



## Performance Analysis of Statistical Time Division Multiplexing Systems

Johnson Adegbenga AJIBOYE<sup>1</sup> and Yinusa Ademola ADEDIRAN<sup>2</sup>

<sup>1</sup> *Dept. of Electrical and Electronics Engineering, Federal University of Technology, Minna, Nigeria. Phone: +2348035532375, +2348055252189*

<sup>2</sup> *Dept. of Electrical Engineering, University of Ilorin, Ilorin, Nigeria. Phone: +2348035903902, +2348054954200*

E-mail(s): [ajiboye2003@yahoo.com](mailto:ajiboye2003@yahoo.com), [yinusaade@yahoo.com](mailto:yinusaade@yahoo.com)

### Abstract

Multiplexing is a way of accommodating many input sources of a low capacity over a high capacity outgoing channel. Statistical Time Division Multiplexing (STDM) is a technique that allows the number of users to be multiplexed over the channel more than the channel can afford. The STDM normally exploits unused time slots by the non-active users and allocates those slots for the active users. Therefore STDM is appropriate for bursty sources. In this way STDM normally utilizes channel bandwidth better than traditional Time Division Multiplexing (TDM). In this work, the statistical multiplexer is viewed as M/M/1 queueing system and the performance is measured by comparing analytical results to simulation results using Matlab. The index used to determine the performance of the statistical multiplexer is the number of packets both in the system and the queue. Comparison of analytical results was also done between M/M/1 and M/M/2 and also between M/M/1 and M/D/1 queue systems. At high utilizations, M/M/2 performs better than M/M/1. M/D/1 also outperforms M/M/1.

### Keywords

Statistical Time Division Multiplexing (STDM); Time Division Multiplexing (TDM); Queueing System.

## **Introduction**

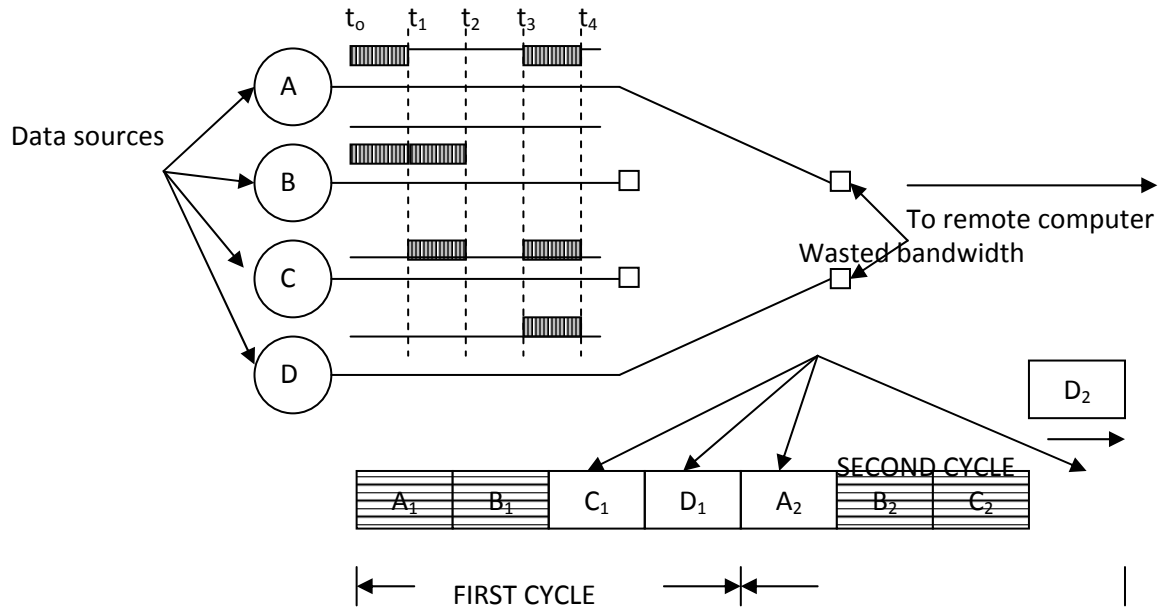
There has been a strong demand to create a singular network technology that is able to support different kinds of data types simultaneously. Packet Switching Network utilizes the advantage of statistical multiplexing to ensure that communication resources are efficiently used. Therefore, packet-switched networks are becoming increasingly popular because of their ability to handle integrated services such as voice, video and data. This has resulted in high network congestion and packet losses due to buffer delays. This study discusses the concept of using Statistical Time Division Multiplexing (STDM) for packet switching network.

The technique of multiplexing has been developed to increase channel utilization. A commonly used multiplexing technique is the Synchronous Time Division Multiplexing in which each terminal is given fixed time duration to send or transmit messages. When this time has elapsed, the channel is made available for the next user. With a well-designed synchronous operation, arrival of data and also its distribution to the Input/Output lines is predictable. It is therefore known in advance which terminal is to transmit; hence, address information is not usually required.

