MODELING AND ANALYSIS OF TRAFFIC IN HIGH SPEED NETWORKS

ACTS ATM Internetwork Project

Information and Telecommunication Technology Center
University of Kansas.

Organization

- Motivation

  Failure of Classical Models.

- Introduction.

  Long Range Dependence.

- Analytical Model.

  Macro-dynamics

  Micro-dynamics

- Performance Analysis Methodology.

  Mean Cell delay.

  Cell loss probability.

- Data collection process and the AAI Network.

Organization (contd.)

- Simulation Model.
  - Model Description.
  - Validation.
- Experimental Evaluation.
  - Mean Cell delay and Cell Loss Probability results.
  - Traffic Micro-dynamics.
  - Sensitivity of number of phases.
  - Second-order statistics.
- Conclusions.
- Future Work.

Motivation

- Traffic studies.
- Queueing performance of a process with infinite variance.
- Switch buffers get filled up faster than those predicted by conventional models
- Example:

    Mean cell delay:

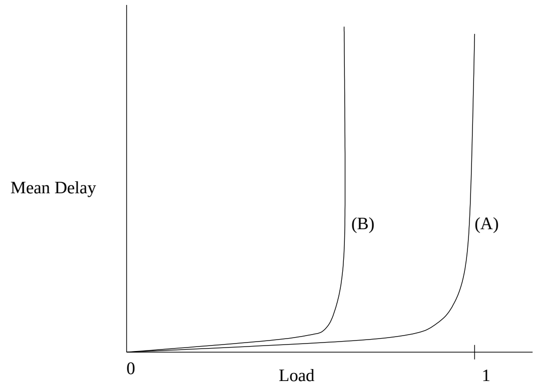


Figure 1: Typical Delay curves for long-range dependent model (B) and conventional model (A).

Motivation (contd.)

- Example:

  Probability of Cell Loss:

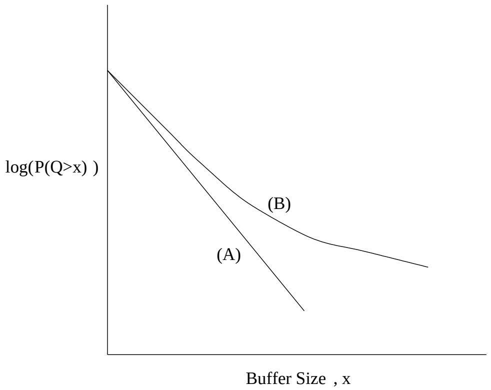


Figure 2: Typical loss curves for long-range dependent model (B) and conventional model (A).

- Traffic modeling and performance prediction required for efficient operational algorithms.

Introduction

- Self similar traffic models:

  FBM, ARIMA, Modeling with Chaotic Maps.

- Definition of self similarity:

  Let $X = (X_t,\ t = 0,\ 1,\ 2...)$ be a covariance stationary process

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

  Exactly:

$$r^{(m)}(k) = r(k), k \geq 0 \tag{2}$$

  Asymptotically:

$$r^{(m)}(k) \to r(k), m \to \infty \tag{3}$$

Introduction (contd.)

- Ramifications

  Non-summable autocorrelation function.

  Divergent Power-Spectrum at the origin.

- Implication

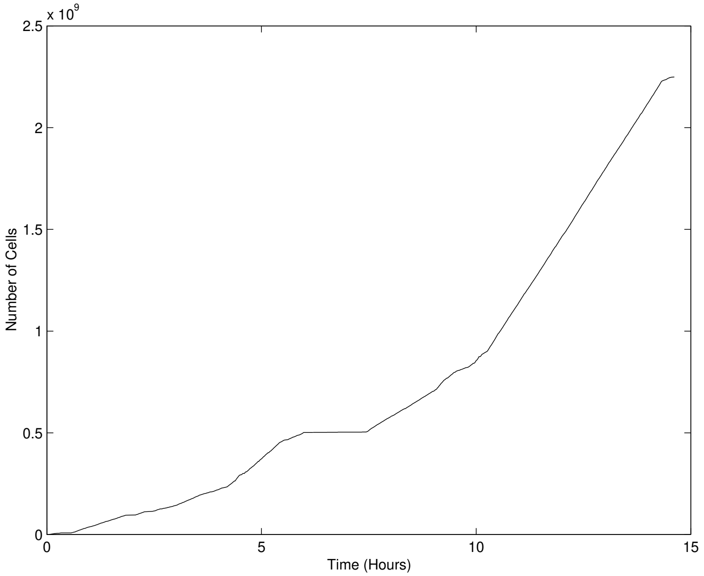
  Burstiness of Aggregate traffic.

- Origin

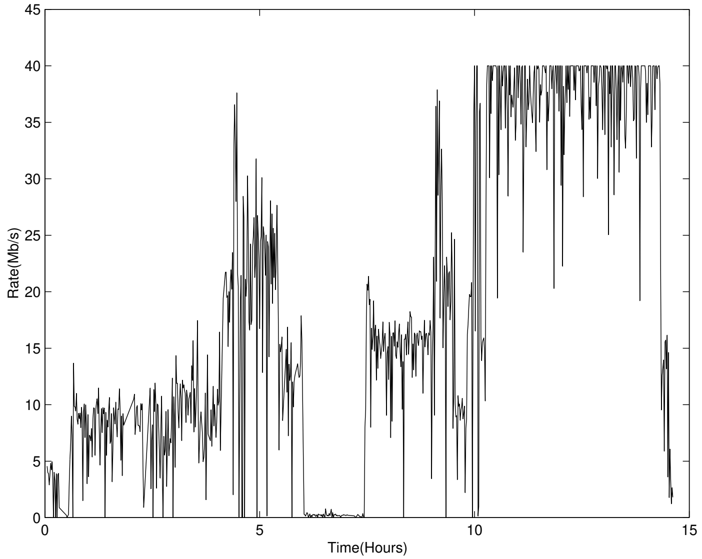
  ON-OFF source frame-work with ON and OFF periods that follow a distribution with infinite variance.

Traffic Model

- Definition: Rate Process: A random process $R(t)$, which is the short-term time average of the arrival process $A(t)$.



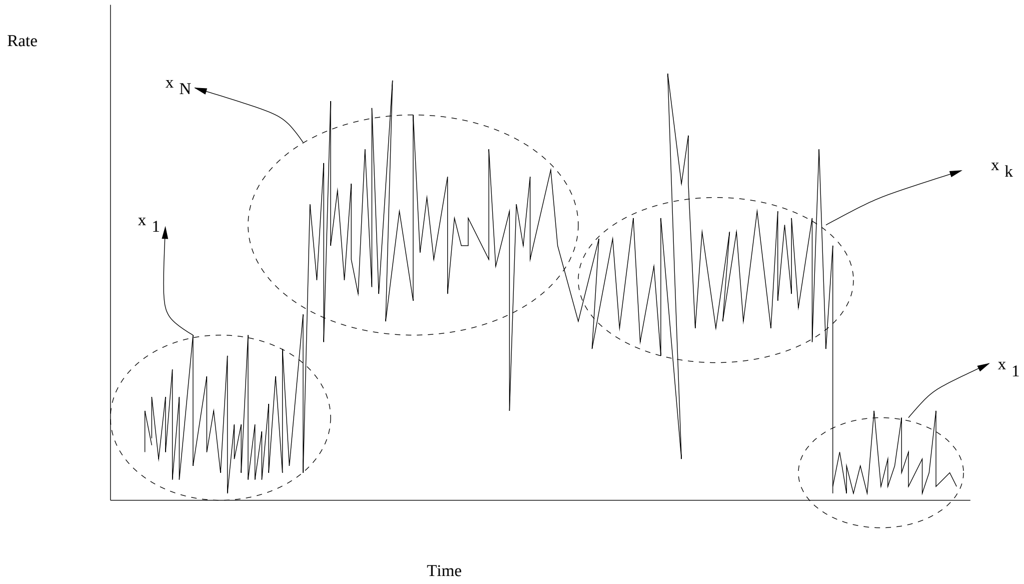
Cell count process



Rate process.

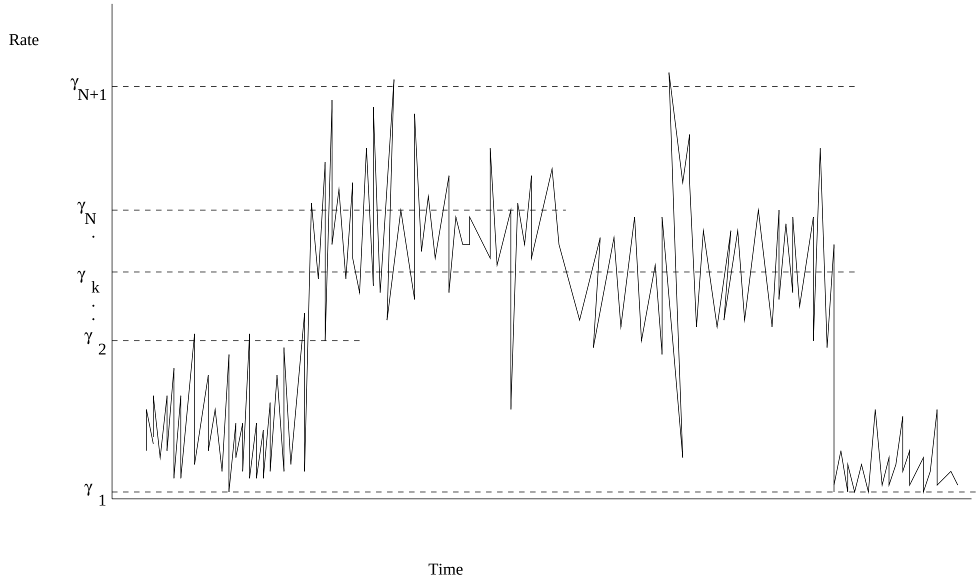
Traffic Model (contd.)

- Model the Macro-dynamics and Micro-dynamics of the Rate Process.
- Rate process is modeled as a non-Markovian , stationary and ergodic phase-process having a finite state space $\mathbf{S}=\{x_1, x_2, ...x_N\}$.



Traffic Model (contd.)

- The rate vector $\bar{\gamma} = [\gamma_1, \gamma_2, ..., \gamma_{N+1}]$ represents the boundary rates for the states and $\gamma_1 \leq \gamma_1 \leq .... \leq \gamma_{N+1}$.



Traffic Model (contd.)

- The probability vector $\bar{\pi} = [\pi_1, \pi_2, ..., \pi_N]$ denotes the steady-state probability vector of the phase process.
- The steady state phase probabilities of the phase process are assumed to follow a distribution with a heavy tail.
- The rates in $\bar{\gamma}$ and the probability vector $\bar{\pi}$ define the macro-dynamics.

Traffic Model (contd.)

- $\bar{\gamma}$ is obtained by partitioning the range $[\gamma_{min},\gamma_{max}]$ where γ_{min} and γ_{max} represent the minimum and maximum rates respectively.
- The rate vector $\bar{\gamma}$ for N phases is:

$$\gamma_1 = \gamma_{min} \tag{4}$$

$$\gamma_i = \gamma_1 + \frac{\gamma_{max} - \gamma_{min}}{N}, i = 2, 3, 4, ..., N \tag{5}$$

$$\gamma_{N+1} = \gamma_{max}. \tag{6}$$

Traffic Model (contd.)

- $\bar{\pi}$ is obtained from the CDF of $R(t)$.
- The CDF of $R(t)$ can be obtained from a theoretical infinite variance distribution. Here, a theoretical Pareto Distribution was used.

  A Pareto Distribution can be given as:

$$F_X(x) = 1 - Kx^{-\beta}, \;\; \beta > 1, x > K. \tag{7}$$

The maximum likelihood estimate of the shape parameter β, for a given set of data samples $D = \{d_1, d_2, .., d_M\}$ is given as

$$\beta = \frac{1}{\frac{1}{M}\Sigma_{i=1}^{M} \, ln(d_i)} \tag{8}$$

Traffic Model (contd.)

- The CDF can also be obtained from an empirical CDF of the rate process.
- The steady-state probability π_i for state x_i, which is bounded by rates γ_i and γ_{i+1} is computed as:

$$\pi_i = P[X \leq \gamma_{i+1}] - P[X \leq \gamma_i], \;\; i = 1, 2, 3 \ldots N - 1$$

$$= F_X(\gamma_{i+1}) - F_X(\gamma_i), \;\; i = 1, 2, 3 \ldots, N - 1 \tag{9}$$

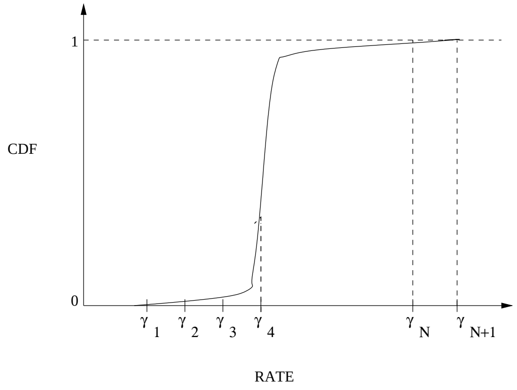


Figure 3: Quantizing the CDF of $R(t)$.

Traffic Model (contd.)

- Within each state x_i, with associated probability π_i, the arrival process is modeled as a point process with a with finite mean γ_{i+1} and finite variance.
- The distribution function defines the traffic micro-dynamics of that state.
- Assuming γ_{i+1} as the mean rate in state x_i is a conservative assumption as it is the upper bound on the rates in each state.

Performance Analysis Methodology

- Let Z denote the random variable associated with a performance parameter of a queue with $R(t)$ as the input process.

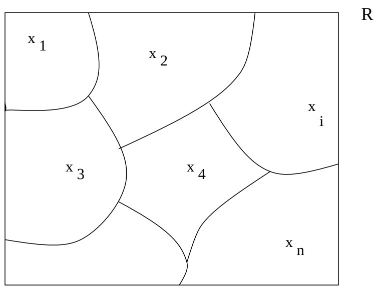


Figure 4: Concept for performance analysis methodology.

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

$$E[Z] = \sum_{i \in S} \pi_i E[Z|S = x_i] \tag{10}$$

- In the above equation the traffic macro-dynamics are described by the values of $\bar{\pi}$ and the micro-dynamics are represented by $E[Z|S = x_i]$.

