Correlated Interarrival Time Poisson Process (CIPP) for Modeling Broadband Teletraffic

Thesis

Submitted in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

R. Manivasakan

Department of Electrical Engineering
INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
October, 1999

Approval Sheet

Thesis entitled: *Correlated Interarrival time Poisson Process (CIPP) for Modeling Broadband Teletraffic* by **R. Manivasakan** is approved for the degree of **DOCTOR OF PHILOSOPHY**.

Examiners

____

____

____

____

Supervisor(s)

____

____

Chairman

Date : ________________________ ________________________

Place : ________________________

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY, INDIA.

Certificate of Course Work

This is to certify Mr.`R. Manivasakan` was admitted to the candidacy of the Ph.D. Degree on `10.06.99` after successfully completing all the courses required for the Ph.D. Degree programme. The details of the course work done are given below.

Sr. No.	Course No.	Course Name	Credits
1	EE608	Adaptive Signal Processing	6
2	EE1698	Computer Vision	6
3	EE2698	Artificial Neural Network	6
4	EES801	Seminar	4

I.I.T. Bombay

Dy. Registrar (Academic)

Date :

Abstract

In this thesis, the problem of broadband teletraffic modeling has been investigated. This work is based on the basic philosophy that, models that capture the correlations in interarrivals are more appropriate in the context of modeling broadband teletraffic than models that capture the correlation in the count sequence as done conventionally. In light of the above, a new process for modeling the broadband teletraffic is presented. Specifically, we have generalized the Poisson process. The generalized Poisson process - herein called as Correlated Interarrival time Poisson Process (CIPP) - is parameterized by a correlation parameter 'ρ' which represents the degree of correlation in adjacent interarrivals and 'λ', the intensity of the process. In particular, we derive an expression for $P_n(t)$ (the probability of n arrivals in the interval $(0, t]$)). This derivation is accomplished in two steps: first, random arrival instants are generated on the real line, such that the interarrival times are correlated with exponential marginals (here, the correlated exponential sequence is generated by the first order exponential auto-regressive (EAR(1)) process) and second, a counting process over this point process is constructed such that the resultant counting process is stationary in the strict sense.

It has been shown empirically, that the CIPP models the VBR video traffic as well as the composite Ethernet traffic quite well. The autocorrelation function, Index of Dispersion for Intervals (IDI) and Index of Dispersion for Counts (IDC) are used as statistical measures for fitting the above model to the data. The expression for burstiness factor is evaluated. The self-similar property of CIPP is explored using the time domain methods and wavelet techniques. The superposition of two independent CIPP processes are also studied.

The stationary distributions for buffer occupancy and the waiting time in a CIPP/M/1 queue are derived. The analytical expressions for various queueing performance measures namely, mean packet waiting time, mean number of packets in the system, packet delay variation and overflow probability corresponding to CIPP/M/1 queue are obtained. We verify the above theoretical results by simulating the CIPP/M/1 queue.

Although, we were successful in deriving the approximate analytical expressions for the buffer occupancy distribution in a CIPP/D/1 queue, the expressions involved pose a numerical problem for large values of ρ. Hence, we estimate the above expressions, by suitable simulations.

As the CIPP/M/1/K queue is analytically intractable, we undertake the simulation based study of the same, where the performance measures used are packet loss probability and conditional packet loss probability. We demonstrate that the buffer occupancy distribution in a CIPP/M/1/K queue (with high ρ) is consistent with those observed in literature for correlated arrivals.

We undertake the performance modeling of statistical multiplexer with *real-world* composite Ethernet traffic using CIPP/M/1 queue. The exponential distribution for packet lengths (thus, for the service times) and infinite capacity for output queue in

a statistical multiplexer are assumed here to use the theoretical results for CIPP/M/1 queue.

By empirical study, we establish that, the finite capacity CIPP/M/1/K queue with exponential services models well, the queueing situation as encountered by *real-world* composite Ethernet traffic in a statistical multiplexer with finite capacity output buffer. Queueing performance measures considered here are, packet loss probability and conditional packet loss probability. We then compare the buffer occupancy distribution of the finite capacity queue with Bellcore Ethernet traffic and the "best fit" CIPP/M/1/K queue with K arbitrary.

We have introduced the Markov modulated CIPP (MMCIPP) to overcome a limitation of CIPP. We show by time domain methods and by wavelet analysis (all pertaining to estimation of Hurst parameter) techniques that, MMCIPP is self-similar to a wider time scale than MMPP. We also establish the above fact by analyzing the queueing performance with the synthetic traces generated by them as input.

We undertake the performance modeling of statistical multiplexer with *real-world* MPEG-1 VBR video traffic input and exponential services using CIPP/M/1 queue. For this purpose, we use a variation of *deterministic smoothing*, which generates packets with variable sizes. In particular, the method generates packets with exponentially distributed size from the given VBR frame size trace. We present the simulation based performance modeling study of finite capacity statistical multiplexer with *real-world* MPEG-1 VBR video traffic using CIPP/M/1/K queue. Queueing performance measures considered are packet loss probability and conditional packet loss probability. Buffer occupancy distribution for 'best fit' CIPP/M/1/K queue and the above *real-world* queue are also compared. We then consider the fluid-flow approximation of queues with CIPP arrivals to study the behavior of statistical multiplexer with VBR video traffic input. Comparisons with other models like, Markov chain, histogram model are also undertaken.

Finally, we give an example where the empirical relationship of the correlation parameter ρ of CIPP to the Hurst parameter, is exploited to study the effect of variability of the traffic on various network performance measures. More specifically, we present a simulation based performance evaluation study of MPEG-2 video when subjected to interference effects by a cross traffic process (background network traffic) modeled using the CIPP model.

Key Words:-
Broadband teletraffic modeling, Correlated Interarrival time Poisson Process (CIPP), interarrival time correlations, index of dispersion for counts (IDC), index of dispersion for intervals (IDI), self-similarity, Hurst parameter, long-range dependence (LRD), short-range dependent (SRD) models, exponential auto-regressive (EAR(1)) sequence, burstiness factor, queueing performance measures, CIPP/M/1 queue, CIPP/D/1 queue, CIPP/M/1/K queue, performance modeling, composite Ethernet traffic, VBR video traffic, MMCIPP, MPEG-2

Contents

List of Figures

List of Tables

List of Symbols

$A(t)$	Traffic process at the interval in $(0, t]$.
D	Decoder buffer size.
H	Hurst parameter.
$\hat{H}_k$	Estimate of Hurst parameter at scale 'k'.
ϵ_n	Innovation sequence.
$F(.)$	Probability distribution function.
$f(.)$	Probability density function.
γ	Load (ratio of arrival rate to service rate) in a queueing system.
K	Storage capacity of a finite capacity queueing system or threshold in an infinite capacity queueing system.
k_m	Number of wavelet coefficients at scale m.
λ	A parameter of CIPP denoting arrival rate.
μ	Service rate in a queueing system.
$\nu(t)$	Number of fixed packet arrivals in $(0, t]$.
$L_n(t)$	Forward recurrence time.
M	Multiplexer buffer size.
$N(t)$	Number of counts in the interval $(0, t]$.
$N_{T_f}(i)$	Number of packets in the frame interval T_f in VBR video, after smoothing.
P_{min}	Minimum percentage error.
P_{max}	maximum percentage error.
$P_i(k)$	Group density
P_j	Stationary probability of j number of customers in the queue seen by an arriving customer.
$\phi_i(z)$	Probability generating function of $P_i(k)$.
$P_n(t)$	Probability of n number of arrivals occurring in the interval $(0, t]$.
P_{ij}	Probability of transition from state i to state j.
Q_0	Infinitesimal generator for MMCIPP.
$Q(t)$	Number of customers waiting or being served at time t.