However, in a synchronous time division multiplexing, most of the time slots in a frame are not being utilized; they are thus wasted. It is therefore inefficient in channel utilization. For instance, in a situation where computer terminals are linked together using a shared computer port, even when all the terminals are working, most times there is no data transfer at any particular terminal [1]. Figure 1 shows the inefficiency of the synchronous time division multiplexer in channel capacity utilization. A whole segment of bandwidth is permanently assigned but only utilized a portion of the time. Data gotten from several representatives operating time sharing systems show that, during an average call, 95% of the user-to-computer channel and 65% of computer-to-user channel are idle. Certainly, other users could utilize the idle periods more efficiently. The process of combining two or more communications paths into one is referred to as multiplexing [2].

Figure 1 shows four data sources and the data produced in the time epochs  $t_0$ ,  $t_1$ ,  $t_2$  and  $t_3$ . Data are collected and sent during each time epoch from all the four sources and then transmitted. Only sources A and B have data to send in the first time epoch. Therefore, of the

four time epochs two slots are not utilized and hence empty. A better system that efficiently utilizes the transmission channel is to statistically multiplex the data.



**Figure 1.** Synchronous Time Division Multiplexing [1]

Statistical Time Division Multiplexing (STDM) is a system developed to overcome some inefficiencies of standard time division multiplexing, where time slices are still allocated to channels, even if they have no information to transmit. The statistical multiplexer dynamically allocates time slots on demand [2]. This allocation changes dynamically depending on the traffic pattern [3]. When FDM is used in a communications network, each input signal is sent and received at maximum speed at all times. This is its chief asset. However, if many signals must be sent along a single long-distance line, the necessary bandwidth is large, and careful engineering is required to ensure that the system will perform properly [4]. Additionally, an STDM may offer the advantages of data compression, error detection and correction, and reporting of traffic statistics [5]. STDM relies on statistic to define which channel needs more dedicated time to get more bandwidth [6].

### ***The Aim of Research***

The multiplexer, viewed as a resource allocator that controls a large bandwidth and allocates access to the link in an efficient manner, will be analyzed as an M/M/1 queue (single

server which is the case of random or Poisson arrivals and exponential service times) and M/D/1 queue (single server which is the case of random or Poisson arrivals and constant or fixed service time). The model will be analyzed to give some insight into the network performance of statistical TDM and the nature of the trade-off between system response time and the speed of the multiplexed channel. Comparison is then made between results obtained for M/M/1 and the counterpart M/D/1 queue models.

### ***Statistical Time Division Multiplexing (STDM)***

Statistical TDM can accommodate variable-length time slots. Stations transmitting at a faster data rate can be given longer slots. Managing variable-length fields requires that control bits be appended to the beginning of each time slot to indicate the length of the coming data portion. This can increase the overhead further and is only efficient with larger time slots. Each slot in a frame is not dedicated to the fixed device. The number of slots in a frame must not necessarily be equal to the number of input devices. More than one slot in a frame can be allocated for an input device. STDM allows maximum utilization of the link and a number of lower speed input lines can be multiplexed to a single higher speed line.

According to Buchanan [4], fast packet switching attempts to solve the problem of unutilized slots of synchronous TDM using the STDM technique. Applications such as voice traffic, which requires a constant data transfer, are allowed to pass through safe routes through the network. Applications that are bursty are also assigned to the same link with the hope that statistically they will not generate bursts of data at the same instance of time. For those that generate bursts of data simultaneously, their data could be buffered and sent later.

STDM can recognize active versus inactive devices as well as priority levels. Further, they can invoke flow control options that can cause a transmitting terminal to cease transmission temporarily, in the event that the STDM's internal buffer or temporary memory is full. Flow control also can restrain low - priority transmissions in favor of higher priority transmissions. Additionally, an STDM may offer the advantages of data compression, error detection and correction, and reporting of traffic statistics [5].

In STDM, individual channels can be grouped to yield higher transmission rates for an individual bandwidth-intensive communication such as a videoconference. The individual channels also can be subdivided into lower speed channels to accommodate many more, less bandwidth-intensive communications, such as low-speed data. In STDM the unused portion

of the bandwidth assigned to each source might be used to service packets generated by other sources. Homayoun [6] proposed the use of water-filling approach to assign available server bandwidths among sources. The server scheduling scheme partitions the available server bandwidth of a shared buffer among sources according to their traffic generation patterns. The queue size of source  $i$  at time  $k+1$  is given by Eq(1):

$$B^{(i)}[k+1] = B^{(i)}[k] + Q^{(i)}[k] \quad (1)$$

where  $B^{(i)}[k]$  is the number of queued packets of the  $i$ -th source at time  $k$  and;  $Q^{(i)}[k]$  is the queuing rate of the  $i$ -th source at time  $k$ .

The loss rate of the source  $I$  at the end of the epoch is given as

$$L(i) = \sum_{k=1}^K I^{(i)}[k] - \sum_{k=1}^K O^{(i)}[k] \quad (2)$$

where  $I^{(i)}[k]$  is the input rate of the  $i$ -th source at time  $k$ ;  $O^{(i)}[k]$  is the output rate of the  $i^{\text{th}}$  source at time  $k$

Each source only utilizes an aggregate portion of the server bandwidth when it has actual data to be transmitted [7]. STDM relies on statistic to define which channel needs more dedicated time to get more bandwidth.

