Modeling and Analysis of Traffic in High Speed Networks

by

Soma S. Muppidi

B.Tech., Jawaharlal Nehru Technological University, India, 1995

Submitted to the Department of Electrical Engineering and Computer Science and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements
for the degree of Master of Science.

_____
Professor in Charge

_____

_____
Committee Members

_____
Date thesis accepted

Dedicated to

My Parents, Suryanarayana Murty and Satyavani for their infinite love, encouragement and support and to my sisters Padma and Uma for all their affection.

Acknowledgments

I would like to thank Dr. Victor Frost, my advisor and committee chair for his patience and encouragement throughout my study here at KU. His guidance and advice have been invaluable throughout the course of this work. I would also like to thank him for giving me the opportunity to work on the AAI project. It has been a tremendous learning experience and I am grateful for it. I would also like to thank Drs. David W. Petr and Joseph B. Evans for serving on my masters committee and for their guidance during the course of my study.

I thank my colleagues on the project especially, Roel Jonkman and Georgious Lazarou. I enjoyed working with them and it has benefited me greatly. I also thank Sampath for his suggestions in improving the draft.

Abstract

Recent traffic studies of high-resolution, high quality traffic measurements have revealed the phenomenon of long-range dependence in network traffic. The implication of the presence of long-range dependence in traffic is that actual network traffic exhibits more bursty behavior compared to traditional ”Poisson-like” models. Traffic modeling and performance prediction in networks with long-range dependent flows is required for the design of efficient congestion control, routing and other network management algorithms. Here a non-Markovian phase process is used to model the traffic process. The phase process captures the macro-dynamic properties of the traffic. The traffic dynamics within each phase i.e., the traffic micro-dynamics, are then described by a random process with finite mean and variance. Network performance using this model has been evaluated in the form of delay vs load curves and cell loss ratio vs buffer size. By capturing the burstiness of the network traffic in the form of traffic macro-dynamics, the model developed here predicts ATM queue performance in network flows that may be inherently long-range dependent in nature. Extensive simulations based on traces of traffic collected from a wide area ATM network has been used to validate the developed model and performance analysis methodology. Also the effects of traffic micro-dynamics on performance has been investigated within the frame work of the developed model. It has been observed that the mean delay is not sensitive to the specific nature of the micro-dynamics while the cell loss is affected by the traffic micro-dynamics.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Broadband networks of the future will carry audio,video and data traffic from many diverse applications. Such networks will need to meet a variety of traffic and performance requirements. Asynchronous Transfer Mode (ATM) has been chosen as the technology to implement B-ISDN. Details about ATM will not be discussed here but can be found in [34]. ATM offers many advantages like bandwidth on demand, Quality of Service (QoS) guarantees etc., which facilitate many new application like multi-media. The integration of traffic from many different types of applications is expected to generate traffic having complex temporal characteristics. In particular broadband network traffic has a complex correlation structure that spans often a wide range of time -scales. Such long-range dependence is not taken into consideration in traditional Markovian models. Thus, conventional traffic models yield system performance predictions that are significantly different from what would actually be obtained in a real networking environment.

Several recent traffic studies [7], [8], [9], have reported that correlation in the traffic arrival process has a significant impact on the network resource management and network performance evaluation. Particularly it has been observed that long range

dependence in the input traffic process dominates the queueing performance. In the context of network traffic, long-range dependence implies that even if the number of packet arrivals is aggregated over larger and larger time-scales, the aggregated process will not smooth out as is expected with finite variance processes like Poisson process or short-range memory processes like Markov-Modulated Poisson Processes.

The benefits of ATM like QoS guarantees, scalability come at a price. Contrary to common misconceptions, ATM is a very complex technology, perhaps the most complex ever developed by the networking industry [22]. ATM signaling protocols and routing protocols indicate the complexity involved with the technology. The increased complexity of operation of ATM networks is expected to generate a significant amount of management traffic in these networks. In this context, accurate modeling of traffic and studying its performance on queueing is an issue of interest.

1.1 Traffic Models

Network traffic can be characterized by a rate process. The rate process is a random process $R(t)$, defined as the short term time average of the random process $A(t)$, which represents the arrival process. The arrival process $A(t)$ denotes the number arrivals at a point in the network, at time t. The rate process can be expressed as:

$$R(t) = \frac{[A(t + \Delta t) - A(t)]}{\Delta t} \tag{1.1}$$

where,

$A(t)$ is the cell count at time t.

$R(t)$ is the throughput at time t.

Traffic modeling aims at developing mathematically tractable models of network traffic with a view to capture important parameters that have impact on the queueing performance. The queueing performance for broadband networks is usually quantified in terms of the mean delay experienced by cells arriving at the input to the queuing system and the cell loss probability which represents the percentage of cells that were not transmitted successfully from the queueing system. As shown in Equation (1.2) the cell loss probability P_l in a finite buffer queueing system of size N is often approximated by $P(Q > N)$ in the corresponding infinite buffer system [33].

$$P_l \approx P(Q > N) \tag{1.2}$$

Q is a random variable representing the queue length.

N is the size of the finite buffer.

Note that $P(Q > N)$ represents the tail of the queue length distribution and hence provides a conservative estimate of the upper bound for the cell loss probability. Figures 1.1 and 1.2 show the typical mean cell delay curves and cell loss probability curves respectively, that would be obtained from a short-range dependent model and that from a model that takes into account long-range dependence. In Figures 1.1 and 1.2 curve (A) is the performance profile predicted by a conventional traffic model and curve (B) shows the corresponding profile that would typically be obtained from a long range dependent model. Note that the delay curve obtained from a conventional traffic model predicts useful operation for utilization values close to 1. Similarly, in Figure 1.2 the curve representing the logarithm of the cell loss ratio plotted against the buffer size bears a non-linear relationship in the case of long range dependence. Since many recent traffic studies like [7], [8], [9] have confirmed the presence of long-range dependence

in network traces, performance curves shown by (B) in Figures 1.1 and 1.2 can be expected to characterize mean cell delay and cell loss probability observed in operational broadband networks. Figures 1.1 and 1.2 clearly indicate that conventional traffic models cannot be used for performance evaluation of real network traces and there is a need to develop alternative traffic models for performance evaluation.

Clearly, traffic modeling and performance prediction is required for the design of efficient congestion control, routing and other network management algorithms. Traffic modeling and performance prediction is in general a hard problem and involves computationally intensive procedures. The traffic arrival processes cannot be matched exactly. Models developed in literature aim at capturing the first order and second order statistics of the rate process (defined by Equation (1.1)). Some of the approaches available in literature for modeling network traffic are Markov Modulated Poisson process (MMPP) [26], Circulant Modulated Rate process (CMRP), Auto Regressive Moving Average (ARMA) process. The above mentioned models can be considered traditional in the sense that these models assume the arrival rate process to have a finite mean and finite variance. Recent traffic studies have convincingly established the presence of long-range dependence in network traffic. Notable among the studies available in literature are the studies by [7], [8], [9]. The ratio between the maximum rate and average rate is measure of the traffic burstiness. The implication of the presence of long-range dependence in traffic is that actual network traffic is burstier than that predicted by traditional "Poisson-like" models. Burstiness impacts many aspects of congestion control and traffic performance. Contrary to the behavior assumed by standard traffic models, aggregate traffic in networks with long-range dependent flows, becomes burstier as the number of active sources increases. Switch buffers having a long-range dependent arrival process as input get filled up faster than those predicted by assuming a Poisson like arrival process. Common approaches for modeling long-range dependent traffic

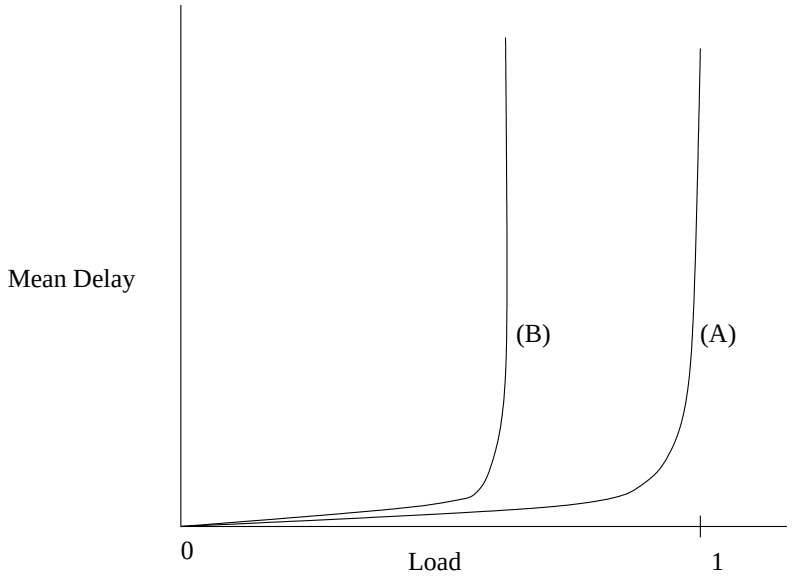

Figure 1.1: Typical Delay curves for long-range dependent model (B) and conventional model (A).

include Fractional Brownian Motion (FBM) [24], Fractional auto regressive integrated moving average processes (ARIMA), chaotic maps [25].

The origin of long-range dependence has been explained in [27] in the frame work of the familiar ON-OFF source model. Real traffic sources are different from existing models as the sojourn times in the ON and OFF states of real sources, follows a long-range dependent process like Pareto etc., contrary to that assumed in standard statistical ON-OFF models. The difference in the sojourn times, results in traffic that is considerably more bursty compared to traffic generated by a conventional ON-OFF source.

1.2 Definition and Characteristics

As long-range dependence has been demonstrated to be an important issue in performance evaluation and traffic modeling in networks, this section deals with, the defini-

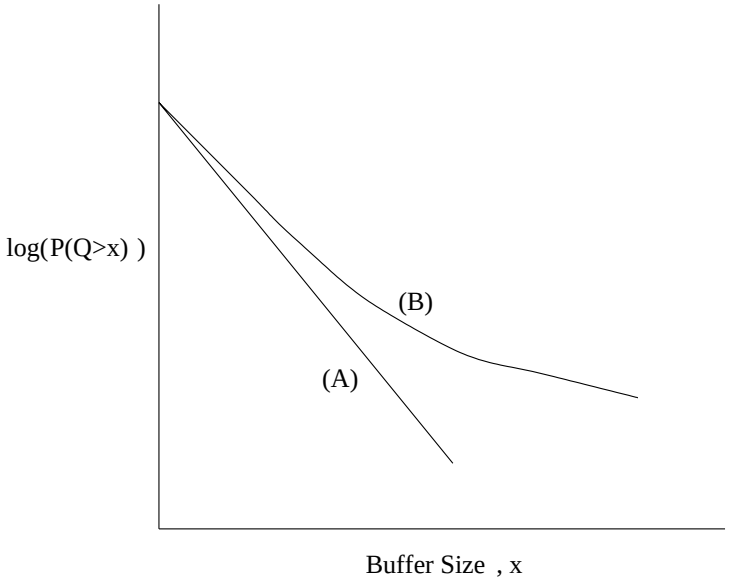

Figure 1.2: Typical loss curves for long-range dependent model (B) and conventional model (A).

tion and the main characteristics of a long-range dependent process.

Let X = (X_t, t= 0, 1, 2...) be a covariance stationary process with mean μ, variance σ^2 and autocorrelation function $r(k)$, $k > 0$. Further suppose that $X^{(m)}$ = ($X_k{}^{(m)}$, k= 1, 2, 3,...; m= 1, 2, 3, ...) be a derived covariance stationary process which is obtained by averaging X over non-overlapping blocks of size m. Therefore, according to the definition, $X^{(m)}$ is given by:

$$X_k{}^{(m)} = \frac{1}{m} \sum_{i=0}^{m-1} X_{km-i} \tag{1.3}$$

Let $r^{(m)}$ be the auto-correlation function of $X^{(m)}$. The process X is called exactly self-similar if

$$r^{(m)}(k) = r(k), k \geq 0 \tag{1.4}$$

the process X is called asymptotically self similar if for large values of k

$$r^{(m)}(k) \rightarrow r(k), m \rightarrow \infty \tag{1.5}$$

Mathematically, self-similarity manifests in the following equivalent ways [9].

The autocorrelation function of the rate process decays slowly, that is the auto-correlation function decays hyperbolically rather than exponentially, implying a non-summable auto-correlation function, which in turn indicates that the process has infinite variance. The auto-correlation function of a general long-range dependent process can be expressed as:

$$r(k) \approx k^{-\beta} G(t), k \rightarrow \infty \tag{1.6}$$

where $0 < \beta < 1$ and $G(t)$ is a slowly varying function of time. A slowly varying function of time can be defined as:

$$\lim_{t \rightarrow \infty} \frac{G(tx)}{G(t)} = 1, x > 0 \tag{1.7}$$

Equations 1.6 and 1.7 imply that the auto-correlation function of the rate process is non-summable i.e.,

$$\sum r(k) = \infty$$

In the frequency domain the power spectral density function obeys a power law near the origin. If $P(\omega)$ denotes the power spectral density, then $P(\omega) \rightarrow \infty$ as $\omega \rightarrow 0$.

Several methods are used to detect self-similarity and estimate the degree of self similarity. Popular techniques include the time domain analysis based on the R/S statis-

tic, analysis of the variances of the aggregate processes $X^{(m)}$, Whittle's periodogram based analysis in the frequency domain. Detailed definitions and tests for self-similarity can be found in [9] and [28].

1.3 Contributions of this Thesis

In this section an outline of the model developed is given and the main contributions of this research are listed. In this thesis a new model as well as a new performance evaluation methodology are developed which accurately capture the queueing performance of observed traffic in a wide area high speed ATM network. The model is a non-Markovian phase modulated process(representing the macro-dynamics of the process) where the dynamics of the arrival process in each phase are described by a random process which has a finite mean and finite variance (which characterize the micro-dynamics of the process). The steady state probabilities of being in each state are determined from an appropriate long-range dependent distribution for the rate process. The rate process is the short term time average of the number of cell arrivals. More details about the rate process are included on page 2. The characteristics of the long-range dependent component of this model may be obtained in several ways:

1. A form for the distribution may be assumed and its parameters estimated or

2. An estimated histogram may be used.

Both these methods will be used here.

The durations of each states of the phase process are proportional to the steady state probability of being in that state. Network performance using the model is presented in the form of delay vs load curves and logarithm of the cell loss ratio vs buffer size. The knee of the delay curve is important from a traffic engineering perspective as it

represents a bound on the useful utilizations in the queueing system. The model developed in this thesis is shown to accurately represent the behavior of the knee of the delay curve. The main contributions of this study are given below:

- Many traffic studies have convincingly established the non-Poisson nature of network traffic in many networks. This research establishes the non-Poisson nature of traffic on an early national scale ATM network, the AAI network. AAI stands for ACTS ATM Internetwork.
- A simple traffic model is developed for traffic in high speed wide area ATM networks.
- A performance analysis technique is developed to estimate:
    1. the mean cell transfer delay and
    2. cell loss ratio at a queueing system.
- The developed performance methodology is validated using traffic measurements from the AAI wide area network.
- The effects of traffic micro-dynamics on performance in terms of mean cell delay and cell loss ratio are investigated in the frame work of the developed model.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows:

1. In Chapter 2 the analytical traffic model that is developed is presented. The different network traffic parameters that define the model along with the method for estimating the parameters is also presented.

2. Chapter 3 discusses the performance evaluation methodology developed. Results from the theory of discrete queues are used to evaluate the performance of a queueing system in terms of the mean cell delay and cell loss ratio.

3. Chapter 4 deals with the AAI network and the data collection process being performed on the AAI network. Detailed configuration of the AAI network, the sites involved in the monitoring process and information on the tools used for data collection is presented. Some observations of the traffic profiles are also given in this chapter.

4. Chapter 5 describes the simulation model developed in this research. This simulation model was used measure the queueing performance based on the collected trace data. Information on the queue simulator which was developed and validation of the simulator are presented. The validation is done by obtaining mean cell delay values for an exponential source using the simulator and comparing the values obtained with standard theoretical results. A similar methodology has been used for validating the cell loss ratio results obtained from simulations based on collected network traces, using the developed simulator.

5. Experimental evaluation of the model and the performance evaluation technique are done in Chapter 6. Experimental queueing results from simulations and analytical results from the developed model are presented in this chapter. The results are presented in the form of delay and loss profiles of the appropriate queueing system. In addition the delay and loss curves are obtained by assuming the Cumulative Distribution Function of the rate process follows Pareto distributions. Chapter 6 also examines the effect of traffic micro-dynamics on queueing performance.

6. Several conclusions from this research are discussed in Chapter 7. Some recommendations for future work are also included.

Chapter 2

Analytical Model

2.1 Background

In this chapter an analytical model for traffic in high speed wide area ATM networks is proposed. The traffic is modeled as a phase modulated process in which the long-range dependence, is captured in the state probabilities of the phase process. Many recent traffic studies [8] have convincingly established the presence of long range dependence in network traffic. A common approach in literature for modeling long range dependency is through the use of self-similar processes, e.g., fractional Brownian motion. However, such models are complex and queueing response cannot be obtained effectively using such models [17], [26], [24], [25]. In this thesis results from the theory of discrete queues, are used to develop an analytically tractable model for modeling traffic in high speed networks.

Formally (as given in [13]) the first-order statistics of a long range dependent process can be defined as follows:

Definition 1 A distribution function F_X of a random variable X on $[0, \infty)$ is called

long-tailed if

$$\lim_{x\to\infty} \frac{1 - F_X(x-y)}{1 - F_X(x)} = 1, y \in R.$$

From the definition and from the main characteristics of long-range dependent processes that were described in Chapter 1, it can be seen that long range dependent processes decay slower than exponential processes. The long tailed random processes are associated with infinite variance and either finite or infinite mean. The processes are also referred to as $Non - Cramer$ type [11], [12]. The queueing techniques developed under the general assumptions of a $Cramer - type$ [11], [12] process are based on the Moment Generating Function techniques [1], [2], [3] or Matrix-Geometric Techniques [4], [5]. Each of these techniques has its own advantages. The moment generating function of a random variable X is defined as:

Definition 2 If X is a random variable then the moment generating function (MGF), $M(\theta)$ of X is defined as $M(\theta) = E[e^{\theta x}]$.

From Definitions 1 and 2 it can be seen that the MGF of a long-tailed random process is infinity. Because of the non-existence of $M(\theta)$ for long-range dependent processes, regular queueing techniques cannot be applied to study the dynamics of these processes. A model is proposed here such that, the dynamics of the rate process in each state are under the general class of $Cramer - type$ processes, but the probabilities of being in a given $Cramer$ state is obtained from the inherent characteristics of the long range dependent traffic. Such a model, as demonstrated here is useful for modeling long-range dependence in network traffic. The proposed model is also seen to capture the second-order statistics of the rate process. An excellent treatment of queueing theory for a discrete time queue is given in [1], on which most of the results used in this

report are based on.

2.2 Traffic Model

In this section the analytical model and the procedure for obtaining the model parameters is presented. Let the random process $R(t)$ denote the rate process of the traffic, which is defined as the short term time average of the random variable representing the number of cell arrivals at a specific point in the network. The detailed definition of the rate process and an explanation about its measurement from cell counts is given in Section 4.2. Results from queueing theory are then employed in obtaining analytical expressions to estimate the performance metrics like the mean cell transfer delay and cell loss ratio at a queueing system whose input process is the random process $R(t)$, the process being modeled. Here the arrival rate process is modeled as being modulated by a phase process. Assume that the phase-process is stationary and ergodic having a finite state space **S**=$\{x_1, x_2, ...x_N\}$. This phase-process captures the macro-dynamic properties of the rate process. Within each state $x_i \in S$, the arrival process is assumed to follow a point process with a distribution with a finite mean and variance. The distribution function of the rate process associated with each state defines the traffic micro-dynamics of that state.

The steady state phase probabilities of the stationary phase process are assumed to follow a distribution with a heavy tail i.e., infinite variance. The variables which characterize the process considered here are:

1. The steady state probability vector $\bar{\pi} = [\pi_1, \pi_2, ..., \pi_N]$, where $\pi_i = P(S = x_i)$.

2. The rate vector $\bar{\gamma} = [\gamma_0, \gamma_1, ..., \gamma_N]$ where $\gamma_1 \leq \gamma_2 \leq .... \leq \gamma_{N+1}$ represent the boundary rates (in cells/sec) for each state $x_1, x_2, ...x_N$

Here we assume that the mean sojourn time in each state of the phase process is proportional to the steady state probability of being in that state. Within each state say state x_i the arrival process is assumed to have $Cramer - type$ characteristics, example, exponential, uniform etc., with mean arrival rate equal to γ_{i+1}. Note that γ_{i+1} is the upper bound on the rates for state x_i and is thus a conservative assumption for performance modeling. The state probabilities themselves are obtained from a $non - Cramer$ type distribution having infinite variance. In this work a Pareto distribution is used, however, any infinite variance process may be assumed to obtain the state probabilities. In particular rate histograms estimated from network trace data have also been used in this work to determine the appropriate state probabilities. It should be noted that in [7], the authors report from empirical findings the usefulness of Pareto Distribution in modeling collected network data that is inherently long range dependent in nature. The results reported here will compare trace based simulations to performance predictions obtained using Pareto and histogram derived state probabilities.

2.2.1 Procedure for obtaining the Rate vector

The vector $\bar{\gamma}$ as defined above is obtained by partitioning the range of possible rates into 'N' sets or levels. The range of possible rates is lower bounded by γ_{min} and upper bounded by γ_{max}. These are often the minimum and maximum rates observed in a measured network trace. These values may also be derived by experience. This is equivalent to quantizing the range of rates into N levels.

Mathematically a partition is defined as given in Definition 3.

Definition 3 A partition Γ on a range Σ is defined as a collection of mutually exclusive subsets σ_i of Σ whose unions equals Σ.

Applying the definition of a partition as given by Definition 3 to $R(t)$, partitioning the random process results in quantizing $R(t)$ into N levels, where each level represents a $'state'$. Each partitioned set represents the set of rates associated with a given state.

The vector $\bar{\gamma}$ for a given data trace is computed as follows. As mentioned before, let the peak cell rate, minimum cell rate of the traffic be denoted as γ_{max} and γ_{min} respectively. Given γ_{max} and γ_{min} the rate vector $\bar{\gamma}$ is chosen such that:

$$\gamma_1 = \gamma_{min} \tag{2.1}$$

$$\gamma_i = \gamma_1 + \frac{\gamma_{max} - \gamma_{min}}{N}, i = 2, 3, 4, ..., N \tag{2.2}$$

$$\gamma_{N+1} = \gamma_{max}. \tag{2.3}$$

Note that Equations (2.1), (2.2) and (2.3) define a uniform quantizer on the random process $R(t)$ and also satisfy the definition of a partition given in Definition 3. The inter-arrival times in each state which represent the traffic micro-scale dynamics of the associated state are chosen to follow a fixed distribution with finite variance and mean γ_{i+1}, while in state x_i. In this analysis it is assumed that the mean arrival rate associated with state x_i is γ_{i+1}, which is the upper bound on the rates in that state. Note that the assumption is conservative and results in an upper bound on any performance parameter estimated using the model.

2.2.2 Procedure for obtaining the Steady State Probability vector

Having obtained $\bar{\gamma}$, the procedure used for obtaining the steady state probability vector $\bar{\pi}$ is now discussed. The vector $\bar{\pi}$ is obtained by quantizing the CDF of the rate process into levels. Two different ways of getting the CDF of the input process have been considered as explained below:

In the presence of long-range dependence, the CDF of the rate process will be similar to that of a heavy-tailed distribution F(.), like Pareto, Log-normal etc., with appropriate parameters. Taking into account the versatility of a Pareto Distribution in modeling network traffic, in this study a Pareto distribution is chosen to obtain an estimate of the state probability vector $\bar{\pi}$, for each of the states defined by the rate vector $\bar{\gamma}$ to obtain the stead state probability vector.

In the most general form, the Cumulative Distribution Function of a Pareto distribution can be given as

$$F_X(x) = 1 - Kx^{-\beta}, \quad \beta > 1, x > K. \tag{2.4}$$

In Equation (2.4), K is the location parameter and β is the shape parameter describing the CDF of the Pareto distribution. The maximum likelihood estimate [37] of the shape parameter β, for a given set of data samples $D = \{d_1, d_2, .., d_M\}$ is given by Equation (2.5) as

$$\beta = \frac{1}{\frac{1}{M}\sum_{i=1}^{M} ln(d_i)} \tag{2.5}$$

The value of the location parameter K, is then determined such that the area under the probability density function of the Pareto distribution equals unity. The shape parameter can also be estimated empericaly form experience.

The state probability vector, $\bar{\pi}$, may also be obtained from an empirical CDF of the rate process obtained from collected network trace data. Suppose that the random process $R(t)$ which denotes the rate process of the collected trace, has a CDF such as that shown in Figure 2.1. This CDF is quantized into N levels at the rate values defined by $\bar{\gamma}$, to obtain the steady state probability vector.

After obtaining the CDF of the rate process either from the theoretical Pareto distribution or from the histogram, the steady state probability π_i for state x_i, which is bounded by rates γ_i and γ_{i+1} can be computed as:

$$\pi_i = P[X \leq \gamma_{i+1}] - P[X \leq \gamma_i], \quad i = 1, 2, 3..., N-1 \tag{2.6}$$

$$= F_X(\gamma_{i+1}) - F_X(\gamma_i), \quad i = 1, 2, 3..., N-1 \tag{2.7}$$

Observe that from Equations (2.6) and (2.7) the information in the CDF of the rate process is now captured in the state probability vector $\bar{\pi}$. A small set of parameters can be used in this model to describe the traffic flow. The number of states, the server rate μ, minimum and maximum rates ($\gamma_{min}, \gamma_{max}$) and the shape parameter β are needed to model the flow. Of these parameters, μ can be fixed at xx, and thus only three statistics must be estimated to predict the delay and loss performance of the queue.