q_{ij}	Probability mass function of group arrivals in ij transitions.
$r(k)$	Autocorrelation function at lag k.
r_{ij}	Probability mass function of renewal transitions from state i to state j.
ρ	A parameter of CIPP denoting interarrival correlation.
S_j	Service time for jth arrival.
$S(j)$	Size of jth packet.
$S^2(n)$	Sample variance of sample size n.
τ	Packet delay variation (Cell delay variation).
t	Time parameter.
T_{GOP}	GOP interval in an encoded video trace.
T_f	Frame interval in an encoded video trace.
T_n	nth arrival instant.
$u_T(i)$	Counts sequence, denoting the number of arrivals in the ith slot slot of consecutive fixed length interval T on the real axis.
V_t	Virtual waiting time.
$V_i(j)$	jth packet (of size $S(j)$) corresponding to ith frame interval in a VBR video trace.
W_n	Waiting time of nth arrival.
W	Stationary waiting time.
x_n^m	nth wavelet coefficient at scale m.
$\mathbf{X}^{(m)}$	Aggregated version of the process $\mathbf{X}$ at level m.
$\bar{X}$	Mean of the random variable X.
X_n	Interarrival time between nth arrival and $(n-1)$th arrival instant.
$Y(t)$	A stochastic process.
$\zeta_T(.)$	Correlation coefficient of the counts sequence $u_T(i)$.

List of Acronyms

AAL2	ATM Adaptation Layer
AR	Auto-Regressive
ABR	Available Bit Rate
ARMA	Auto-Regressive Moving Average
ARIMA	Auto-Regressive Integrated Moving Average
ATM	Asynchronous Transfer Mode
BD	Birth Death Process
CIPP	Correlated Interarrival time Poisson Process
CBR	Constant Bit Rate
cdv	cell delay variation
CPLP	Conditional Packet Loss Probability
CPD	Correlated Poisson Distribution
CTDS	Continuous Time Discrete State (Process)
CCITT	*Comite Consultatif International Telegraphique et Telephonique*
D-BMAP	Discrete-Batch Markovian Arrival Process
DAR	Discrete Auto-Regressive (Model)
DFT	Discrete Fourier transform
EAR(1)	First Order Exponential Auto-Regressive Process
EMA(1)	First Order Exponential Moving Average Process
fBm	fractional Brownian motion
fGn	fractional Gaussian noise
FCFS	First Come First Serve
FDM	Frequency Division Multiplexing
FTP	File Transfer Protocol
GAR(1)	First Order Gamma Auto-Regressive Process
GPD	Generalized Poisson Distribution
GOP	Group of Pictures

iid	independent and identically distributed
IDI	Index of Dispersion for Intervals
IDC	Index of Dispersion for Counts
IPP	Interrupted Poisson Process
LAN	Local Area Network
LUF	Link Utilization Factor
LRD	Long-Range Dependence
LST	Laplace-Stieltjes Transform
MAN	Metropolitan Area Network
MC	Markov Chain
MF	Modified Finch Process
MP	Markov Process
MMPP	Markov Modulated Poisson Process
MMCIPP	Markov Modulated Correlated Interarrival time Poisson Process
MLE	Minimum Least Squares (Estimate)
MPEG	Moving Average Expert Group
NNTP	Network News Transfer Protocol
pdv	Packet Delay Variation
PDU	Program Data Unit
PES	Packetised Elementary Streams
QoS	Quality of Service
RW	Random Walk
RW	Renewal Process
R/S analysis	Rescaled Adjusted Range Plot
SMP	Semi-Markov Process
SMTP	Simple Mail Transfer protocol
SRD	Short-Range Dependent (Models)
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TELNET, RLOGIN	TCP protocols
TS	Transport Streams
T-STD	TS - System Target Decoder
VBR	Variable Bit Rate
WAN	Wide Area Networks
WWW	World Wide Web

Chapter 1

Introduction

1.1 Introduction

Integrated broadband networks are expected to support various traffic types such as data, voice, image and video. Traffic generated from these services is substantially different in their statistical characteristics and quality of service[1] (QoS) requirements. For example, voice traffic has a bandwidth requirement of several Kpbs and is delay sensitive, while high speed data traffic used for file transfers or LAN interconnection is hundreds of Mbps and is loss sensitive [1].

Asynchronous transfer mode (ATM) is one of the transport mechanisms accepted by CCITT to implement the requirements of integrated broadband networks. In an ATM network, all information such as voice, data, video is conveyed using fixed size packets[2] (called *cells* in ATM terminology). To achieve multimedia transport, an ATM network adopts a simplified transport protocol based on hardware cell switching with no flow control or retransmission inside the network. The ATM network takes advantage of the bit rate variations of individual sources by *statistical multiplexing*. In this way, individual sources share a link of capacity less than the sum of individual peak rates, achieving significant *multiplexing gain*. This multiplexing gain is restricted by the QoS requirements of the sources. Thus, statistical multiplexing could lead to more efficient use of network resources, but, may require new kinds of bandwidth management, QoS provisioning and traffic control.

Statistical multiplexing involves N input streams (which are asynchronous and statistical) and N output streams, so that traffic from any of the N input streams can be

[1] Quality of Service is a set of parameters (often statistical bounds), that the network agrees to maintain throughout the session of the call, agreed during the call set up phase. The attributes of QoS may be packet loss ratio, packet delay variation, maximum packet transfer delay and mean packet transfer delay.

[2] In specific sections (where the situation demands), we use the term *cell* to denote a fixed length ATM packet of size 53 byte. Otherwise we use the term *packet* to denote a more general variable sized information unit in broadband networks. This terminology is used throughout this thesis.

switched over to any of the N output streams. Now, each of the N output streams can be modeled as a single server queue, with the arrival streams forming the superposition of the N input streams in general. Accordingly, this single server (with such arrival streams) has been of interest to telecommunication researchers over the years. Hence, understanding the nature of traffic in such heterogeneous multiservice networks is essential for engineering, operations and performance evaluation of these networks. Moreover, to facilitate the mechanization of multiplexing, routing and control, it is imperative to appropriately model and characterize the input traffic, and to map the input source parameters (could be mean bit rate, peak bit rate or it could be even Hurst parameter (H) which measures the degree of self-similarity) into network parameters (could be mean cell delay, cell loss rate and delay jitter). In light of the above, an important question often raised by network planners and designers is: "Is there an accurate and useful traffic model in the form of a simple stochastic process with minimal parameters and, when fed into the single server queue, gives the same queueing performance as a real traffic stream?". Such a traffic model will be very useful in network design tools or in tools supporting real-time traffic management. In this thesis, we propose a model for broadband teletraffic. We begin by giving a brief overview of the literature on teletraffic modeling from an historical perspective.