### ***Impact of Statistical Multiplexing on Voice Quality***

Statistical multiplexing over the air interface is very interesting and challenging. The radio spectrum is a very expensive and scarce resource and there is need for it to be efficiently utilized. According to Enderes et al [8], whenever statistical multiplexing is applied, situations of temporary overload may occur, which have to be taken into account. In CDMA systems, a peak only degrades the carrier-to-interference ratio, which might be acceptable to some point. For TDMA systems, such as general packet radio service (GPRS), a number of parallel and independent voice flows share a common pool of channels (time slots); and, due to temporary lack of channels it may happen that some voice packets are dropped or delayed.

The impact of statistical multiplexing on QoS was examined by Endres et al. Simulation of the speech patterns for the users sharing the same wireless facilities (channels) becomes was done. A discrete-time version of Brady's Markovian two-state model of speech was implemented to simulate the packet generation characteristics of mobile users. Based on the model, a user can be either in talkspurt state  $S_t$  or in a silent state  $S_s$ .

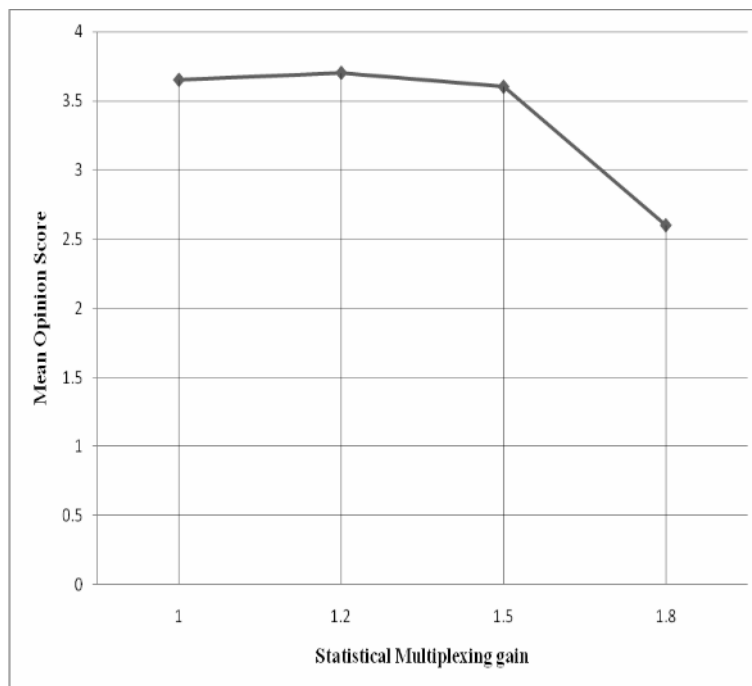
Talkspurt and silence lengths are assumed to be exponentially distributed and the transition probabilities are given by

$$T (s \rightarrow t) = 1 - \exp (-T_{\text{off}}/T) \quad (3)$$

$$T (s \rightarrow t) = 1 - \exp (-T_{\text{on}}/T) \quad (4)$$

where  $T = T_{\text{off}} + T_{\text{on}}$ .

Enderes et al simulated different statistical scenarios in order to assess the effect of front end clipping on voice quality [8]. They implemented a simulation in a real time demonstration platform utilized to acquire subjective indicators of voice quality by performing a Mean Opinion Score (MOS) Test. For the test, 6 females and 7 males were selected. They are all students living in Britain aged between 18 and 24 years. Each of them participated in the MOS test in two sessions, one with a female and the other one with a male partner, except for the first student and the last one, which took part only once. In each session, the four scenarios of which the sequence is selected randomly are graded. The MOS for each scenario is averaged over 24 grades. Fig. 2 shows the results of this experiment.



**Figure 2.** Statistical multiplexing gain versus Mean opinion score [8]

From the result, MOS drops with the increase in the statistical multiplexing gain, besides the slight increase between 1.0 and 1.2. The differences in subjective conversation quality between the line-switched case and the statistical multiplexing gain of 1.2 and 1.5 are not significant. For statistical multiplexing gain values above 1.5, the impact of front-end clipping starts becoming significant. In order to obtain a better understanding of that “good

quality” region (up to statistical multiplexing of 1.5) more points, i.e. more subjects and more tests, are required.

### *Statistical Multiplexing of Digitized Speech*

Bandwidth consideration is a major issue in determining the effectiveness and efficiency of communication systems. Several methods have been adopted in order to make a more economical use of speech transmission media and thereby reducing the bandwidth required to transmit a given speech information. A statistical multiplexer has an aggregate transmission bandwidth that is less than the sum of channel bandwidths because the aggregate bandwidth is used only when there is actual data to be transported from I/O ports.