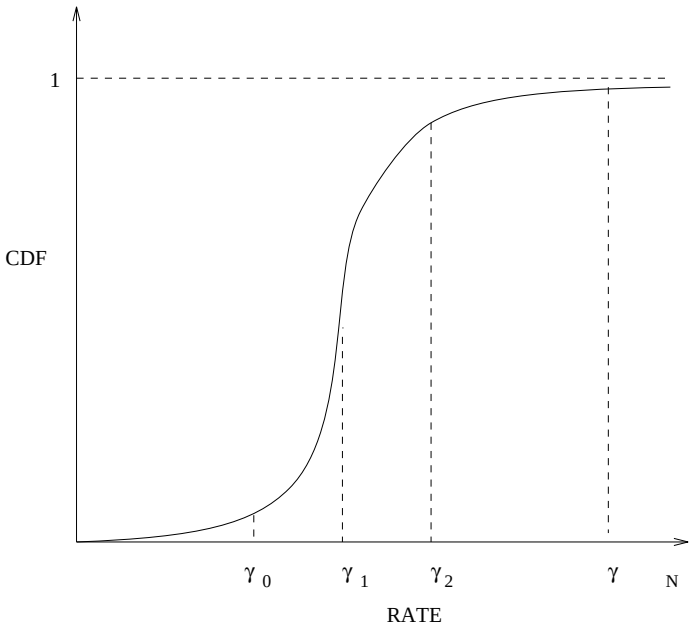

Figure 2.1: Partitioning the Rate Process into states.

Chapter 3

Performance Analysis Methodology

In this chapter a performance analysis technique is developed based on the model derived in the previous chapter. The performance metrics of interest here are mean cell delays and cell loss ratio. Before the performance method is explained, the analytical model developed in Chapter 2 is summarized here as follows: Obtain the CDF of the rate process. This is done in two ways: The CDF can be obtained as a theoretical distribution eg., a Pareto or can be calculated from a given collected data trace and based on the resulting empirical histogram. Quantize the CDF into 'N' levels, where level i is represented by rate γ_{i+1} in the vector $\bar{\gamma}$. Define the i^{th} state x_i, to be represented by levels i and $(i+1)$. The probability of being in the state x_i is given by the element π_i in the vector $\bar{\pi}$ as π_i = F(γ_{i+1})-F(γ_i). The performance evaluation methodology is systematically developed below. Let Z denote a random variable associated with a performance parameter of the queue, having $R(t)$ as the input process. Here Z is delay or loss probability. The objective of the performance analysis technique is to characterize the random variable Z in terms of the properties of $R(t)$. In each of the states the arrival process is assumed to follow a fixed $Cramer - type$ distribution with a given mean arrival rate γ_{i+1} for state x_i i.e.,

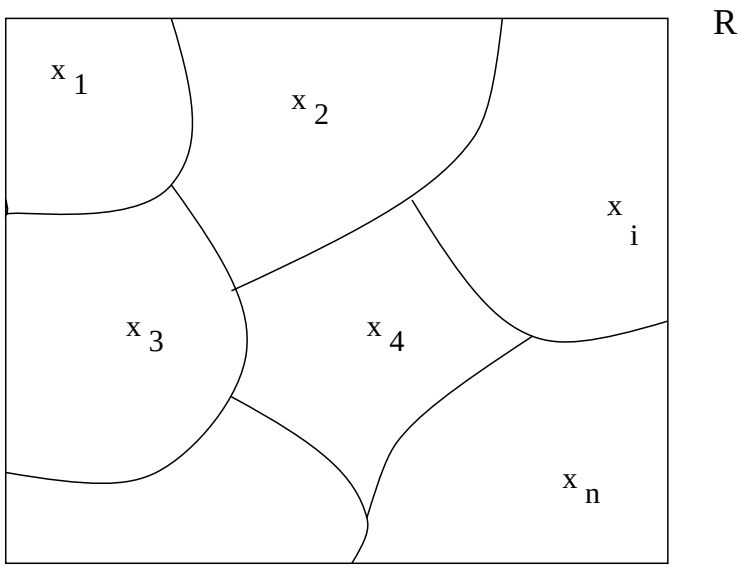

Figure 3.1: Concept for performance analysis methodology.

$$E[R(t)|i^{th}\,state] = \gamma_{i+1} \tag{3.1}$$

Within any given state of the input process say, x_i, the queueing system behaves as a slotted-G/ D/ 1 type of queue with mean arrival rate given by Equation (3.1). Figure 3.1, illustrates the rate process in relation to the disjoint state space of the set **S**=$\{x_1, x_2, ...x_N\}$.

Given the linearity property of expected value, the expected value of the random variable Z can be written as

$$E[Z] = \sum_{i \in S} \pi_i E[Z|S = x_i] \tag{3.2}$$

Note that Equation (3.2) represents the effect of traffic macro-dynamics as well as the traffic-micro dynamics. The traffic macro-dynamics are described by the values of

$\bar{\pi}$ and the micro-dynamics are represented by $E[Z|S = x_i]$. The macro-scale dynamics capture the burstiness in the input trace.

In the remaining part of this chapter, the effect of micro-dynamics will be considered. Consider a queueing system in which the time is slotted. Assume the server to the queue has a deterministic service time so that at the at the end of each slot a single cell in the queue is served. Let n denote the number of cells in the system at a given time and $P_n^i(j)$ represents the probability that there are n cells in the queue at the end of the j^{th} slot, given that the input process is in state x_i. During each slot there are a random number of arrivals to the queueing system. Let p_k^i denote the probability that there are k arrivals to the system when the input process is in state x_i.

In order that the system may have n cells at the end of the $(j+1)^{st}$ slot, there should be k, $0 \leq k \leq n+1$ cells, present at the end of the j^{th} slot and $n-(k-1)^+$ arrivals during the $j+1$ slot, where $(x)^+$ represents the maximum of $\{0, x\}$. As the queue depletes at a constant rate, the relationship between the cell arrivals to the system and the number of cells in the system can be written as a recursive relation. The relationship is commonly referred to in literature as Lindley's recursion [14] and is given in Equation (3.3). Finding the system occupancy involves solving the recursive relation. An efficient solution technique is obtained by noting that a slotted queueing system with a random number of arrivals in each slot falls under the M/ G/ 1 paradigm of Nuets described in [5]. The PGF method, similar to that in [1] is used for solving the equation.

$$P_n^i(j+1) = \sum_{k=0}^{n+1} P_k^i(j) p_{n-(k-1)^+}^i \tag{3.3}$$

3.1 Analysis of Mean Delay

Let Q be the random variable representing the length of a queue with $R(t)|S = x_i, i = 1, 2, ..., N$ as the input process. Suppose Q_i is the random variable associated with the length of the queue given that the input process is in state x_i. In this section the analytical expression for predicting the queue length is obtained. The analysis given in [1] is used here. The well known Little's result is then used for obtaining a delay estimate. Define $G_i(z)$ to be the Probability Generating Function (PGF) of the number of cells in the system, given that the input process is in state x_i.

$$G_i(z) = \sum_{n=0}^{\infty} z^n P_n^i(j) \tag{3.4}$$

Similarly define $F_i(z)$ to be the PGF of the number of arrivals to the system during a slot in which the arrival process is in state x_i . Therefore

$$F_i(z) = \sum_{k=0}^{\infty} z^k p_k^i \tag{3.5}$$

From the above two equations as $j \to \infty$, $G_i(z)$ can be expressed as

$$G_i(z) = (z-1)p_0^i F_i(z)[zI - F_i(z)]^{-1} \tag{3.6}$$

Differentiating equation(3.6) we get:

$$G_i^{(1)}(z)[zI - F_i(z)] + G_i(z)[I - F_i^{(1)}(z)] = p_0^i[F_i(z) + (z-1)F_i^{(1)}(z)]$$

$$(3.7)$$

where $H^{(n)}(z)$ denotes the n^{th} derivative of function $H(z)$ with respect to z

Differentiating equation(3.6) again

$$G_i^{(2)}(z)[zI - F_i(z)] + 2G_i^{(1)}(z)[I - F_i^{(1)}(z)] - G_i(z)F_i^{(2)}(z) = p_0^i[2F_i^{(1)}(z) + F_i^{(2)}(z)(z-1)] \quad (3.8)$$

Taking the limit as $z \to 1$ in Equations (3.7) and (3.8), the expression for $G_i^{(1)}(1)$ can be obtained as

$$G_i^{(1)}(1) = \frac{1}{1 - F_i^{(1)}(1)}\{\frac{1}{2}F_i^{(2)}(1) + (2p_0^i - 1)F_i^{(1)}(1) + [F_i^{(1)}(1)]^2\} \quad (3.9)$$

Note that $G_i^{(1)}(1)$ represents the mean number of cells in the system and Equation (3.9) expresses the system occupancy in terms of known quantities with the exception of p_0^i. Note that p_0^i represents the probability that the system is empty and is therefore equal to the utilization of the system at the given load subtracted from unity. Also note that since $F_i(z)$ is the PGF of the arrival process within a given slot, $F_i^{(1)}(1)$ is the mean of the arrival process in cells/slot and $F_i^{(2)}(1)$ can be expressed in terms of the mean and variance of the arrival process as given in Equation (3.10). In order to ensure numerical stability for an infinite buffer queue, $F_i^{(1)}(1)$, $i \in S$, is assumed to be less than unity.

$$F_i^{(2)}(1) = \sigma_i^2 + \lambda_i^2 - \lambda_i \quad (3.10)$$

where

σ_i is the variance of the arrival process, given that the arrival process is in state x_i.

λ_i is the mean of the arrival process in cells/slot, in state x_i.

The result of the analysis is stated as below: For a slotted-G/ D/ 1 queue the mean number of cells in the queue can be described in terms of the arrival process as

$$E[Q_i] = G_i^{(1)}(1) = \frac{1}{1 - F_i^{(1)}(1)}\{\frac{1}{2}F_i^{(2)}(1) + (2p_0^i - 1)F_i^{(1)}(1) + [F_i^{(1)}(1)]^2\} \quad (3.11)$$

Applying Little's result and simplifying using Equation (3.10), final result of the performance analysis is stated as:

$$E[D] = \frac{1}{\lambda}\sum_{i=1}^{N}\frac{\pi_i}{1 - \rho_i}\{\frac{1}{2}\sigma_i^2 + \frac{3}{2}\rho_i^2\} \quad (3.12)$$

where

λ=$\sum_{i=1}^{N} \pi_i\gamma_{i+1}$ is the mean arrival rate.

D is the random variable representing the delay experienced by cells arriving at the input of the queueing system.

ρ_i, is the probability that the system is occupied i.e., the utilization given $S = x_i$

3.2 Analysis of Cell Loss

In this section analytical expression for the estimating the cell loss ratio is reviewed. This analysis is based on [3]. Unlike the case for estimating the mean cell delays, the expressions for the calculation of cell loss in a deterministic server queueing system are considerably more complex. The cell loss ratio in a finite buffer queueing system is approximated by the tail of the queue length distribution of the infinite buffer counterpart. Here we assume that the number of arrivals to the queueing system in

a slot follows a Poisson process. That is, for estimating cell loss we assume that the micro-dynamics are Markovian. This theoretically limits loss prediction methodology to Poisson micro-dynamics. The practical implications of this assumption will be addressed emperically in Chapter 6. That is, for estimating cell loss we assume that the micro-dynamics are Markovian. Exact expressions for cell loss in terms of the ergodic occupancy distribution of an M/ D/ 1 system are now presented. The cell loss estimate for the arrival process is then obtained as sum of loss estimates obtained from individual states weighted by the probability of being in that state.

The expressions for the M/ D/ 1 queue are obtained as a special case of the M/ G/ 1 analysis by using a recursive numerical technique due to Kielson and Servi and given in [3]. Consistent with the notation used in the previous section, let P_k^i denote the probability that there are k cells in the queue given that the input process is in state x_i i.e., P_k^i = $P[Q = k|S = x_i]$. Here the cell loss probability estimate given that the input process is in state x_i, is approximated as the probability $P[Q > K|S = x_i] \approx P_L[K|S = x_i]$, for a given system size of 'K'. Let the utilization of the system given that the input process is in state x_i, be ρ_i. The utilization of a queueing system is defined as the ratio of the mean arrival rate to the queueing system to the service rate. For stability we require $0 \leq \rho_i < 1 \forall i$. Note that $P_0^i = 1 - \rho_i$ given in the previous section. P_k^i, $P_L[K|S = x_i]$ are given below as :

$$P_k^i = \frac{(1-\rho_i)p_k^i + \sum_{n=1}^{k}\{1 - \sum_{m=0}^{n} p_m^i\}P_{k-n}^i}{p_0^i} \quad k = 1, 2, 3, ...K \tag{3.13}$$

$$P_L[k|S = x_i] \approx P[Q > K|S = x_i] = 1 - \sum_{k=0}^{K} P_k^j \tag{3.14}$$

In general the probabilities $p_{j,n}$ are state dependent and can be obtained from the relation :

$$F_{x_i}{}^*(\rho_i[1-z]) = \sum_{j=0}^{\infty} z^j p_n^i \tag{3.15}$$

where

In Equation (3.15), $F_{x_i}{}^*(z)$ denotes the Laplace-Stieltjes transform of the random variable x_i. The Laplace-Stieltjes transform can be defined below:

Definition 1 Let X be a non-negative random variable with distribution $F_X(x)$. Then

$$F_X{}^*(s) = \int_0^{\infty} e^{-sx} dF_X(x)$$

is called the Laplace-Stieltjes transform of the random variable X.

For the case of a deterministic service Equation (3.15) evaluates to

$$e^{-\rho_i[1-z]} = \sum_{j=0}^{\infty} z^j p_n^i \tag{3.16}$$

Upon matching the coefficients of z^j in Equation (3.16), p_n^i evaluates to

$$p_n^i = \frac{e^{-\rho_i} {\rho_i}^n}{n!} \tag{3.17}$$

The final result for obtaining the cell loss probability estimate can be stated as

$$P_L(K) = \sum_{i=1}^{N} \pi_i (1 - \sum_{j=0}^{K} P_j^i) \tag{3.18}$$

$$= \sum_{i=1}^{N} \pi_i (1 - \sum_{k=0}^{K} \frac{(1-\rho_i)p_k^i + \sum_{n=1}^{i} \{1 - \sum_{m=0}^{n} p_m^i\} P_{k-n}^i}{p_0^i}) \tag{3.19}$$

It is should be noted that, using the performance analysis technique described above produces a non-linear relation between cell loss as the function of the buffer size, which is typical in the case of long range dependent traffic. Similarly, this analysis produces delay curves typical characteristics of delay curves that can be expected from long range dependent network traffic.

Chapter 4

Data Collection and the AAI

4.1 Configuration of the AAI Network

The performance analysis in this research is based on the trace data collected from the ACTS ATM Internetwork (AAI) network [21]. This chapter describes the method used for collecting the traces, the configuration of the AAI wide area ATM network and information on the data collection process. As a part of this research the AAI network was extensively monitored by the University of Kansas. Data totaling to about 1.8 Gbytes has been collected from the AAI network in the interval starting from January '96 to July '97. A number of ports on the ATM switches at various locations in the AAI network were continuously monitored. Data from the selected ports on the switches was collected using periodic Simple Network Management Protocol (SNMP) requests. SNMP is a standard protocol used to monitor switches and the networks to which they attach [35]. The data traces consist of switch buffer cell counts collected with a sampling period of approximately 60 seconds. The collected data traces are resampled so that the granularity between successive cell counts is exactly equal to 60 seconds. The tools and methods used for collecting the measurements and obtaining the plots are

Table 4.1: Site involved in the collection process

Site	Location
Naval Research Lab (NRL)	Washington, D.C.
Army Research Lab (ARL)	APG, Maryland
Naval Command Control and Ocean Surveillance Center (NCCOSC)	San Diego, California
Corps Of Engineers Waterways Experiments Station (CEWES)	Vicksburg, Mississippi
Naval Research Lab, Stennis Space Center (NRLSSC)	Stennis Space Center, Mississippi
Government Systems Division (GSD)	Kansas City, Missouri
EROS data center (EDC)	Sioux Falls, South Dakota
Sprint Technology Integration And Operations Center (TIOC)	Overland Park, Kansas
University of Kansas (KU)	Lawrence, Kansas

discussed in this chapter.

Starting from January 96 specific AAI edge switches and MAGIC switches were monitored continuously. The sites involved in the AAI network are given in Table 4.1. The VPI assignments between the sites are fixed. The configuration of the AAI network is shown in Figure 4.1.

At each of the above mentioned sites at least one port on the selected switch was being sampled. All the data was collected from the edge switches which are FORE switches. Switches inside the AAI cloud are not monitored as a part of this study.

4.2 Data Collection

In this section the methods and tools which have been used for obtaining the measurements are discussed. Data was collected in the form of switch buffer cell counts, using SNMP. The method adopted relies on querying the SNMP agent that is main-

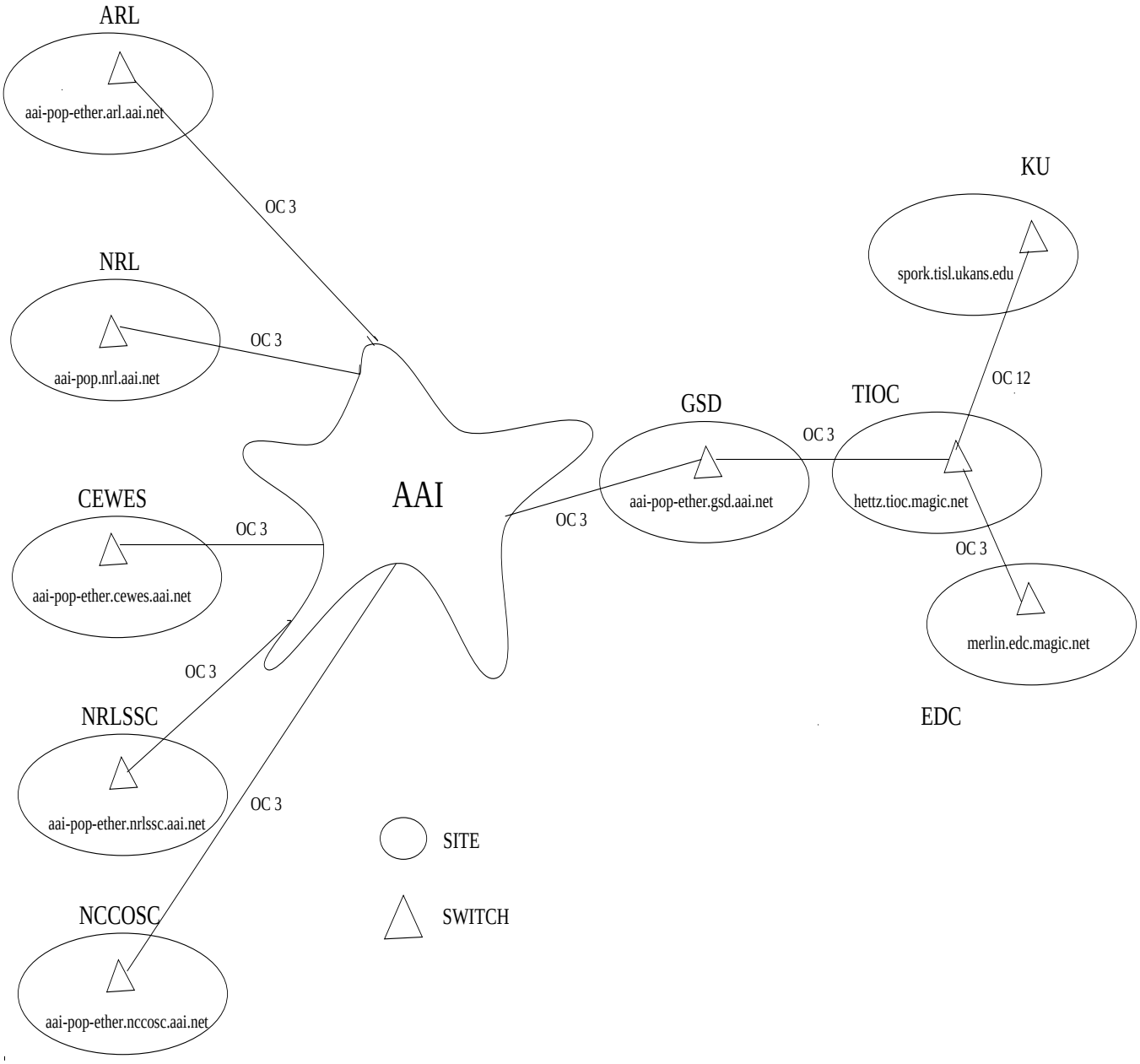

Figure 4.1: Connections of the sites and switches being sampled.

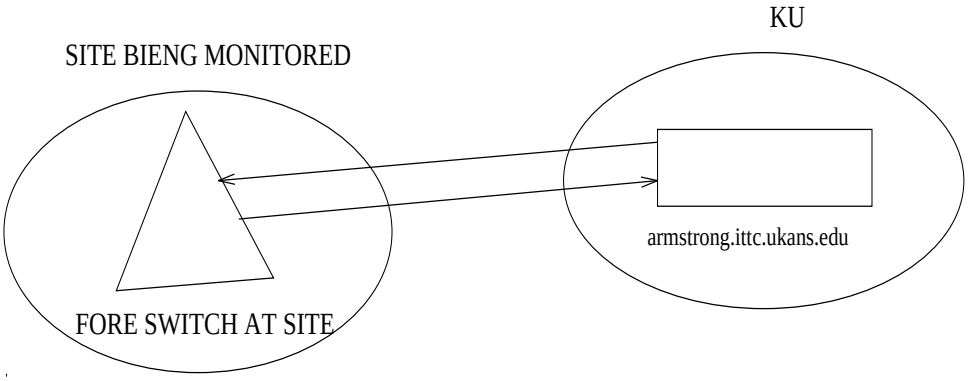

Figure 4.2: SNMP requests and responses.

tained on the FORE switches in the network. The SNMP agents are maintained by the SNMP daemon running on each switch. Agents support a Management Information Base (MIB) containing information about the operation of each switch. Among other things the MIB supports counters for the number of cells transferring over a port, a particular virtual path, or a particular virtual channel. Traffic flowing out of a port cannot be monitored on a virtual channel basis as the FORE MIB does not support the MIB element. The information as needed can be obtained by querying the appropriate MIB variable. SNMP requests are sent to each AAI switch approximately every 60 seconds from a KU SPARCcenter 2000 computer named $armstrong.ittc.ukans.edu$. The requested cell count is returned as response to the request. The cell count is time-stamped when it is successfully received.

The Tricklet library [36] has been used for making the SNMP requests. The Tricklet library offers a convenient interface for obtaining network management information using SNMP. In particular the SNMP-GET, SNMP-GET NEXT, SNMP-TBL and SNMP-SET requests are implemented in the Tricklet library. The SNMP-GET requests the value of the specified MIB variable. SNMP-GET NEXT retrieves the next MIB variable in the hierarchy to the specified variable. SNMP-TBL is implemented as a series of GET and GET NEXT requests. It returns a group of MIB variables which

form a subtree with the requested variable as the root. The cell counts obtained from the responses to the SNMP requests and their corresponding time-stamps are continuously archived. Successive time-stamps in the archived data differ by approximately 60 seconds, as the SNMP requests are sent out at approximately 60 second intervals with a variation of a few seconds. The data collected is resampled so that the time between successive cell counts is equal to 60 seconds. The re-sampling is done through linear interpolation of raw cell counts. The process of interpolation is shown in Figure 4.3. The linear interpolation technique used for resampling the raw cell counts can be expressed as Equation 4.1. In Equation 4.1, for any given value of k, t_1 represents the largest time in the trace that is less than $k\Delta t$ and t_2 represents the smallest time in the trace that is greater that $k\Delta t$.

$$c(k\Delta t) = c(t_1) + \frac{c(t_2) - c(t_1)}{t_2 - t_1} * (k\Delta t - t_1), \quad k = 1, 2, 3, ... \tag{4.1}$$

where

$c(k\Delta t)$ is the cell count obtained by interpolation.

Δt is the time interval after resampling and is 60 seconds.

$c(t_2)$ and $c(t_1)$ are cell counts from collected data which are sampled at approximately 60 seconds.

From the resampled data the throughput for each sampling period is calculated as the time average of the number of cells transferred during this period. The length of the sampling period is equal to 60 seconds. The random process $R(t)$, representing the rate process can be expressed as:

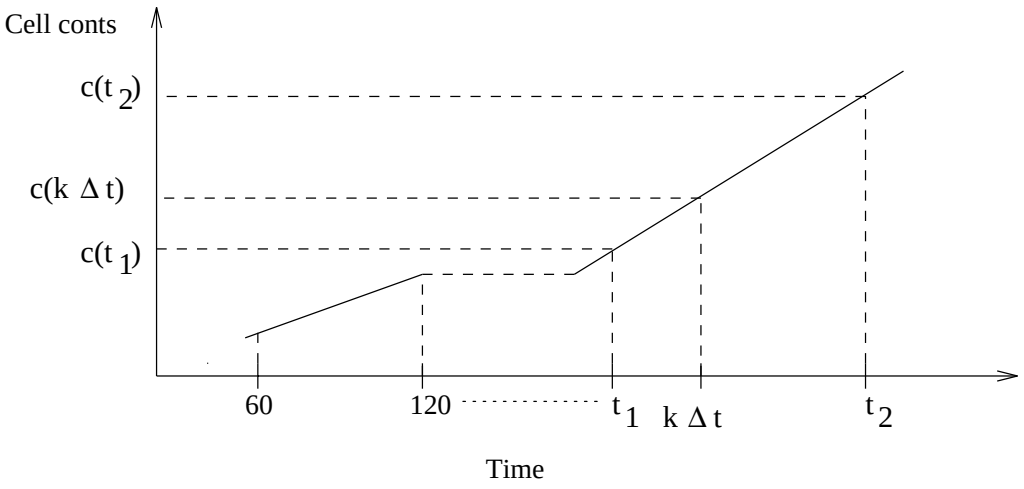

Figure 4.3: Re-sampling by linear interpolation of cell counts.

$$R(t) = \frac{[A(t+\Delta t) - A(t)] * 424}{\Delta t} b/s, \tag{4.2}$$

where,

$A(t)$ is the cell count at time t.

$R(t)$ is the throughput at time t.

In the above equation Δt is 60 seconds as the actual values of Δt are obtained from the resampled cell counts. As explained in Chapter 3, the model developed is based on uniformly quantizing the range of the random process $R(t)$ into 'N' levels. Note that the range of the random process, $R(t)$ is the range of rates present in the given trace file and is thus bounded by γ_{min} and γ_{max}.

In order to observe events at various parts of the network, a tool was developed to summarize each day's of data collected in the form of thumb-nail gifs plotted on a per site basis, for every port being monitored. Information related to the status of the ARP server for AAI and the connectivity status of the various AAI hosts is also monitored.

The tool developed can be used for viewing and analyzing the traffic profiles generated by experiments being done on the network.