Performance Analysis Methodology (contd.)

- Performance prediction in terms of Mean Cell Delay and Cell loss Probability.
- Notation:
  Let

  Slotted system, infinite buffer.

  n denote the number of cells in the system at a given time.

  $P_n^i(j)$ 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.

  p_k^i denote the probability that there are k arrivals to the system when the input process is in state x_i.
- Queue dynamics are described by:

$$P_n^i(j+1) = \sum_{k=0}^{n+1} P_k^i(j) p_{n-(k-1)^+}^i \qquad (11)$$

- $(x)^+$ represents the maximum of $\{0, x\}$.

Performance Analysis Methodology (contd.)

- Mean-Delay:

  Slotted-G/ D/ 1 system, analysis based on Probability Generating Function (PGF).

  Average cell delay given as:

$$E[D] = \frac{1}{\lambda}\sum_{i=1}^{N}\frac{\pi_i}{1-\rho_i}\{\frac{1}{2}\sigma_i^2 + \frac{3}{2}\rho_i^2\} \tag{12}$$

where

$\lambda = \Sigma_{i=1}^{N} \pi_i \gamma_{i+1}$ 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 $S = x_i$

Performance Analysis Methodology (contd.)

- Cell loss Probability

  Let $P_L[K|S = x_i]$ denote cell loss probability in a finite buffer of size K, given that the input process is in state x_i.

- $P_L[K|S = x_i]$ is approximated as $P[Q > K|S = x_i]$

$$P_L[K|S = x_i] \approx P[Q > K] \tag{13}$$

- Analysis of Cell loss for a General arrival process to a slotted system is hard. Assume that the micro-dynamics are exponential.

$$P_L(K) = \sum_{i=1}^{N} \pi_i (1 - \sum_{k=0}^{K} \frac{(1-\rho_i)p_k^i + \Sigma_{n=1}^{i} \{1 - \Sigma_{m=0}^{n} p_m^i\} P_{k-n}^i}{p_0^i}) \tag{14}$$

$$p_n^i = \frac{e^{-\rho_i} {\rho_i}^n}{n!} \tag{15}$$

Summary of Methodology

- Obtain $(\gamma_{min}, \gamma_{max})$.
- Obtain the CDF of the rate process. The CDF can be obtained as a
  - theoretical distribution, for example using β, the shape parameter of a Pareto distribution.
  - or can be calculated from a given collected data trace.
- Quantize the CDF into 'N' levels, where level i is represented by rate γ_{i+1} in the vector $\bar{\gamma}$.
- 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).
- Given the linearity property of expected value $E[Z] = \Sigma_{i \in S}\, \pi_i E[Z|S = x_i]$

Data Collection Process and the AAI.

- Performance Analysis is based on Trace data collected from the AAI network.
- Configuration:

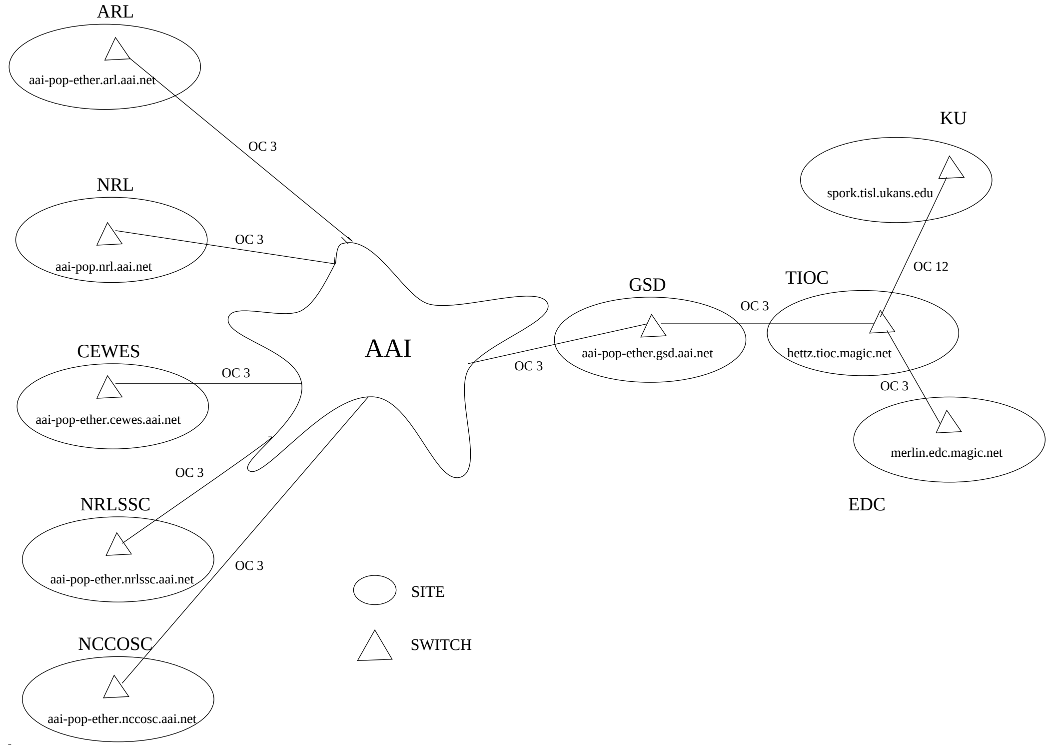


Figure 5: Connections of the sites and switches being sampled.

Data Collection Process and the AAI(contd.)

- Data collected from several edge switches using Simple Network Management Protocol (SNMP).
- Total Data collected: $\approx$ 1.8 Gbytes.
- Sampling interval is approximately 60 seconds.
- Data was re-sampled using linear interpolation so that sampling interval is exactly 60 seconds.

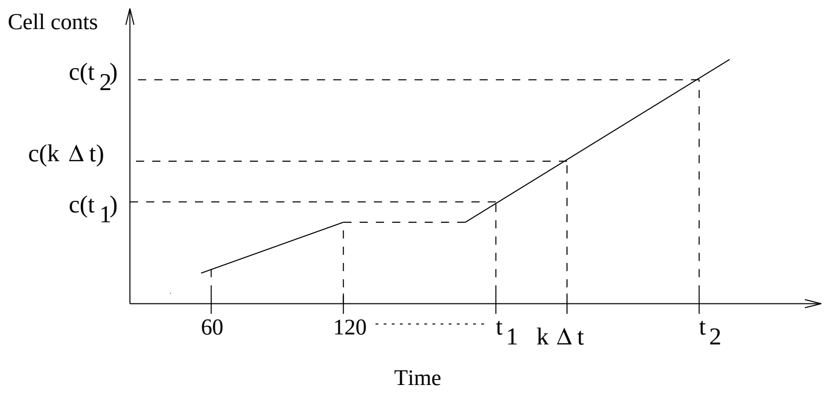


Figure 6: Re-sampling by linear interpolation of cell counts.

Data Collection Process and the AAI (contd.)

- Re-sampling:

$$c(k\Delta t) = c(t_1) + \frac{c(t_2) - c(t_1)}{t_2 - t_1} * (k\Delta t - t_1), \quad k = 1, 2, 3, ... \tag{16}$$

  where

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

  Δt is the time interval after re-sampling and is 60 seconds.

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

- The rate process $R(t)$ can now be obtained as:

$$R(t) = \frac{c(t+60) - c(t)}{60} \tag{17}$$

Simulation Model

- Queueing Model is an infinite buffer with a deterministic server.
- Time is slotted with a single cell served at the end of the slot.
- Simulation based on calculating the "unfinished work" in the queue at the end of a slot, from the equation:

$$n(k+1) = max(0, n(k) + a(k) - 1). \tag{18}$$

where

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

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

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

Simulation Model (contd.)

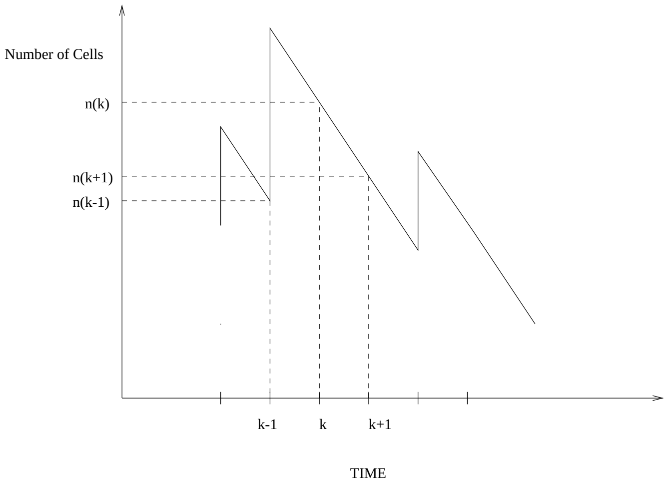


Figure 7: Delay estimation from cell counts

- Mean Cell delay is given by:

$$\bar{\tau} = \frac{1}{M} \sum_{k=1}^{M} \frac{n(k)}{\mu} \qquad (19)$$

Simulation Model (contd.)

- In Equation (19)

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

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

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

- Probability of Cell loss:

  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.

$$P_L(x) = P[n(k) > x] \tag{20}$$

Simulation Model (contd.)

- Validation: Simulator validated for Mean Delay and Cell loss.
  - By comparing simulation results with those predicted by standard theoretical results.
  - Source which generates cells with exponentially distributed inter-arrival times with a known average rate has been constructed.
  - Exponential inter-arrivals generated using Transform Method.

Simulation Model (contd.)

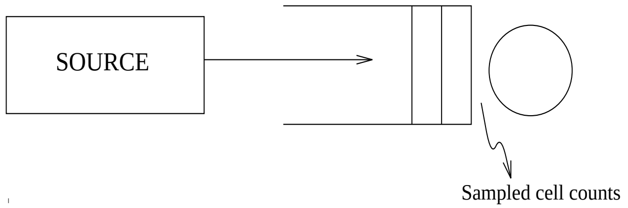


Figure 8: Validation of Model

- Output of the source sampled at constant time intervals of T_S seconds - Simulates the data collection process.
- T_S=60 seconds as in the case of collected data.
- Queueing system is a M/ D/ 1 system for which theoretical results are available.
- Considering the statistical nature of the simulation, simulation results closely agree with the experimental results.

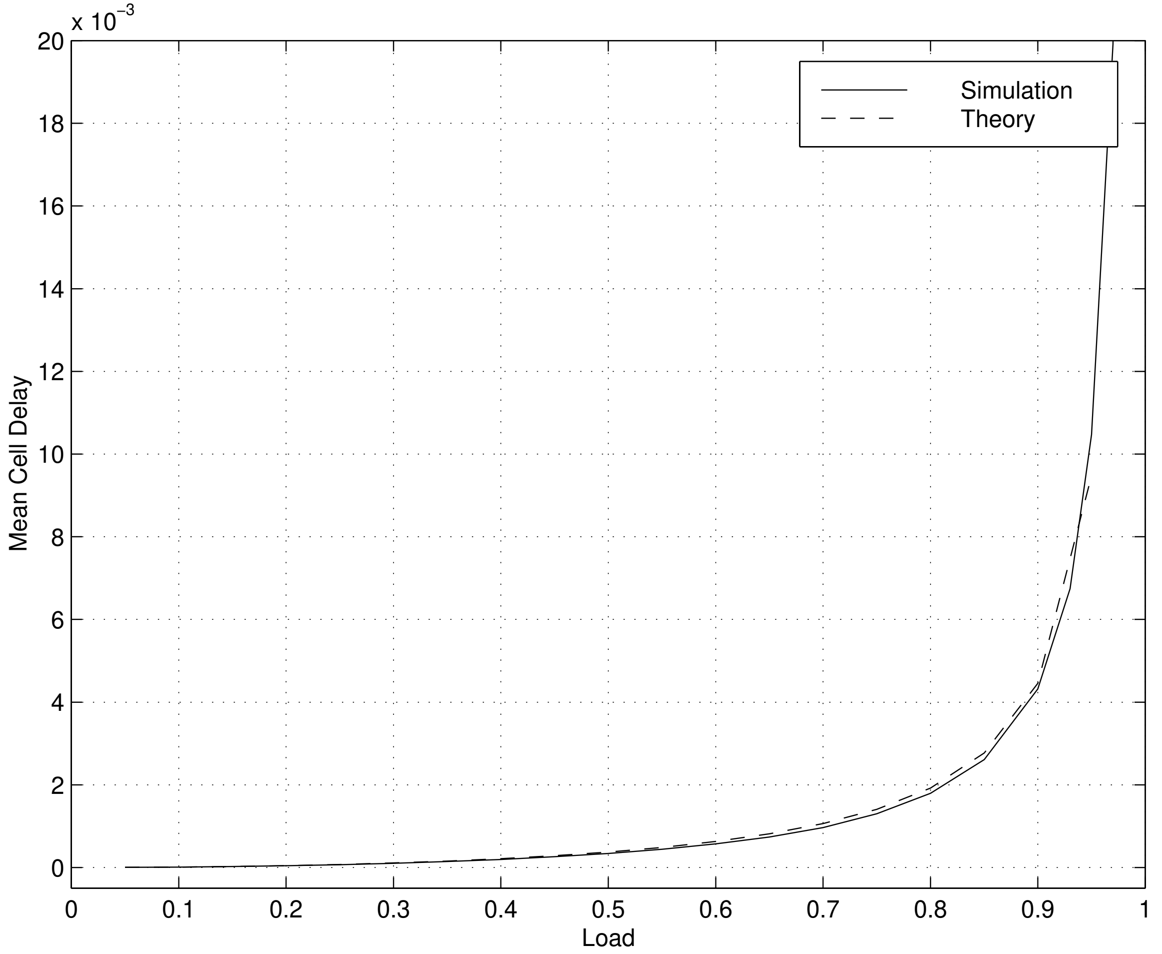


Figure 9: Validation of the simulator for mean cell transfer delay results.