1.2 Historical Perspective of Teletraffic Modeling

The problem of teletraffic modeling first arose in the context of telephone traffic congestion. The pioneer investigator was the Danish mathematician Erlang, who, in 1909, published *The Theory of Probability and Telephone Conversations*. His work dealt with the characterization of telephone communications at the call level, i.e., the expected number of calls per hour and their distribution, holding time, determination of busy hour, etc. The statistical study of arrival of telephone calls and their holding times was motivated by the need to determine the optimum number of telephone lines to handle the prescribed call frequencies and to provide a certain QoS. (In plain old telephone system (POTS), QoS parameters were, essentially a small blocking probability of a call and a small call set-up delay. Predominantly, Poisson process was used, as a model for telephone call arrivals, as there was enough empirical evidence that the call arrivals follow Poisson process. Work on the application of queueing theory to telephony continued even after Erlang. In 1927 Molina published, *Applications of Probability to Telephone Trunking Problems*, which was followed by a year later by Fry's *Probability and its Engineering Uses*, which extended much of Erlang's earlier work. In 1973, Kuczura [3] proposed interrupted Poisson process (IPP) to model the traffic overflow of a large trunk group in a telephone exchange. In IPP, the Poisson process is alternatively turned ON and OFF for an exponentially distributed time. The IPP provided a simple approximate description of the overflow traffic and consequently facilitated analytical studies.

With the advent of computers in 1970's, new kinds of problem arose, namely, queueing problems associated with synchronous transfer mode (STM) based time-sharing systems (say, a time division multiplexing (TDM) system) [4], delay analysis in store and forward networks (as in packet switched networks) [5], throughput problems in multi-access protocols (channel allocation problem as in ALOHA [6]). All these topics, resulted in the re-birth of teletraffic engineering in a different form.

As the gap between telecommunications and data communications kept on reducing in late 80's and early 90's, the concept of integrated broadband networks carrying multimedia traffic emerged. Eventually, ATM was accepted by CCITT as a transport mechanism to carry heterogeneous traffic in such networks. In ATM networks, all information is carried by fixed length packets called "cells". Here, teletraffic engineering took the latest form. This encompasses, modeling the superposed cell arrivals in a statistical multiplexer (which is a principal component in broadband networks), performance evaluation of the statistical multiplexer using the above model and use the findings of above studies for engineering the switches, multiplexers and real-time control of these networks. It is *this* scenario of teletraffic engineering, that is of much relevance to this thesis.

In the telephone era, teletraffic engineering was concerned with the statistical behavior of call arrivals (and their holding times) only; and a call required fixed amount of bandwidth, and as a result, knowledge of the number and duration of calls was sufficient to determine the needed resources to provide a certain QoS (mainly, blocking probability). In contrast, multimedia traffic is characterized by a high variability (during the call) in its bandwidth needs. Hence, one is *not* only interested in the number and duration of calls, but also the statistical properties of the information flow during the call (arising out of variable bit rate (VBR) nature or bursty nature of the cell arrival stream), in order to make efficient use of the resources while guaranteeing a high QoS [7]. This calls for a more complicated statistical model for the cell arrival process. We review some of the teletraffic models addressed in literature.

The versatile Markovian arrival process [8] (later called as *Neuts* process [9]) and the associated matrix-geometric methodology [10, 11] developed by Neuts is a boon to the teletraffic modeling specialists. Many of the traffic models in current literature, namely Markov modulated Poisson process (MMPP), Markov arrival process (MAP), discrete-batch Markovian arrival process (DBMAP) are special cases of Neuts process. Since this versatile class of processes is closely related to finite state Markov processes, its theoretical analysis is elementary and its relation to finite systems of differential equations guarantees that, at least for a wide range of parameter values, the resulting equations may readily be solved numerically. The queueing theory for the Neuts process is developed in [9], wherein the author has generalized the Pollaczek-Khintchin formula for the Laplace-Stieltjes transform (LST) of the virtual time of the M/G/1 queue to N/G/1 queue and the resulting Volterra system of integral equations. The matrix-analytic methodology developed by Neuts and later used to its full potential by Ramaswamy [9], is used even today by teletraffic researchers in situations, where analytic results are sought. The main disadvantage of this method is that, it is computationally intensive due to the iterative algorithms which are used to compute the stationary probabilities of interest.

In 1986, a Markov modulated model was proposed in [12] for the superposition of packetised voice and data sources. Starting with the source model for voice and data, their superposition is approximated by MMPP, by matching some of its moments. Then, the matrix analytic methodology developed by Neuts and Ramaswamy was applied to obtain the analytical solutions for stationary distributions. These results are used for performance evaluation of related statistical multiplexer.

While the above model considers packet arrivals as distinct entities, the fluid flow model proposed in [13] assumes input traffic as fluid. In that model, they consider a buffer receiving messages from a finite number of statistically independent and identical information sources that asynchronously alternate between exponentially distributed periods in the "ON" and "OFF" states. While ON, a source transmits at a uniform rate. The buffer depletes through an output channel with a given maximum rate of transmission. They analyze a set of differential equations which describes the equilibrium buffer distribution. The mathematical results render the computation of this distribution and its moments trivial and so also the waiting time moments. The methodology in [13] found a natural application in analyzing statistical multiplexing of voice sources, and Daigle and Longford [14] compare it favorably against other methods of greater numerical complexity. A differential approach was taken by Stern [15], in which the binomial model is used to represent voice sources but treats voice and data as discrete packets rather than assuming a fluid flow model.

In Maglaris *et al* [16], queueing models for statistical multiplexing of VBR video traffic are presented. Two correlated Markov processes are shown to model the experimental data well. The first model is discrete time, continuous state Markov process, and is basically a first order auto-regressive process. The authors have obtained the parameters of the model from a real-world video telephony trace. The second model is a continuous time discrete state Markov chain which was used with a flow equivalent approximate queueing analysis to obtain the common buffer length distribution. The binomial source model is a special case of the above discrete state model. The experimental results showed that the synthetic data (bit rate) generated by the models are consistent with the experimental data and also predicts the queueing performance reasonably well.

In Blondia *et al* [17], a discrete-time batch Markovian arrival process (D-BMAP) is presented to model a large class of VBR sources and their superpositions. D-BMAP is the discrete-time analogue of the versatile MAP introduced by Neuts [8]. It is shown that this rich class of discrete-time arrival process is computationally tractable and it gives accurate solution for the various performance measures of the related statistical multiplexer: buffer occupancy distribution, cell loss probability, conditional cell loss probability. Moreover, it is shown that the output process of multiplexer belongs to the same class of processes.

In Heyman *et al* [18], it is shown that the number of bits per frame of a VBR video traffic trace is gamma (or negative binomial) distributed. In the same, a simple AR(2) model is proposed for modeling the VBR video bit rate sequence. The authors find that although AR models capture the autocorrelation very well, the marginal distributions of interest in real-world data are poorly modeled by them. This marginal distribution of bit rate is captured by multi-state Markov chain reasonably well. In addition to that, the queueing behavior of VBR video traffic is also seen to be captured by multi-state Markov chain. Also discrete auto-regressive (DAR(1)) model [19] captures reasonably well (although not as good as the Markov chain) the performance of statistical multiplexer with VBR traffic input. These models lack analytical tractability.

In a different direction, Addie [20] considered a statistical multiplexer where, the net input process forms a stationary ergodic Gaussian discrete-time stochastic process. In this case, the authors show that it is possible to derive analytically, a simple approximate formula for tail distribution of waiting time in closed form, using a method based on Wiener-Hopf factorization technique. The resulting formula is in terms of three parameters of the net input process: the mean, the variance and the auto-covariance sum. One important advantage of the above method is that, it provides a solid theoretical basis for traffic characterization by these parameters although it is approximate.