4.3 Observations

As a part of the AAI project, traffic flows at various locations in the AAI network were extensively monitored. In this section some initial observations about the flow patterns observed in the AAI network are presented. Observations from the traffic flows are important as they indicate the types of traffic that are prominent in a network focus the problem of modeling to include these traffic types. As expected of traffic in a broadband network, the AAI traffic flow patterns show significant burstiness. Apart from this, EMMI traffic for a significant amount of time has been observed in the traffic flows in the network. EMMI is the Efficient Multi Media Interface for transporting video, voice and data over ATM networks. The EMMI device is attached externally to a workstation which hosts the driver and application software. The hardware of the device is workstation-independent. The EMMI device provides user-selectable quantization factors for compressing video data. The value of the quantization factor effects the quality of the video image that is transmitted and determines the bandwidth required for the image transmission. The bandwidth for EMMI traffic varies from 1.85 Mb/s - 85 Mb/s depending on the appliation.

Apart from EMMI traffic, observations from the traffic profiles indicate the presence of a significant amount of background traffic. Typically the background flows have been observed to have a peak magnitude of 1-2 Mb/s. There are many existing techniques in literature to model specific kinds of application traffic. However, there are no models for modeling background traffic. Therefore this study also aims at modeling the background traffic and study its queueing performance. One of the factors

contributing to the significant amount of background traffic is the complexity involved in the operation of an ATM WAN. The routing protocols in ATM for example, can be expected to generate some amount of background traffic as the switches exchange routing tables and other update messages periodically.

Chapter 5

Simulation Model

5.1 Simulation Model Description

In this chapter the simulation model that was developed is described and validated. The simulator is used for obtaining estimates of the mean cell transfer delay and the cell loss ratio of the queueing system. The basic queueing model that is simulated is an infinite buffer queue with a deterministic server. The model is implemented by keeping track of the number of cells in the queue at all times. By knowing the capacity of the server, along with the number of cells in the system, the delay incurred in the queue for an arriving cell can be estimated. In addition an estimate for the cell loss probability for a fixed buffer size is obtained by approximating it with the tail of the queue length distribution. Given a buffer size 'N', the cells loss probability is estimated as the probability $P(Q > N)$. As mentioned in Section 1.1, the value thus obtained is a conservative estimate for the cell loss ratio. The assumption of a deterministic server is consistent with fixed cell size in ATM networks. In the simulation model the slot size is fixed as the reciprocal of the server capacity expressed in $cells/sec$. Therefore at the end of each slot a single cell is served if the queue is non-empty. Within each slot cells

arrive to the queueing system with inter-arrival times that follow a known statistical distribution having a finite mean and variance. In the case of trace driven simulations the number of cells that arrive to the queueing system are obtained from a collected data trace.

Since the queue has a deterministic server, the dynamics of the queueing system defined by the the number of cells in the queue at any instant of time can given by :

$$n(k+1) = max(0, n(k) + a(k) - 1). \tag{5.1}$$

where

$n(k)$ is the number of cells in the system at the end of the k^{th} and

$a(k)$ is the number of cells that arrived during the k^{th} slot.

$max(a, b)$ represents the minimum of the quantities a and b.

The different quantities in Equation (5.1) which are used for obtaining cell delay estimates from cell counts are shown in Figure 5.1. In the simulations the number of cells in the queue is sampled at the end of every time slot.

Within each sampling interval of T_s seconds the arrival process is assumed to follow a point process. The number of arrivals with in a cell slot, given by $n(k)$, is obtained from the assumed point process for the cell arrivals. The specific point process chosen in the simulation model, simulates the traffic microdynamics component of $R(t)$ as described in the traffic model.

The deterministic rate of the server is varied in order to get the delay values for different values of load on the link. After the number of cells in the queue is determined using Equation (5.1) the cell delay is obtained as:

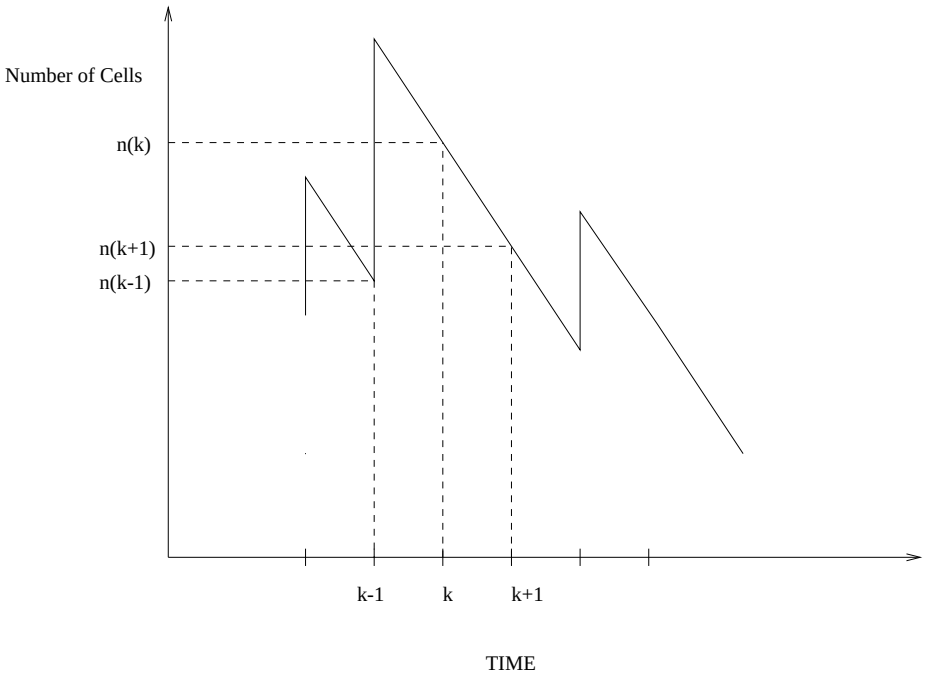

Figure 5.1: Delay estimation from cell counts

$$\hat{\tau}(k) = \frac{n(k)}{\mu} \tag{5.2}$$

where

μ is the rate of the deterministic rate of the server in cells/sec.

$\hat{\tau}(k)$ is the delay estimate for the cell at the end of the k^{th} slot.

As can be seen from Equation (5.2), the delay estimate $\hat{\tau}(k)$ is a function of slot number. Therefore the average cell delay is for the whole trace is calculated using the relation:

$$\bar{\tau} = \frac{1}{M} \sum_{k=1}^{M} \hat{\tau(k)} \tag{5.3}$$

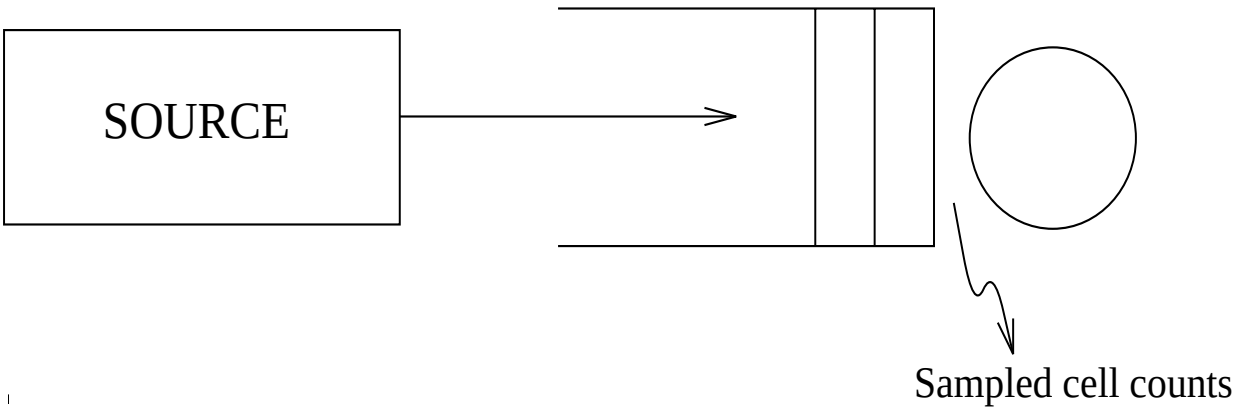

Figure 5.2: Validation of Model

where

$\bar{\tau}$ is mean cell delay over the whole trace and

M is the total number of slots for the given trace.

The cell loss probability for a buffer size x denoted as $P_L(x)$, is obtained by counting the relative number of times, the queue length $n(k)$ exceeds the value x.

$$P_L(x) = P[n(k) > x] \tag{5.4}$$

5.2 Validation of Model

The simulation model is validated by performing simulations with standard sources described in terms of known parameters and comparing the simulation results with those predicted by standard theoretical results.

For validating the simulation model a source which generates cells with exponentially distributed inter-arrival times with a known average rate has been constructed. A data trace is obtained by sampling the output of the known source at constant time intervals of T_S seconds. Exponential inter-arrival times for the cells in the constructed source are generated based on the Transform Method described in [16]. The Transform

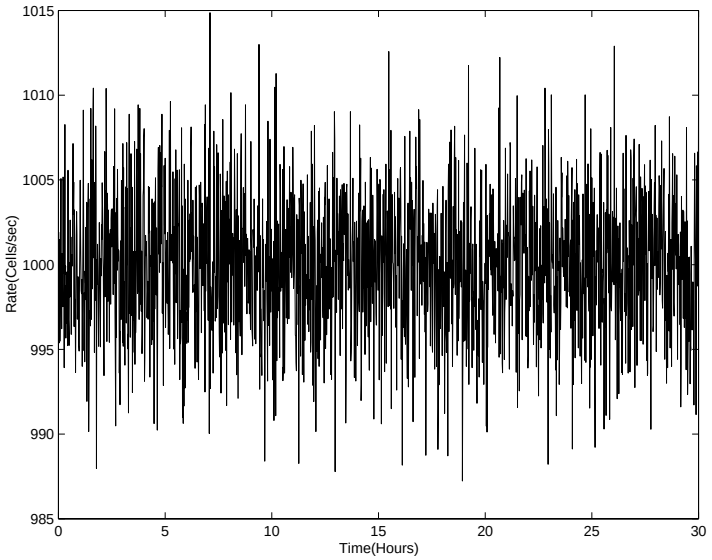

Figure 5.3: Generated trace with exponentially distributed inter-arrival times.

Method is based on the principle that a random variable with any arbitrary probability density function (pdf) can be generated, by applying a simple transformation to a Uniform Random variable. The Transform Method utilizes the fact that the Cumulative Distribution Function of an arbitrary random variable has a Uniform Distribution in the interval [0,1]. Refer [16] for the detailed derivation.

The synthesized trace is used in the algorithm for the cell-level simulation described in Section 5.1. The process of obtaining the trace which is used for validation of the simulation model is shown in Figure 5.2. The method of obtaining the synthesized trace by sampling the output of the source at T_S seconds is analogous to the method by which traces are collected from the network by sending periodic SNMP requests. Figure 5.3 shows the synthesized trace obtained by sampling the source at T_S = 60 seconds. The trace consists of inter-arrival times that are exponentially distributed with a mean of 1 ms.

The values of the estimates of the mean-cell delay and cell loss ratio obtained from the simulation results are compared to the theoretical values. The theoretical values of

the mean queueing delay are obtained from the well known Pollaczek-Kinchin relation and Little's result [3] as applied to an M/D/1 system which is given as

$$D = (\frac{1}{\lambda})[\rho + \frac{\rho^2}{2(1-\rho)}] \tag{5.5}$$

where

λ is the average arrival rate to the queue and

ρ is the input load to the system.

D is the mean cell delay.

The validation results for the mean cell delays are plotted in Figure 5.4. From Figure 5.4 we see that the results obtained from the simulator match the theoretical values very well even at high values of load. Considering the *statistical nature* of the simulation, the comparison shows that the developed simulator and the simulation model work correctly.

The simulator is now validated in terms of the cell loss ratio estimate which is approximated by the tail of the queue length distribution. Unlike the case for estimating the mean cell delays, the expressions for the calculation of cell loss in a deterministic server queueing system are considerably more complex. For this reason, the loss calculations for an M/ D/ 1 system are often approximated with those of an M/ M/ 1 queueing system, which requires less numerical calculation. Exact expressions for cell loss in terms of the ergodic occupancy distribution of an M/ D/ 1 system are presented as a special case of an M/ G/ 1 system as explained in Chapter 3.

Based on the M/ D/ 1/ K model, the theoretical values of the logarithm of $P[Q > x]$ is plotted against the buffer size in Figure 5.5. In the same figure simulation results obtained from the sampled simulated trace are also shown. The analytical equations

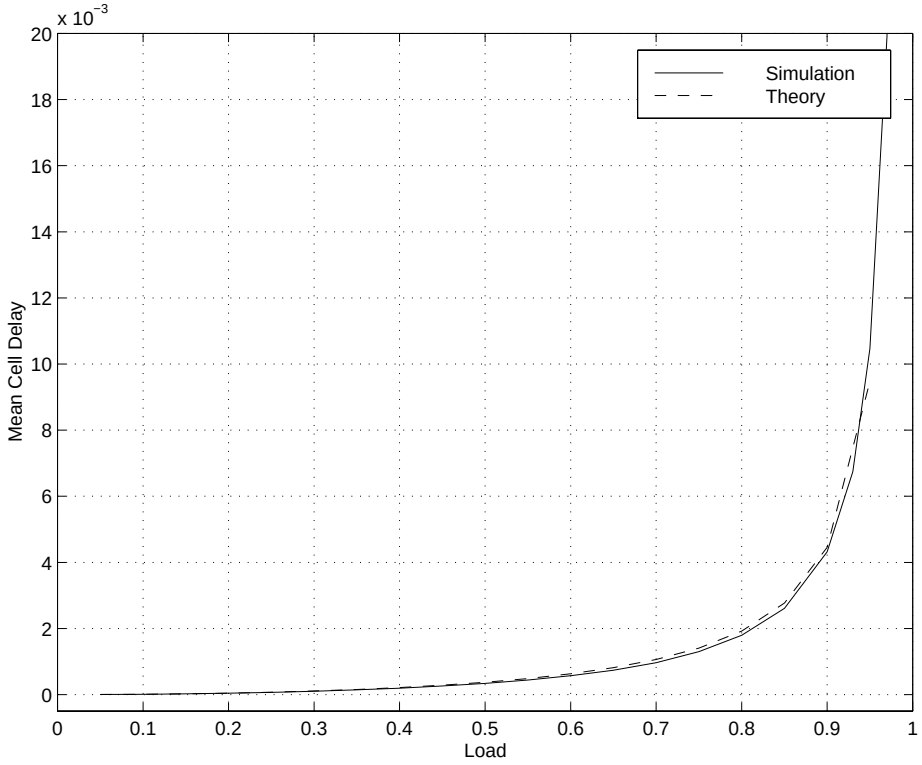

Figure 5.4: Validation of the simulator for mean cell transfer delay results.

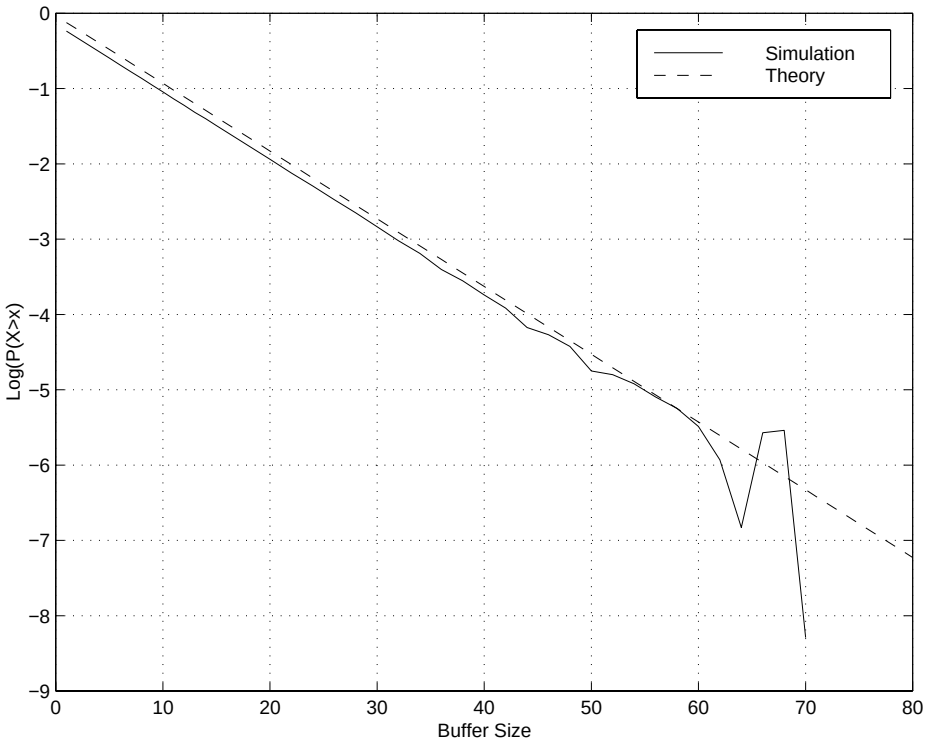

Figure 5.5: Validation of the simulator for Cell loss ratio results.

used for obtaining the theoretical values are given in Section 3.2. From Figure 5.5 it can be seen that the estimates of cell loss ratio obtained from trace based simulation results matches the theoretical values up to a value of 10^{-6}. Note that cell loss ratios less than 10^{-6} cannot be estimated accurately as the variance of the estimator is high in the loss ranges greater than 10^{-6} because of the limited number of cells in the trace.

Chapter 6

Experimental Validation of the proposed Analysis Performance Methodology

6.1 Traffic Traces Used in Model Evaluation

In this chapter the model and the performance evaluation technique described in Chapters 2 and 3 respectively are evaluated experimentally. The evaluation is done by comparing the performance predicted by the model, with the corresponding performance obtained from simulation based on collected network traces. Several network traces collected from the AAI network are employed here. The traces are collected based on the method described in Chapter 4. The traces used for the experimental evaluation in this chapter are listed below in Table 6.1.

The traces listed in Table 6.1 were collected with a sampling interval of approximately 60 secs and were resampled so that the sampling interval is exactly equal to

Table 6.1: Data traces used for model validation

Trace Name	Duration (Hours)	# cells	Characteristic
TIOC	48	2.4896e08	Background
NCCOSC	25	3.02697e08	Background
NRL1	6	5.1054e08	EMMI
Phillips	15	2.227e09	SC '95
GSD	6	2.23624e08	SC '95
KU	3	9.23146e08	EMMI
NRL2	12	1.8645e08	EMMI

60secs. The traces are chosen such that traffic with different rate levels ranging from low (approximately 1 Mb/s peak rates) to high (Approximately 70 Mb/s peak rates) is taken into consideration for experimental evaluation of the model. The traces also represent traffic with different traffic types. The traffic type represents the main characteristic of the flow in the trace. The specific traffic flow types used here are :

1. Background: Significant complexity is involved in the operation of a national scale ATM network. Background traffic represents management flows, routing message updates etc., can be expected to be generated in operational ATM WANS. The traces labeled 'TIOC' and 'NCCOSC' shown in Figures 6.1 and 6.4 respectively represents typical background traffic flow.

2. EMMI: The EMMI traffic flows are the multimedia traffic profiles obtained from the EMMI system described in Chapter 4. Multimedia application profiles in future networks will be similar to the EMMI flows used here. The traces labeled 'NRL1', 'KU' and 'NRL2', shown in Figures 6.7, 6.16 and 6.19 respectively are the EMMI flows. As mentioned above, the traffic type of the trace represents the main characteristic of the traffic in the trace. However as aggregate traffic from a port is being modeled, note that other types of traffic can also present along with the main characteristic flow. For example, along with the EMMI traffic the

'NRL1' and 'KU' traces contain bursts of short duration, that are typically FTP flows from the collected port.

3. SC '95 flows: The SC '95 flows are the traffic flows collected by KU during Supercomputer'95 (SC'95) in December 1995. The SC '95 flows represent bursty application traffic profiles typical in wide area ATM networks. The traces labeled 'Phillips' and 'GSD', shown in Figures 6.10 and 6.13 respectively are the SC '95 flows. Note that in December 1995 the access link to the Phillips Labs was a DS-3 link. The Phillips data trace was collected with a sampling interval of approximately 10 secs and was resampled so that the sampling period is exactly 60 seconds which is same as the sampling interval in other traces used here.

For each of the traces shown in Table 6.1 performance is evaluated in terms of mean cell transfer delay and the probability of cell loss through a switch output queue. The results for the mean delay are given here as delay vs load curves while the cell loss probability estimates are given as the logarithm (to base 10) of the probability of the queue length exceeding a given finite buffer size plotted against the buffer size in cells.

For each of the traces in Table 6.1, the performance results (mean delay and probability of cell loss) are obtained from:

1. The model, with the CDF of the input process obtained from a theoretical infinite variance distribution. A Pareto distribution with the shape parameter estimated from the traffic flow is used here.

2. The model, with the CDF of the input process obtained from the histogram of the collected network trace data.

3. Direct simulation based on the collected network trace data.

In the figures showing the results of the experimental evaluation, the performance prediction values obtained from (1) are labeled as 'Model' and the corresponding results obtained from (2) and (3) are labeled as 'Histogram' and 'Simulation' respectively.

6.2 Results

In this section the results of experimental evaluation are discussed. Note that in order to demonstrate the usefulness of the non-Markovian model and the performance evaluation technique developed in this study, traces with varying amount of burstiness, for different periods of time were considered in this study. The traffic profiles that are considered in this chapter for experimental evaluation of the model can be typically expected in typical wide area ATM networks.

The results for average cell delay are shown in Figures 6.2, 6.5, 6.8, 6.14, 6.11, 6.17, 6.20. The load was varied by changing the cell service time in the simulation model. Observe that mean delay values of up to .5 seconds have been shown in the delay curves, to demonstrate that for high delay values, the analytical and simulation results match reasonably, considering the statistical nature of the simulation. The knee of the delay curve is important from a traffic engineering view point. The knee of the delay curve is the region of the delay profile where the delay values start increasing sharply with an increase in the load. The knee of the delay curve therefore represents a trade off between delay and load. In the next section, the knee of the delay profiles are discussed. Also note from Figures 6.2, 6.5, 6.8, 6.14, 6.11, 6.17, 6.20 that unlike the case of a short range dependent model, the model and the perform analysis method that are developed here, captures the burstiness in the input trace and predict delay profiles that are similar to those obtained from real network trace data obtained from a high speed wide area ATM network.

The results for the cell loss probability estimates are shown in Figures 6.3, 6.6, 6.9, 6.12, 6.15, 6.18, 6.21. Each of the figures shows the cell loss profiles from the model and simulation based on collected traces. From Figures 6.3, 6.6, 6.9, 6.12, 6.15, 6.18, 6.21 it can be seen that the cell loss probability estimates of unto 10^{-6} are accurately modeled. Also from the Figures 6.3, 6.6, 6.9, 6.12, 6.15, 6.18, 6.21 note that the model captures the non-linear dependence between logarithm of cell loss and buffer size. Cell loss probabilities less that 10^{-6} cannot be accurately obtained because of the limited number of cells in the trace, because of which the variance of the estimator increases. Note that since the simulations performed are at the cell level, it takes very long time to obtain cell loss probability estimates less that 10^{-6}.

All the figures in this section were obtained by assuming that the traffic micro-dynamics in a state follow an exponential distribution. In the next section the effects of traffic micro-dynamics on the mean cell delay and cell loss ratio estimate are investigated. In the next section the effect of load on the cell loss in a queueing system is also investigated.

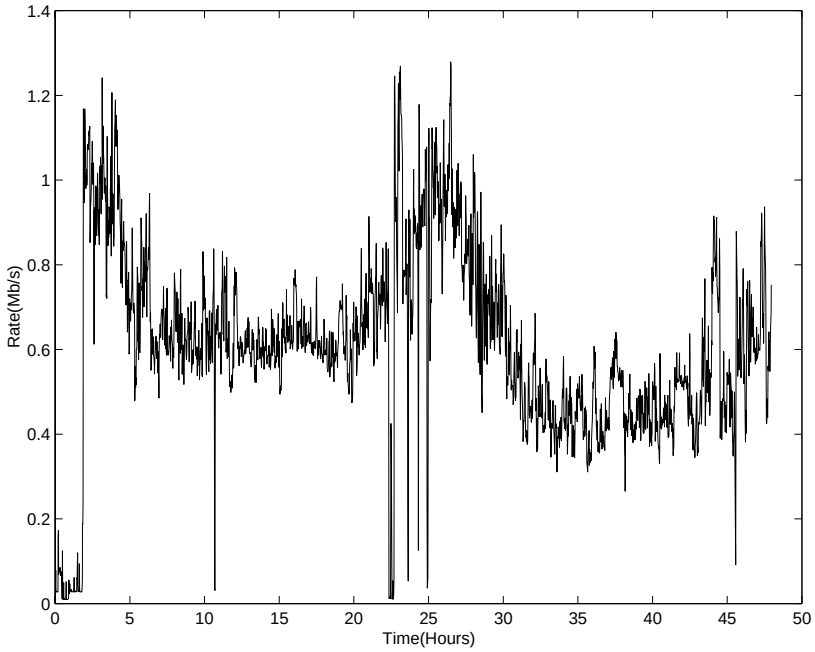

Figure 6.1: Data trace Collected from the TIOC site.

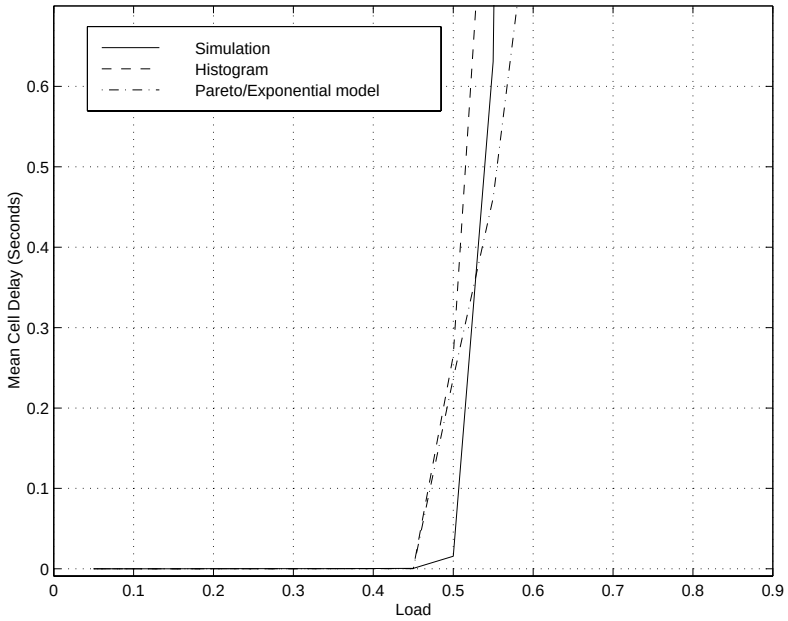

Figure 6.2: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 5.6$), histogram and simulation of the collected data trace labeled 'TIOC' and shown in Figure 6.1 using $N = 15$ input phases.