The silent intervals that separate energy bursts in normal speech sounds can be judiciously utilized. In this approach, the Time Assignment Speech Interpolation (TASI), which represents an analog STDM scheme, was devised. The speech information is interpolated into the silent intervals so that greater information is carried in a given frequency bandwidth. In normal telephone conversations, a majority of time is spent in a latent (idle) state. TASI trunks allocate snippets of voice from another channel during this idle time. As digital speech processing became more common, TASI systems called Digital Speech Interpolation (DSI) were created. These had analog inputs and digital outputs. The DSI gain is defined as [9]

$$g = (n-k)/(c-k) \quad (5)$$

where:  $n$  = the number of subscriber channels;  $c$  = the number of transmission channels, and  $k$  = the number of data channels

### *STDM Performance Models*

Rajabi et al [10] studied time constraint of an M/M/1 queue. They simulated a proposed model for the multiplexing of a virtual channel. Assuming packet P1 is to be routed from node N1 to N2 and the destination is N3. If this packet is blocked in node N3, the entire resource is blocked in nodes N1 and N2. This implies that packet transfer through N1 or N2 is blocked. A virtual channel will resolve this problem. The Markov chain shown in fig. 3 illustrates the behavior of virtual channel occupancy of messages taking into account the timing constraint of these messages.

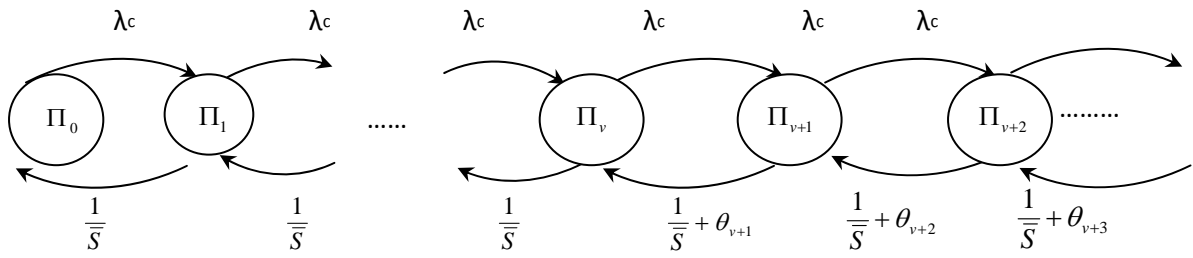


Figure 3. Markov chain for virtual channel [11]

State  $\Pi_n$  corresponds to  $n$  virtual channels being requested, when  $n \in [0, v-1]$ , arriving messages are able to find a free virtual channel. However, when  $n \in [v, \infty]$  an arriving message will find all virtual channels busy and will be forced to join the queue and will only be serviced when the virtual channel becomes free and provided the deadline time has not been exceeded. This system can be expressed analytically as shown in equations (6) and (7)

$$\Pi_n = (\lambda_c S)^n \Pi_0 \quad (n \leq v) \quad (6)$$

$$\Pi_{v+i} = (\lambda_c \bar{S})^v \Pi_0 \left( \frac{\lambda_c}{\theta} \right)^i / \Pi_{j=1}^i \left( \frac{1}{S\theta} + j \right) \quad (i \geq 1) \quad (7)$$

Heffes et al. [12] worked on Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance. They studied the performance of a statistical multiplexer whose inputs consist of a superposition of packetized voice sources and data. The performance analysis predicts voice packet delay distributions as well as data packet delay distributions. The superposition is approximated by a correlated Markov Modulated Poisson Process (MMPP), which is chosen such that several of its statistical characteristics identically match those of the superposition. Matrix analytic methods are then used to evaluate system performance measures. In particular, they obtained moments of voice and data delay distributions and queue length distributions.

### The M/M/1 Queue Model

In the M/M/1 queuing model it is assumed that both arrivals and service times are negative exponentially distributed (or inter-arrival distribution or service time distribution is Markovian) with a single server. It is the most widely used queuing system and it is a good



approximation for a large number of queuing systems. The simplest queuing system consists of two components; the queue and the server. The M/M/1 queuing system is that with Poisson arrival process (M), exponential service times (M), single server (1) and an infinite number of waiting positions. The state of the system is defined as the total number of customers in the system (either being served or waiting in queue) [13]. A sequence of random events can be modeled by a Poisson arrival process. The model for a link transmission system which has a Poisson arrival process and an exponential service distribution is called an M/M/1 queue. The M/M/1 queue is the most basic and important queuing model.

According to Rajabi et al [10], the exponential service distribution is given as

$$S = \mu e^{-\mu t} \quad (8)$$

where  $\mu$  = service rate.

A very important parameter,  $\rho$ , of queuing systems is the ratio of the arrival and the service rate. The parameter is called traffic intensity, traffic rate, utilization factor or occupancy. It is defined as the average arrival rate,  $\lambda$ , divided by the average service rate,  $\mu$ . The value of  $\rho$  shows how busy the server is. For a stable system, the average service rate will be higher than the average arrival rate; hence,  $\rho$  will be less than unity ( $\rho \leq 1$ ). Otherwise, the queues will grow permanently, infinitely and boundless. This is for the average rates. The instantaneous arrival rate may exceed the service rate. However, over a long period of time, the service rate should always exceed arrival rate.