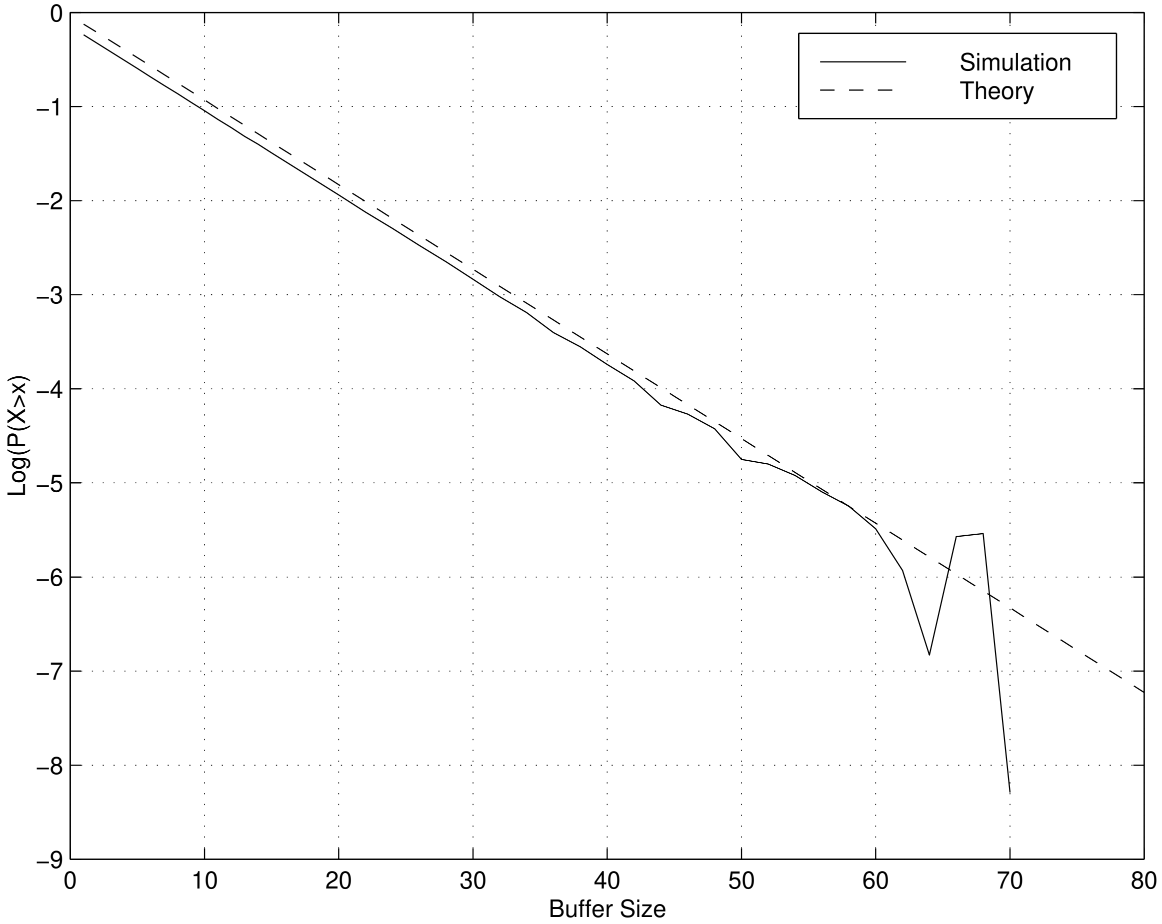


Figure 10: Validation of the simulator for Cell loss ratio results.

Experimental Evaluation

- Experimental Evaluation done by comparing the performance predicted by the model, with the simulation results obtained from the AAI network.
- Traces used:

Table 1: Data traces used for model validation

Trace Name	Duration (Hours)	# cells	Characteristic	Mean Rate(Mb/s)
NCCOSC	25	3.02697e08	Background	1.55795
Phillips	15	2.227e09	SC '95	20.3064
NRL	6	5.1054e08	EMMI	10.307321

- Traces chosen such that different traffic levels are used.
- Trace Types:
  - Background: Management Flows, routing updates etc.
  - EMMI: Multimedia traffic profiles.
  - SC '95 flows: Bursty application traffic profiles typical in wide area ATM networks.

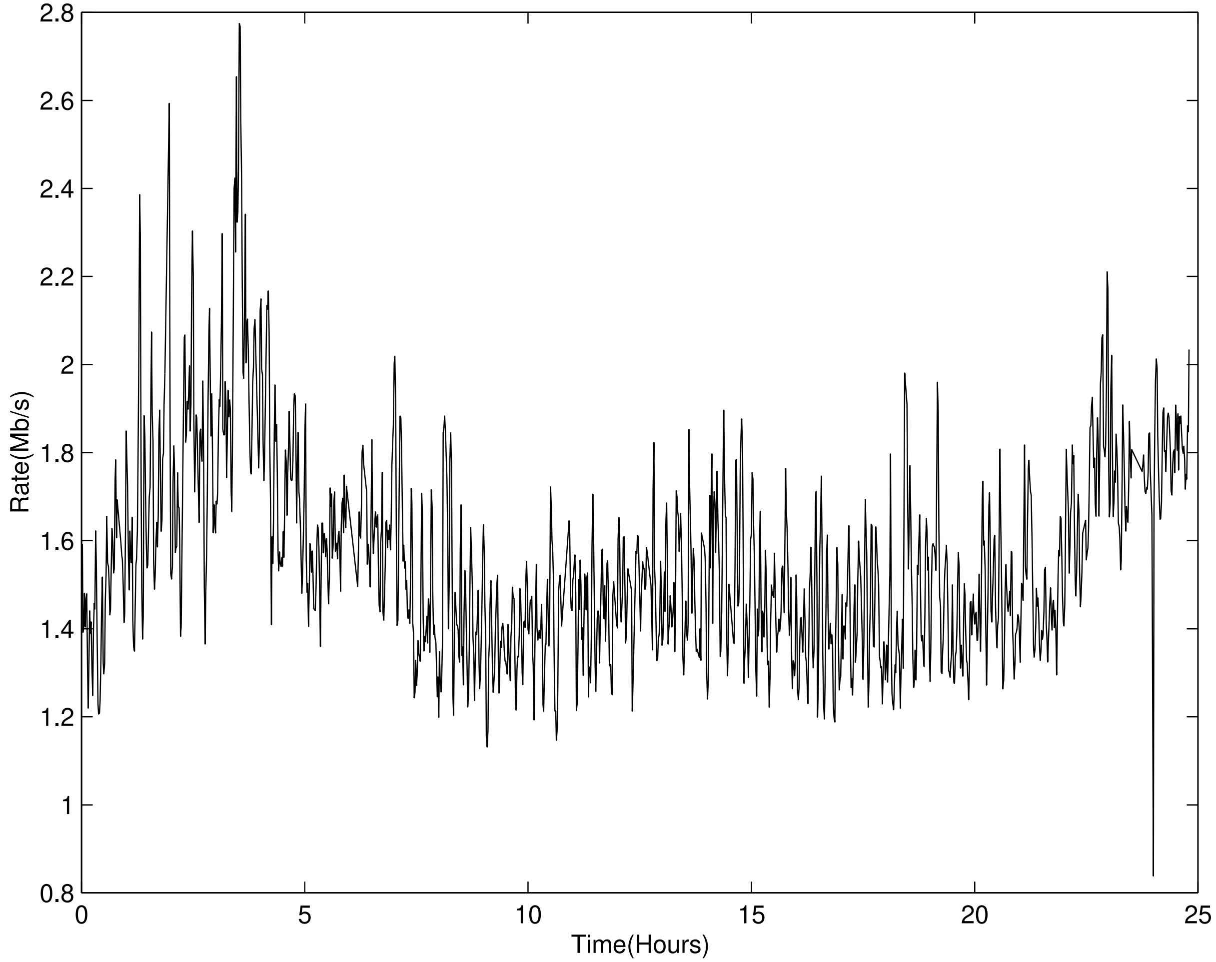


Figure 11: Data Collected from NCCOSC site.

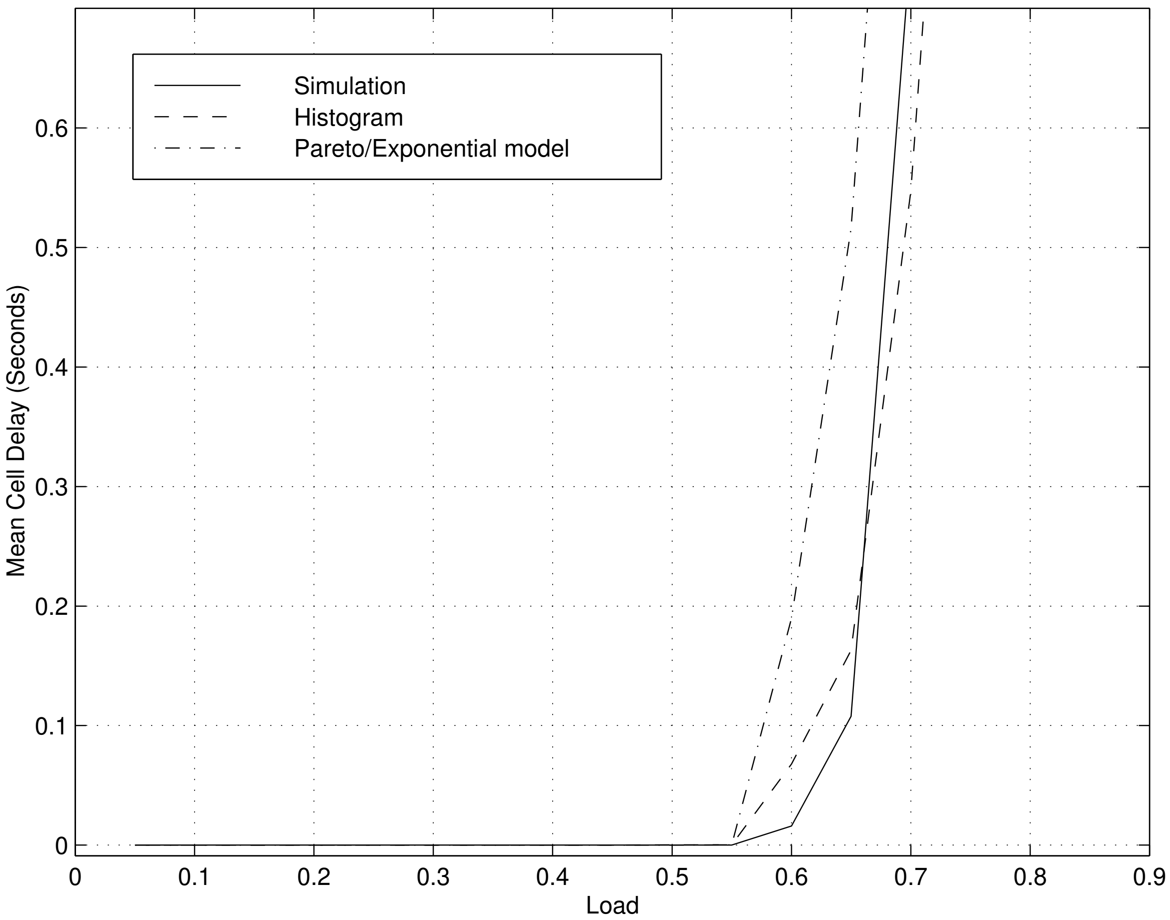


Figure 12: Mean Cell Delay estimate obtained from theory ($\beta = 8.2$), histogram and simulation for the trace labeled 'NCCOSC' and shown in Figure 11, using $N = 15$ input phases.

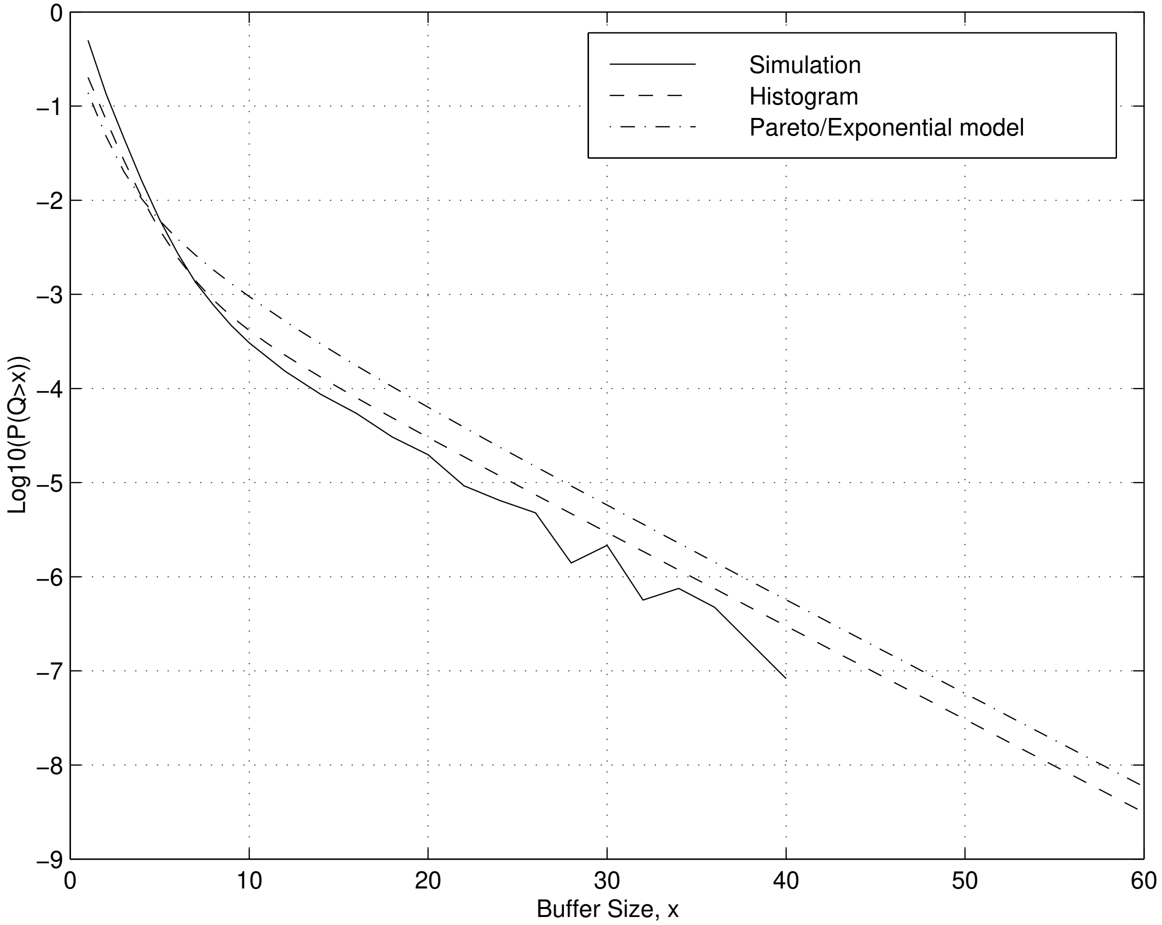


Figure 13: Comparison of cell loss probability estimates obtained from theory ($\beta = 8.2$), histogram and simulation of the collected data trace labeled NCCOSC and shown in Figure 11, using $N = 15$ input phases. $\rho = .4$.