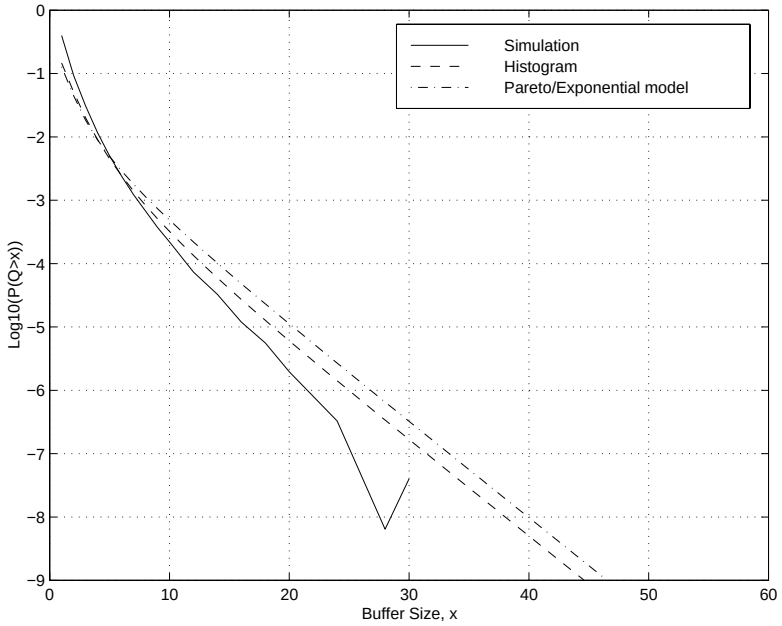

Figure 6.3: Comparison of cell loss probability estimates obtained from theory ($\beta = 5.6$), histogram and simulation of the collected data trace labeled 'TIOC' shown and in Figure 6.1 using $N = 15$ input phases. $\rho = .4$.

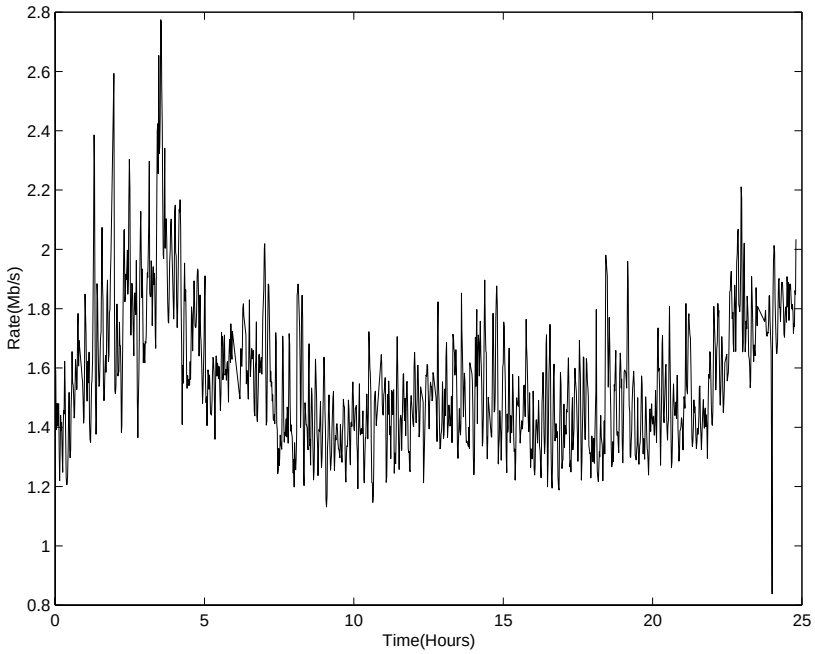

Figure 6.4: Data trace Collected from the NCCOSC site.

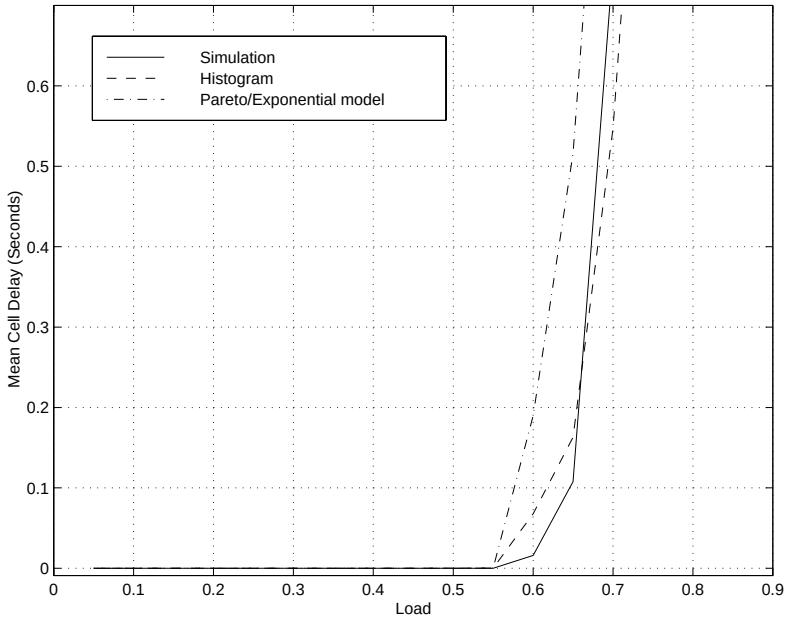

Figure 6.5: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 8.2$), histogram and simulation of the collected data trace labeled 'NCCOSC' and shown in Figure 6.4 using $N = 15$ input phases.

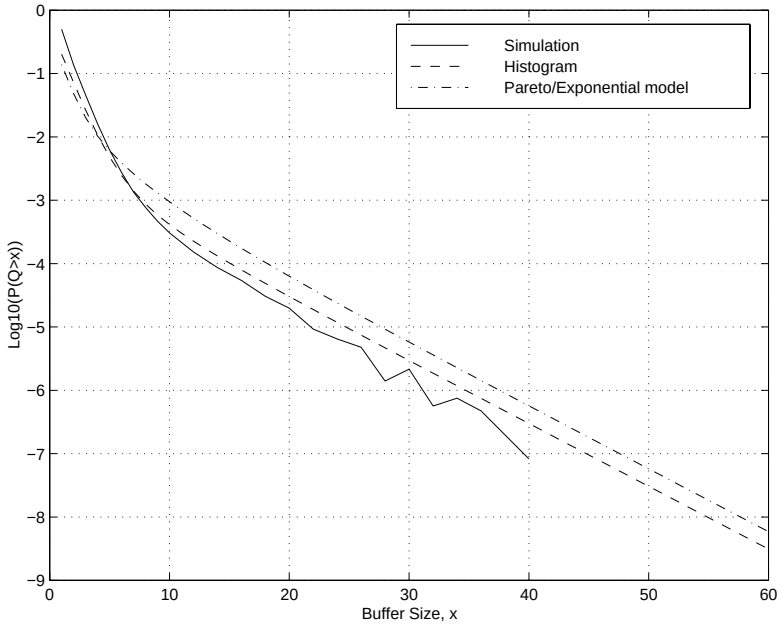

Figure 6.6: Comparison of cell loss probability estimates obtained from theory ($\beta = 8.2$), histogram and simulation of the collected data trace labeled 'NCCOSC' shown in Figure 6.4 using $N = 15$ input phases. $\rho = .4$.

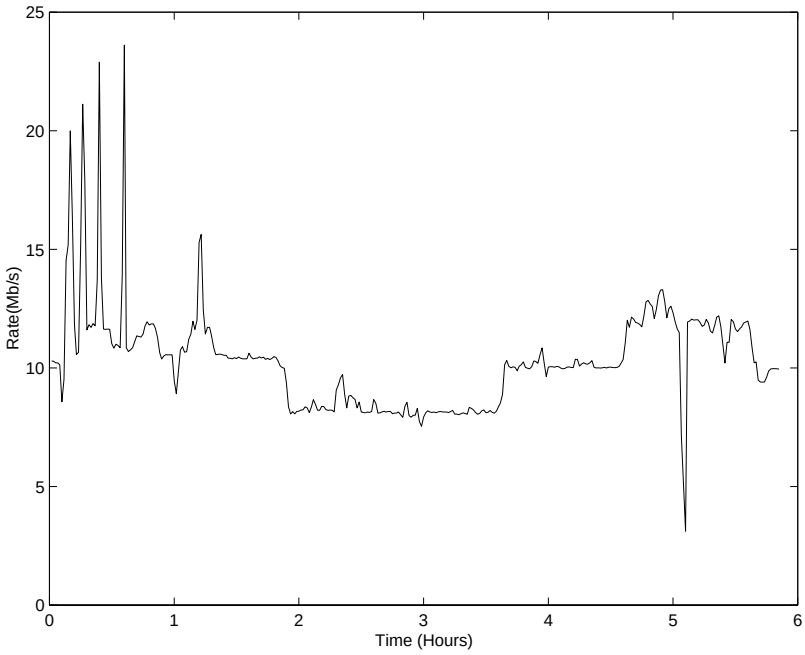

Figure 6.7: Data trace Collected from the NRL site.

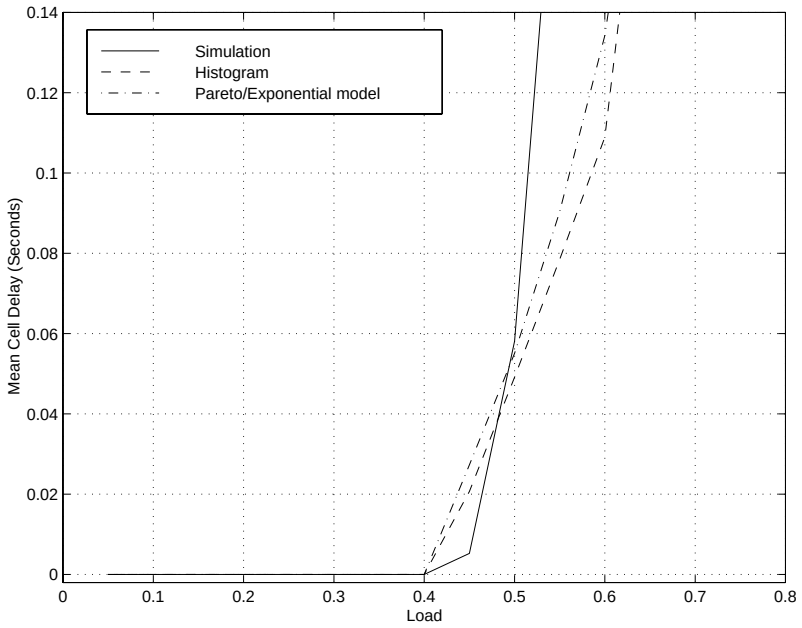

Figure 6.8: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 2.2$), histogram and simulation of the collected data trace labeled 'NRL1' and shown in Figure 6.7 using $N = 15$ input phases.

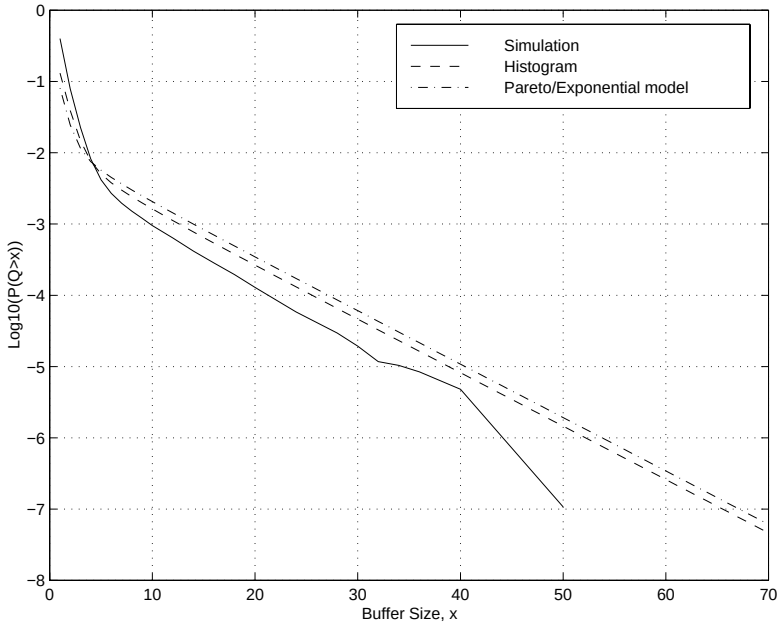

Figure 6.9: Comparison of cell loss probability estimates obtained from theory ($\beta = 2.2$), histogram and simulation of the collected data trace labeled 'NRL1' and shown in Figure 6.7 using $N = 15$ input phases. $\rho = .4$.

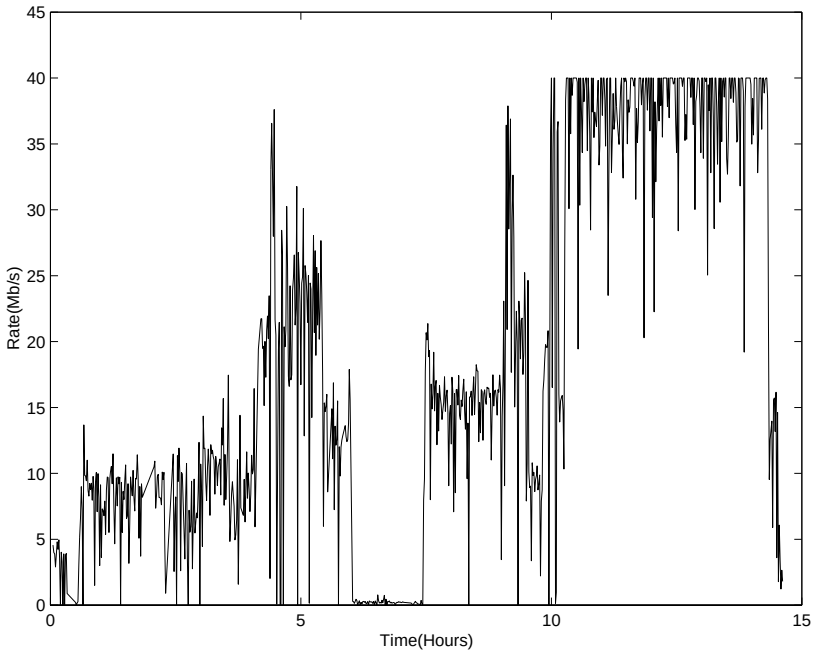

Figure 6.10: Data trace Collected from the Phillips site.

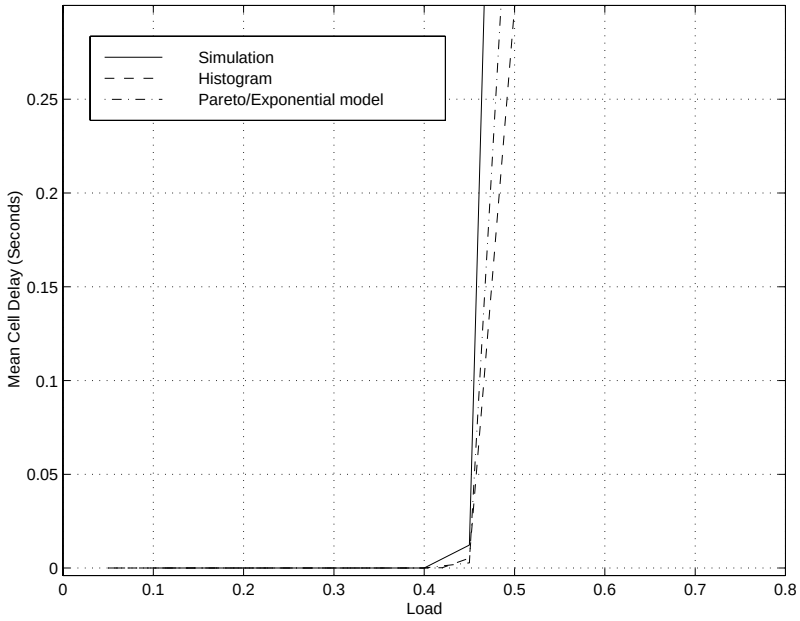

Figure 6.11: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 1.2$), histogram and simulation of the collected data trace labeled 'Phillips' and shown in Figure 6.10 using $N = 15$ input phases.

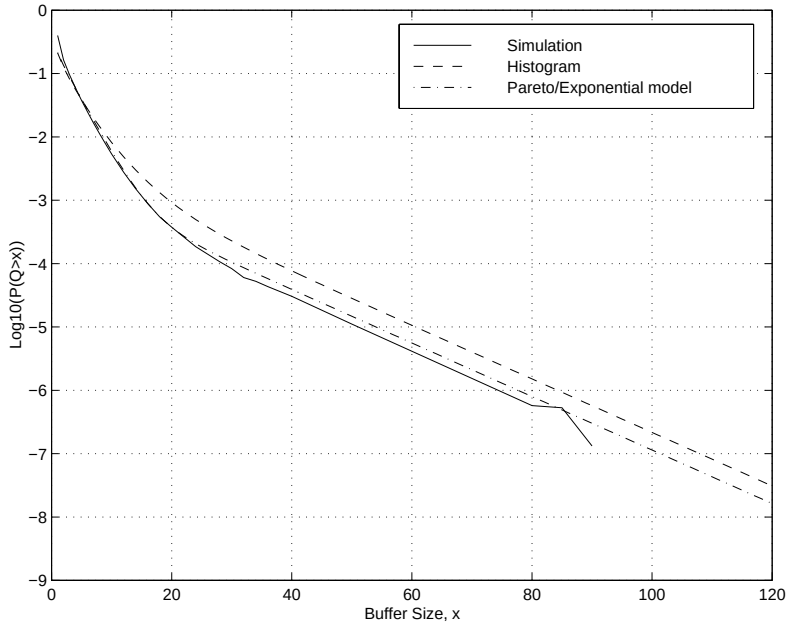

Figure 6.12: Comparison of cell loss probability estimates obtained from theory ($\beta = 1.2$), histogram and simulation of the collected data trace 'Phillips' and shown in Figure 6.10 using $N = 15$ input phases. $\rho = .4$.

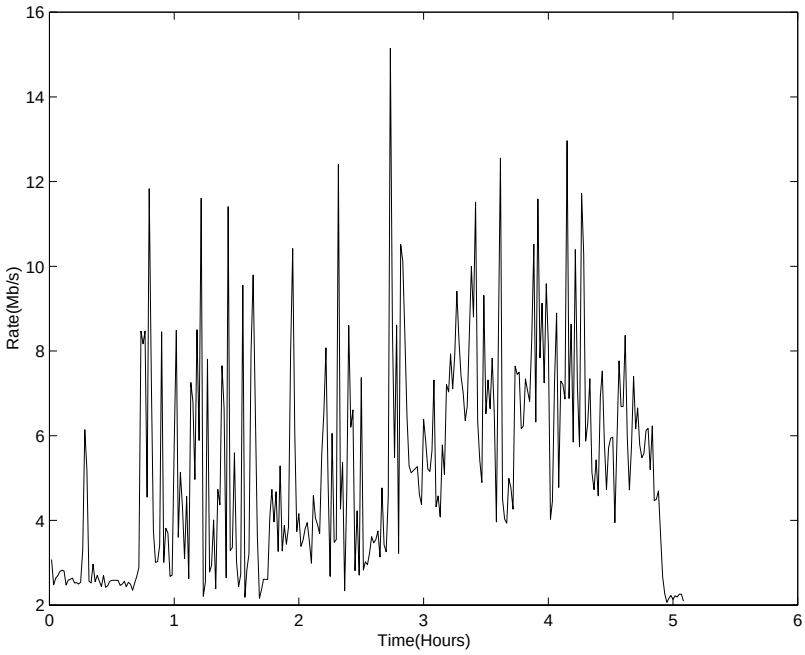

Figure 6.13: Data trace collected from the GSD site.

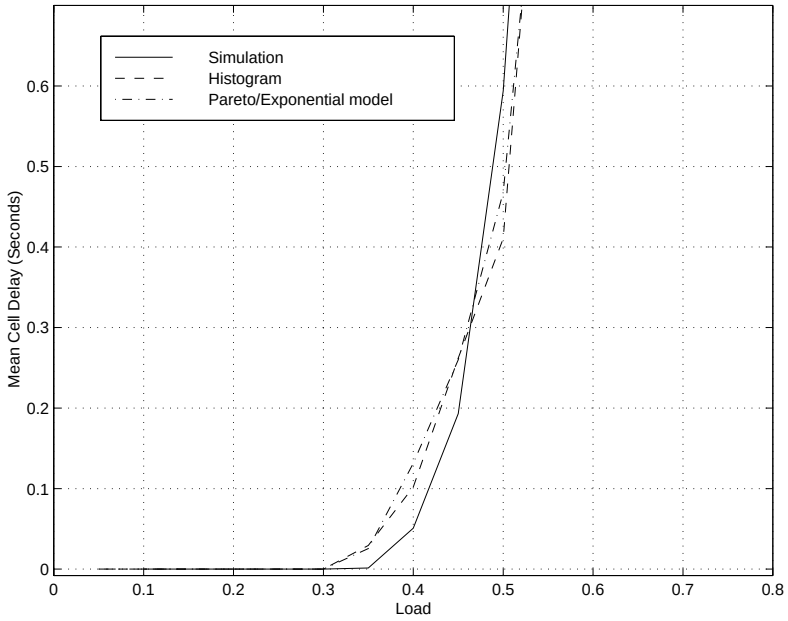

Figure 6.14: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 1.4$), histogram and simulation of the collected data trace labeled 'GSD' and shown in Figure 6.13 using $N = 15$ input phases.

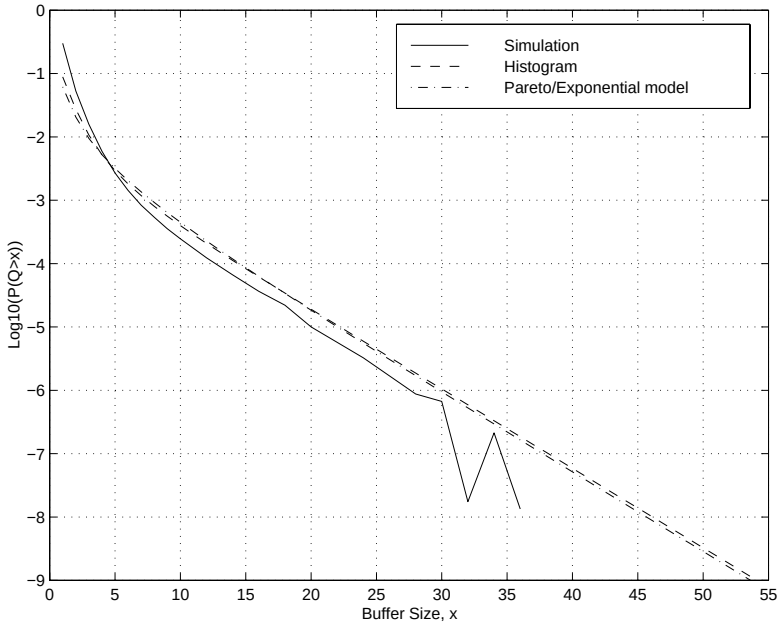

Figure 6.15: Comparison of Cell loss probability estimates obtained from theory ($\beta = 1.4$), histogram and simulation of the collected data trace labeled 'GSD' and shown in Figure 6.13 using $N = 15$ input phases. $\rho = .4$.

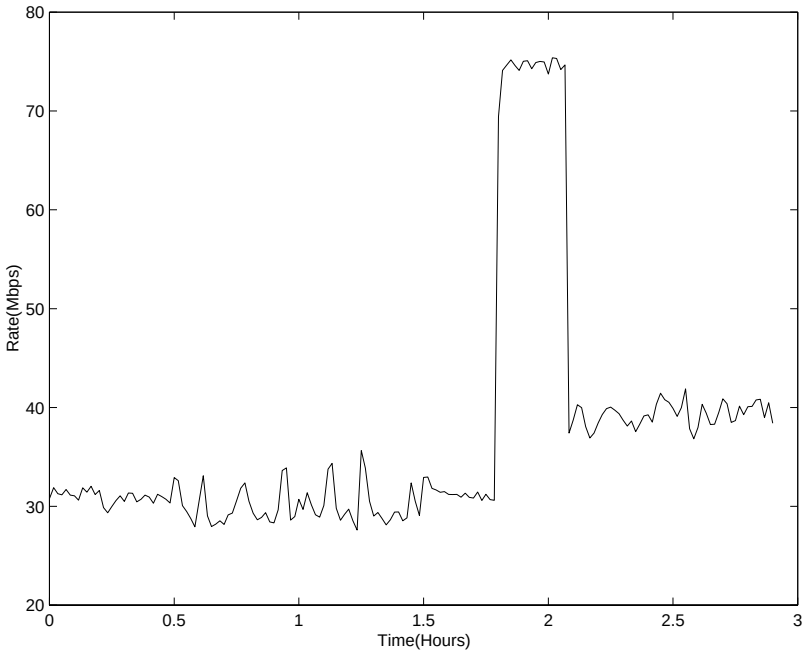

Figure 6.16: Data trace collected from the KU site.

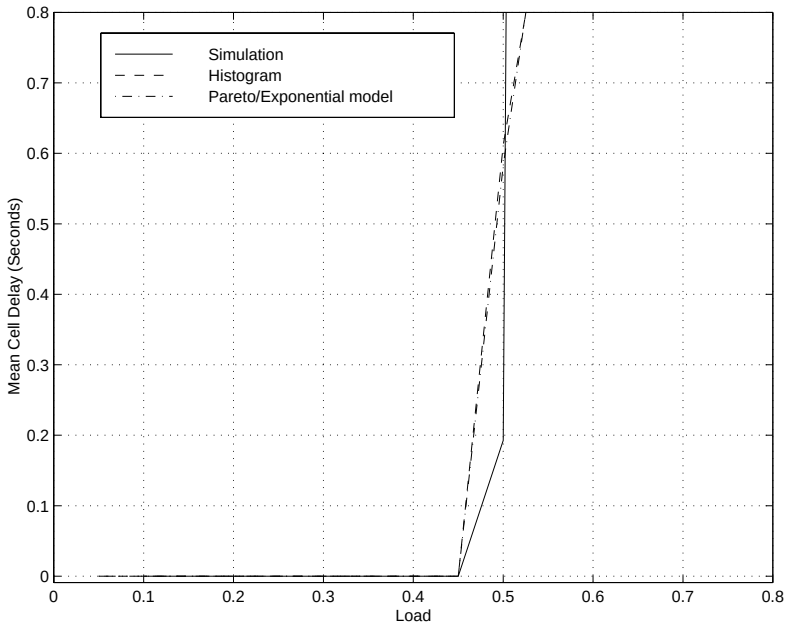

Figure 6.17: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 2.1$), histogram and simulation of the collected data trace labeled 'KU' and shown in Figure 6.16 using $N = 15$ input phases.