A general solution technique for queueing analysis (in discrete time frame) of multimedia traffic on ATM using generating function approach is given in Li [21]. Here, the arrival process is modeled by superposition of a number of independent Markov chains, while each characterizes the stochastic properties of different type. He uses a discrete-time two state Markov chain as a basic element to construct models for different traffic types. The Markov chain considered here, alternates between the ON and OFF states with the holding time in each state geometrically distributed. While in the ON state, a constant length message is periodically generated. He considers the deterministic service intervals and infinite buffer capacity. With these assumptions the author carries out a discrete-time queueing analysis for the integration of multiple traffic types, to obtain the general expressions for first few moments of the queue size. Numerically it has been shown that, slowly varying traffic types exert the largest effect on the queueing process. This has been further corroborated by Li and Hwang in [22] (and in its companion paper [23]), but in a different context. It has been concluded that the input power spectrum is most essential to queueing analysis as compared to bispectrum, trispectrum and the input rate steady state distribution [22] [23]. Furthermore, input power in the low frequency band has a dominant impact on queueing performance, whereas high frequency power to a large extent can be neglected.

Another significant contribution in the area of teletraffic modeling is the discovery of long-range dependence (LRD) in Ethernet Local Area Network (LAN) traffic in Bellcore [24]. Leland *et al* in their landmark paper [24] demonstrate that the Ethernet LAN traffic is statistically self-similar. The implications of this remarkable findings are

1. Conventionally used models will be unable to capture this fractal like behaviour as they are basically short range dependent.

2. Commonly used measures for "burstiness" such as the index of dispersion (for counts), the peak-to-mean ratio, or the coefficient of variation (for interarrival times) are no longer meaningful parameter for self-similar traffic, but can be replaced by the Hurst parameter.

3. The nature of congestion produced by self-similar network traffic models differs drastically from that predicted by standard formal models. (To be precise, the the congestion produced by self-similar traffic is worse than the one produced by models considered in literature.)

4. Contrary to commonly held belief that multiplexing traffic streams tends to produce smoothed out aggregate traffic with reduced burstiness, aggregating self-similar traffic streams can actually intensify burstiness rather than diminishing it.

Moreover, further studies have confirmed that metropolitan area network (MAN) traffic [25], wide area network (WAN) traffic, and variable bit rate (VBR) video traffic [1, 26, 27], also exhibit LRD (or self-similar or fractal) characteristics.

Tsybakov *et al* [28], consider a large class of asymptotically second-order self-similar processes, for modeling the composite traffic input to an ATM node. For this class of models, the lower bounds on overflow probability and cell loss probability have been obtained. They also derive stationary distribution of the cell delay in an infinite capacity queue. They give some results on the merging of second order self-similar processes apart from stationary distribution of cell delay, which are of more practical significance.

In [29], the authors demonstrate empirically that, apart from its statistical significance in *Bellcore* Ethernet traffic measurements, LRD or self-similar behavior has phenomenal impact on queueing performance. This has direct implications in engineering and operations of integrated broadband networks. Considering shuffling experiments that make the packet interarrival times independent, or that retain short range correlations or long range correlations, they show that, LRD is not only relevant for queueing performance but, also plays a crucial role in several teletraffic engineering issues, such as dimensioning of buffers and real-time bandwidth allocation.

1.3 Discussions

In the above sections, we have briefly surveyed a variety of teletraffic models which researchers have proposed. There are enough empirical evidence that, the individual source traffic namely, voice and video have non-zero correlation [16, 30], (for non-zero lags) in their rate or (counts) process (in fact, they exhibit LRD or self-similar behavior). This is true for composite traffic also. Thus, the Poisson model is ruled out, despite its analytical simplicity. Hence, researchers were looking for models, which have correlations in increments on the one end, and on the other hand, are analytically tractable. The correlation in increments could be brought in a number of ways, namely, considering correlations in the counts sequence by say a discrete auto-regressive (DAR(1)) model, or by considering correlations in rate process by an auto-regressive (AR) model or by introducing correlations by modulating Markov chain or finally by considering *correlations in the interarrival* sequence itself. In literature, researchers have considered models which capture correlation in the counts sequence given as cells per unit time (as in DAR(1) model [18]), or in the (traffic) rate process (as in AR model [16] or self-similar models [24, 28]) or the correlation introduced by the modulating Markov chain as in Markov modulated models [31]. Note that, it has been observed by Paxson *et al* [32], that there is correlation in interarrivals in Ethernet traces. All of the models given in the previous section, are intended to capture the correlation in the increments, by considering correlation in counts sequence or in rate process or introducing correlation in the modulating Markov chain.

In light of the above, the following questions are pertinent: "Can one define a process with correlations in interarrivals, still retain all nice properties of the Poisson process? That is, is it possible to get a closed form expression for $P_n(t)$ which denotes the number of arrivals in the interval $(0, t]$? Will it be self-similar at least over a range of time scales of interest? Is it possible to get the expressions for various queueing performance measures, if we consider a queue with the above process as input? What are all its implications on the queueing performance with respect to the degree of correlation in interarrivals?". These are the questions which this thesis addresses.

1.4 Contributions of the Thesis

In this thesis, the problem of broadband teletraffic modeling has been investigated. The basic philosophy over which this work stands is that, considering models that capture correlations in interarrivals is more appropriate in the context of modeling broadband teletraffic than considering models that capture the correlation in the counts sequence as done conventionally in literature. In short, the major contributions of this thesis are the following: *A counting process with correlated interarrivals has been derived and justified for modeling VBR video traffic and composite Ethernet traffic. Its statistical properties and properties relevant for broadband traffic modeling have been investigated. Then, stationary distributions of interest for the queue with this process as input and exponential services*[3] *are derived. Consequently, the performance modeling of statistical multiplexer with composite Ethernet traffic and VBR video traffic input using the queues with above process has been undertaken. An example illustrating the applicability of the relation between the degree of interarrival correlation and the degree of self-similarity is presented.*

A new process for modeling the broadband teletraffic is proposed. Specifically, we have generalized the Poisson process. The generalized Poisson process - herein called as Correlated Interarrival time Poisson Process (CIPP) - is parameterized by a correlation parameter 'ρ' which represents the degree of correlation in adjacent interarrivals in addition to 'λ' the intensity of the process. After developing queueing theory for

[3] By "exponential services", we mean that the service times are *iid*'s with exponential distribution.

the CIPP/M/1 queue, analytical expressions for various queueing performance measures have been derived. These analytical measures are used for studying the performance modeling of statistical multiplexer with composite Ethernet traffic and VBR video traffic input using the CIPP/M/1 queues. More specifically, the contributions arising out of the thesis can be formally summarized in nutshell as follows.

- We generalize the Poisson process, which is referred herein as Correlated Interarrival time Poisson Process (CIPP). To be more precise, we derive an expression for $P_n(t)$, which denotes the probability of n arrivals in the interval $(0, t]$.

- Most of the traffic models in current literature characterize the correlation in the sequence of number of packet (cell) arrivals in the fixed size, disjoint, consecutive slots in the time axis or introduce correlation by the Markov chain modulating a micro process (usually the Poisson process) as in Markov modulated processes. We contend that, a model capable of capturing the correlation in the interarrival sequence is a more appropriate candidate in the context of modeling teletraffic data, rather than a model capturing correlations in the counts sequence (as done in [18, 33] for VBR modeling) or capturing the correlation by the modulating Markov chain as in short-range dependent (SRD) models. We show empirically, that the CIPP models quite well the VBR video traffic as well as the composite Ethernet traffic [34] with respect to the following second order measures: (1) autocorrelation of counts sequence (2) index of dispersion for intervals (IDI) and (3) index of dispersion for counts (IDC).

