysis, Kwame Nkrumah University of Science and Technology (KNUST)

[3] Handbook of Teletraffic Engineering First Edition, Geneva, January 2005

[4] Theodore S Rappaport, Wireless Communication principles & Practice, Prentice Hall Inc.

[5] Ulrich et al, Advanced Stochastic Modelling Applied to Telecommunication Networks and Distributed Systems, Dagstuhl Seminar 9813 (March 30, - April 4, 1998)

[6] N. Gautam, Stochastic Models in Telecommunications for Optimal Design, Control and Performance Evaluation, The Pennsylvania State University

Simulation Tool: **MATLAB**