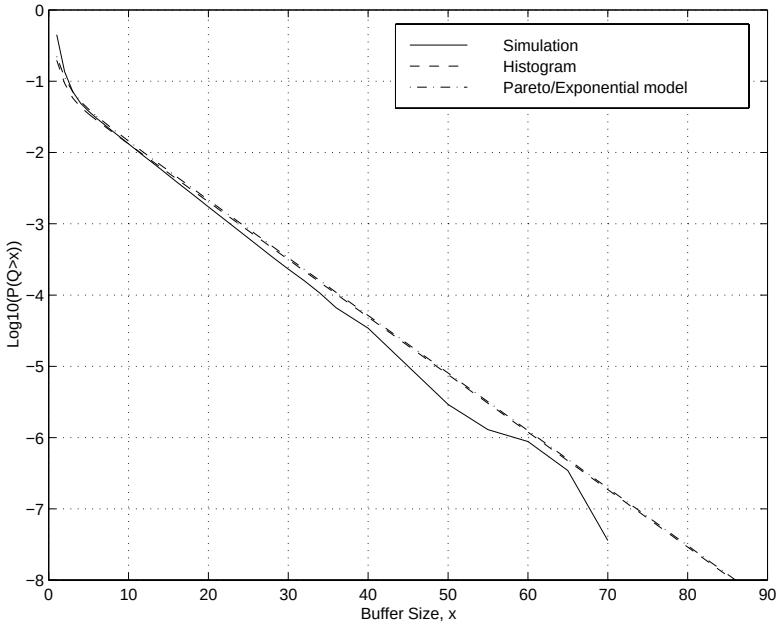

Figure 6.18: Comparison of Cell loss probability estimates obtained from theory ($\beta = 2.1$), histogram and simulation of the collected data trace labeled 'KU' and shown in Figure 6.16 using $N = 15$ input phases. $\rho = .45$

6.3 Traffic Microdynamics

In this section the effect of traffic micro-dynamics on the queuing performance is investigated. The effect of traffic micro-dynamics on queue dynamics has been a point of contention of many traffic studies. In traffic studies like [8], it has been pointed out that long-range dependence is the dominant traffic characteristic that determines the queuing performance. However in [20], it has been argued that traffic dynamics at the micro level do influence the queuing performance.

It should be recalled that in the context of the model developed, the traffic characteristics within a given phase of the input process are defined as the traffic micro-scale dynamics. The micro-scale dynamics define the cell inter-arrival times within a particular phase. In order to study the effect of micro-dynamics, different distributions of

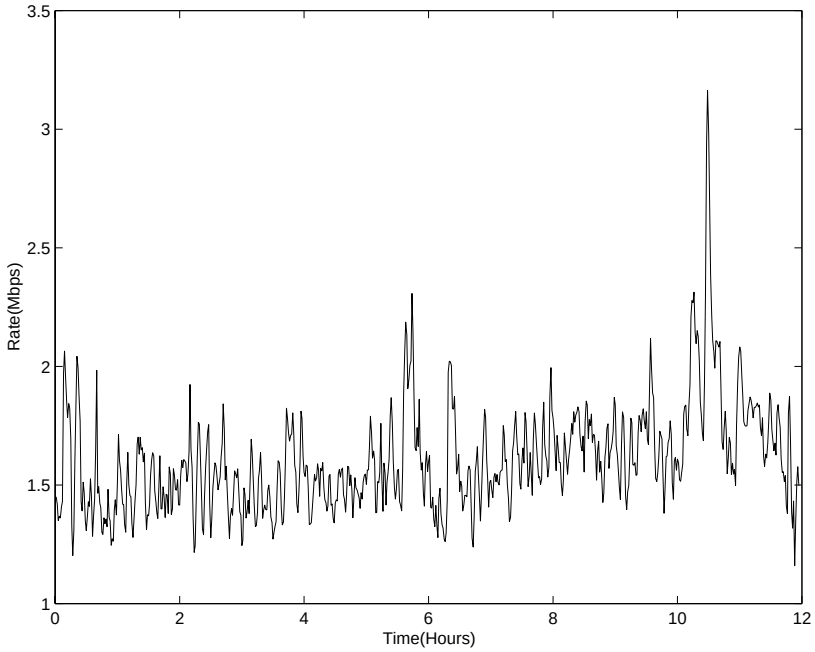

Figure 6.19: Data trace collected from the NRL site.

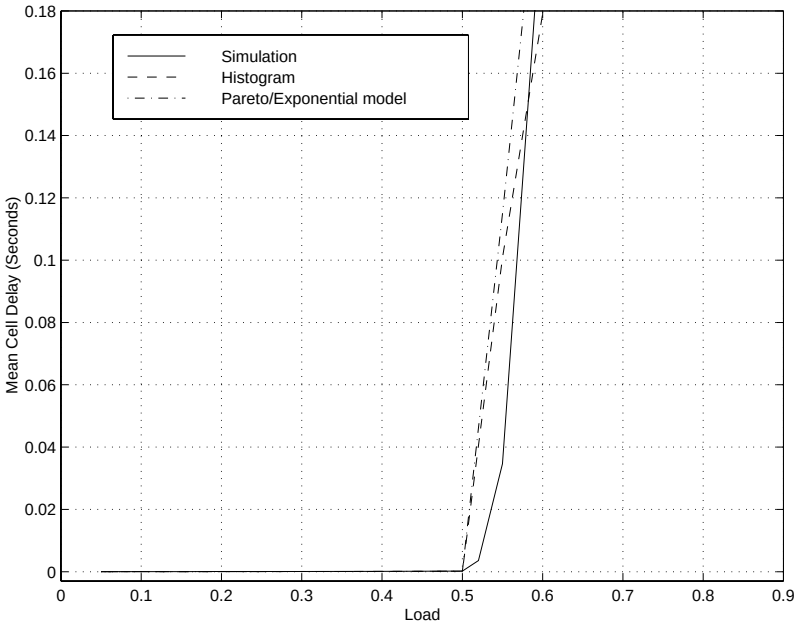

Figure 6.20: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 3.4$),histogram and simulation of the collected data trace labeled 'NRL2' and shown in Figure 6.19 using $N = 15$ input phases.

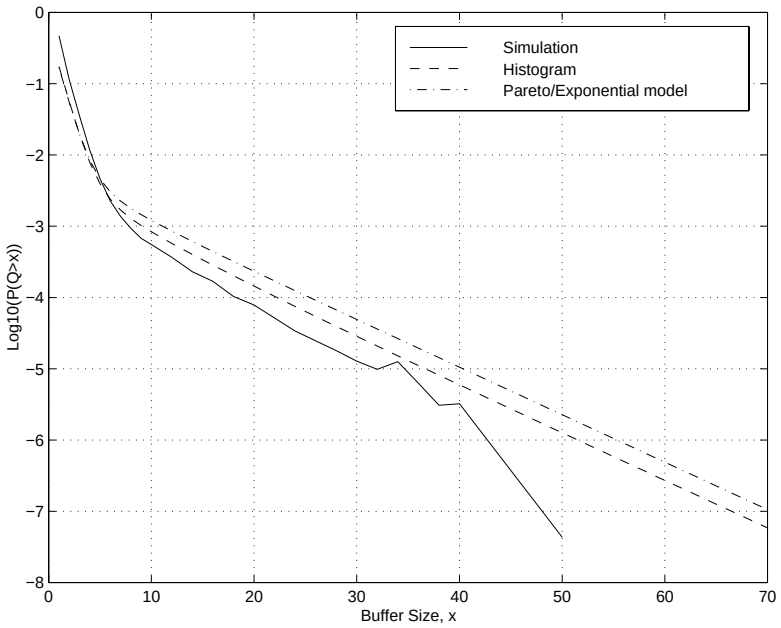

Figure 6.21: Comparison of Cell loss probability estimates obtained from theory ($\beta = 3.4$), histogram and simulation of the collected data trace labeled 'NRL2' and shown in Figure 6.19 using $N = 15$ input phases. $\rho = .45$.

micro-dynamics are employed and the queueing performance is studied both in terms of the mean cell delay and cell loss ratio. In particular, exponential and uniform distributions are considered here for defining the micro-scale dynamics with in a given phase. Figures 6.22, 6.26, 6.29, 6.32, 6.35, 6.38, 6.41 show the delay profiles predicted by the model by employing exponential and uniform distributions for the traffic microscale-dynamics. The results are presented at a finer time scale compared to the previous results to focus on the performance near the ”knee” of the delay vs load curve. These results demonstrate that the delay curves predicted by the model using different micro-dynamics are similar. Also it can be seen that the effect of micro-dynamics on the average cell transfer delay at a queueing system is not significant, in the presence of long-range dependence. Note that the model can be used to effectively analyze the different micro-dynamics because of the G/ D/ 1 analytical technique used for computing the mean queue lengths during each of the sojourn periods of the phase process. The simulation model described in Chapter 5 is now used to verify the effect of micro-dynamics on delay that was predicted by the model. Figures 6.23, 6.27, 6.30, 6.33, 6.36, 6.39 , 6.42 the delay curves obtained from simulations based on the collected traces listed in Table 6.1. The delay curves obtained from the simulations are similar to those predicted by the model and help establish the relative insensitivity of average cell delay to micro-dynamics in the presence of long-range dependence.

Apart from demonstrating the effect of traffic micro-dynamics using analysis, a qualitative explanation to the effect in terms of the power spectrum of the rate process can also be provided. It has been shown that traffic micro-dynamics contribute a white noise term to the power spectrum of the rate process. The amount of white noise contributed depend on the amount of randomness associated with the micro-dynamics. In

the power spectrum, long-range dependence manifests itself as a divergent spectrum:

$$\lim_{\omega \to \infty} P(\omega) \approx \infty$$

As the low frequency domain characteristics determines the queuing performance, long-range dependence predominately dominates the queuing response. However, in the absence of long-range dependence the white-noise contribution of traffic micro dynamics to the low-frequency spectrum cannot be ignored.

The equations for estimating the cell loss probability in a queueing system having a general phase dependent process as the input are complex and analytically difficult to obtain. The influence of micro-dynamics on cell loss are investigated through extensive trace driven simulations based on the traces listed in Table 6.1. As employed in the case of delay results, exponential and uniform distributions are considered for the micro-dynamics. Figure 6.24 shows the cell loss curves at loads of .4 and .6, using both exponential and uniform micro-dynamics. The Figure 6.24 indicates that effect of micro-dynamics on cell loss depends on the load of the queueing system. Similar results were obtained from simulations based on other collected traces. Therefore, Figures 6.25, 6.28, 6.31, 6.37, 6.34, 6.40, 6.43 show the results of the simulations for estimating cell loss probability as the load and the micro-dynamics change. Each of the plots shows the cell loss probability estimate obtained at a given load, using exponential and uniform microscale dynamics. The simulation results indicate that unlike the case of average delay, the cell loss probability values are sensitive to the micro-dynamics, at low loads. As the load increases, the influence of micro-dynamics on cell loss probability reduces. At high loads the cell loss in the queue is mainly determined by the bursts in the arrival process, which is modeled by the macro-dynamic structure of the input process, where as at low loads, the dynamics of cell arrivals represented by the micro-

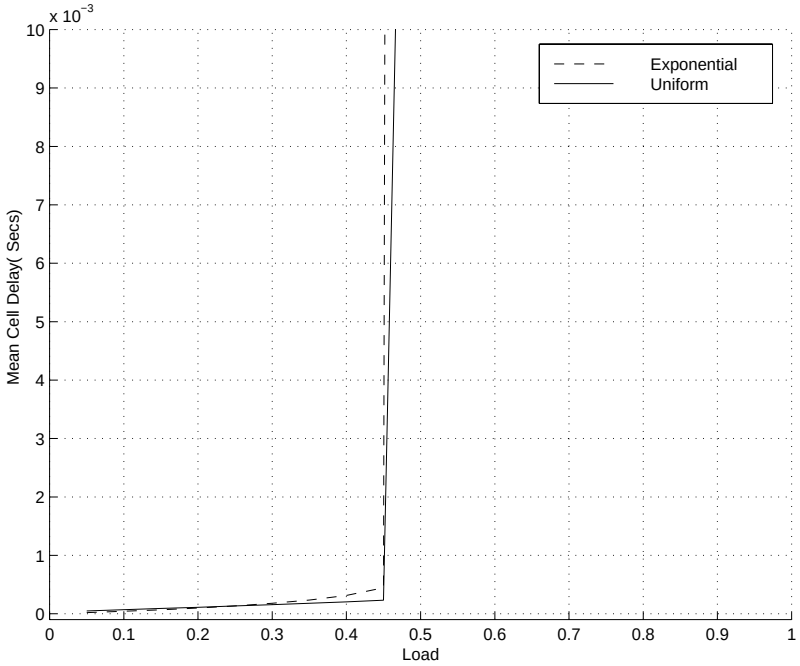

Figure 6.22: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 5.6$), for the trace labeled 'TIOC' and shown in Figure 6.1 using $N = 15$ input phases.

dynamics of the input process dominate the cell loss in the queueing system. Also the loss estimates obtained with exponential micro-dynamics is always greater than the loss probability obtained assuming uniformly distributed microscale dynamics. Figure 6.44 shows the relation between logarithm of the cell loss probability plotted against the load for a trace with exponentially distributed cell inter-arrivals. The generated trace with exponential cell inter-arrival times is the trace used for validating the simulation model was used for obtaining the curve in Figure 6.44. One interesting observation form Figure 6.44 is that unlike the corresponding non linear curves shown in Figures 6.25, 6.28, 6.31, 6.37, 6.43 for traffic with long-range dependence, the plot of logarithm of cell loss vs load is linear as expected.

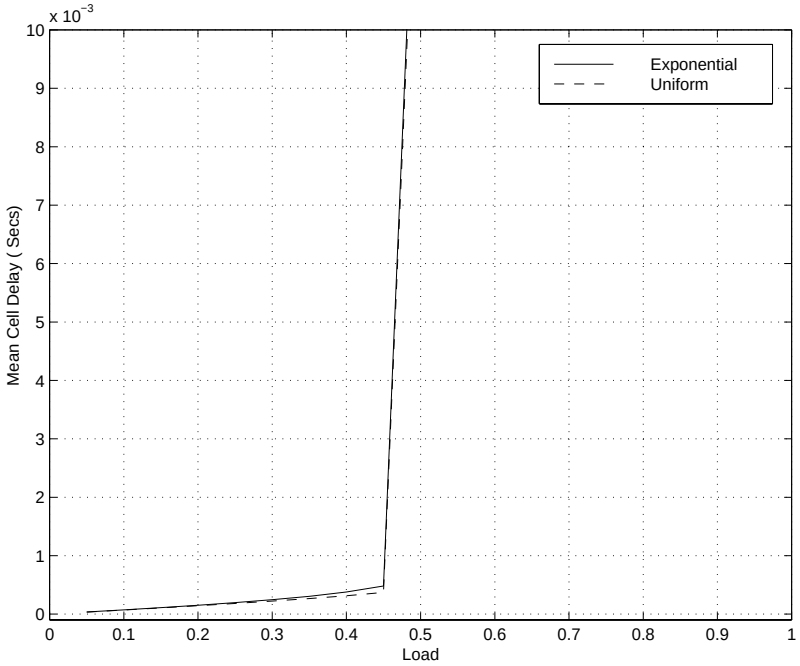

Figure 6.23: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'TIOC' and shown in Figure 6.1.

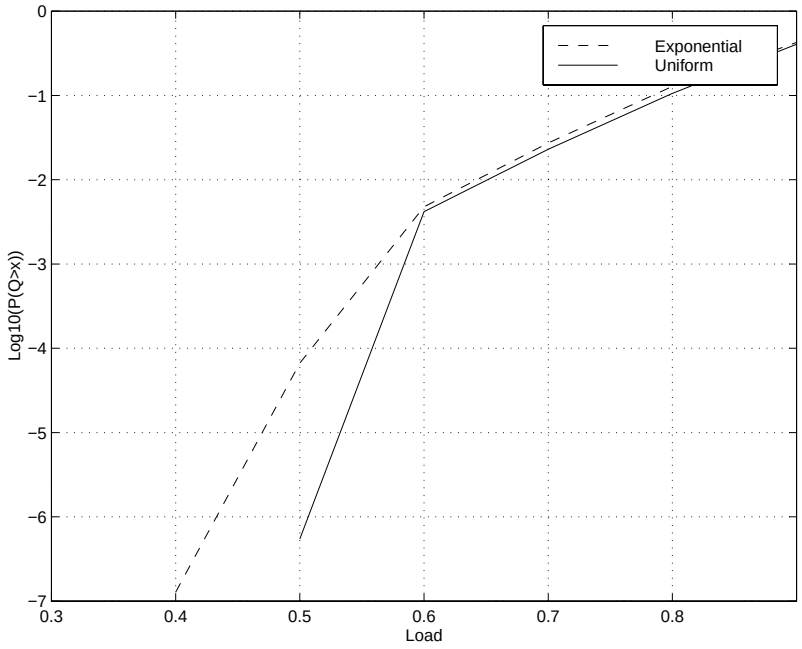

Figure 6.24: Effect of traffic micro-dynamics on Cell loss probability at different values of load, obtained from simulation of the trace labeled 'TIOC' and shown in Figure 6.1 using $N = 15$ input phases.

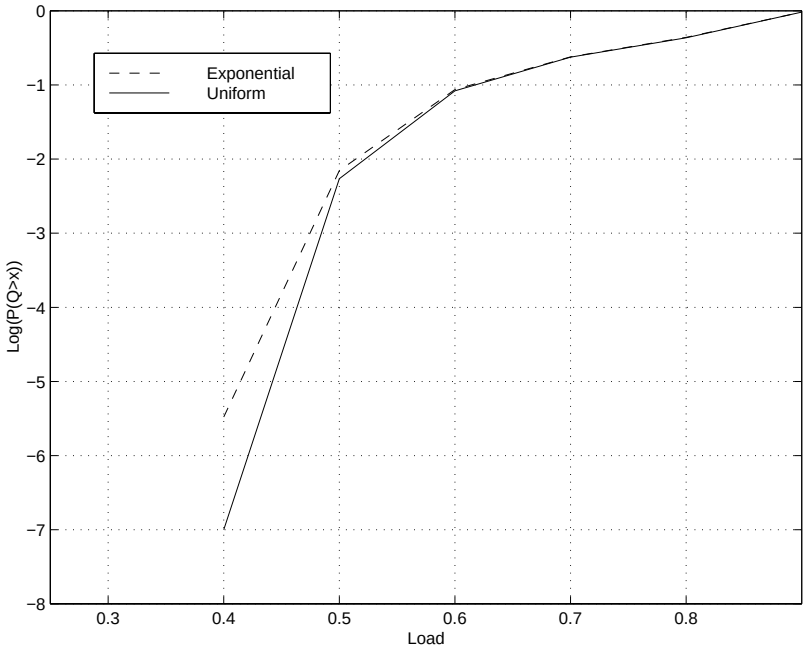

Figure 6.25: Effect of traffic micro-dynamics on Cell loss probability obtained for a fixed buffer size of 20 cells from simulation of the trace data trace labeled 'TIOC' and shown in Figure 6.1.

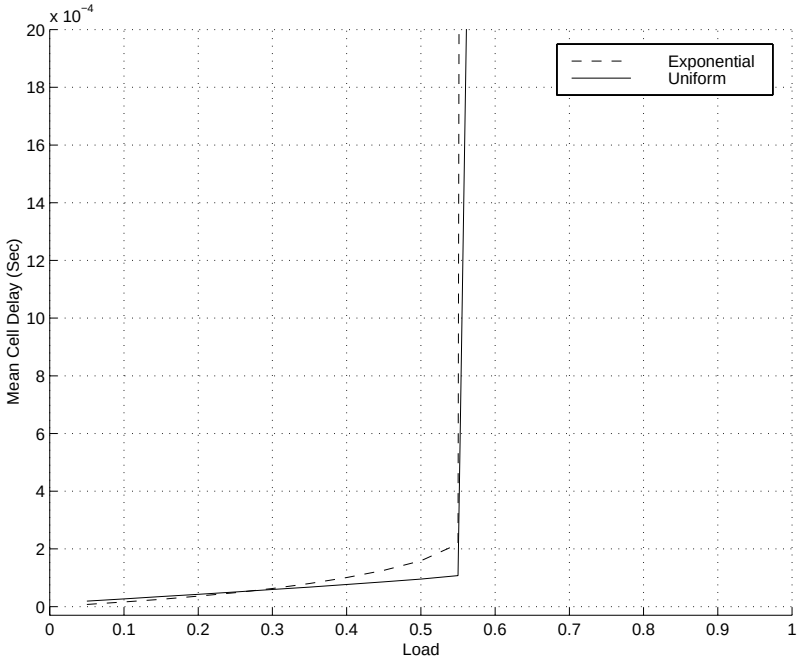

Figure 6.26: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 8.2$), for the data trace labeled 'NCCOSC' and shown in Figure 6.4 using $N = 15$ input phases.

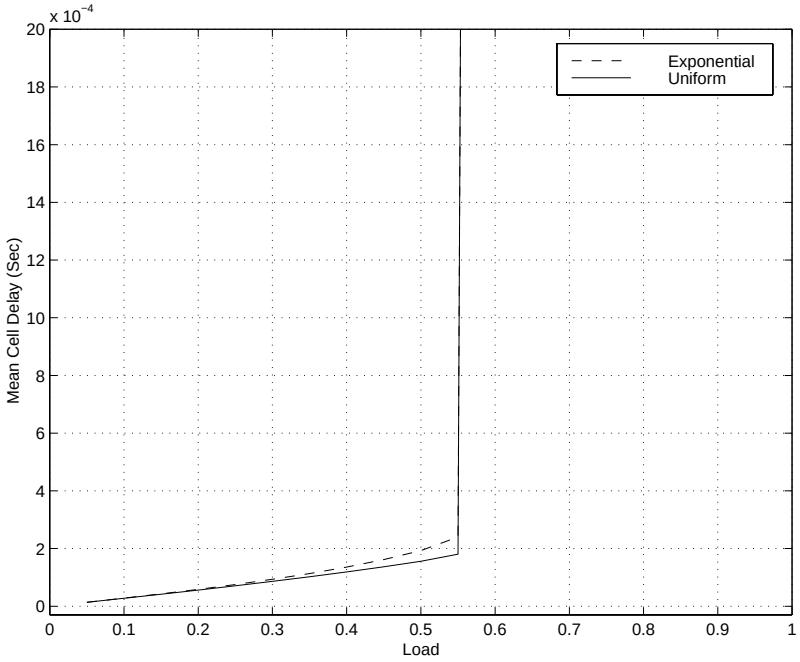

Figure 6.27: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the data trace labeled 'NCCOSC' and shown Figure 6.4.

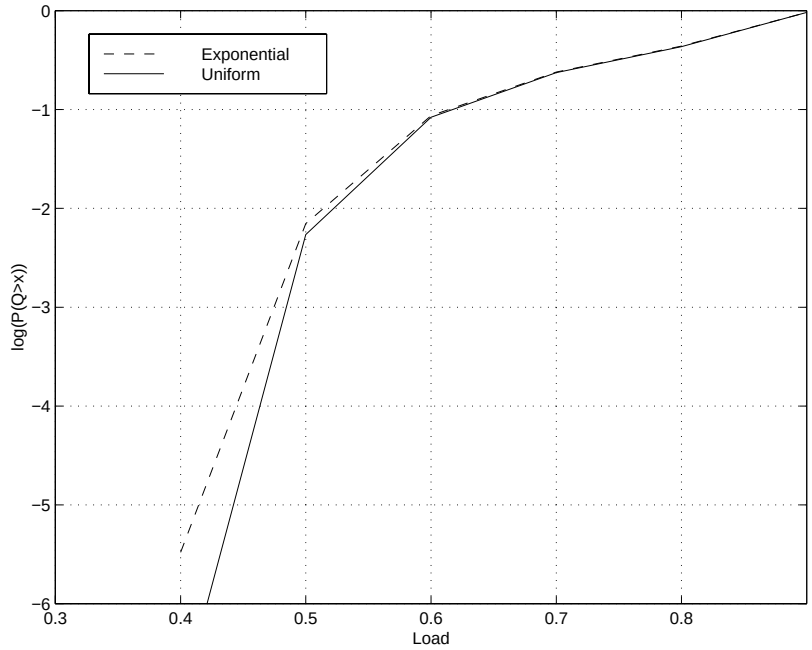

Figure 6.28: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a fixed buffer size of 25 cells from simulation of on the trace labeled 'NCCOSC' and shown in Figure 6.4s.

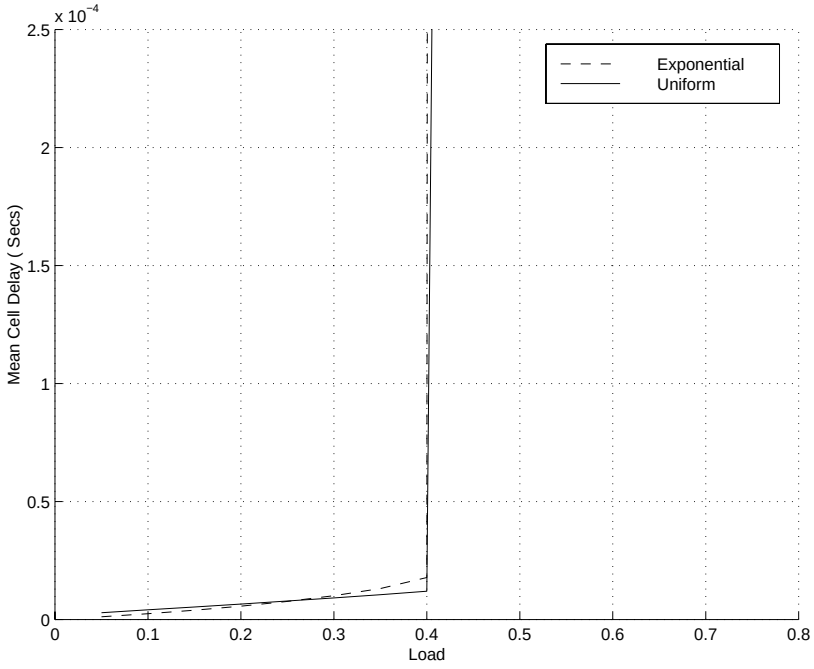

Figure 6.29: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 2.2$), for the trace labeled 'NRL1' and shown in Figure 6.7 using $N = 15$ input phases.