The utilization of the multiplexer is given by [10, 14]:

$$\rho = \lambda/\mu \quad (9)$$

The mean number of packets in the system for M/M/1 queue is given by [14]

$$E[N] = \lambda/(\lambda - \mu) = \rho/(1 - \rho) \quad (10)$$

The mean number of customers in the queue is given by

$$E[W] = \lambda^2/[\mu(\mu - \lambda)] = \rho^2/(1 - \rho) \quad (11)$$

From the equation (11), it is clear that as  $\rho$  approaches unity, the number of customers becomes very large. This can be easily justified intuitively. The utilization,  $\rho$ , will approach 1 when the average arrival rate,  $\lambda$ , starts approaching the average service rate,  $\mu$ . When this happens, the server would always be busy, hence leading to a queue build up (i.e.  $E[N]$  becomes large). The average time in queue is

$$E[W_q] = (\rho/\mu) * [1/(1 - \rho)] = \lambda/[\mu(\mu - \lambda)] = \rho T_s/(1 - \rho) \quad (12)$$

where  $T_s$  is the average service time for each arrival. The average time of the customer waiting in the system (or spent in the system) is also known as the response time and it is sum of average time in queue and service time.

$$E[W] = E[W_q] + 1/\mu$$

$$E[W] = \lambda/[\mu(\mu - \lambda) + 1/\mu] = 1/(\mu - \lambda) = T_s/(1 - \rho) \quad (13)$$

From equation (13) it can be deduced that as the mean arrival rate  $\lambda$  approaches and equals the mean service rate,  $\mu$ , the waiting time becomes very large.

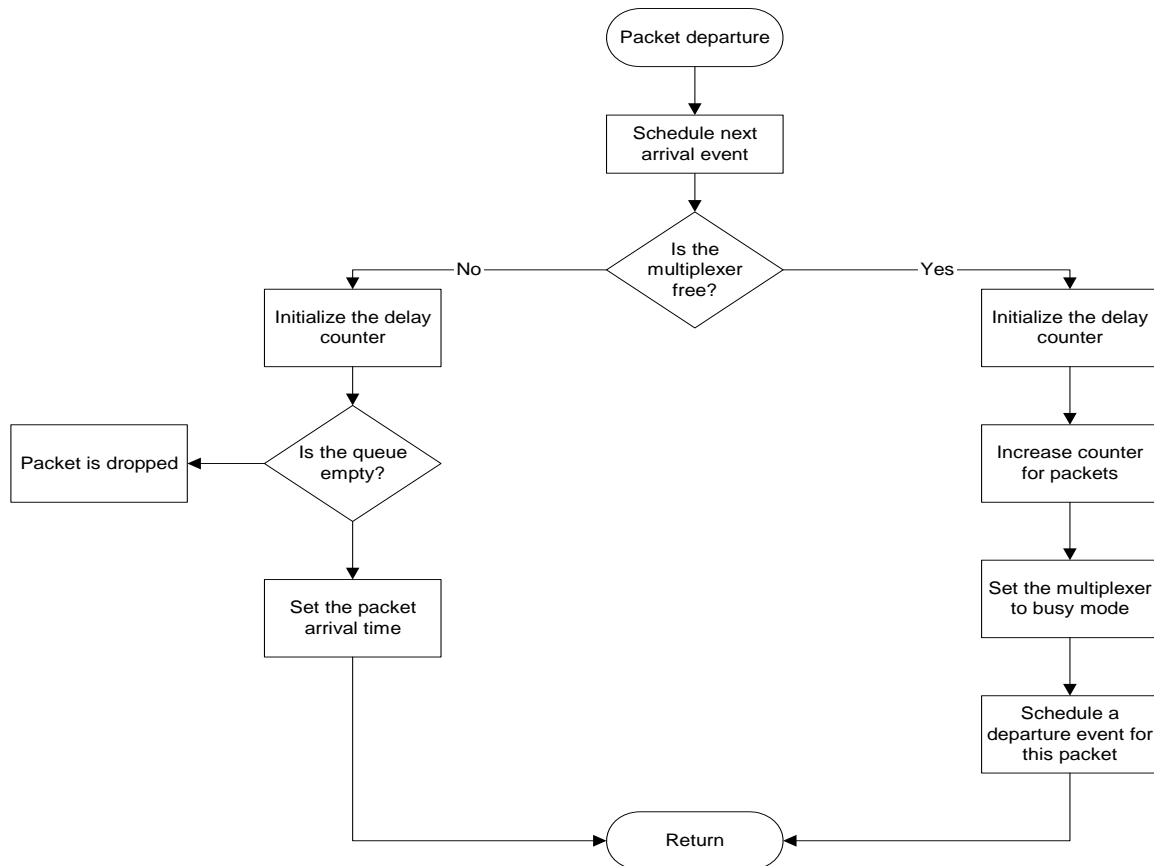
### ***The Multiplexer Model***

In the M/M/1 queue model the server provides service to the arriving packets. When a packet arrives and the server is idle the packet is served immediately, otherwise the packet joins a queue. When the server has served the packet, the packet departs and if there are packets waiting in the queue, one is immediately dispatched to the server. Therefore, the model consists of two events:

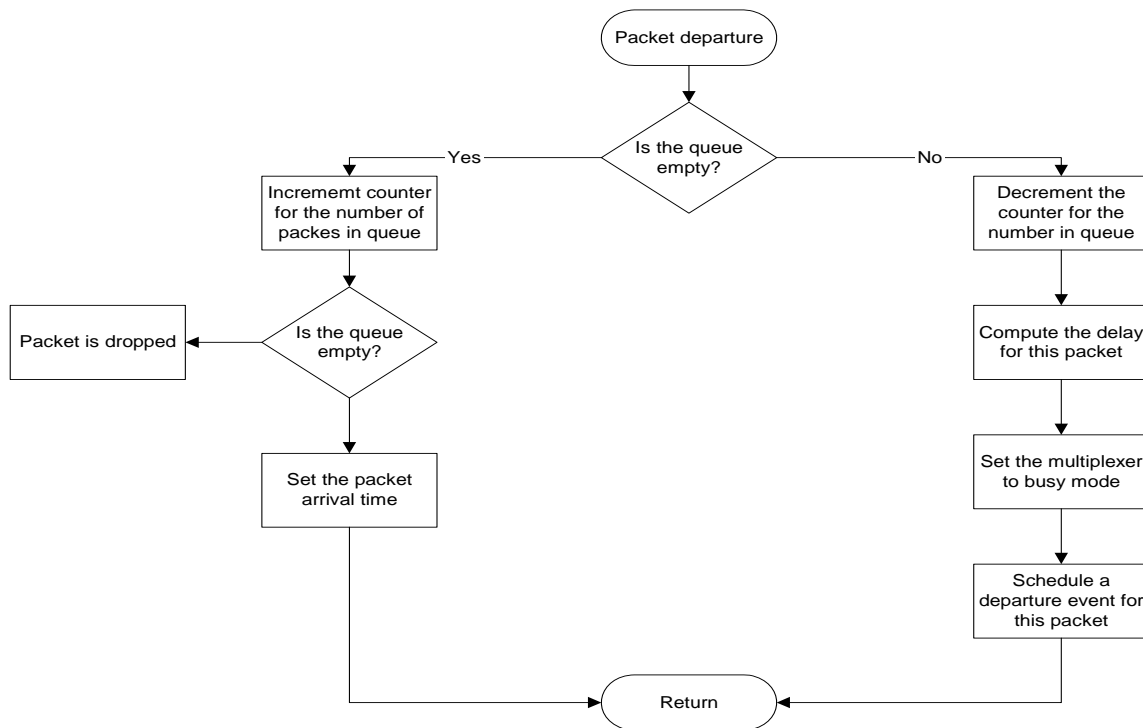
- Packet arrival
- Service completion

Figures 4 and 5 show the flowchart of the multiplexer arrival and departure routines respectively for this model. When there is packet arrival and the server is idle the packet is immediately served and then a departure event is scheduled to coincide with the service completion. However, when there is packet arrival and the server is busy then the packet is queued. From such model, it is possible to estimate the following outcomes:

- The average packet response time (time of arrival to time of departure)
- The throughput rate (packets served per unit time)
- The server utilization (percentage of elapsed time that the server is busy)
- The average queue length (number of packets at the facility)