- We evaluate the properties of CIPP, both statistical properties as well as properties which are of relevance to modeling broadband teletraffic. In particular, we explore burstiness factor [35] and self-similar behavior. We investigate the self-similar property of CIPP using time domain methods and wavelet analysis. We study the superposition of two independent CIPP processes.

- We consider single server FCFS queue with CIPP arrivals where the service times are exponential (CIPP/M/1 queue). We derive the stationary distributions for buffer occupancy and waiting time [36].

- We obtain the analytical expressions for various queueing performance measures namely, mean packet waiting time, mean queue length, variance of packet delay variation and overflow probability [37] corresponding to CIPP/M/1 queue. We verify the above theoretical results by simulating a CIPP/M/1 queue.

- We consider queues with CIPP arrivals and with *deterministic* service times (CIPP/D/1 queue) to take into account the deterministic switching time of a broadband switch. Although, we have approximate analytical expressions for the buffer occupancy distribution, it poses a numerical problem for large values of ρ. Hence, we resort to estimate the above expressions, by suitable simulations.

- We consider finite capacity queues (with capacity K packets) with CIPP arrivals and exponential service times (CIPP/M/1/K queue). We resort to the simulation study of the CIPP/M/1/K queue, as it is analytically intractable. The queueing measures under study are packet loss probability and conditional packet loss probability. We demonstrate that the buffer occupancy distribution for the CIPP arrival case, is consistent with those observed in literature for correlated arrivals [38, 39].

- We show that the infinite capacity CIPP queue with exponential services, models well the queueing situation as encountered by *real-world* composite Ethernet traffic in a statistical multiplexer. We assume that the output queue of the statistical multiplexer has infinite capacity and the service times are exponentially distributed. The exponential distributions are assumed here to use the theoretical results for CIPP/M/1 queue developed above.

- By empirical study, we establish that the finite capacity CIPP queue with exponential services models well, the queueing situation as encountered by *real-world* composite Ethernet traffic in a statistical multiplexer with finite capacity output buffer. Queueing performance measures considered here are, packet loss probability and conditional packet loss probability. Also, we study the corresponding buffer occupancy distributions.

- Although the *phase plot* of CIPP, captures the diagonal lines as in the *phase plot* of Bellcore Ethernet traffic, it fails to capture some other features. This limitation is overcome by the Markov modulated CIPP (MMCIPP). We show by time domain methods and by wavelet analysis (all pertaining to estimation of Hurst parameter) techniques that, MMCIPP is self-similar to a wider time scale than CIPP. We also establish the above fact by analyzing the queueing performance with the synthetic traces generated by them as input. We also show that phase plot of MMCIPP captures features of the *real-world* Ethernet traffic, better than CIPP.

- We undertake the performance modeling of statistical multiplexer with *real-world* MPEG-1 VBR video traffic input and exponential services using CIPP/M/1 queue.

We use the queueing performance measures given above for our performance evaluation study [40].

- By empirical study, we establish that, finite capacity CIPP queue with exponential services models quite well, the queueing situation as encountered by *real-world* MPEG-1 VBR video traffic in a statistical multiplexer [41]. Queueing performance measures considered are packet loss probability and conditional packet loss probability. Buffer occupancy distribution is also compared.

- We undertake the empirical study of the fluid flow approximation of queues with CIPP input, to show that it models well the VBR video traffic, better than MC and histogram models reported in the literature.

- The MPEG-2 standard assumes that the end-to-end delay from an MPEG-2 system's encoder to decoder is constant to ensure that the encoder and decoder clocks operate at the same frequency and the decoder buffer does not overflow nor underflow. But, an ATM network introduces jitter while transporting the corresponding cells from source to destination, which may cause problems in decoder operations. The main source of the jitter is the statistical nature of the background traffic which are basically, multiplexed version of other highly variable individual component traffic. Thus, statistical characteristics of the MPEG coded cell stream of the VBR video traffic under study, depends on how variable (or precarious), the background network traffic process is, for the same mean rate of the background network traffic. More precisely, the jitter suffered by cells belonging to video traffic under study, largely depends on the degree of self-similarity of the background network traffic for deterministic service systems. Since there exists an empirical relationship between the ρ of CIPP and the estimated Hurst parameter H (a degree of self-similarity), CIPP forms a natural choice for the modeling the background network traffic. We undertake the simulation study of end-to-end transport of *real-world* MPEG-2 video traffic using CIPP network traffic [42]. The high ρ in CIPP model here accounts for the self-similar nature of background traffic. We show the interactions among ATM multiplexer buffer size, decoder buffer size, the degree of self-similarity of background traffic process in the network and the various performance measures including packet loss ratio.

1.5 Organization of the Thesis

The thesis is organized into 9 chapters. Here, an outline of the thesis is given.

Chapter 2 provides the necessary mathematical background material, needed to understand the methods used in later Chapters. To start with, an overview of the classification of stochastic processes which are of interest to queueing theorists is given. This allows one to understand and appreciate the dependence structure in the process, namely CIPP, addressed in this thesis and the non-renewal queueing theory presented in Chapter 5. An overview of self-similar processes and their wavelet analysis is given. A section is exclusively devoted for inference of self-similar processes, wherein, wavelet methods are also indicated. Some general comments about the subject of the chapter is included in the section on discussions.

Chapter 3 mainly concentrates on the analysis and estimation tools required for performance evaluation, current traffic models and other technical details necessary for broadband teletraffic modeling. A careful examination of *real-world* traffic traces would provide an intuitive feeling about the stochastic features we would like to capture, which is necessary for model building. For this reason, we first present the broadband teletraffic characteristics including the VBR video traffic characteristics and the WAN traffic characteristics. Then, we discuss, traffic modeling from a practical perspective, where a measurement based tool is described. We then give an overview of conventional traffic models from Poisson process, to self-similar models. We also discuss queueing performance measures which are all of practical interest. The chapter ends with discussions.

In Chapter 4, the Correlated Interarrival time Poisson Process (CIPP) is defined and its distribution is derived. It is justified for the VBR traffic and the composite Ethernet traffic. For that purpose, we have used the following measures based on second order properties, namely, (1) Autocorrelation of counts $\{u(i)\}$ (2) Index of dispersion for intervals (IDI) and (3) Index of dispersion for counts (IDC). Out of the above three, the CIPP has analytical expression only for IDI. Both the statistical properties of CIPP and properties which are relevant to modeling broadband teletraffic are evaluated. The expression for burstiness factor is evaluated for CIPP. The self-similar property of CIPP is explored. The empirical study of superposition of independent CIPP process is carried out, since the superposition process is analytically intractable. We have also derived the interval stationary version of CIPP.