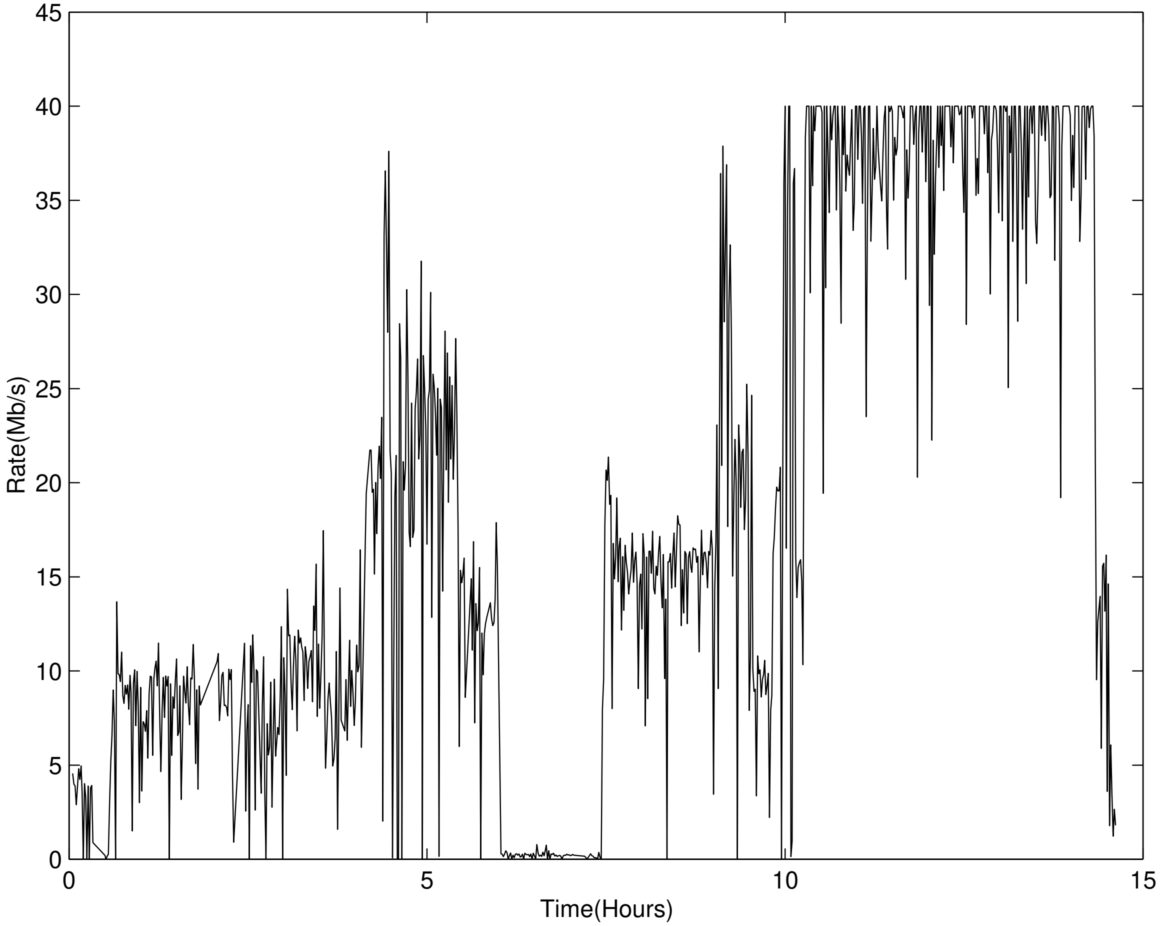


Figure 14: Data trace Collected from the Phillips site.

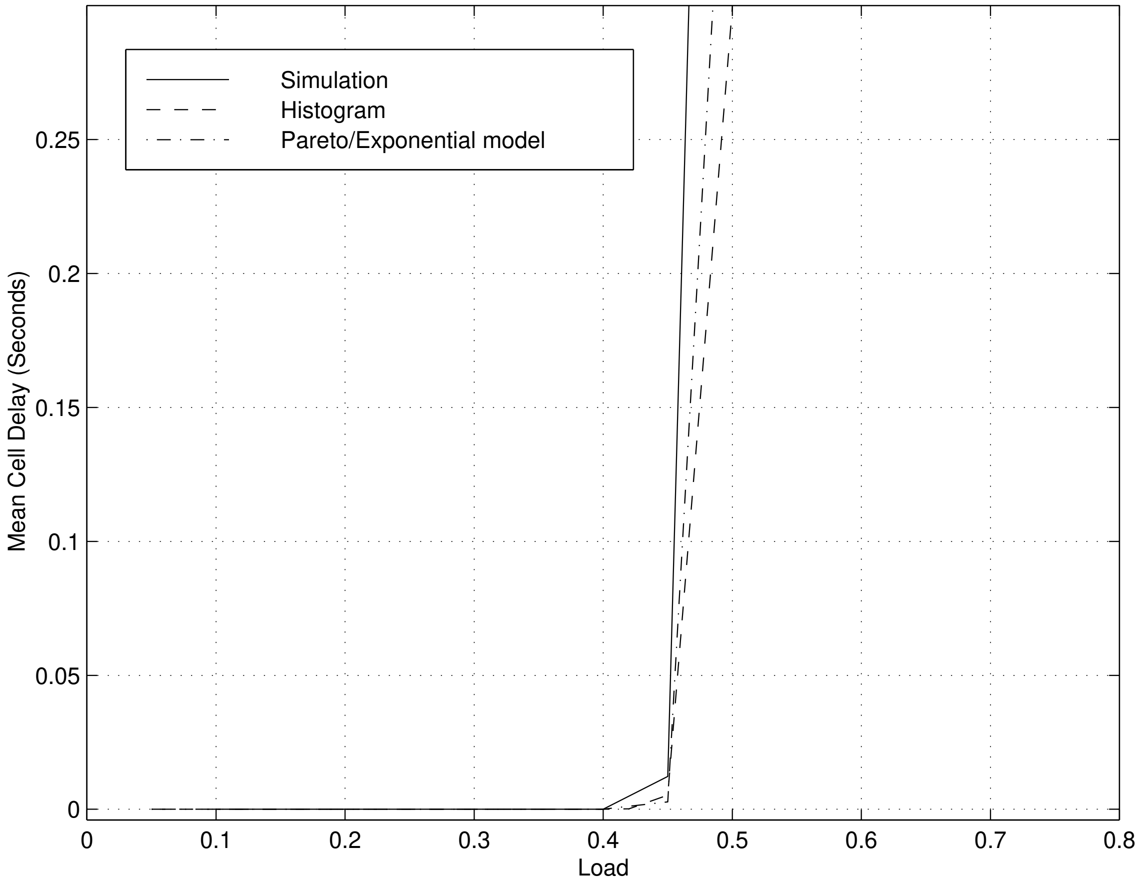


Figure 15: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 1.2$), histogram and simulation of the collected data trace labeled Phillips and shown in Figure 14, using $N = 15$ input phases.

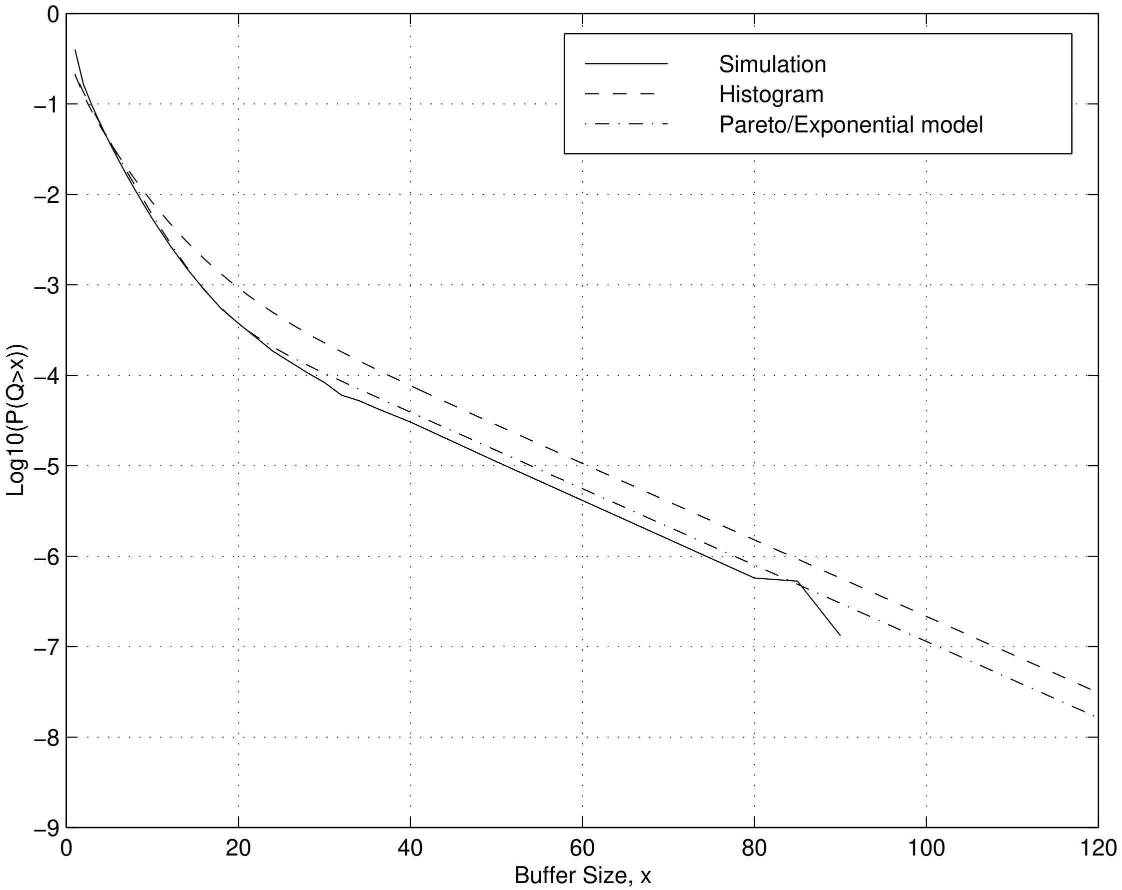


Figure 16: Comparison of cell loss probability estimates obtained from theory ($\beta = 1.2$), histogram and simulation of the collected data trace labeled Phillips and shown in Figure 14, using $N = 15$ input phases. $\rho = .4$.

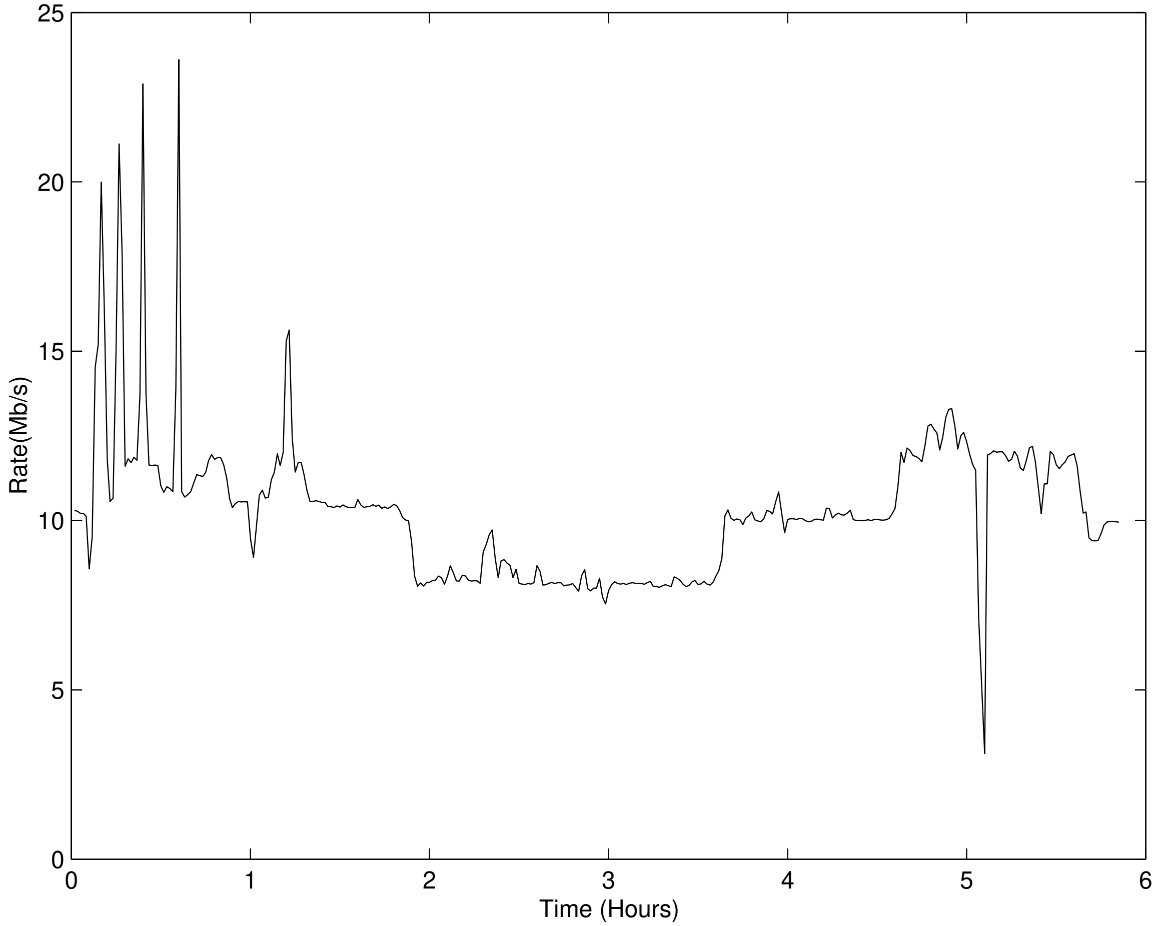


Figure 17: Data trace Collected from the NRL site.