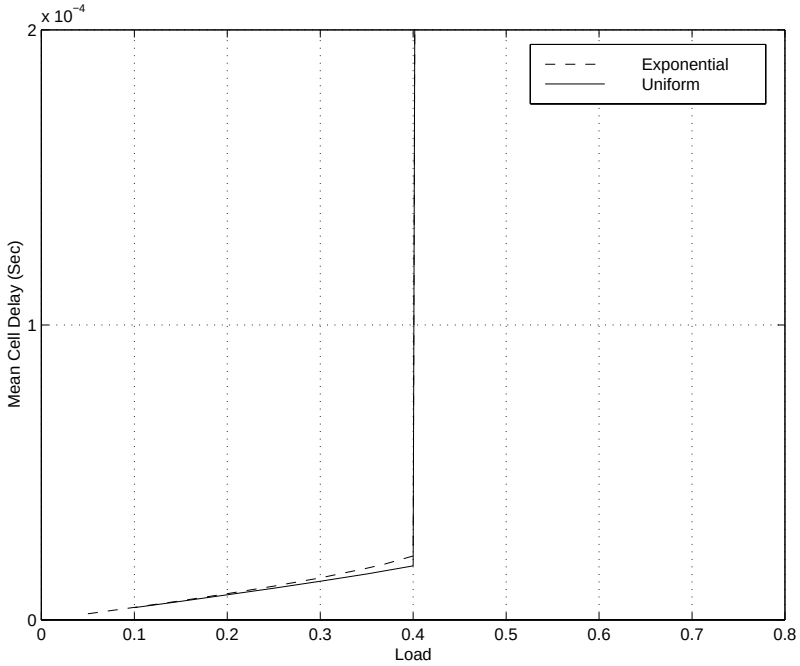

Figure 6.30: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'NRL1' and shown in Figure 6.7.

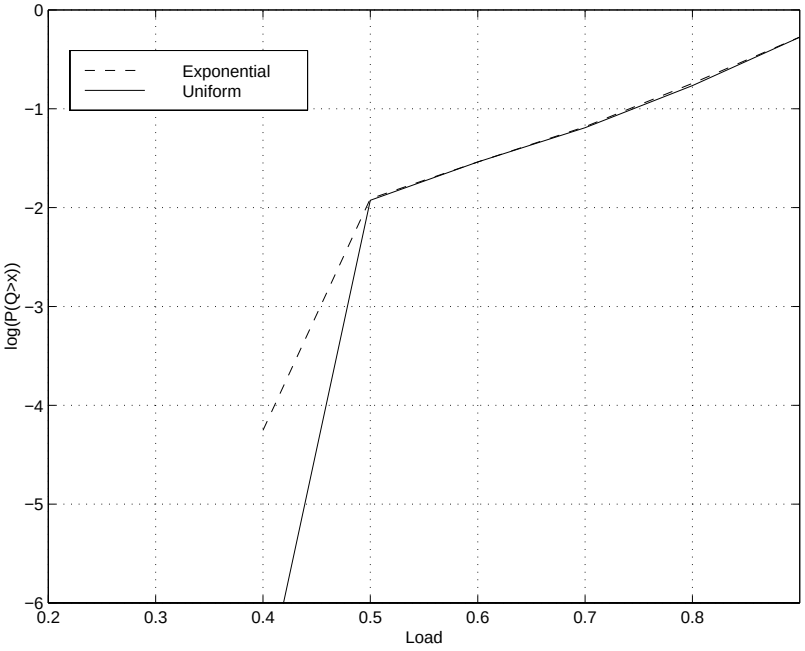

Figure 6.31: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a buffer size of 25 cells from simulation of the trace labeled 'NRL1' and shown in Figure 6.7.

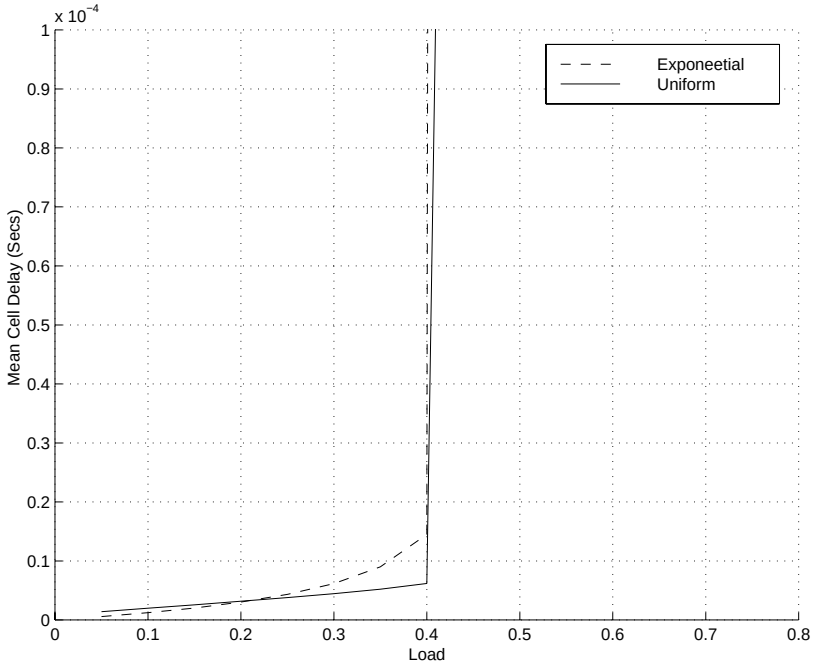

Figure 6.32: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 1.2$), for the trace labeled 'Phillips' and shown in Figure 6.10 using $N = 15$ input phases.

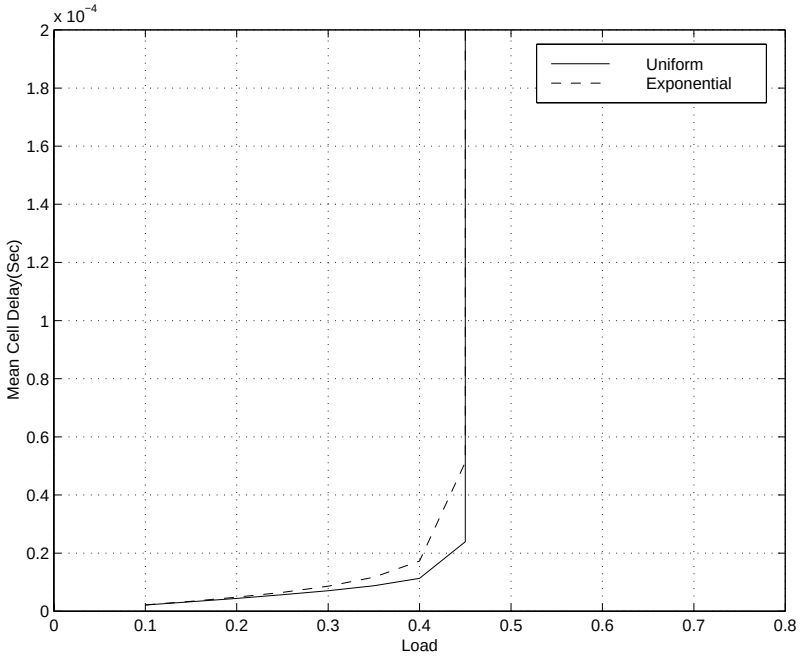

Figure 6.33: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'Phillips' and shown in Figure 6.10.

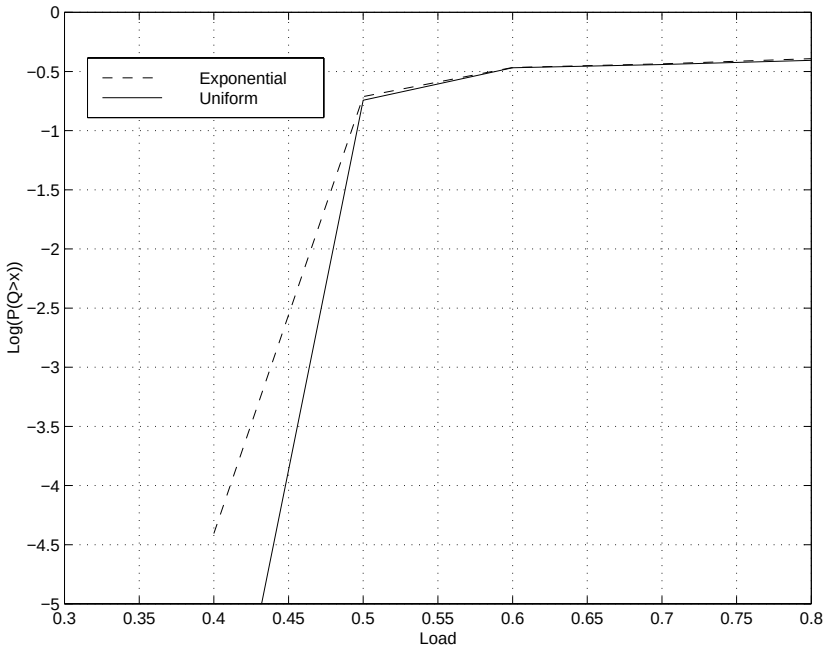

Figure 6.34: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a buffer size of 30 cells from simulation of the trace labeled 'Phillips' and shown in Figure 6.10.

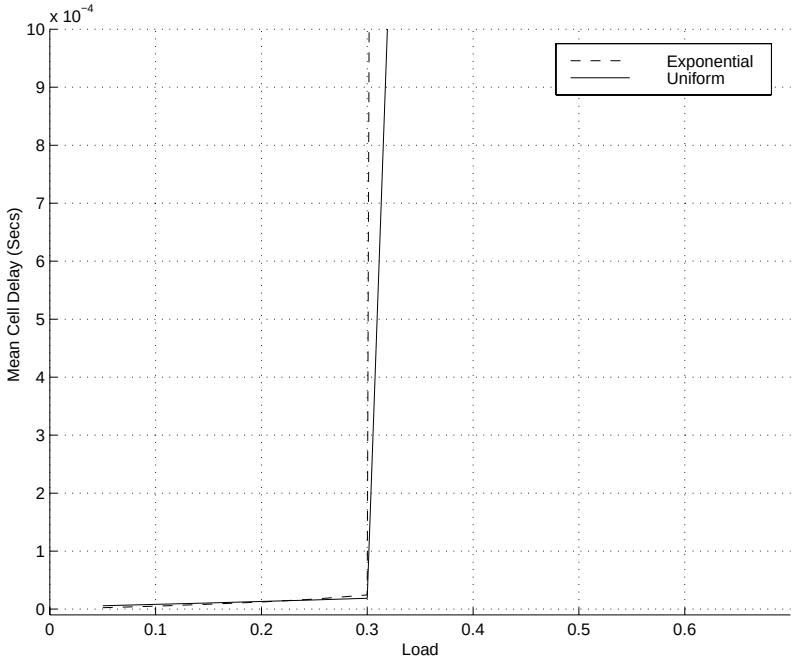

Figure 6.35: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 1.4$), for the trace labeled 'GSD' and shown in Figure 6.13 using $N = 15$ input phases.

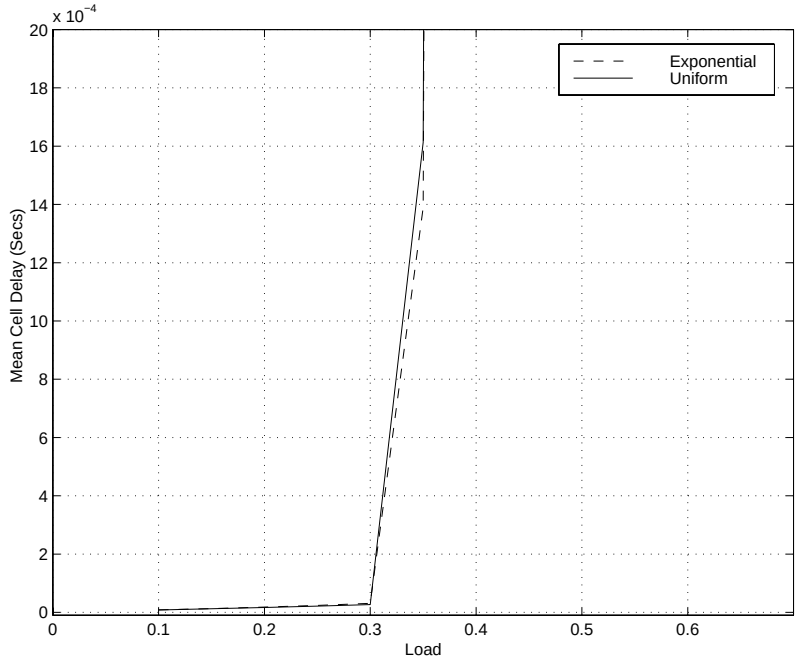

Figure 6.36: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'GSD' and shown in Figure 6.13.

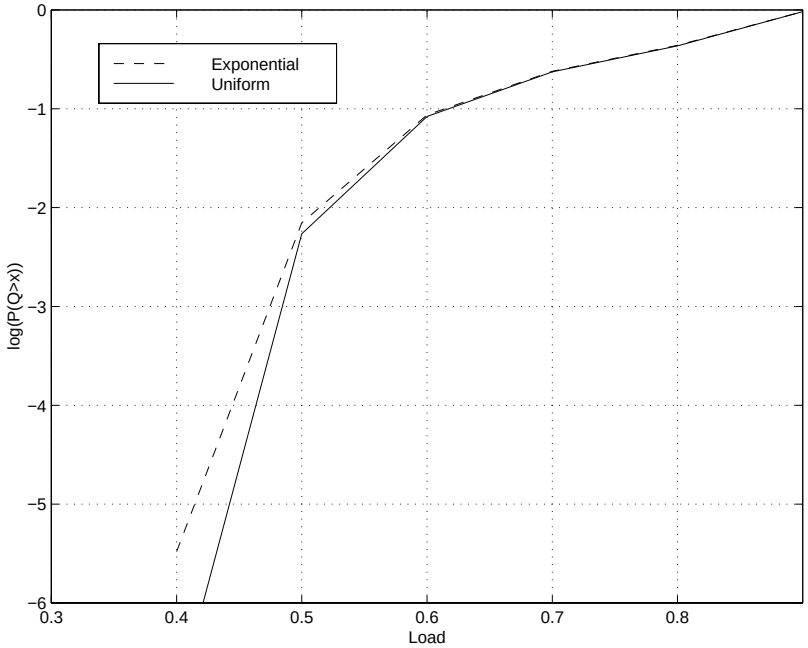

Figure 6.37: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a buffer size of 25 cells from simulation of the trace labeled 'GSD' and shown in Figure 6.13.

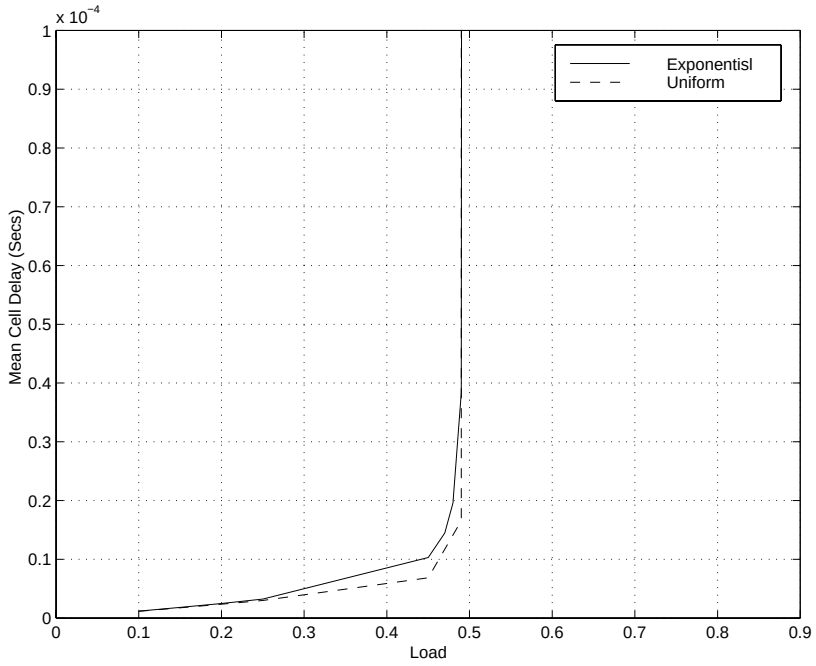

Figure 6.38: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 2.1$), for the trace labeled 'KU' and shown in Figure 6.16 using $N = 15$ input phases.

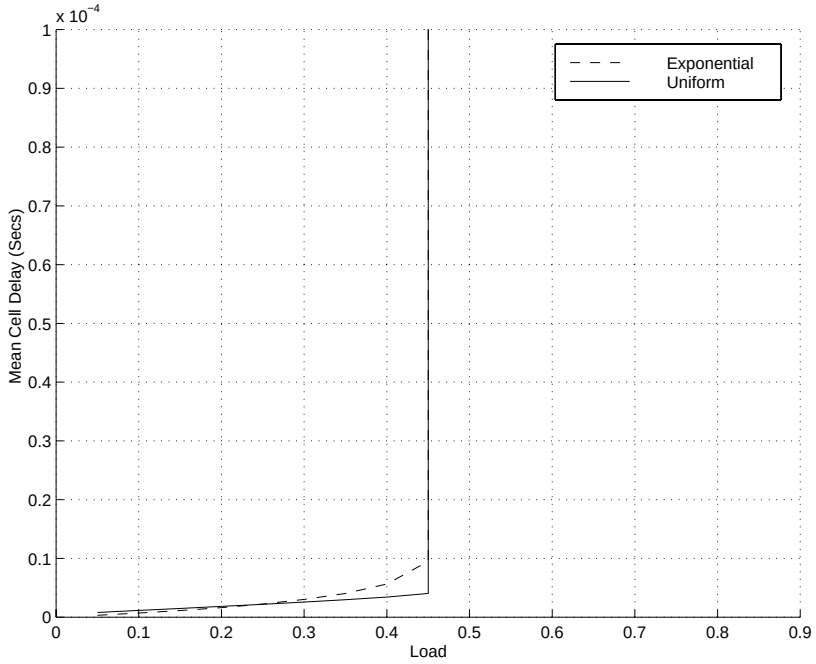

Figure 6.39: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'KU' and shown in Figure 6.16.

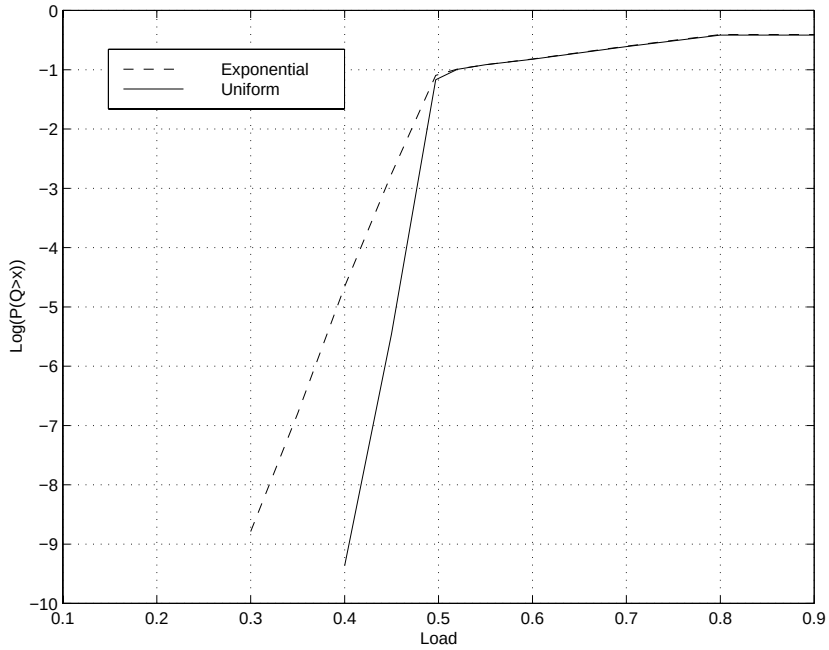

Figure 6.40: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a buffer size of 20 cells from simulation of the trace labeled 'KU' and shown in Figure 6.16.

6.4 Sensitivity of the Model to number of phases

As mentioned, the results in the Section 4.2 were obtained using exponential micro-dynamics and $N = 15$. In Section 4.3 the effects of micro-dynamics were studied. In this section the dependence of mean delay and cell loss results on the number of phases, N, is considered. Performance results are estimated for $N = 10, 15, 25$ and 35 phases. The results for cell delay sensitivity are shown in Figures 6.45, 6.47, 6.49, 6.51, 6.53, 6.55 and 6.57. The corresponding cell loss results are shown in Figures 6.46, 6.48, 6.50, 6.52, 6.54, 6.56 and 6.58. From the results it can be seen that the performance prediction results are not sensitive to high values of N. The results in these figures show that using fifteen phases in the arrival process is sufficient to produce accurate results.

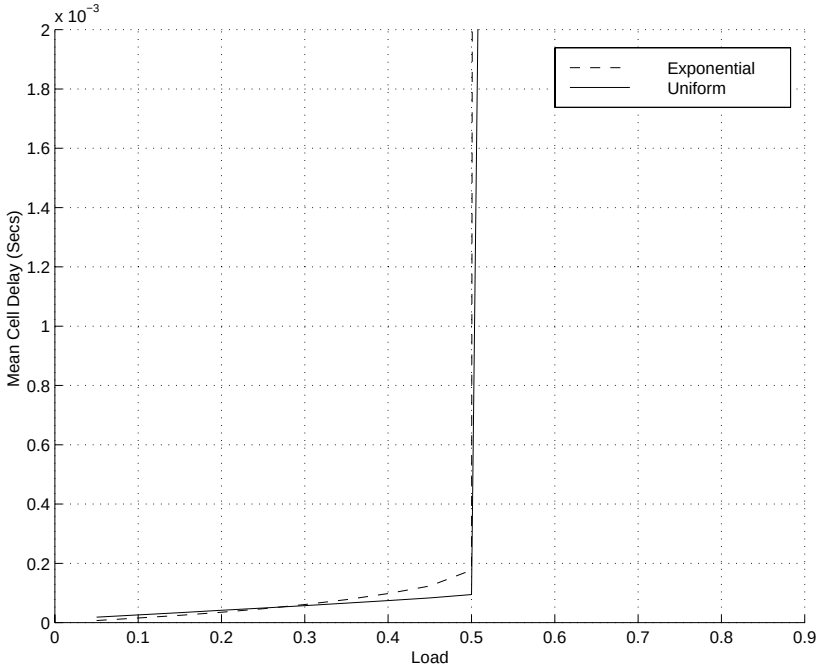

Figure 6.41: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 3.4$), for the trace labeled 'NRL2' and shown in Figure 6.19 using $N = 15$ input phases.

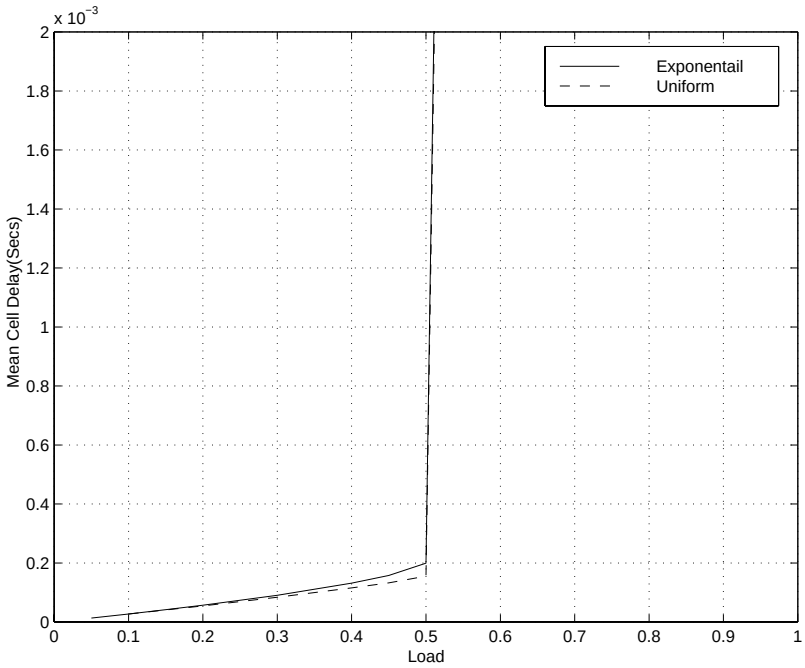

Figure 6.42: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'NRL2' and shown in Figure 6.19.

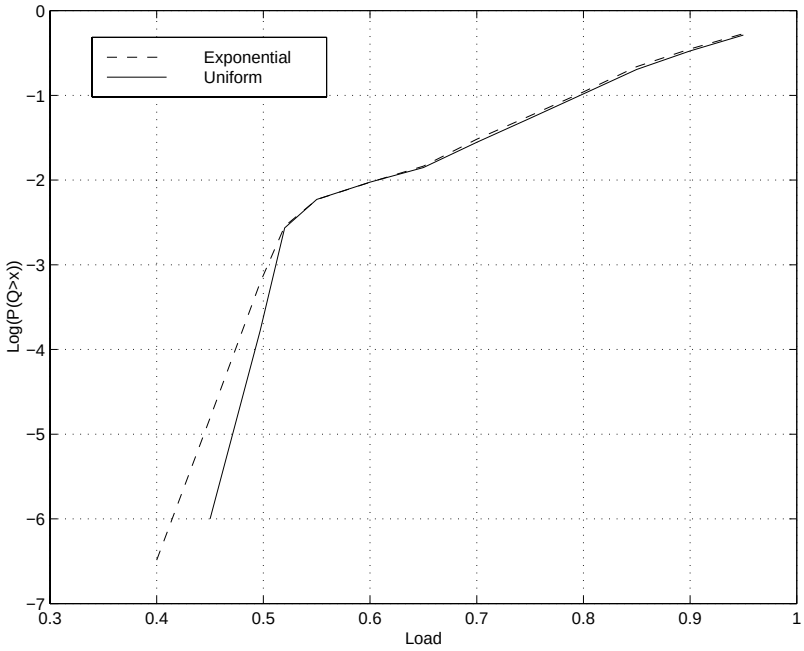

Figure 6.43: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a buffer size of 30 cells from simulation of the trace labeled 'NRL2' and shown in Figure 6.19.