The analysis of queues with CIPP arrivals has been considered in Chapter 5. First we consider single server FCFS queue with CIPP arrivals and exponential services (CIPP/M/1 queue). We derive the analytical expressions for stationary distributions for buffer occupancy and waiting time for the above infinite capacity queue. We derive the analytical expressions for the queueing performance measures, namely, mean queue length, mean waiting time, packet delay variation and overflow probability. We verify these theoretical results by simulating the CIPP/M/1 queue. Implications of the above theoretical results are discussed. To take into account the cell independent switching time of real-world situation of ATM switch, we consider the CIPP queues with deterministic service intervals (CIPP/D/1 queue). We use Benes approach, to obtain the approximate expression for buffer occupancy distribution. The (approximate) analytical expression for the buffer occupancy distribution poses a numerical problem for large values of ρ. Hence, we estimate the expressions for stationary distribution, by suitable simulations. Since all realistic systems have finite storage, we consider finite capacity CIPP queues. We carry out a simulation study of the CIPP/M/1/K queue, since the corresponding analysis, seems intractable. The performance measures investigated are packet loss probability and conditional packet loss probability. We demonstrate that, the buffer occupancy distribution for CIPP case, is consistent with those observed in literature for correlated arrivals. We give a numerical example to demonstrate the effect of correlation in interarrivals on the queueing performance.

Chapter 6 mainly concentrates on the problem of performance modeling of statistical multiplexer with broadband teletraffic input (in particular, the composite Ethernet traffic), using CIPP queues. Specifically, we apply the queueing theory results developed in Chapter 5 to the problem of performance modeling of statistical multiplexer with broadband composite teletraffic input using CIPP/M/1 queue. For performance evaluation purposes, we use the queueing performance measures (corresponding to CIPP/M/1 queue) namely, the mean delay, the mean number of packets in the buffer, the packet delay variation (*pdv*) and the overflow probability. Then, we undertake the performance modeling of finite capacity (capacity K units) statistical multiplexer[4] with *real-world* composite traffic using CIPP/M/1/K queue. We use the simulation approach here, since, CIPP/M/1/K queue is analytically intractable. Queueing performance measures considered here are, packet loss probability, conditional packet loss probability. Buffer occupancy distribution is also compared for a queue with Bellcore Ethernet traffic input and the "best fit" CIPP/M/1/K queue. Although the *phase plot* of CIPP, captures the diagonal lines as in the *phase plot* of Bellcore Ethernet traffic, it fails to capture some other features. This motivates us to define Markov Modulated CIPP (MMCIPP), a vari-

[4]By "finite (infinite) capacity statistical multiplexer", we mean the capacity of output queue of statistical multiplexer is finite (infinite).

ant of CIPP. We show by time domain methods and by wavelet analysis (all pertaining to estimation of Hurst parameter) techniques that, MMCIPP is self-similar to a wider time scale than CIPP or MMPP. We also establish the above fact by analyzing the queueing performance of a queue with many shuffled versions of synthetic traces generated by MMCIPP as input. We show that MMCIPP captures features in the phase plot of Bellcore Ethernet traffic better than CIPP.

Chapter 7 discusses the application of queueing theory results of Chapter 5 for the performance modeling of statistical multiplexer with MPEG VBR video traffic input. Queueing performance measures considered here are mean waiting time, mean number of packets in the buffer, packet delay variation *pdv* and overflow probability. In order to apply the queueing theory developed in Chapter 5 for modeling purposes, one needs the data to be in the arrival instants format. For this purpose, we use the simulation methodology followed in [39] to *smooth out* an MPEG-1 VBR video trace. We carry out the performance modeling of the output queue of ATM multiplexer with a *smoothed* VBR video traffic input by CIPP/M/1/K queue. Finite capacity queue is considered to account for finite capacity of a realistic broadband switch. Queueing performance measures considered here are, packet loss probability, conditional packet loss probability. Buffer occupancy distribution is also compared for a queue with Bellcore Ethernet traffic and the "best fit" CIPP/M/1/K queue. Finally, we consider the fluid-flow approximation with CIPP arrivals to study the VBR trace driven simulations.

We study the effect of degree of self-similarity of the background traffic on a particular traffic stream, by modeling the background traffic as CIPP in Chapter 8. More specifically, we present a simulation based performance evaluation study of MPEG-2 video in ATM networks when subjected to interference effects by a cross traffic process (background network traffic) modeled using the CIPP model [42]. The empirical relationship of the correlation parameter ρ of CIPP to the Hurst parameter is exploited to capture the self-similarity of the background network traffic. Also, we show the interactions among ATM multiplexer buffer size, decoder buffer size, the degree of self-similarity of background traffic process in the network and the various performance measures including packet loss ratio.

The thesis concludes with Chapter 9 wherein suggestions for future work are also indicated.

Chapter 2

Background Material

As the methods and models used for the performance evaluation of integrated broadband networks (used in later chapters), rely heavily upon stochastic processes, a good background of the theory of stochastic processes is necessary. In this chapter, we give some background material on stochastic processes in the context of queueing theory. We also present self-similar processes and wavelets and its role in the inference for self-similar phenomena. Some of the definitions, notations and concepts introduced here, are used in subsequent chapters.

We merely state some fundamental definitions and concepts of a subclass of stochastic processes namely, the continuous time, discrete state (countable - may be infinite) stochastic processes to which counting processes form a subset. These stochastic processes may possess the Markov property throughout their evolution or only at the transition instants. If they satisfy Markovian property throughout, then, the *holding times* (the random time that the process or system stays in a state - also occasionally referred to as *state times*) are independent and exponentially distributed (but may not be *identically* distributed). Such discrete valued processes are called as *Markov chains.* If the discrete valued stochastic processes possess the Markovian property only at the transition instant, still retain the independence among their holding times, then we have *semi-Markov chains.* These semi-Markov chains have their holding times *arbitrarily* distributed. If in addition, these processes possess dependence among their holding times, then we have more general discrete state stochastic processes. The counting process developed in this thesis, forms a special case of such a discrete state, continuous time stochastic process with dependent holding times.

It has been observed that, the *real-world* Ethernet traffic (packets per unit time) exhibits scale invariance or self-similar behaviour [24]. Moreover, metropolitan area network (MAN) traffic [25], wide area network (WAN) traffic, and variable bit rate (VBR) video traffic [1, 26, 27], also exhibit self-similar behavior. Thus, the self-similar stochastic processes are of interest to a teletraffic engineer. The self-similar process was introduced by Mandelbrot and Van Ness [43] in the context of fractional Brownian motions in order to explain the "cumulated water flow" phenomena observed by Hurst [44]. They gave a rigorous mathematical definition for a self-similar process (what are called as *exact* self-similar process in teletraffic literature). But, in teletraffic practice, a specific case of exact self-similar processes, namely the second order self-similar processes are in widespread use.

Wavelet technique is an important analysis and estimation tool for self-similar processes. A result due to Wornell [45] (in a special reference to self-similar processes) renders wavelet techniques, be used for the estimation of Hurst parameter of any given dataset. A wavelet based estimator of the Hurst parameter (a measure of degree of self-similarity) developed by Abry *et al* [46] is shown to be an asymptotically unbiased estimator of the Hurst parameter with the capability of removing the effects of deterministic non-stationarities in the data.

2.1 Preliminaries of Stochastic Processes

In this section, we give a pictorial classification of various stochastic process of interest to a queueing theorist, from pure birth processes to a more general discrete state stochastic process with dependent holding times. This material will provide preliminaries, for understanding a counting process (with dependent interarrival times addressed in this thesis) - a generalization of Poisson process - which forms a subset of a more general discrete state stochastic process with dependent holding times. We will revisit this classification, when we introduce CIPP in Chapter 4.

2.1.1 Definition and Classification of Stochastic Processes

A stochastic process is a family of random variables $X(t)$ where the random variables are "indexed" by the time parameter t. The classification of stochastic processes depends on the following three quantities:

1. *state space*: Could be continuous or discrete. If the state space is discrete, then the process is often called as *chain*.