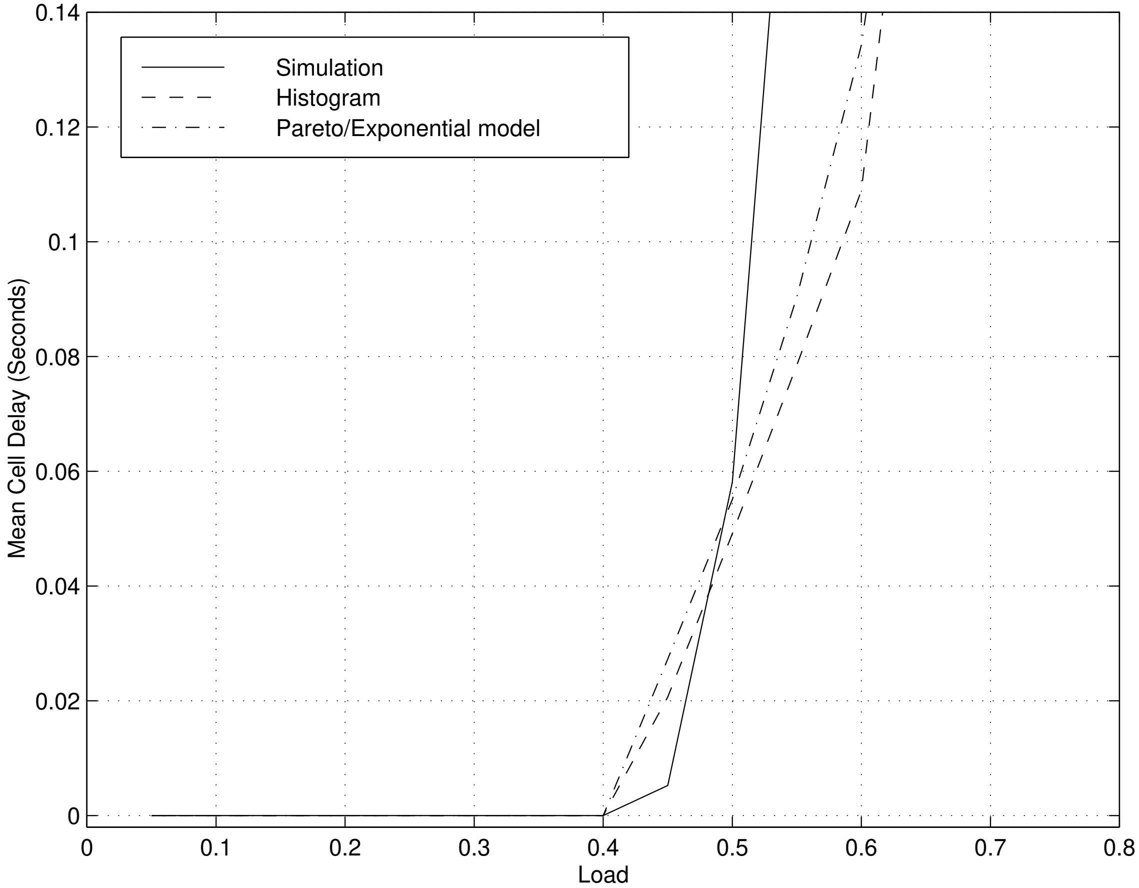


Figure 18: Comparison of Mean Cell Delay estimates obtained from theory ($\beta = 2.2$), histogram and simulation of the collected data trace labeled NRL and shown in Figure17, using $N = 15$ input phases.

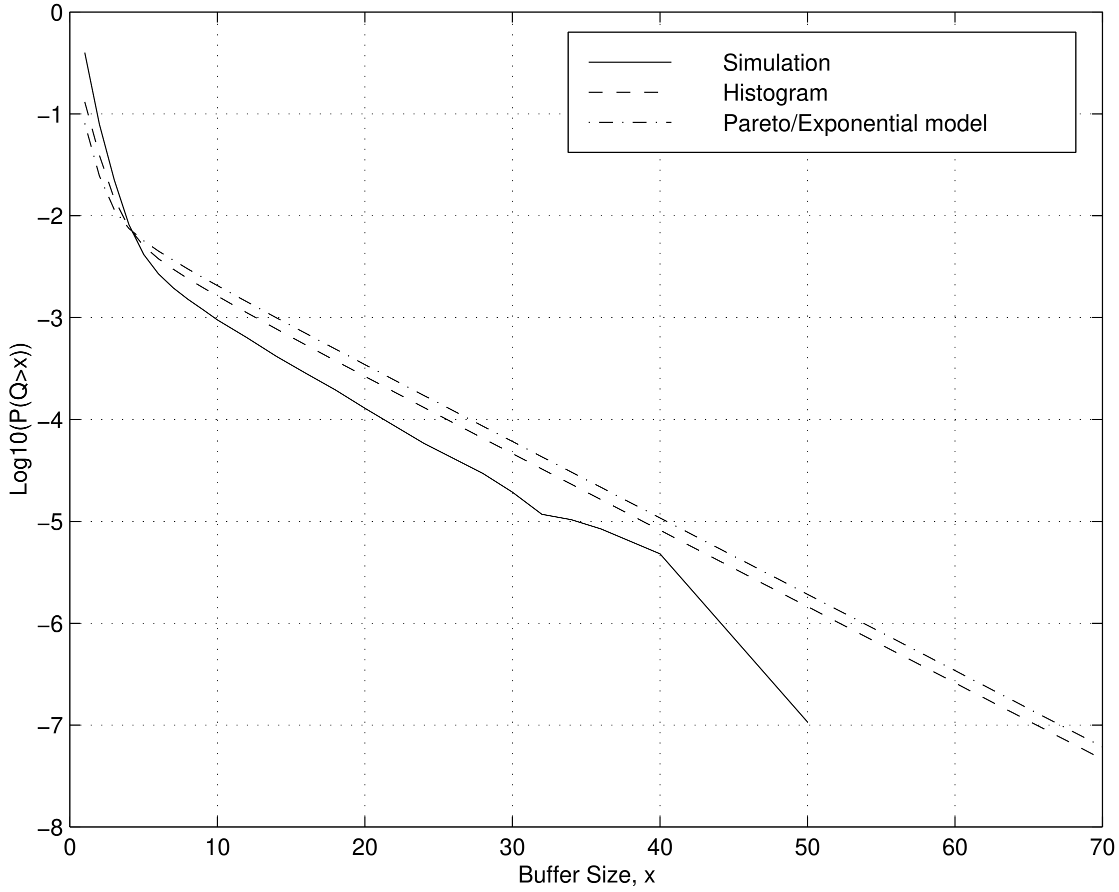


Figure 19: Comparison of cell loss probability estimates obtained from theory ($\beta = 2.2$), histogram and simulation of the collected data trace labeled NRL and shown in Figure 17 using $N = 15$ input phases. $\rho = .4$.

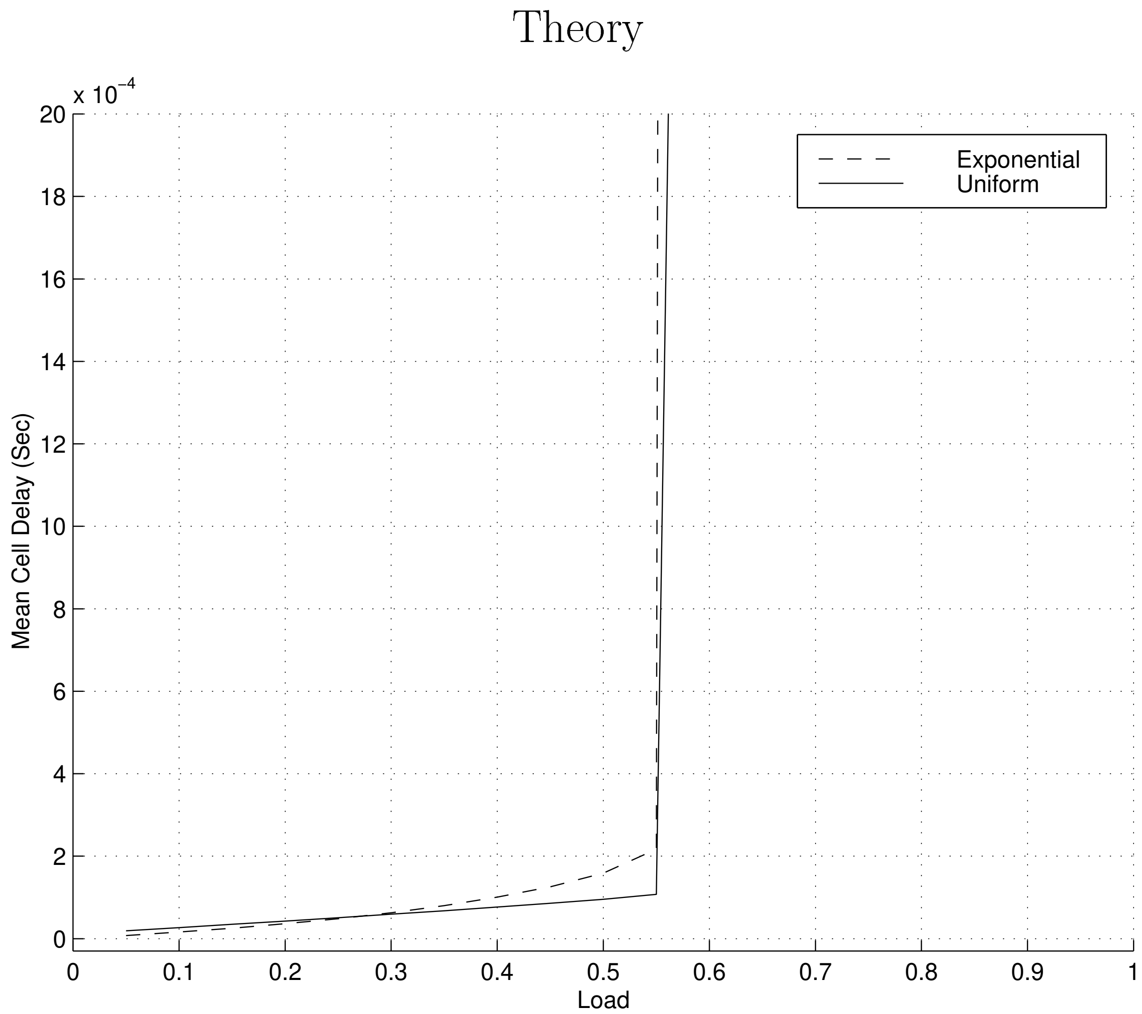


Figure 20: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 8.2$) for the trace labeled 'NCCOSC' and shown in Figure 11, using $N = 15$ input phases.

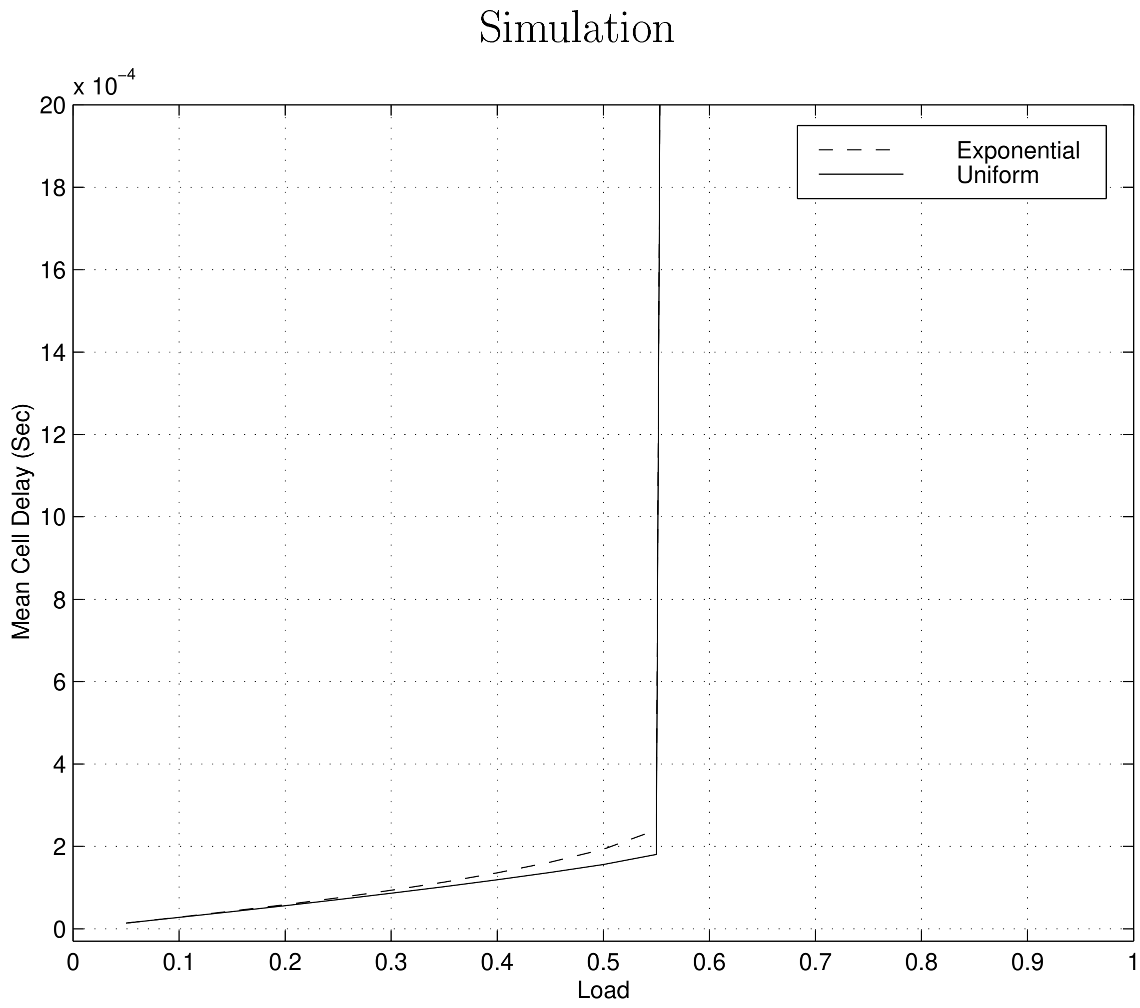


Figure 21: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'NCCOSC' and shown Figure 11.