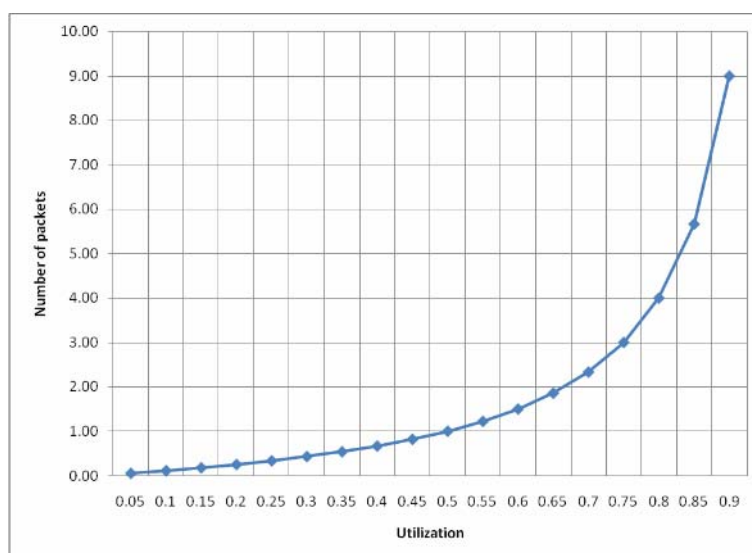
*Figure 4. Packet arrival routine*



*Figure 5. Packet departure routine*

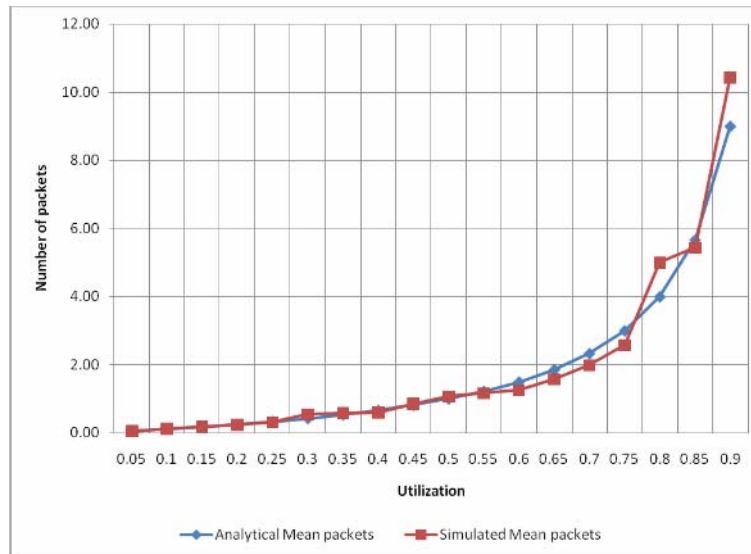
### The M/M/1 Analytical and Simulated Queuing Model Results for Multiplexer System

The result for the mean number of packets in the multiplexer system is shown in Fig.6. The figure shows a plot of the result of the analytical M/M/1 queuing model generated for 0-90% load. It reveals that the number of packets in the multiplexer increases exponentially as the utilization increases. At low workload intensity (low utilization), as expected, an arriving packet meets low competition, hence the multiplexer transmits packets infrequently and therefore because there is no much contention, the number of packets in the entire system is small. However, as the utilization increases above the “knee”, there is more contention and therefore the number of packets in the multiplexer system is significantly increased. The multiplexer reaches saturation at a certain arrival rate, when the utilization is very close to 1, therefore if the link utilization approaches 100%, delay approaches infinity.



**Figure 6.** Utilization vs. average number of packets in an M/M/1 System for offered load of 0-90%

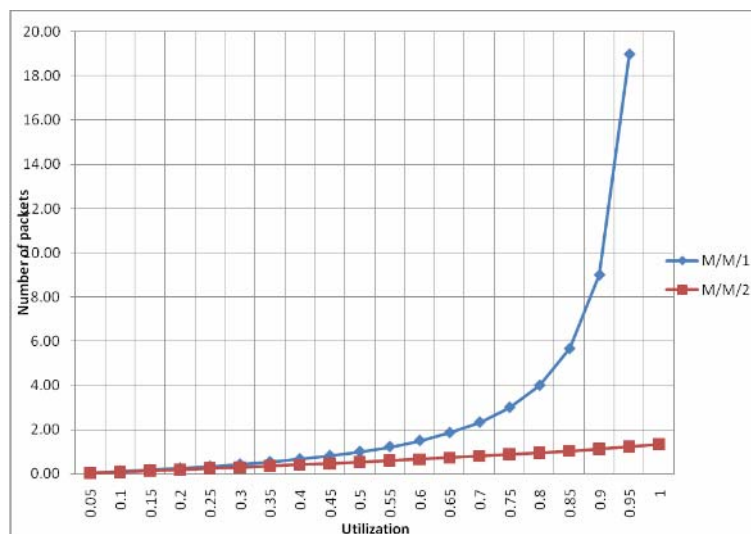
Figure 7 shows the plot of the comparison between the theoretical predictions and the values obtained from the simulation program for Utilizations between 0-0.9. The figure shows that up to utilization of 0.90, simulation results have a very close range with the analytical results. However, at higher utilization unreliable results may be obtained. For the multiplexer to work efficiently the utilization should be kept below 0.8.



**Figure 7.** Comparison of Analytical and Simulated Mean Number of packets in an M/M/1 Multiplexer System for 0-90% offered load

### **The M/M/k Analytical Queuing Model Results**

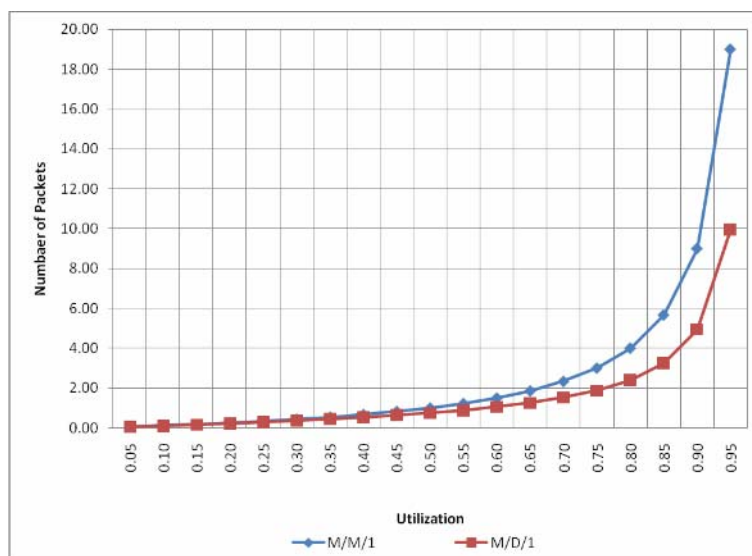
Figure 8 show the analytical results for the mean number of packets in queue for the M/M/k queuing model when the multiplexer is viewed as a 2-server system. The results shows the M/M/2 is twice more scalable than the M/M/1. As seen, utilization can only go beyond 1 in an M/M/2 multiplexer system. Up to a utilization of about 0.4 it is observed that the number of packets in the M/M/1 and M/M/2 are close but above this utilization the values for the M/M/2 are consistently lower.