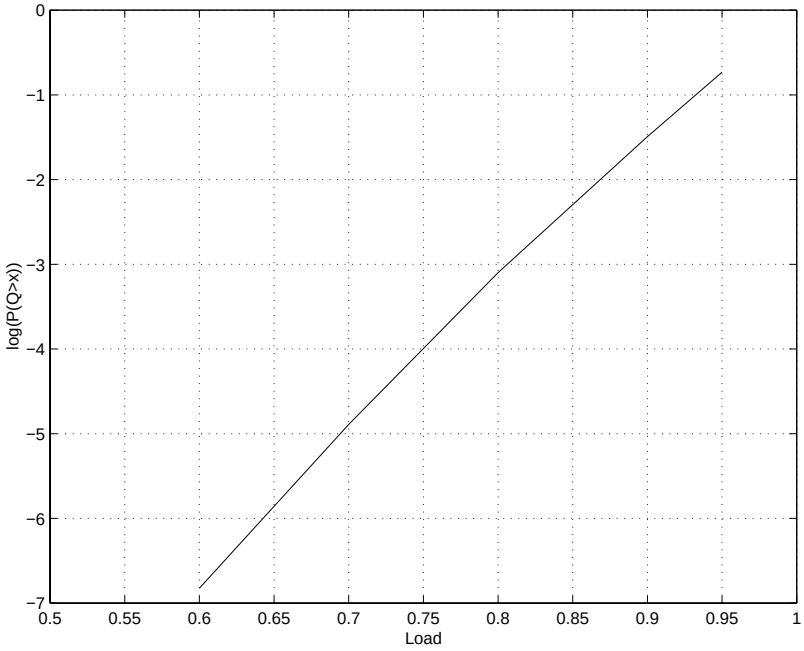

Figure 6.44: Plot of cell loss vs load for a trace with exponential inter-arrival times.

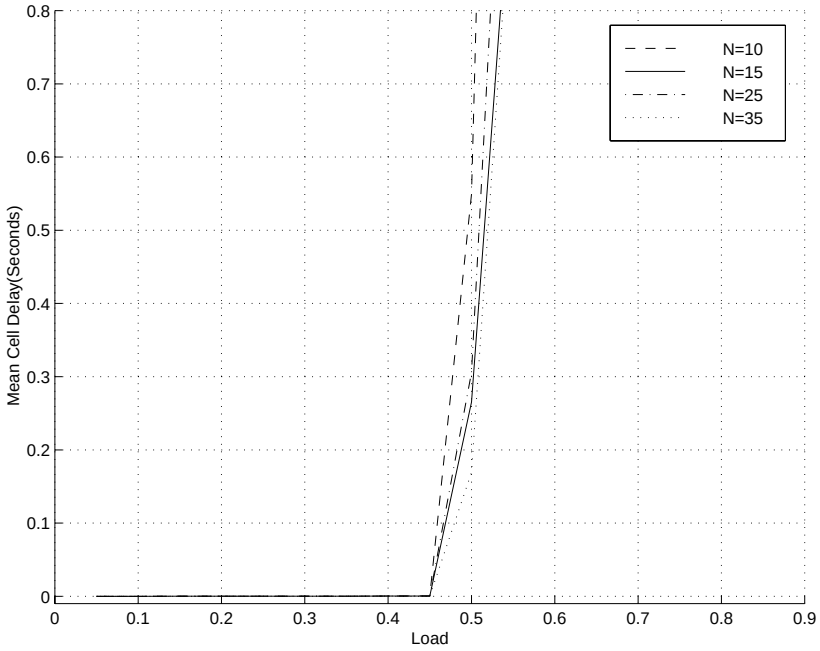

Figure 6.45: Sensitivity of mean cell delay to the number of phases for the trace labeled 'TIOC' and shown in Figure 6.1 using $N = 15$ input phases.

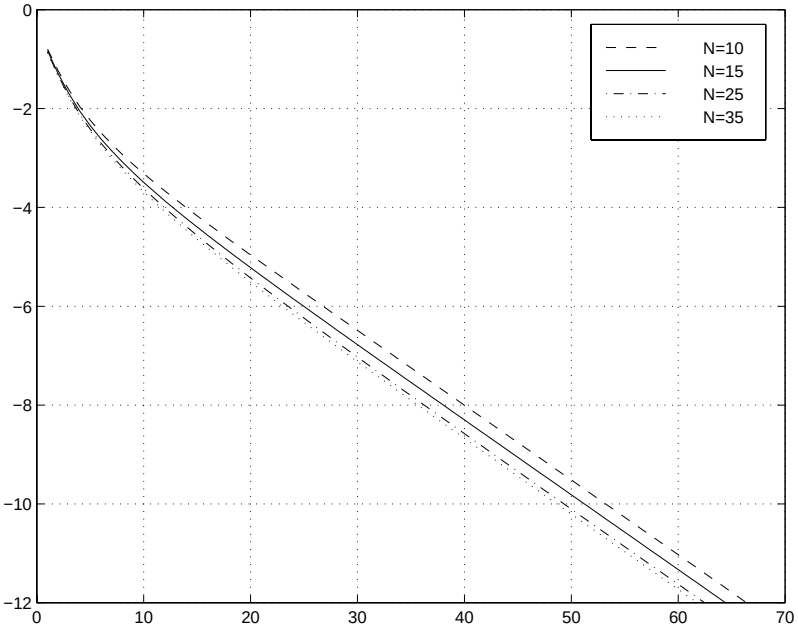

Figure 6.46: Sensitivity of cell loss probability to the number of phases for the trace labeled 'TIOC' and shown in Figure 6.1 using $N = 15$ input phases. $\rho = .4$.

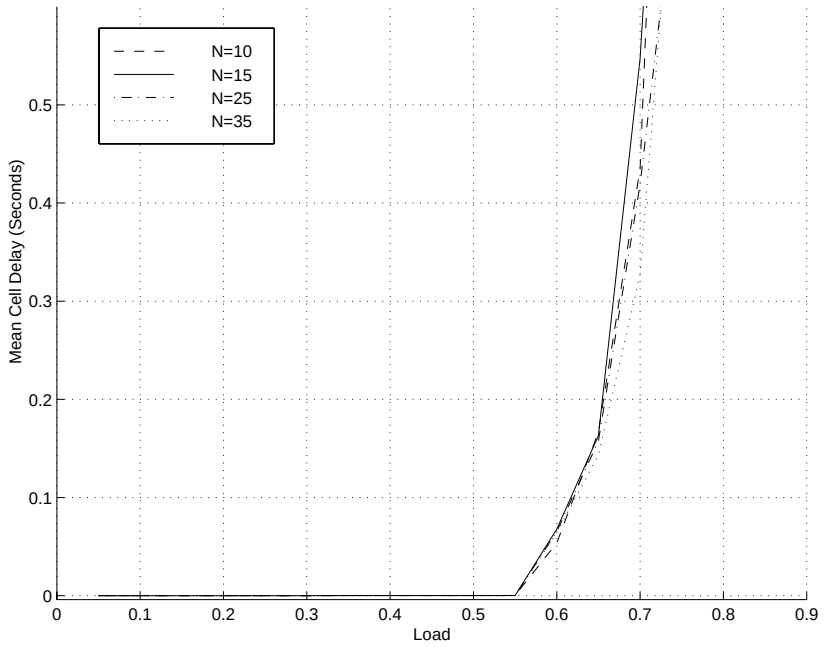

Figure 6.47: Sensitivity of mean cell delay to the number of phases for the trace labeled 'NCCOSC' and shown in Figure 6.4 using $N = 15$ input phases.

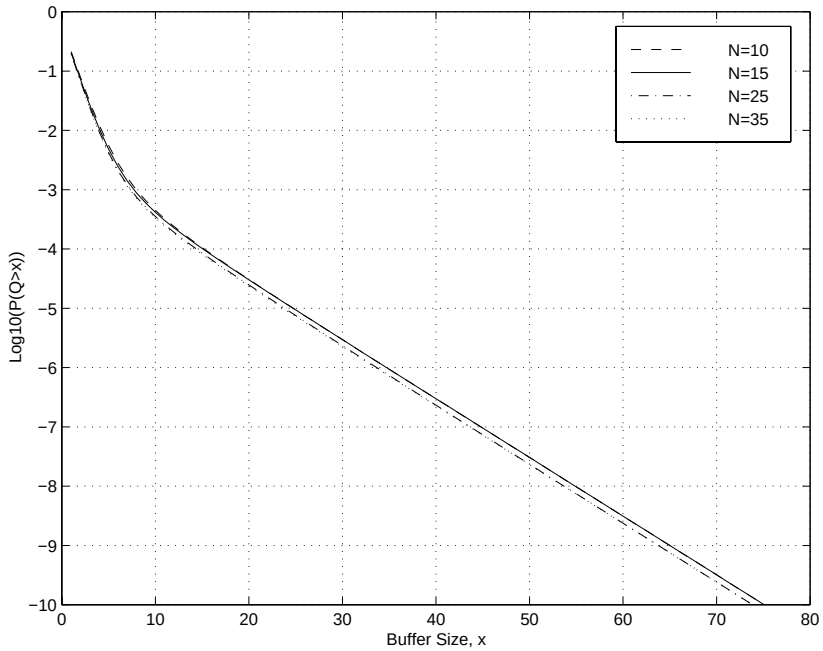

Figure 6.48: Sensitivity of cell loss probability to the number of phases for the trace labeled 'NCCOSC' and shown in Figure 6.4 using $N = 15$ input phases. $\rho = .6$.

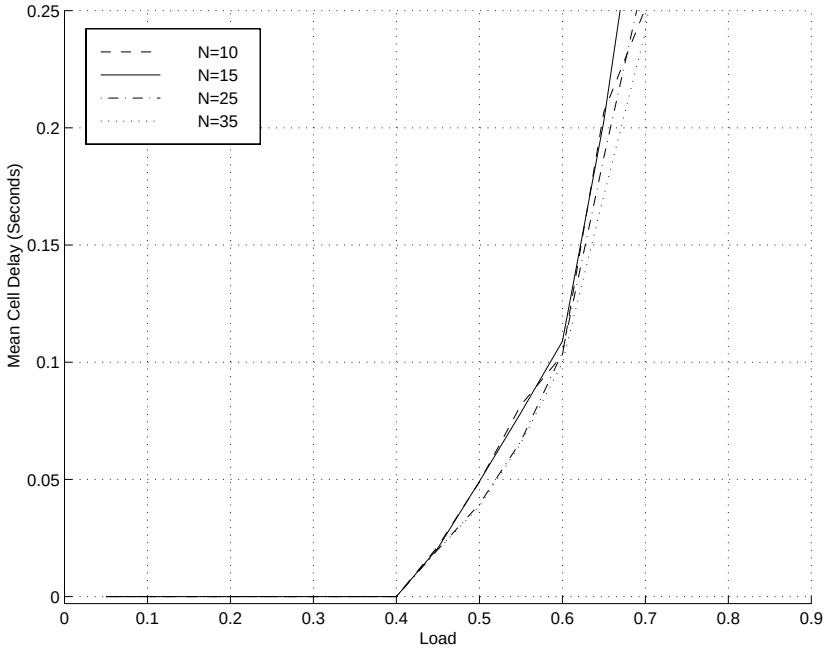

Figure 6.49: Sensitivity of mean cell delay to the number of phases for the trace labeled 'NRL1' and shown in Figure 6.7 using $N = 15$ input phases.

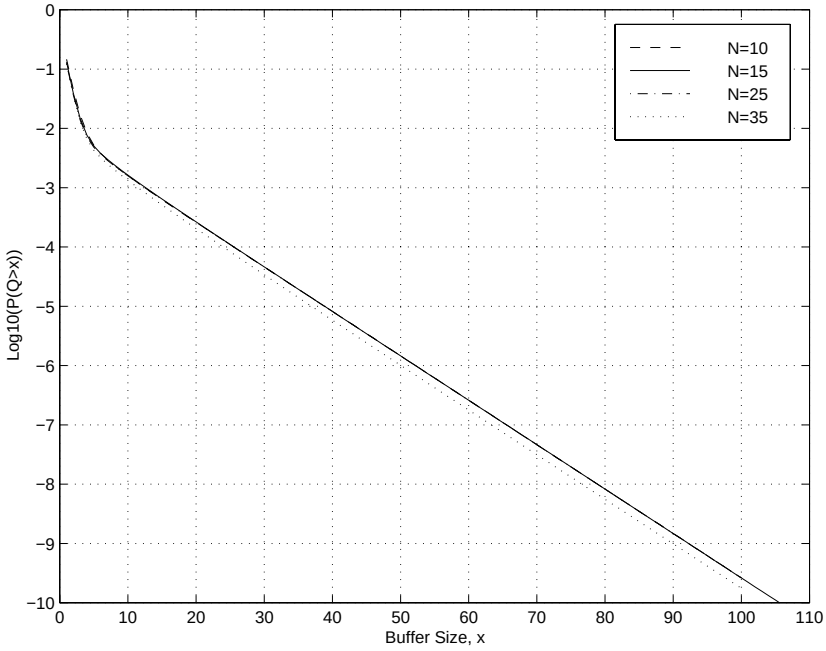

Figure 6.50: Sensitivity of cell loss probability to the number of phases for the trace labeled 'NRL1' and shown in Figure 6.7 using $N = 15$ input phases. $\rho = .45$.

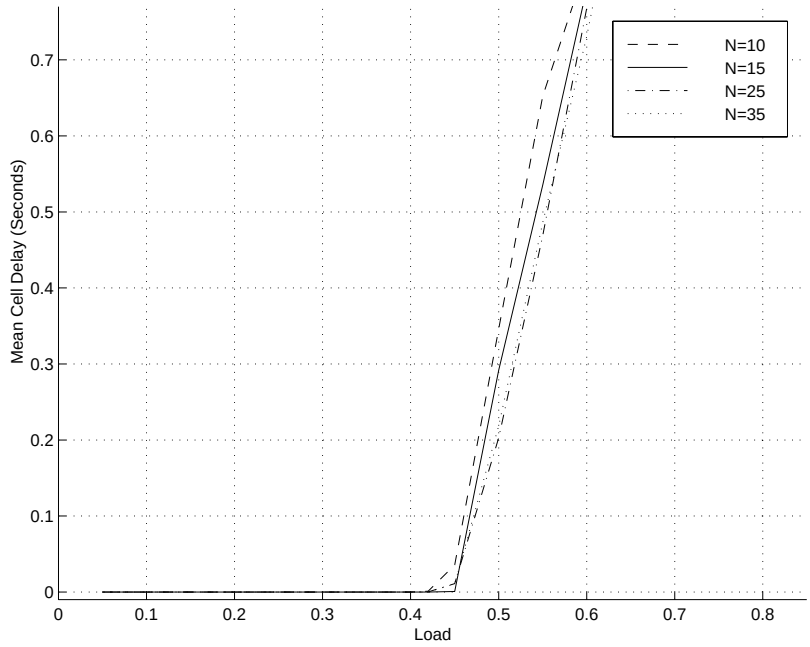

Figure 6.51: Sensitivity of mean cell delay to the number of phases for the trace labeled 'Phillips' and shown in Figure 6.10 using $N = 15$ input phases.

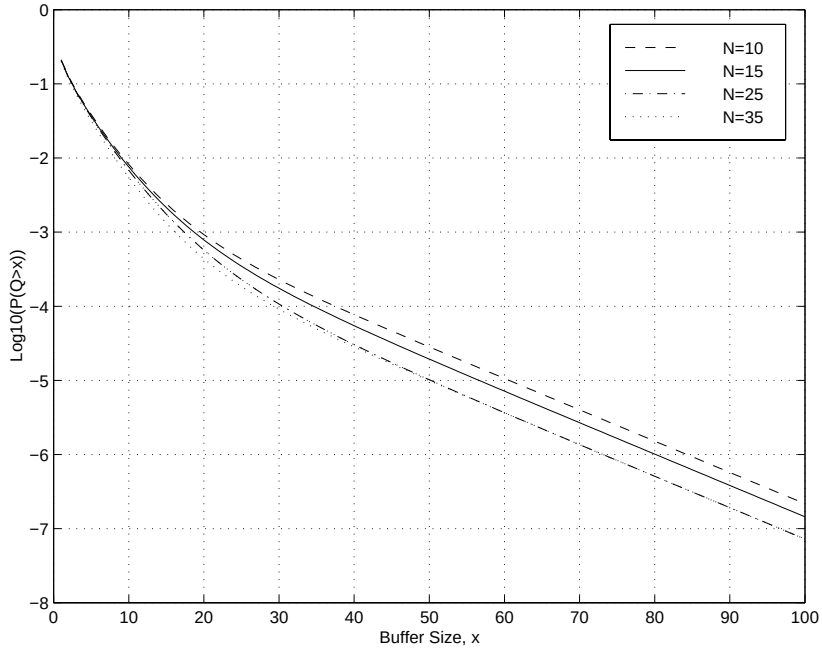

Figure 6.52: Sensitivity of cell loss probability to the number of phases for the trace labeled 'Phillips' and shown in Figure 6.10 using $N = 15$ input phases. $\rho = .4$.

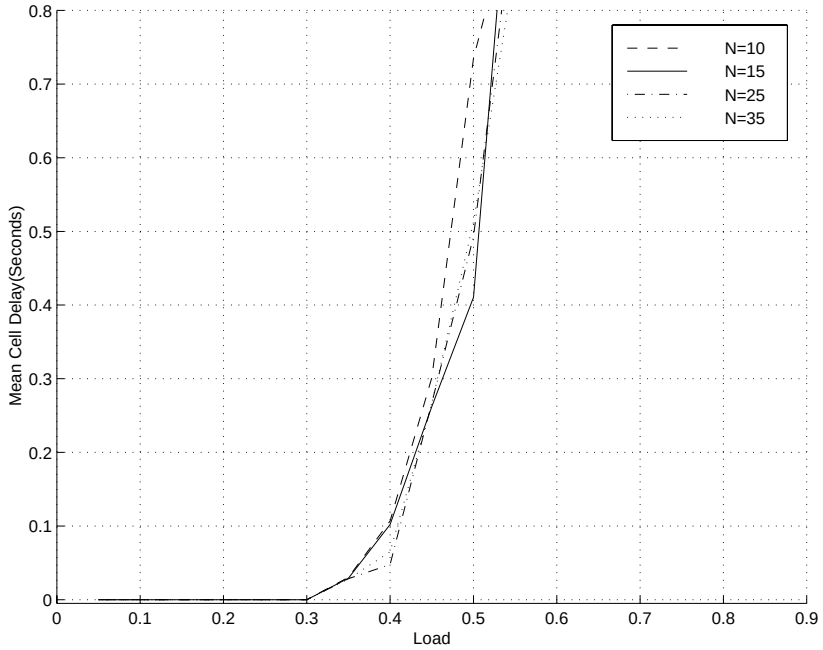

Figure 6.53: Sensitivity of mean cell delay to the number of phases for the trace labeled 'GSD' and shown in Figure 6.13 using $N = 15$ input phases.

6.5 Second-order statistics

In this section accuracy of the model in terms of the second-order statistics of the rate process is demonstrated. In time domain, the second-order statistics of the rate process $R(t)$, are represented by its auto-correlation function. The auto correlation function obtained from collected trace data is compared with that obtained from a Pareto/Exponential model. The sojourn times in each of the phases is assumed to be proportional to the steady state probability of being in that phase i.e., the input process is assumed to be in phase x_i for duration proportional to π_i. In particular for a sampling interval of T_s seconds, the sojourn time in phase x_i is equal to the product of T_s and π. Figures 6.59, 6.60, 6.61, 6.63, 6.64, 6.65 show the autocorrelation function obtained from trace data and that obtained using the model. From these figures we see that autocorrelation functions of the model and real network data match rea-

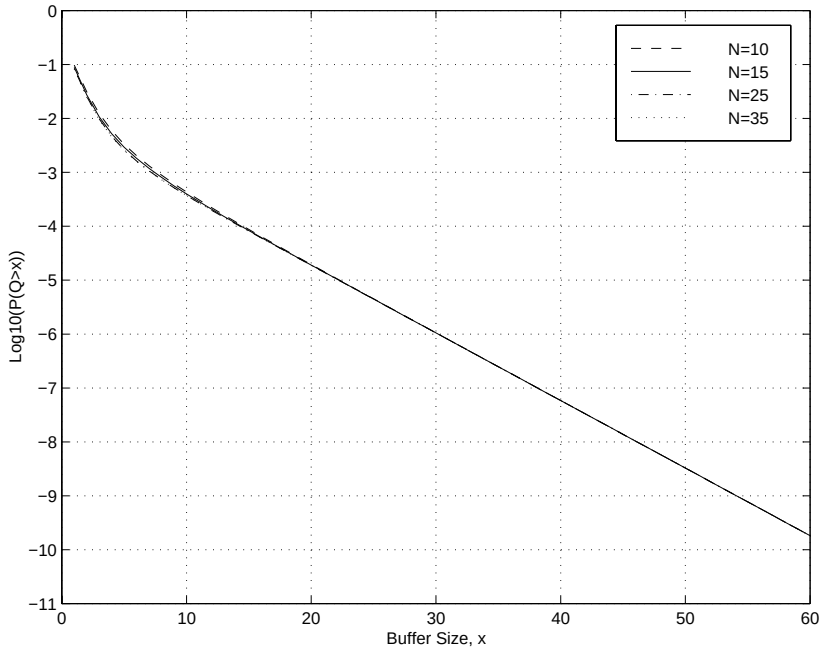

Figure 6.54: Sensitivity of cell loss probability to the number of phases for the trace labeled 'GSD' and shown in Figure 6.13 using $N = 15$ input phases. $\rho = .4$.

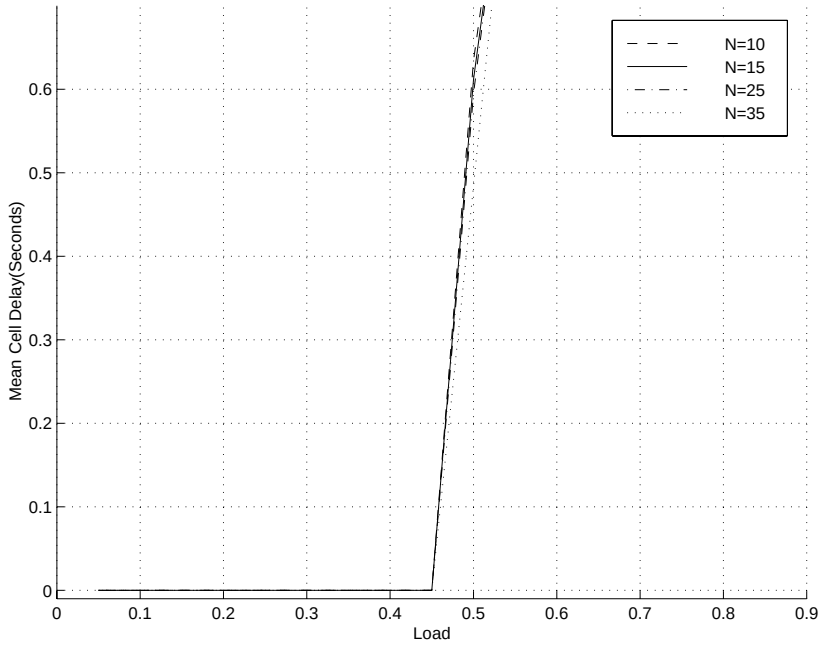

Figure 6.55: Sensitivity of mean cell delay to the number of phases for the trace labeled 'KU' and shown in Figure 6.16 using $N = 15$ input phases.

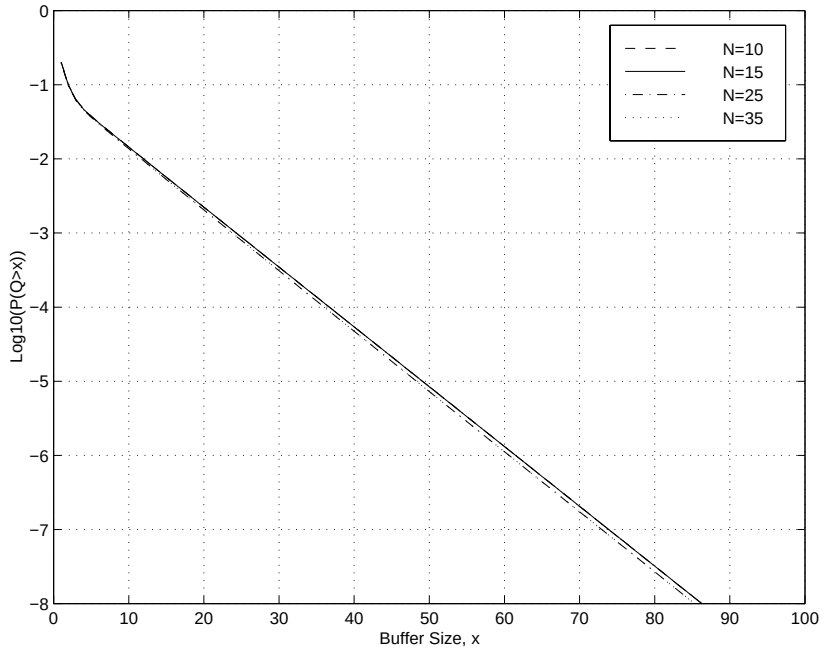

Figure 6.56: Sensitivity of cell loss probability to the number of phases for the trace labeled 'KU' and shown in Figure 6.16 using $N = 15$ input phases. $\rho = .5$.