2. *index (time) parameter*: Could be continuous or discrete. If the transitions are allowed in finite or countable set of time epochs, then we have discrete time process (X_n). If it is continuous, then we have continuous time process $(X(t))$.

3. *dependency:* The statistical dependencies among random variables $X(t)$ for different values of the index parameter t. The truly defining feature of a stochastic process is the relationship of the random variables $X(t)$ or X_n to other members of the same family. One has to specify the complete joint distribution function among the random variables $\mathbf{X} = [X(t_1), X(t_2), \ldots]$ namely,

$$\mathbf{F}_{\mathbf{X}}(\mathbf{x}; \mathbf{t}) \triangleq P\{X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n\} \tag{2.1}$$

for all $\mathbf{x} = (x_1, x_2, \ldots, x_n), \mathbf{t} = (t_1, t_2, \ldots, t_n)$, and n. Note that, the statistical dependencies (within the random variables of a random process), may range from independence (as in white noise), memoryless (as in Poisson processes), with limited memory (as in renewal processes, semi-Markov processes and more general discrete state stochastic processes with dependent holding times), to long-range dependent (LRD), wherein the memory is infinite (as in self-similar stochastic processes).

Below we mention some of the useful types of stochastic processes, that are characterized by different kinds of dependency relations among their random variables. We review only the definitions for processes of our interest. For a more detailed review, see [47].

1. **Independent Processes.** A stochastic sequence $\{X_n\}$ is called independent process, if the random variables involved are independent, i.e.,

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}; \mathbf{t}) &= f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) \\ &= f_{X_1}(x_1; t_1) \ldots f_{X_n}(x_n; t_n) \end{aligned} \tag{2.2}$$

2. **Markov Processes**. Markov processes [48] are a class of stochastic processes, in which the random variables possess a simple form of dependency. A set of random variables $\{X_n\}$ forms a *Markov chain* (a Markov process with discrete state space) if the probability of next value (state) X_{n+1} depends only upon the current value (state) X_n and not upon any of the previous values. Expressed analytically, the *Markov property* may be written as,

$$\begin{aligned} &P\{X(t_{n+1}) = x_{n+1} | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_1) = x_1\} \\ &\qquad = P\{X(t_{n+1}) = x_{n+1} | X(t_n) = x_n\} \end{aligned}$$

where $t_1 < t_2 < \ldots < t_n < t_{n+1}$ and x_i is included in some discrete state space. The holding time in a continuous time Markov chain is exponentially distributed (in addition to being *independent*). This is the consequence of Markovian property [47]. Note that exponential distribution is the *unique* continuous distribution which is *memoryless*. Similarly, in discrete-time Markov chain, the holding time (number of consecutive time slots the process stays) is *geometrically* distributed. Geometrical distribution is the *unique* discrete distribution which is memoryless. This memoryless property is required of all Markov chains and restricts the generality of the processes one would like to consider. Note that, in a general discrete state stochastic process, if the holding times are *not* independent, the Markovian property *does not* hold (and hence, does not form a Markov chain), even though if the holding times are exponential.

3. **Birth-death Processes**. These may be either discrete or continuous-time processes in which the defining condition is that the state transitions take place between *neighboring* states only. If we denote the probability of making a transition to state j given that the process is currently in state i, by P_{ij}, then $P_{ij} = 0$ for $|j - i| > 1$. Note that, the holding times are exponentially distributed and independent, but may not be stationary.

4. **Semi-Markov Processes**. In a semi-Markov process, the holding times are arbitrarily distributed (as opposed to being exponentially distributed). Note that, at the transition instants, the process may behave just like an ordinary Markov chain and we have an *embedded* Markov chain at those instants. Similar definition for a discrete-time semi-Markov Process also holds.

5. **Random Walks**. A sequence of random variables $\{S_n\}$ is referred to as a random walk (starting at the origin) if

$$S_n = X_1 + X_2 + \ldots + X_n \quad n = 1, 2, \ldots \tag{2.3}$$

where $S_0 = 0$ and $X_1, X_2, \ldots$ is an *iid* (independent and identically distributed) sequence. Random walk may be interpreted as the location of a particle in some state space. In continuous time, discrete state random walk (state space is discrete and the common distribution for X_n is discrete), the transition probability p_{ij} of going from one state i to another state j will depend only upon the difference in indices $j - i$ (which we denote by q_{j-i}). The holding times between transition instants $\{\tau_i\}$, is altogether a different process. Here it is assumed that they form an independent sequence, but *not* identically distributed.

6. **Renewal Processes**. Renewal process is a counting process whose interarrival times are *iid*s with arbitrary distribution. We may think of Equation 2.3, as describing a renewal process in which the random variable S_n denotes the n^{th} arrival

instant and X_n denotes the interarrival time between $(n-1)^{th}$ and n^{th} arrival instants.

The relationship among the Markov processes, semi-Markov processes, and their special cases are illustrated for discrete state systems in Figure 2.1. The distribution of time between transitions (or *holding time* distribution), is denoted by f_τ. By saying "f_τ is memoryless", we mean, if it is a discrete-time process, then f_τ is a geometric distribution, whereas if it is a continuous-time distribution, then f_τ is an exponential distribution. Furthermore, it is implied that f_τ may be a function of both of the current state and the next state of the process.

The figure shows that the birth-death processes form a subset of Markov processes, which in turn form a subset of the class of semi-Markov processes. Similarly, renewal processes form a subset of random walk processes which themselves form a subset of semi-Markov processes. Moreover, there are some renewal processes that may also be classified as birth-death processes. Similarly, those Markov processes for which $p_{ij} = q_{j-i}$ (that is, where the transition probabilities depend only upon the difference of the indices) overlap those random walks where f_τ is memoryless. A random walk for which f_τ is memoryless and for which $q_{j-i} = 0$ when $|j-i| > 1$, overlaps the class of birth-death processes. If in addition to this last requirement, our random walk has $q_1 = 1$, then we have a process that lies at the intersection of all the five processes shown in figure. This is referred to as "pure birth" process; though f_τ must be memoryless, it can be a distribution which depends upon the state itself. If f_τ is independent of the state (thus giving the constant "birth rate") then we have a much celebrated process, the *Poisson* process, which plays a crucial role in queueing theory.

Note that, when one moves from center to periphery, the dependency as mentioned above ranges from memorylessness (the most benign situation, in which, when a queue is fed with a counting process from such a class of processes, the corresponding queueing theory is analytically tractable and queueing performance is benign and optimistic. Examples: Poisson process and the corresponding M/M/1 queue), limited dependent (things are little complicated, the stationary probabilities for the corresponding queue are still analytically tractable, but, computationally intensive, often recursive. Examples: renewal process and GI/M/1 queue, GI/GI/1 queue), to long range dependent (things are *most* complicated here, wherein till date no queueing theory results are available except that some bounds are obtained for a queueing measure [28, 49]. Examples: self-similar processes).