**Figure 8.** Mean number in System versus utilization for M/M/1 and M/M/2 analytical queue models

### *M/M/1 and M/D/1 Queuing Result*

Figure 9 shows the result of the mean number of packets versus utilization for M/M/1 and M/D/1 analytical queue model.



**Figure 9.** Mean number of packets versus utilization for M/M/1 and M/D/1 analytical models

### **Discussion**

Simulations were done and data were collected for the graphical plots of utilization against buffer size in order to determine the efficiency of STD. The performance of the statistical multiplexer is described by the number of packets in the system and in queue, average delay and buffer size. Fig. 6 shows how the mean number of packets in system increases with utilization. At low workload intensity (low utilization), an arriving packet meets low competition. Hence, the multiplexer transmits packets infrequently and therefore because there is no much contention, the number of packets in the entire system is small. However, as the utilization increases, there is more contention and more packets in the system. The multiplexer reaches saturation at a certain arrival rate, when the utilization is very close to 1, therefore if the link utilization approaches 100%, delay approaches infinity.

Fig. 8 compares the mean number in queue versus utilization for a single server and 2-server queue. The single server reaches saturation when the utilization is very close to 1 while the 2-server reaches saturation when the utilization is close to 2.

Fig. 9 compares plot values of average queue size versus utilization for both M/M/1 and M/D/1 queue models. The plot result shows that the M/M/1 model is visually above the M/D/1 model, therefore, the poorest performance is exhibited by the exponential service time, and the best by the constant service time. For higher utilization, the variability of the mean queue size and the residence time is greater (than at low offered loads or utilization) for M/M/1 and M/D/1 models. In M/D/1 residence time is lower than that of M/M/1 due to the lower variance in service time. For high utilization, the difference between M/M/1 and M/D/1 is approximately a factor of 2.

### Conclusions

Simulation results to analyses the performance of the statistical multiplexer when viewed as M/M/1 and M/M/k queue model has been presented in this work. Comparison was also done between the M/M/1 model and the M/D/1. Results confirm that increase in the utilization leads also into increase in the number of packets in the queue and in the system. These results also imply a corresponding increase in the buffer size and packets delay. Randomness (in service and arrival) is what causes queuing at buffers. Results show that queues and queuing delay increases dramatically as utilization becomes greater than 90%. Comparison between the M/M/1 and M/D/1 reveals that M/D/1 outperforms M/M/1 to up to 50% at higher utilization.

### References

1. Wesley W. C., *Design Considerations of Statistical Multiplexers*, MSc thesis, Computer Science Department, University of California, Los Angeles, California, 1998.
2. William S., *Data and Computer Communications*, Seventh Edition, Prentice-Hall of India Private Limited, New Delhi-110001, 2004.
3. Regis J. B., Donald W. G., *Voice and Data Communication Handbook*, Fourth Edition, 2000.
4. William B., *Distributed Systems and Networks*, McGraw-Hill Publishing Company, 2001.

5. Ray H., *Telecommunication and Data Communication Handbook*, Copy right c 2007.
6. Homayoun Y., *A Neuro-Forecast Water-Filling Scheme of Server Scheduling*, Department of EECS, University of California, Irvine hyousefi@uci.edu, 2002.
7. Homayoun Y., Edmond A. J., *Dynamic Neural-Based Buffer Management for Queuing Systems with Self-Similar Characteristics*, IEEE Transactions on Neural Networks, Volume 16, No. 5, September, 2005.
8. Enderes T., Khoo S.C., Somerville C.A., Samaras K., *Impact of Statistical Multiplexing on voice quality in cellular networks*, MSWIM 2000 Boston MA USA. Copyright ACM 2000 1-58113-304-9/00/08.
9. David R. S., *Digital Transmission Systems*, 3<sup>rd</sup> Edition, 2002.
10. Ali R., Farhad H., *Time Constraint M/M/1 Queue*, Faculty of Electrical and Computer Eng., School of Engineering, University of Tehran, 2003.
11. Jianwei H., Chee W., Mung C., Raphael C., *Statistical Multiplexing over DSL Networks*, Prentice Hall Publishers, 2000.
12. Heffes H., Lucantoni D., *A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance*, Selected Areas in Communications, IEEE Journal on Volume 4, Issue 6, Sep 1986.
13. Handbook, *Teletraffic Engineering*, worked out as a joint venture between the International Telecommunication (ITU) Union and International Teletraffic Congress (ITC), Geneva, 2005.
14. Brian D. B., *Basic Queuing Theory*, Prentice Hall Publishers, 2000.