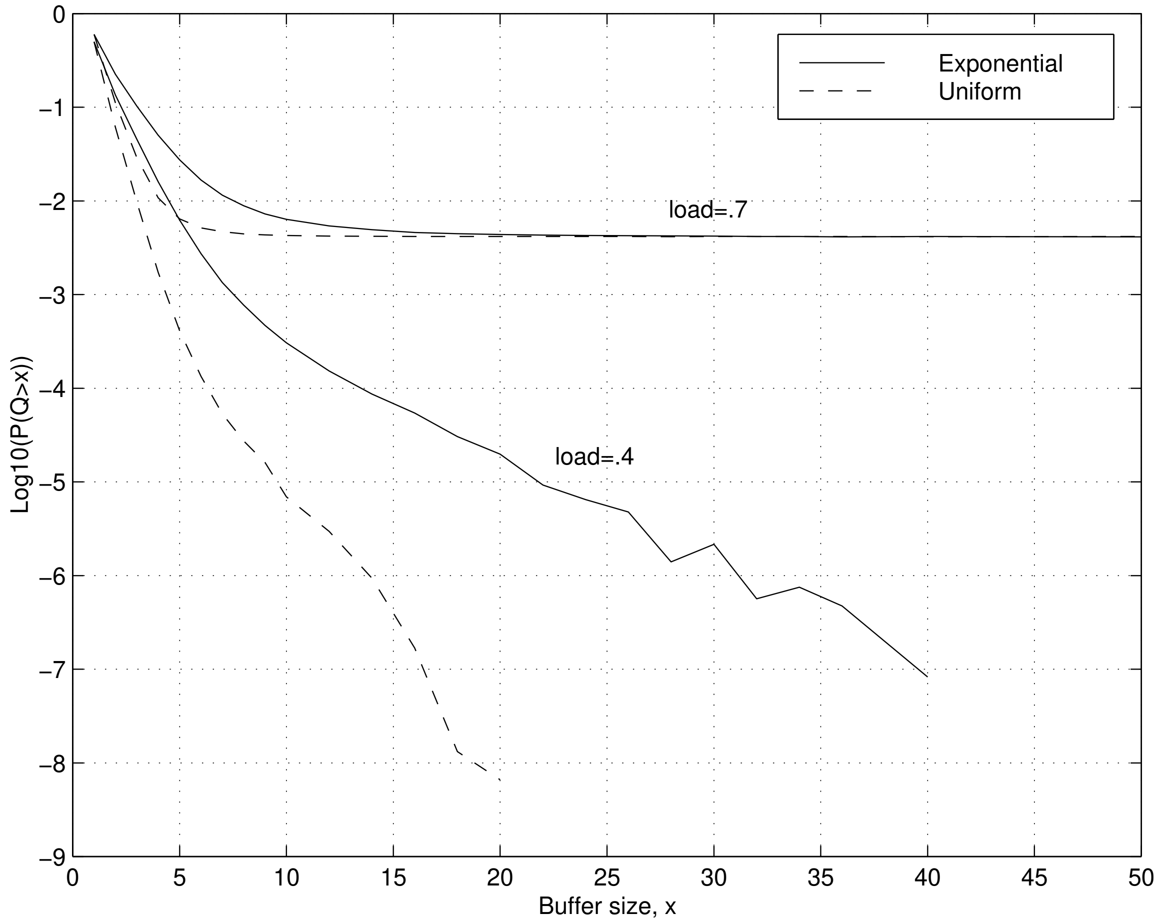


Figure 22: Effect of traffic micro-dynamics on Cell loss probability obtained from simulation of the trace labeled 'NCCOSC' and shown Figure 11.

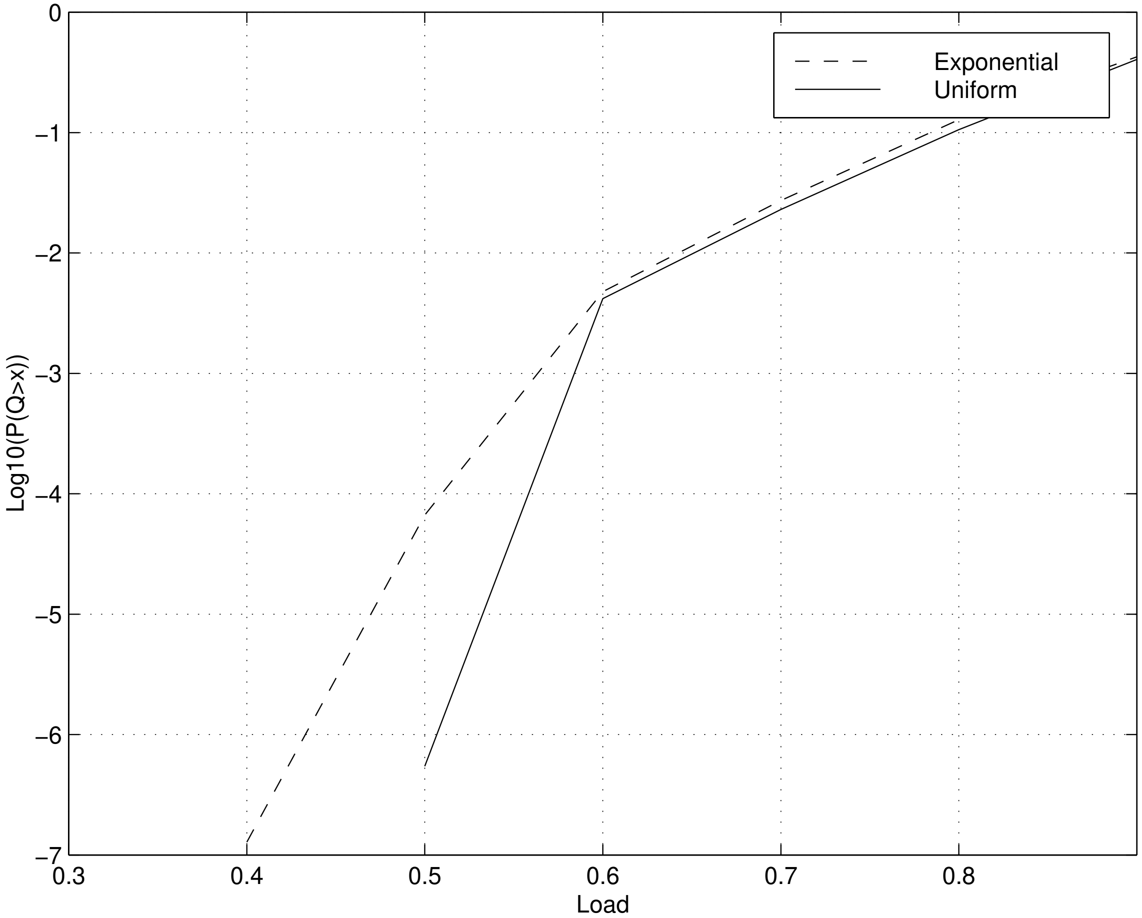


Figure 23: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a fixed buffer size of 15 cells from simulation of on the trace labeled 'NCCOSC' and shown in Figure 11.

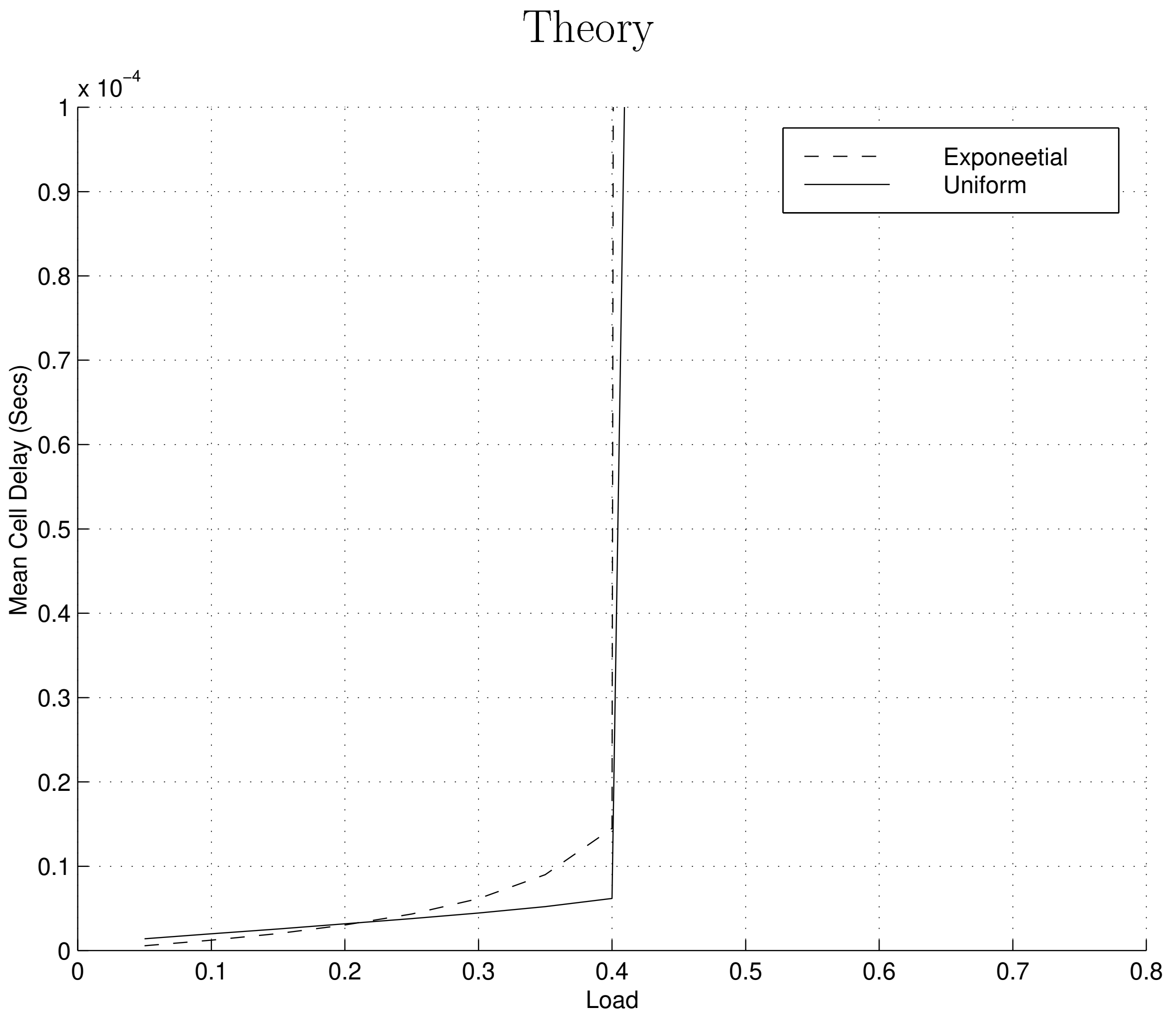


Figure 24: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 1.2$), for the trace labeled 'Phillips' and shown in Figure 14, using $N = 15$ input phases.

Simulation

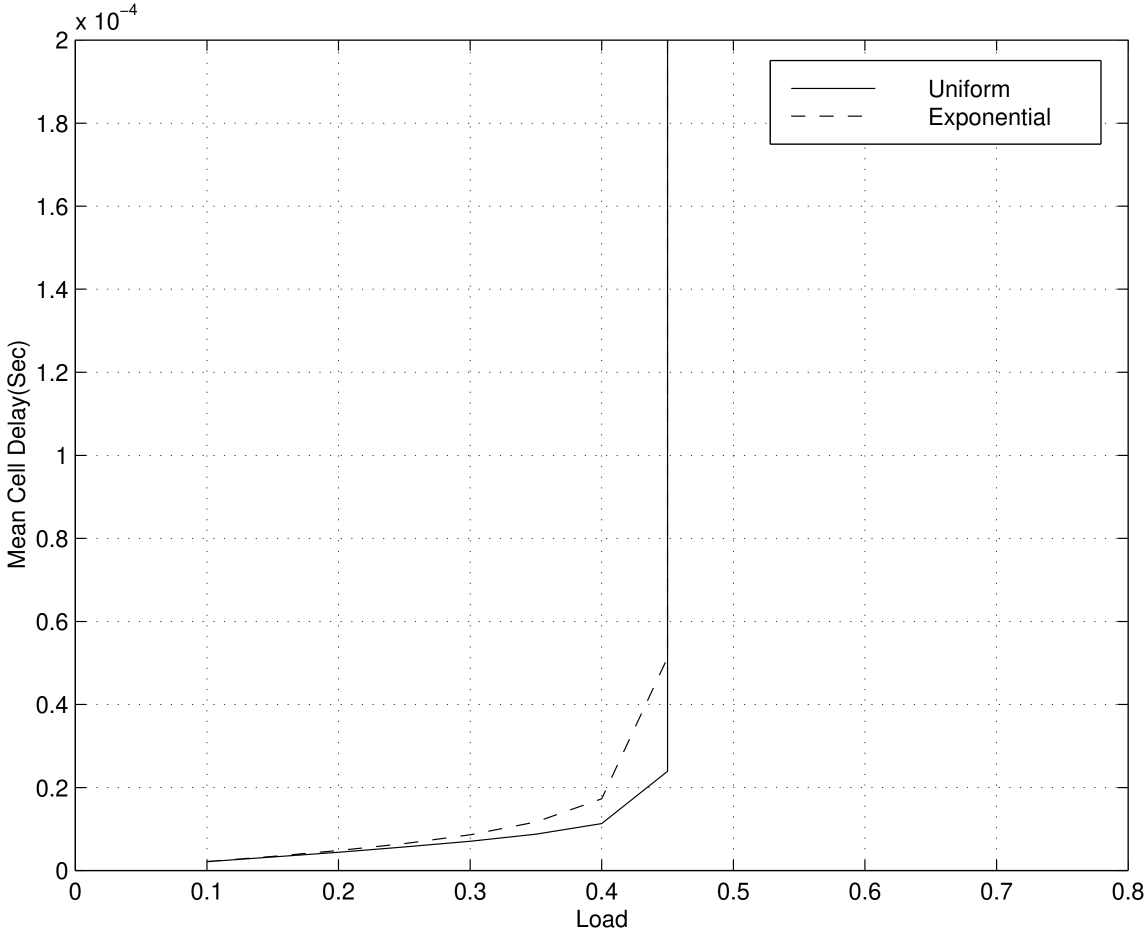


Figure 25: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled 'Phillips' and Figure 14.