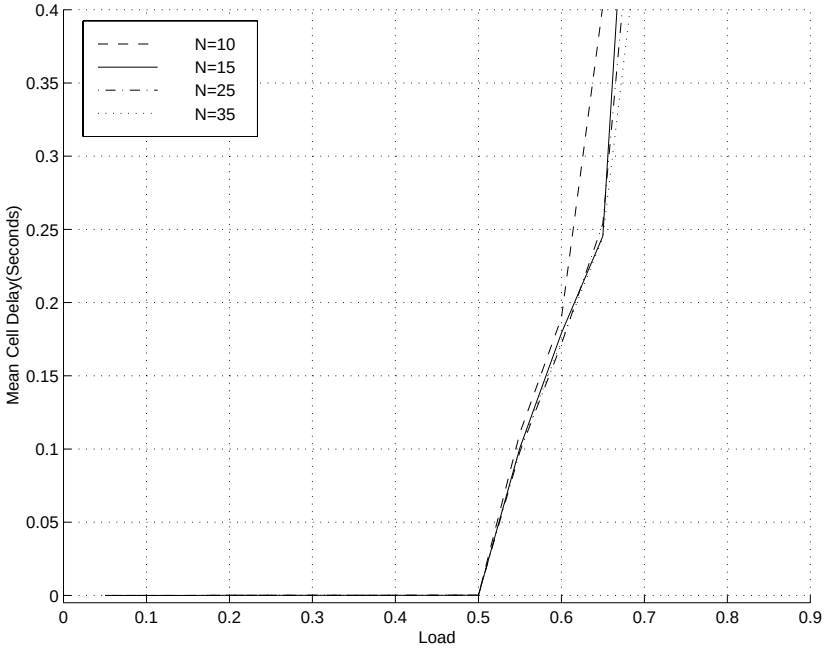

Figure 6.57: Sensitivity of mean cell delay to the number of phases for the trace labeled 'NRL2' and shown in Figure 6.19 using $N = 15$ input phases.

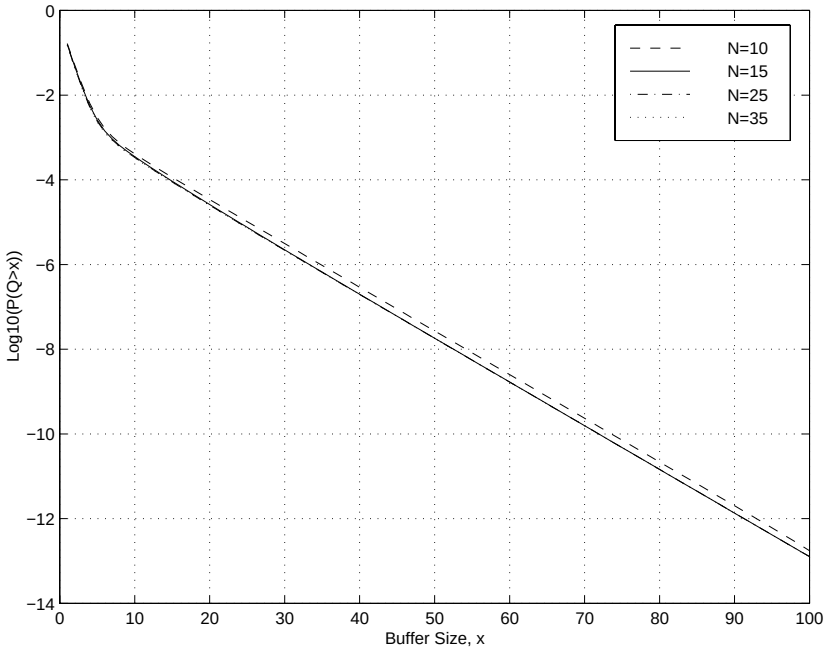

Figure 6.58: Sensitivity of cell loss probability to the number of phases for the trace labeled 'NRL2' and shown in Figure 6.19 using $N = 15$ input phases. $\rho = .5$.

sonably well even for long lags. The accuracy of the model in capturing higher order statistics like the bispectrum and trispectrum is not studied as such statistics are shown to have a marginal influence on the queuing performance [17] and computing higher order statistics is computationally intensive. Also note that further work needs to be done for obtaining analytical expressions for computing the autocorrelation function numerically.

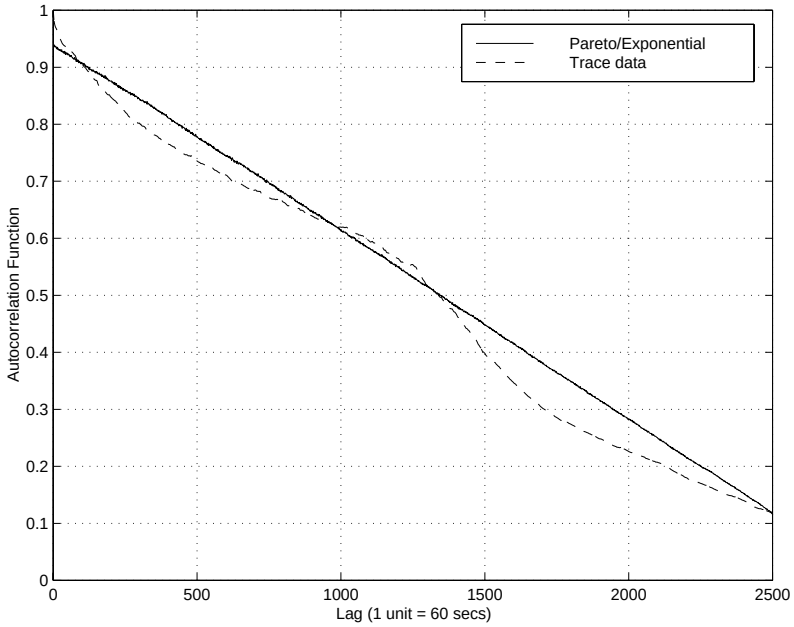

Figure 6.59: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'TIOC' and shown in Figure 6.1.

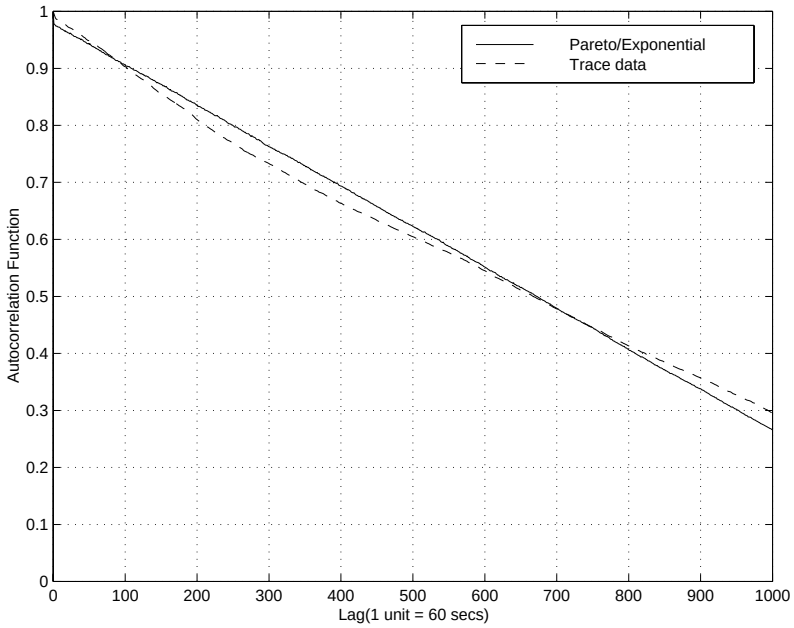

Figure 6.60: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'NCCOSC' and shown in Figure 6.4.

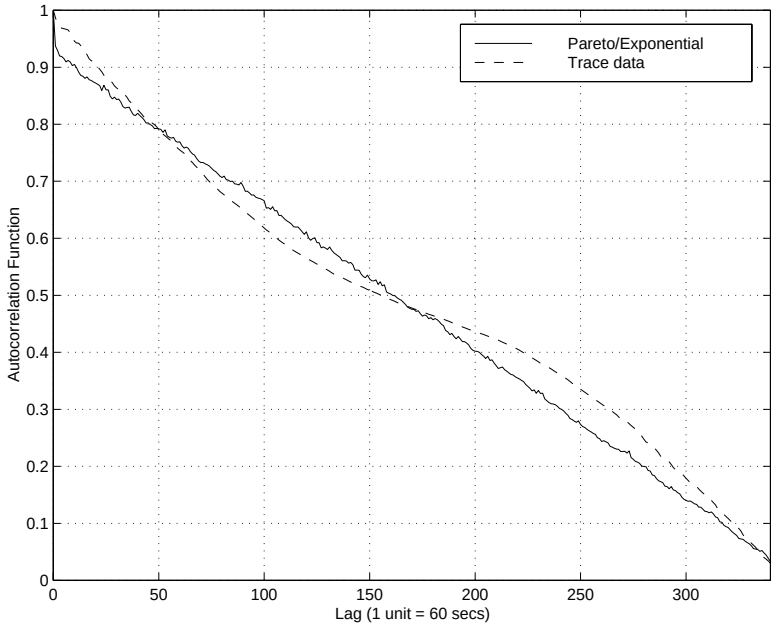

Figure 6.61: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'NRL1' and shown in Figure 6.7.

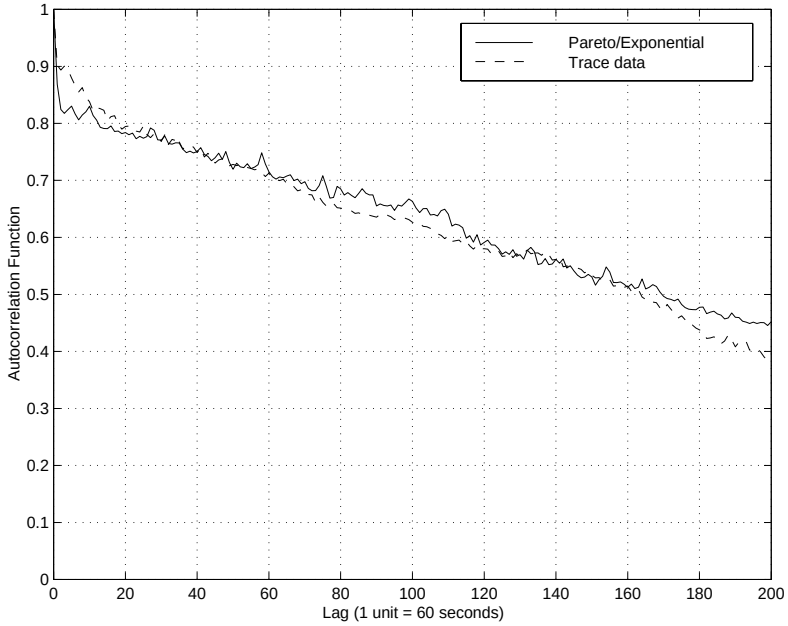

Figure 6.62: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'Phillips' and shown in Figure 6.10.

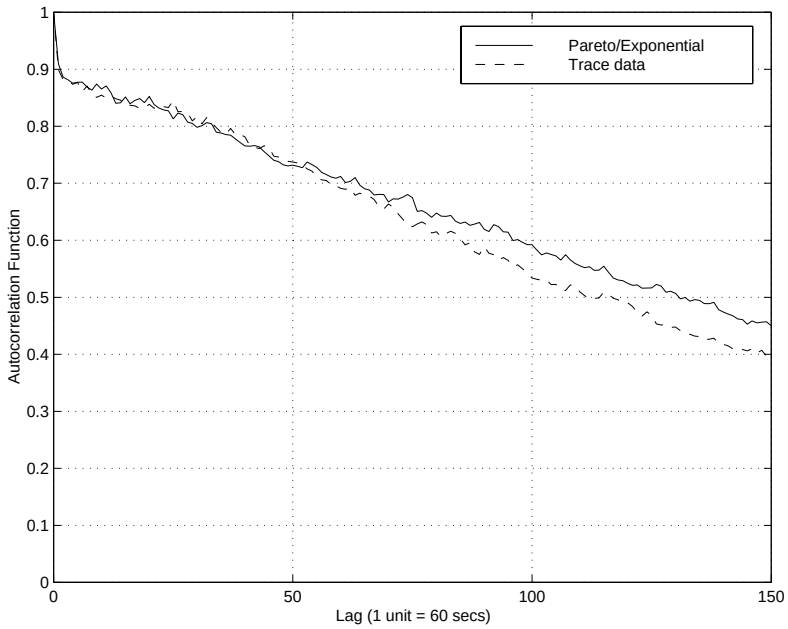

Figure 6.63: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'GSD' and shown in Figure 6.13.

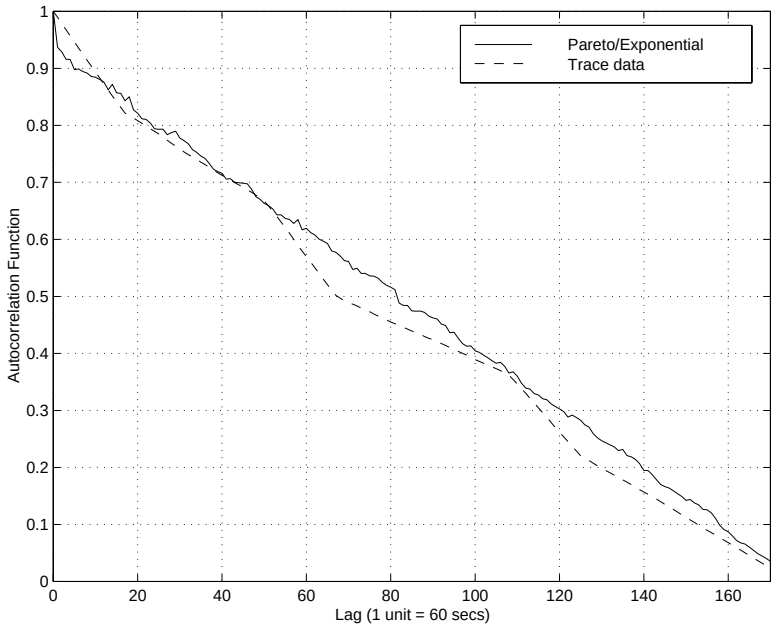

Figure 6.64: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'KU' and shown in Figure 6.16.

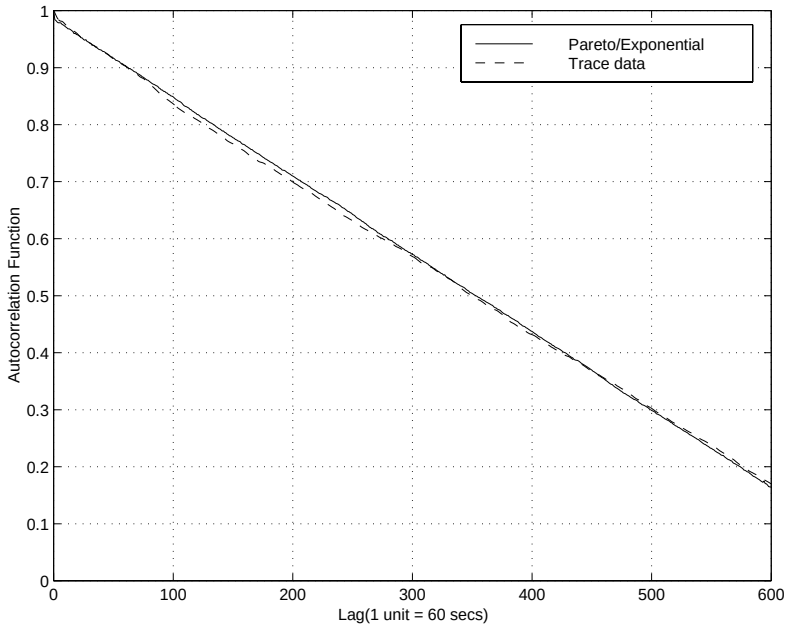

Figure 6.65: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'NRL2' and shown in Figure 6.19.

Chapter 7

Conclusions and Future work

7.1 Conclusions

A simple phase modulated model as well as a performance evaluation technique were developed in this study to predict hte delay and loss performance of ATM queues. One of the ways in which long-range dependence can manifest itself in network traffic is the burstiness associated with the traffic process. By capturing the burstiness of the network traffic in the form of traffic macro-dynamics, the model developed here, can predict ATM queue performance that can be expected in future networks that may be inherently long-range dependent in nature. Indeed, extensive trace driven simulations performed as a part of this study using collected data traces from the AAI ATM WAN, have shown that the simulation results match with the results predicted by the model. The experimental evaluation of the model is done in terms of mean cell delay and cell loss probability.

Using the developed model, the effect of traffic micro-scale dynamics on the queueing performance were investigated. The results obtained from the study of micro-dynamics indicate that the mean delay estimate is relatively invariant to the actual

micro-dynamics, in the presence of long-range dependence. However, the cell loss probability is found to be sensitive to micro-dynamics at low loads. At higher loads, the bursts on the macro-scale, dominate the cell loss in the queue and the micro dynamics play a minor role in the cell loss probability value obtained.

As demonstrated by the results obtained in Chapter 6, a theoretical Pareto distribution, with the appropriate shape parameter is a versatile distribution for modeling traffic in high speed networks.

It was also noted that, the logarithm of the cell loss probability has a linear relationship (see Figure 6.44) to load in the case of exponentially distributed inter-arrival times, but is non-linear in the case of real network trace data.

7.2 Future work

In this concluding section some recommendations for future work are proposed. The model developed here is a simple model which effectively captures the impact of burstiness in long-range dependent network traffic flows. However there is no provision for modeling any periodicity that may be present in the input process. Due to the lack of the periodic component, the model developed in this study is not accurate for modeling periodic data like MPEG encoded video data. The model can be modified in the future with the addition of this aspect.

As mentioned in Chapter 2, the CDF of the rate process in the model can be obtained from any theoretical infinite variance probability distribution. In this study a Pareto distribution was used. Other theoretical long-range dependent distributions should be evaluated.

The study of the influence of traffic micro-scale dynamics on cell loss was based on extensive simulations in this research. It will be of interest to obtain analytical

expressions, for predicting the effect of micro-dynamics on cell loss.

Also 'hybrid models' i.e., traffic models which the distribution of the micro-dynamics is phase dependent, can be easily constructed in the frame work of the model developed here. This will be of interest for cell loss estimates as it has been demonstrated in this study that the micro-dynamics have a significant impact on the cell loss results at low loads.

Appendix A

Simualtion model

In this sectoin the code for the discrete queue simulator is given.

```
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#include<sys/time.h>

void main(int argc, char *argv[]){

int j,flag=0,i;

char a[512];

double servrate=0, sampling_interval=0;
double slot=0,arrivals_pslot=0,slots_pint=0;
double ftime=0,frate=0;
double qlen=0,pqlen=0,aqlen=0,count=0;
```

```
double avrate,load;
double atint=0,tbase,tint,trand,tprev=0,aatint=0;
double no_slots=0,no_cells=0;
double aarrivals_pslot=0;
double des=0,u=0,beta,term;

FILE *fp, *FP, *Fp, *fid;

struct timeval tp;
struct timezone tzp;

/* Randomize based on the time of running the simulation*/
gettimeofday(&tp,&tzp);

fp=fopen(argv[1],"r");
FP=fopen(argv[2],"a");
Fp=fopen(argv[6],"w");
fid=fopen(argv[7],"w");

avrate=atof(argv[3]);
sampling_interval=atof(argv[4]);
load=atof(argv[5]);
servrate=avrate/load;

/* Randomize the seed here */
srand48(tp.tv_sec);
```

```
slot=1/servrate;
slots_pint=sampling_interval/slot;
printf("%g %g %g %g\n",avrate,sampling_interval,load,servrate);

while(fgets(a,512,fp) != NULL){

  sscanf(a,"%lf %lf",&ftime,&frate);

  no_cells=frate*sampling_interval;

  tbase=1/frate;
  beta=tbase/(tbase-1);
  beta= -1/beta;
  for(j=0;j<=slots_pint;j++){
    des=drand48();
       while (atint<=slot){
          u=drand48();
          tint=-log(u)*tbase;
          if(tint <0) puts("1:yes\n");
          atint+=tint;
          if (atint>slot) break;
          arrivals_pslot++;
          aarrivals_pslot++;
          if (aarrivals_pslot >= no_cells) {
             flag=1;
```

```
          break;
       }
       aatint+=tint;
    }
no_slots++;
 if(flag == 1){
    flag=0;
 } else {
    atint-=slot;
 }
 qlen=pqlen+arrivals_pslot;
 count++;
 aqlen+=qlen;
 qlen--;
 if (qlen < 0) qlen=0;
 pqlen=qlen;
      while(atint>slot){
          flag=1;
          atint-=slot;
          if (atint<0) break;
          j++;
          count++;
          aqlen+=qlen;
          qlen--;
          if (qlen < 0) qlen=0;
          pqlen=qlen;
```

```
            no_slots++;
        }
        if(flag==1){
              flag=0;
              atint+=slot;
        }
        /*if(atint<0) puts("yes\n");*/
        arrivals_pslot=1;
        aarrivals_pslot++;

/* This slot ends */
if (aarrivals_pslot>=no_cells || no_slots>=slots_pint){
        if(slots_pint-no_slots>0) {
              term=slots_pint-no_slots;
              count+=term;
              for(i=1;i<term;i++){
                 aqlen+=qlen;
                 qlen--;
                 if (qlen<0) qlen=0;
                 pqlen=qlen;
              }
        } else {
              /* Add the remaining cells, if any at the end*/
              qlen+=no_cells-aarrivals_pslot;
              pqlen=qlen;
        }
```

```
          fprintf(Fp,"%g %g %g %g %g  %g\n",aatint,aarrivals_pslot,no_c
ots,slots_pint,count);
          aarrivals_pslot=1;
          aatint=0;
          no_slots=0;
          break;
     }
   }
}

printf("%g %g %g\n", aqlen, count, aqlen/count);
fprintf(FP,"%g %g %g\n", load,aqlen/count,aqlen/(count*servrate));

fclose(fp);
fclose(FP);

}
```

Bibliography

[1] J. N. Daigle, Y.Lee, ”Discrete Time queues with phase dependent arrivals”, IEEE Trans. Commun., April 1994.

[2] J. N. Daigle and D. M. Lucantoni, ” Queueing system having phase dependent arrivals and service rates”, Dekker 1994.

[3] J. N. Daigle, ”Queueing Theory for Telecommunications” , Addison-Wesley Publishing, 1992.

[4] S. Q. Li, ”A general solution technique for discrete queuing analysis of multi-media traffic in ATM”, IEEE Trans. Commun., July, 1991.

[5] M. F. Nuets, ” Matrix Geometric solutions in Stochastic Models”, The John Hopkins University Press, 1981.

[6] M. F. Nuets, ”Structured Stochastic Matrices of the M/G/1 type and Their Applications”, Marcel Dekker Inc. 1989.

[7] V. Paxson and S. Floyd, ”Wide Area Traffic: The failure of Poisson Modeling”, IEEE/ACM Trans.Networking, June 1995.

[8] A. Erramilli, O. Narayan, W. Willinger, ”Queueing analysis with long-range dependent traffic”, IEEE/ACM Trans.Networking, April 1996.

[9] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, ”On the self-similar nature of Ethernet Traffic”, Proc.Acm Sigcomm’93.

[10] A. Shwartz and A. Weiss, ”Large deviations for Perforamnce Analysis”, Chapman and Hall, 1994.

[11] J. A. Bucklew, ”Large Deviation Techiniques in Decision, Simulation and Estimation”, Wiley, 1990.

[12] P. R. Jelenkovic and A. A. Lazar, ”On the dependence of the Queue Tail Distribution on Multiple Time Scales of ATM Multiplexers”, Proceedings of CISS, John Hopkins University, 1995.

[13] P. R. Jelenkovic and A. A. Lazar, ”Multiplexing On-Off sources with Subexponential On periods: Part I, IEEE INFOCOM ’97, April 1997.

[14] M. W. Garett and M. Vetterli, ”Congestion control Strategies for Packet Video”, Fourth International Workshop on Packet video, Japan, 1991.

[15] T. G. Robertazzi, ”Computer Networks and Systems”, second-edition, Springer-Verlag, 1995.

[16] M. Jeruchim, P. Balaban, K. S. Shanmugan, ”Simulation of Communication Systems”, New York: Plenum, 1992.

[17] S. Q. Li and C. L. Hwang, ”Queue Response to input correlation functions: Discrete spectral analysis”, IEEE/ACM Trans.Networking, 1993.

[18] D. R. Cox, ”Long-Range dependence: A Review”, Iowa State Univ. Press, 1984.

[19] S. Muppidi and V. S. Frost, ”Delay analysis of Network Traffic using collected Data Traces”, ITTC Technical Report, ITTC-TR-10980-17, March 1997.

[20] L. A. Kulkarni and S. Q. Li, ”Measurement Based Traffic Modeling”, TISE Technical Report.

[21] S. Muppidi, V. S. Frost, J. B. Evans, ”Traffic Flow Patterns in the AAI Network: January - June 1996”, TISL Technical Report, TISL-TR-10980-15, August 1996.

[22] Anthony Alles, ”ATM Internetworking”, Engineering InterOp, Las Vegas, March 1995.

[23] A. Papoulis, ” Probability, Random variables, and Stochastic processes”, Third Edition, Mc. Graw Hill, Inc., 1992.

[24] I. Norros, ”Studies on a model for connectonless traffic based on fractional Brownian motion”, 1992.

[25] A. Erramilli and R. P. Singh, ”Chaotic maps as models for packet traffic”, in Proc. 14th ITC, France 1994.

[26] H. Heffes and D. M. Lucantoni, ”A Markov modualted charecterization of packetized voice and data traffic and related statistical multiplexer performance”, IEEE J. Select. Areas Commun., 1984.

[27] W. Willinger, M. S. Taqqu, R. Sherman, D. V. Wilson, ”Self-similarity through high variability: Statistical Analysis of Ethernet traffic at Source level”, Proc. ACM Sigcomm’95, 1995.

[28] G. Samoradnidski and M. S. Taqqu, ”Stable Non-Gaussian Random Processes”, Wiley eastern.

[29] V. Paxson, ”Emperically derived analytical models of wide area TCP connections: Extended report”, IEEE/ACM Transactions on Networking, vol. 2, no. 4, August 1994.

[30] V. Paxson, ”Growth trends in wide-area TCP connections”, available as ftp.ee.lbl.gov:/WAN-TCP-growth-trends.ps.Z.

[31] A. Erramilli and J. Wang, ” Monitoring Packet levels in networks”, in Proc. IEEE Globecom’94, 1994.

[32] A. M. Breiphol and K. S. Shanmugan, ”Random Signals Detection, Estimation and Data Analysis”, New York: Wiley 1992.

[33] N. Shroff and M. Schwartz, ” Improved loss calculations at an ATM Multiplexer”, IEEE Infocom ’96, San Francisco, March ’96.

[34] Martin De Pyker, ” Asynchronus Transfer Mode Solution for Broadband ISDN”, Second Edition, Ellis Horwood, London, 1995.

[35] Douglas E. Comer, ”Internetworking with TCP/ IP, Volume 1”, Prentice- Hall Publications, 1995.

[36] Dirk Wisse, Jan Voorschot, ”Tricklet V6.0”, ftp://dnpap.et.tudelft.nl/pub/btng.

[37] Raj Jain, ” The art of computer system performance evaluation”, Wiley Eastern, 1995.