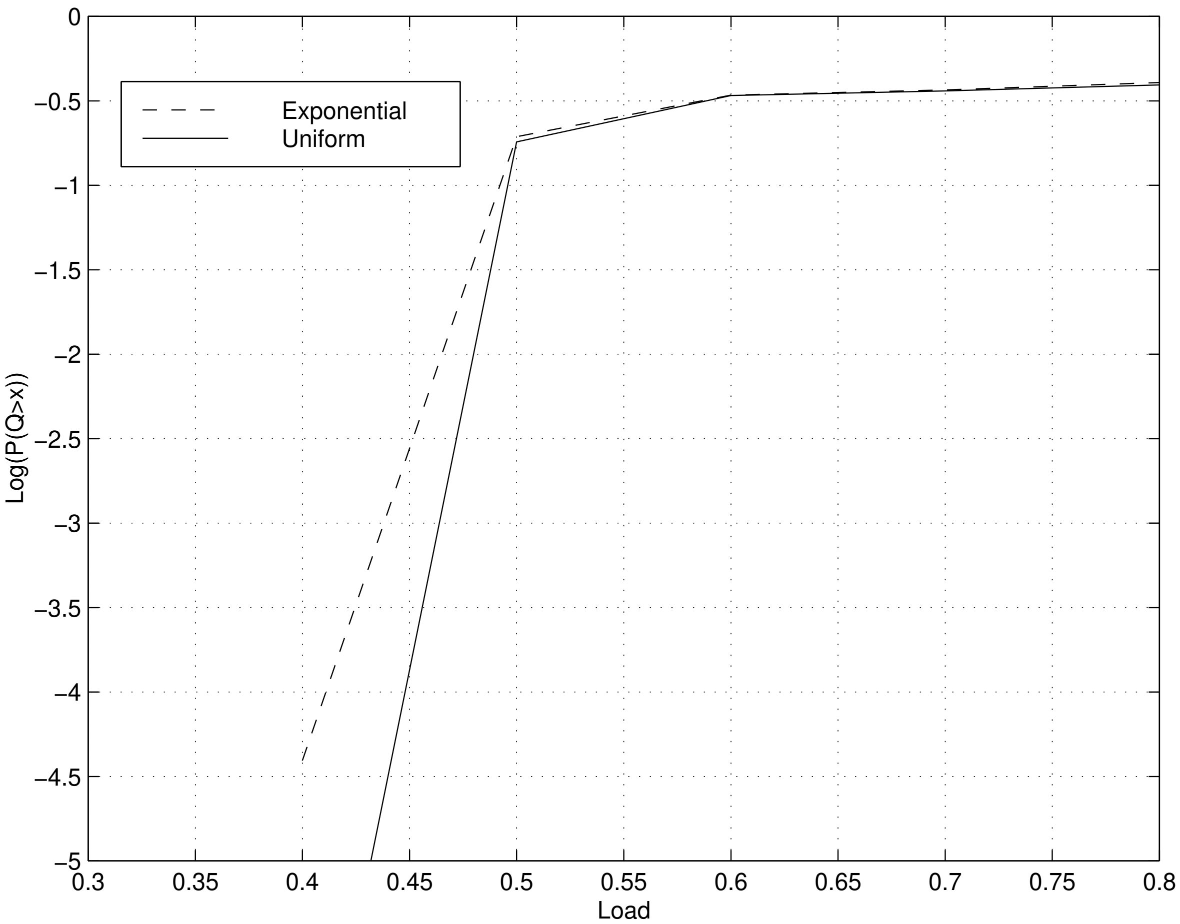


Figure 26: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a fixed buffer size of 30 cells from simulation of the trace labeled 'Phillips' and shown Figure 14.

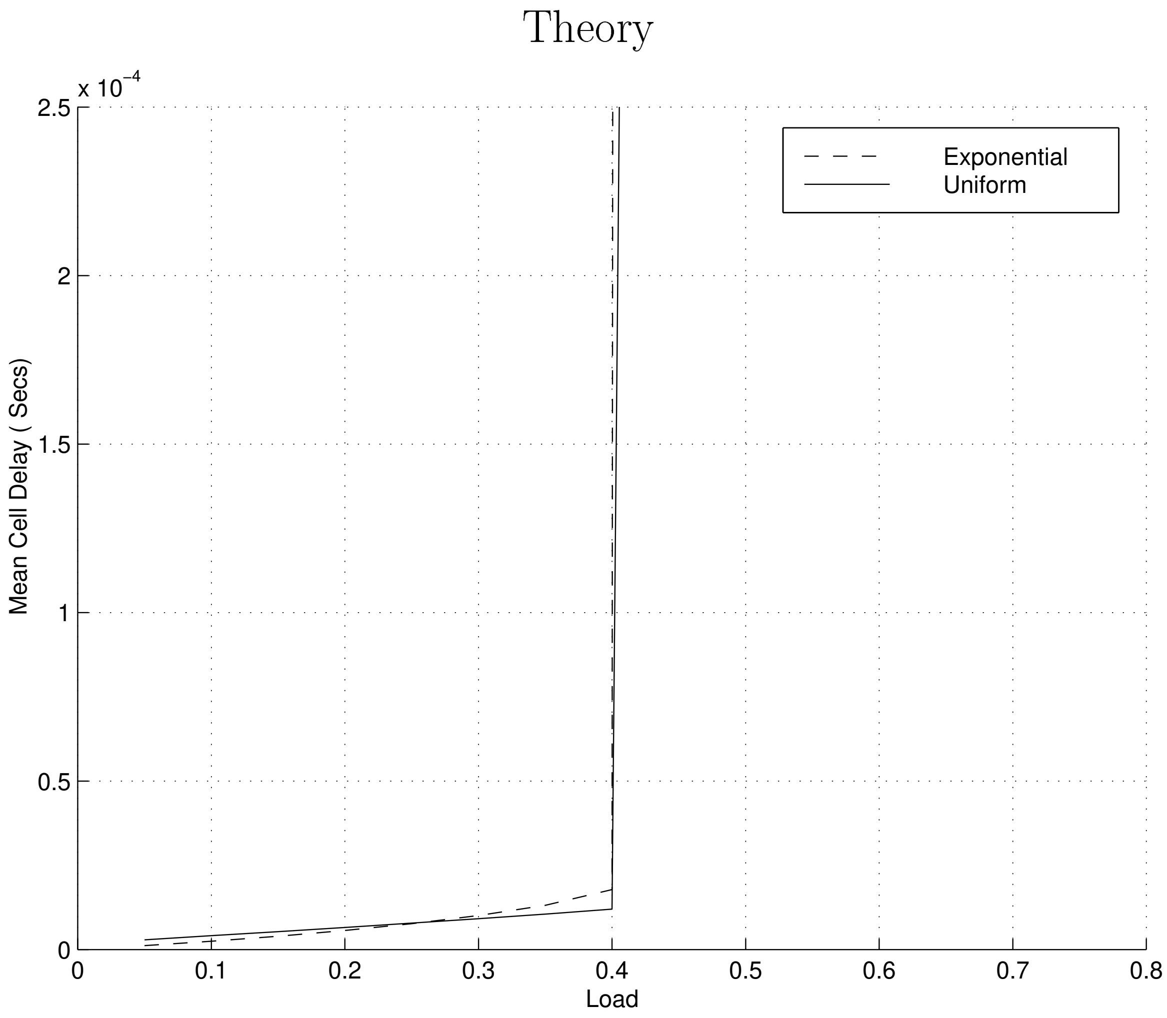


Figure 27: Effect of traffic micro-dynamics on Mean Delay predicted by theory ($\beta = 2.2$) for the trace labeled NRL and shown in Figure 17, using $N = 15$ input phases.

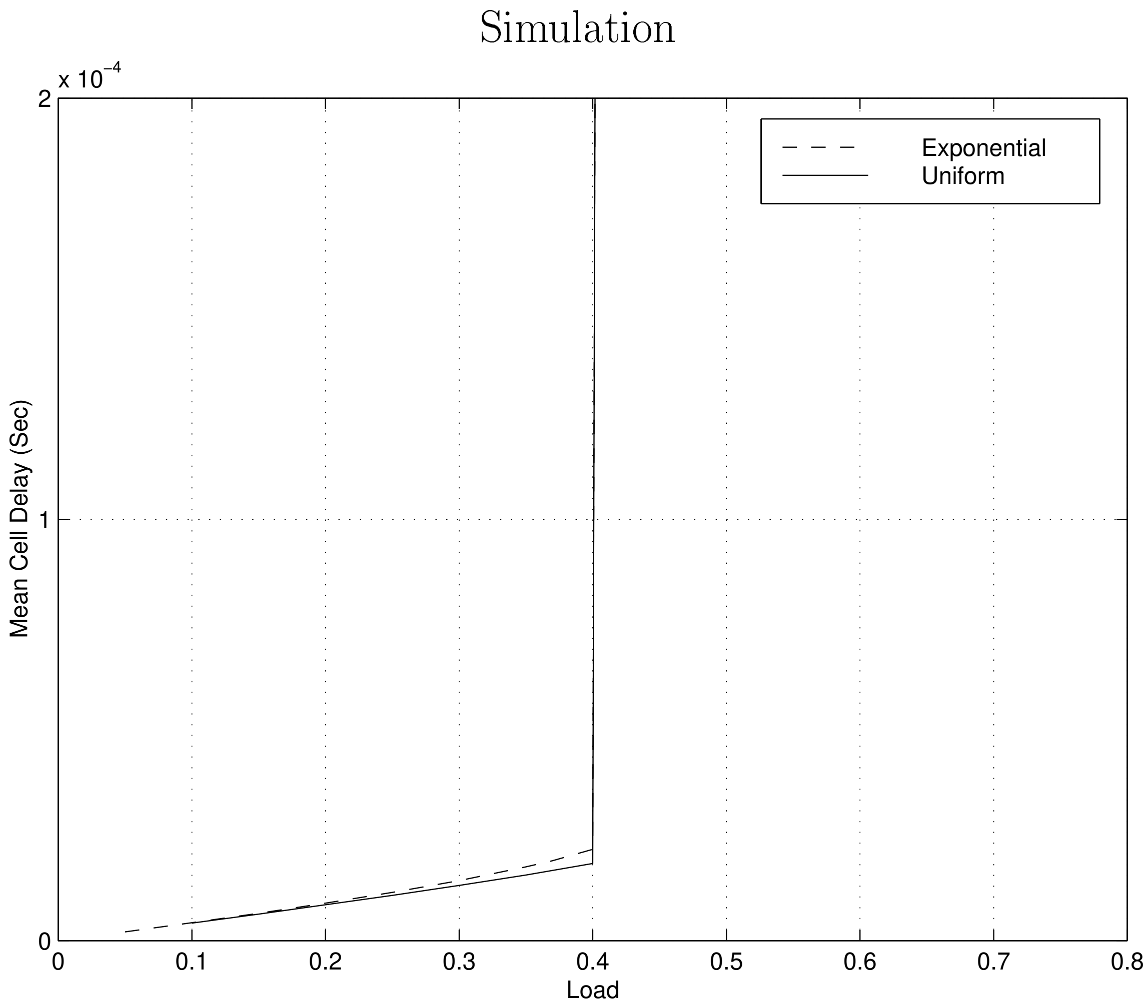


Figure 28: Effect of traffic micro-dynamics on Mean Delay obtained from simulation of the trace labeled NRL and shown in Figure 17.

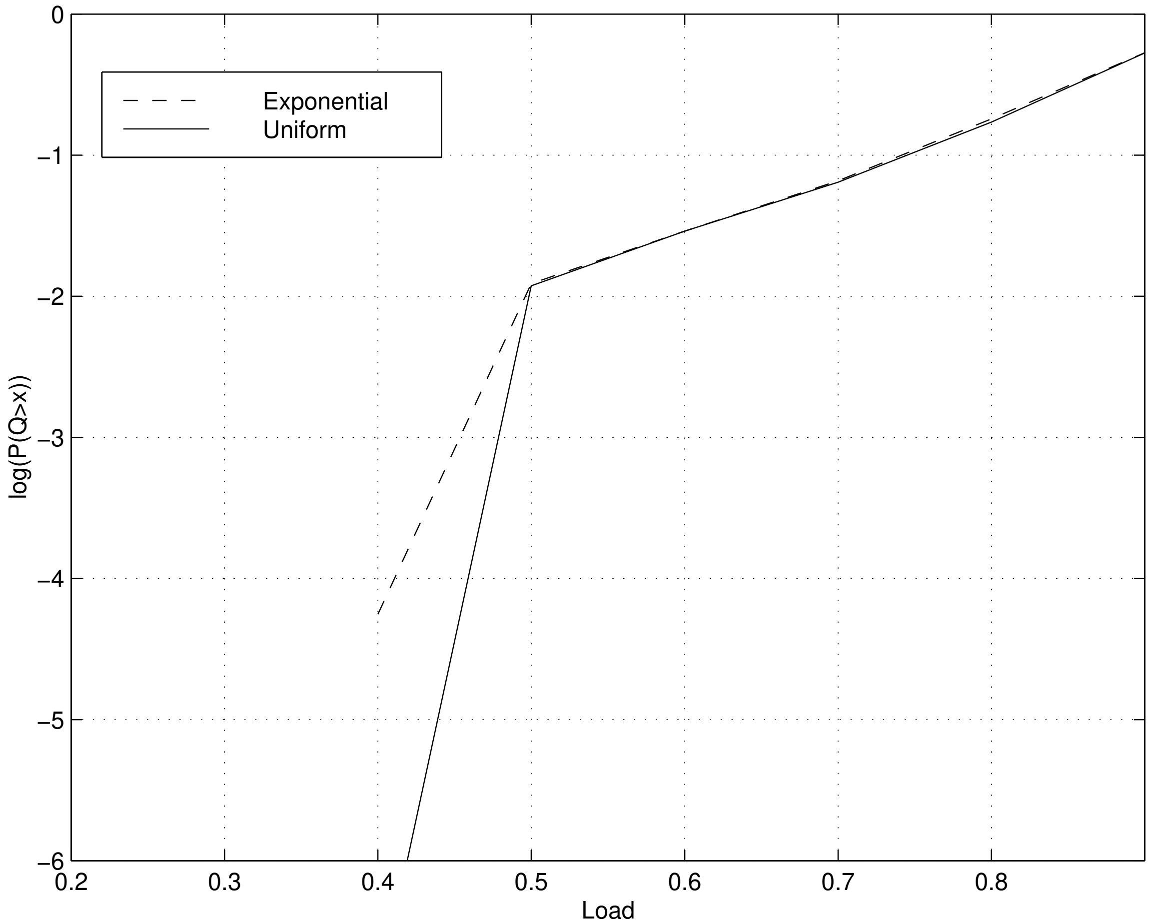


Figure 29: Effect of traffic micro-dynamics on Cell loss probability estimate obtained for a buffer size of 25 cells from simulation of the trace labeled NRL and shown in Figure 17.

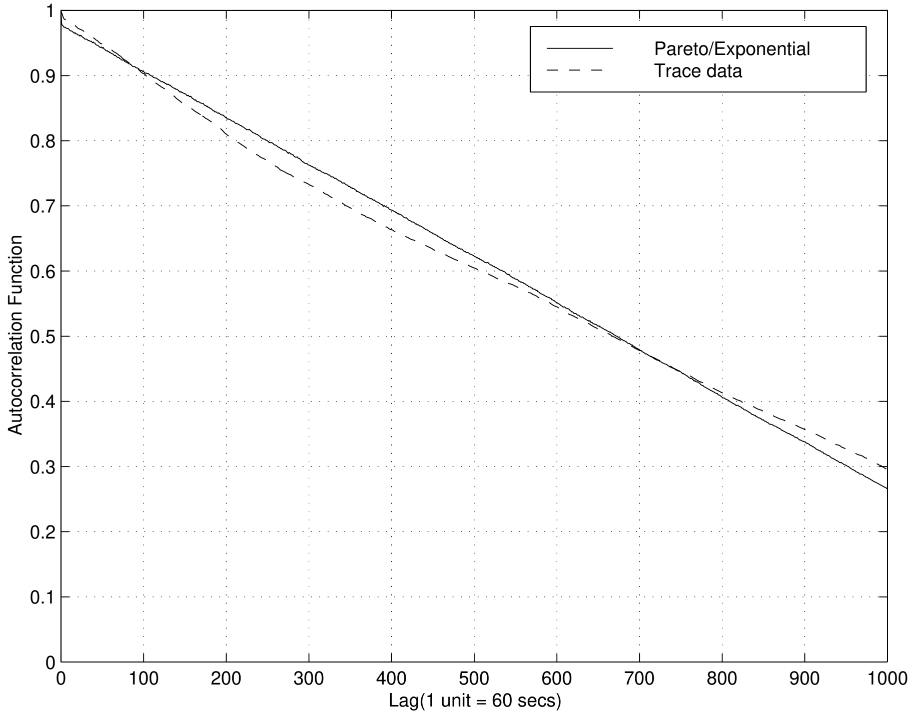


Figure 30: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'NCCOSC' and shown in Figure 11.

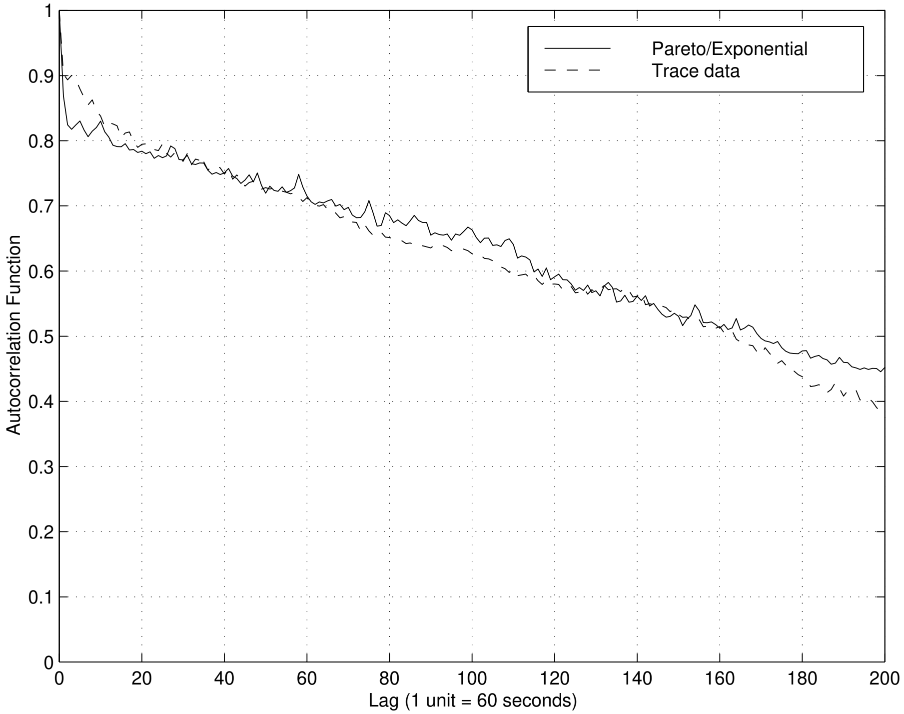


Figure 31: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'Phillips' and shown in Figure 14.

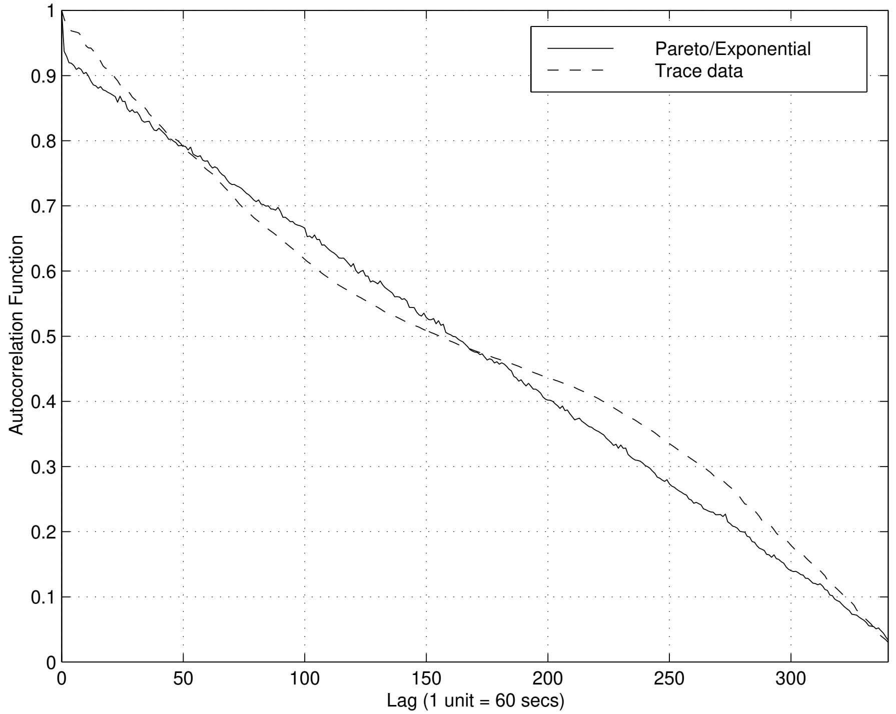


Figure 32: Second order statistics obtained from trace data and Pareto/Exponential model for the trace labeled 'NRL' and shown in Figure 17.

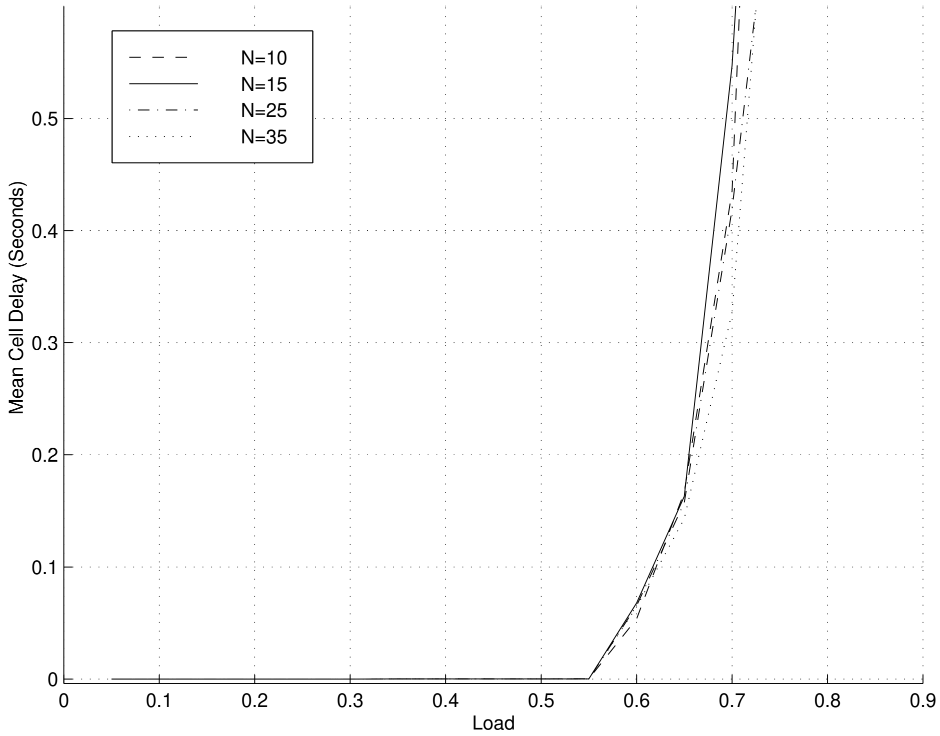


Figure 33: Effect of different number of phases on mean cell delay demonstrated using the trace labeled 'NCCOSC' and shown in Figure 11.

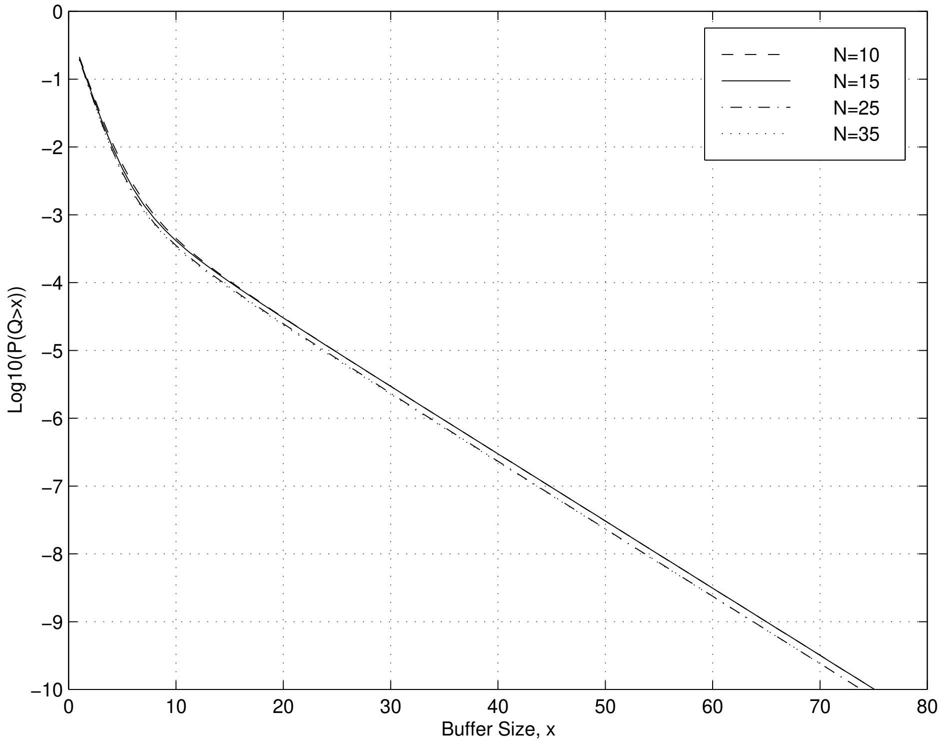


Figure 34: Effect of different number of phases on cell loss probability demonstrated using the trace labeled 'NCCOSC' and shown in Figure 11.

Conclusions

- A simple phase modulated model as well as a performance evaluation technique were developed in this study to model traffic in wide area ATM networks.
- Methodology was validated with extensive trace driven simulations performed using collected data traces from the AAI ATM WAN.
- The experimental evaluation of the model is done in terms of mean cell delay and cell loss probability.
- The study establishes non-Poisson nature of traffic on an early national scale ATM network, the AAI network.
- Effect of traffic micro-dynamics on queueing performance were investigated.
- In the presence of LRD, mean delay estimate is relatively invariant to the actual micro-dynamics.
- A low loads cell loss probability was found to be sensitive to micro-dynamics.
- At higher loads, bursts on the macro-scale, dominate the cell loss in the queue and the micro dynamics play a minor role.

- Logarithm of the cell loss probability has a linear relationship to load in the case of exponentially distributed inter-arrival times, but is non-linear in the case of real network trace data.
- Second order statistics are modeled reasonably.

Future Work

- Simple model - Significant complexity can be added.
  Effects of the UPC scheme and UPC parameters.
- Adding a periodic component to take into account the periodicity of the input process.
- Considering the effects of different theoretical infinite-variance distributions.
- Developing analytical expressions for studying effects of micro-dynamics on cell-loss.
- Developing expressions for computing the autocorrelation function numerically.
- Constructing traffic models with phase dependent micro-dynamic distributions.