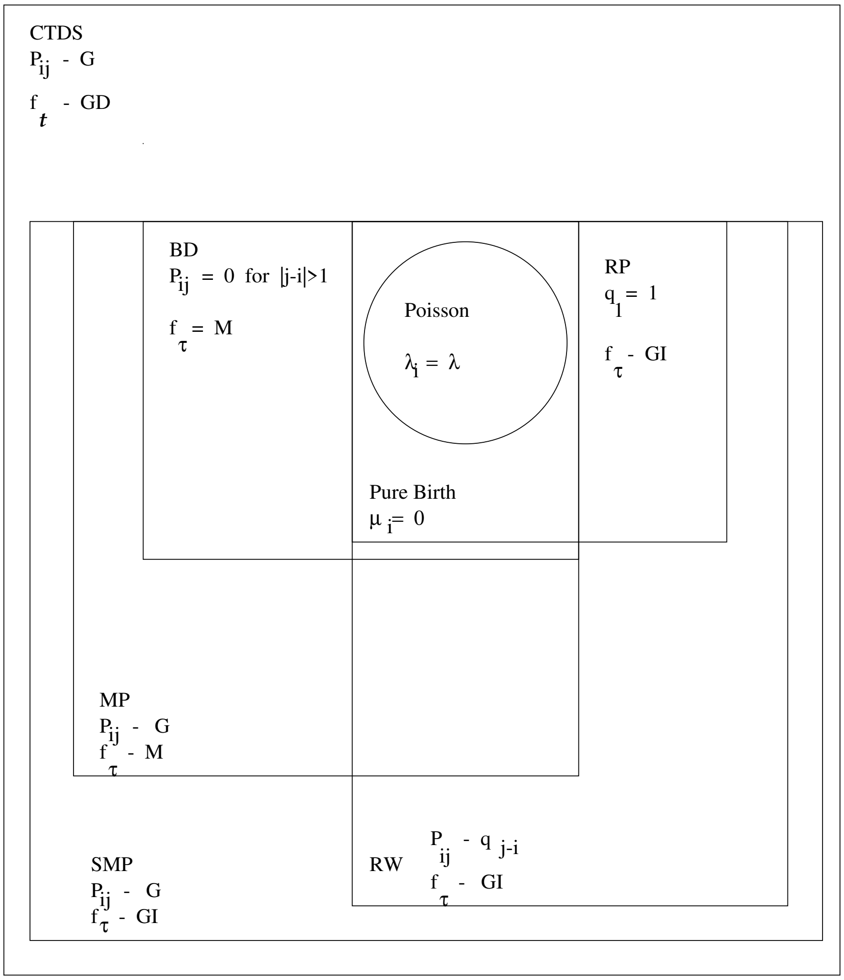


Figure 2.1: Relationship between various stochastic processes of interest to a queueing theorist: **Legend:-** CTDS - Continuous Time Discrete State Stochastic Processes. SMP - Semi-Markov Processes. MP - Markov Processes. BD - Birth-Death Processes. RW - Random Walk. RP - Renewal Processes. P_{ij} - Probability of transition from state i to state j. M - Memoryless distribution. GI - General and Independent. GD - General and Dependent. f_τ - holding time distribution.

2.2 Self-similar Stochastic Processes

Crudely, a stochastic process is said to be self-similar if it behaves in exactly the same way at all time scales, up to a scaling factor. Here, we present only definitions, concepts and their properties. For more details about self-similar processes, see the collection of articles [50] and the monograph [51].

2.2.1 Definitions

We give here the Mandelbrot's rigorous mathematical definition of self-similar processes [43], before introducing second order self-similarity which is of practical importance.

Definition: 2.1 The increments of a stochastic process $\{Y(t); \ -\infty \leq t \leq \infty\}$ is said to be *self-similar* with parameter $H \ (H \geq 0)$ if for any $h > 0$ and any t_0 and $(\tau > 0)$,

$$\{Y(t_0 + \tau) - Y(t_0)\} \ \stackrel{d}{=} \ \{h^{-H}(Y(t_0 + h\tau) - Y(t_0))\} \tag{2.4}$$

where, the notation $\{Y(t)\} \stackrel{d}{=} \{Z(t)\}$, is used to denote the random processes $Y(t)$ and $Z(t)$ have the same finite dimensional distributions. Here, the parameter H is called the Hurst parameter which denotes the degree of self-similarity.

When working with the second order properties only, one needs to deduce an equivalent definition for second order self-similarity. Note that the above definition states that the *increments* of a process $Y(t)$ being self-similar, whereas below, we talk about the self-similarity of the discrete-valued process $\mathbf{X} \triangleq (X_k : \ k = 0, 1, 2, \ldots)$. The process $\mathbf{X}$ can now be thought of as a time series generated from $Y(t)$ by noting down the increments of $Y(t)$ in fixed length, consecutive, disjoint intervals in the time axis. We use the definitions introduced in [28].

Let $\mathbf{X} = (X_k : \ k = 0, 1, 2, \ldots)$ be a *covariance stationary* (also called *wide sense stationary*) stochastic process with mean μ, variance σ^2 and autocorrelation function $r(k), \ \ k \geq 0$.

Definition: 2.2 The process $\mathbf{X}$ is called *exactly second-order self-similar* with Hurst parameter $H = 1 - \beta/2 \quad 0 < \beta < 1$, if its autocorrelation function is of the following form,

$$r(k) = \frac{1}{2}[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}] \tag{2.5}$$

for $k = 1, 2, \ldots$.

For each $m = 0, 1, 2, 3, \ldots$, let $\mathbf{X}^{(m)} = (X_k^{(m)} : k = 1, 2, 3, \ldots)$ denote the new covariance stationary time series (with corresponding autocorrelation function $r^{(m)}$) obtained by averaging the original series $\mathbf{X}$ over non-overlapping blocks of size m. That is, for each $m = \{1, 2, 3, \ldots\}$, $\mathbf{X}^{(m)}$ is given by $X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + \ldots + X_{km}), k \geq 1$. Note, that, every time a new $\mathbf{X}^{(m)}$ is constructed, it represents the increments of the process[1] $Y(t)$.

Definition: 2.3 The process $\mathbf{X}$ is called *asymptotically second-order self-similar* with parameter $H = 1 - \beta/2, \ \ 0 < \beta < 1$, if for all $k = 1, 2, \ldots$,

$$\lim_{m\to\infty} r^{(m)}(k) = \frac{1}{2}[(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}] \tag{2.6}$$

One can verify that [28], under the above definitions, an exactly second-order self-similar process with parameter $H = 1 - \beta/2, \ \ 0 < \beta < 1$ must have $r(k) \sim H(2H-1)k^{-\beta}$, whereas for an asymptotically second-order self-similar process with above Hurst parameter H, $r(k) \sim ck^{-\beta}$ for some constant c (which is not necessarily equal to $H(2H-1)$).

2.2.2 Properties

Mathematically, self-similarity manifests itself in a number of equivalent ways: (i) the variance of the sample mean decreases more slowly than the reciprocal of the sample size *(slowly decaying variances)*, i.e., $var(X^{(m)}) \sim \sigma^2 \ \ m^{-\beta}$, as $m \to \infty$, with $0 < \beta < 1$ and σ^2 is the variance of the original series $\mathbf{X}$ (ii) autocorrelations decay hyperbolically rather than exponentially fast, implying a non-summable autocorrelation function $\sum_k r(k) = \infty$ *(long-range dependence)*, (note that $r(k)$ given by (2.5) is non-summable) and (iii) the spectral density $f(.)$ obeys a power-law near the origin ($1/f$-noise), i.e., $f(\lambda) \sim c_2|\lambda|^{-\gamma}$, as $\lambda \to 0$, with $0 < \gamma < 1$ and $\gamma = 1 - \beta$. c_2 is a constant. In fact, $f(\lambda)$ has the following expression,

$$f(\lambda) = c|e^{2\pi i\lambda} - 1|^2 \sum_{l=-\infty}^{\infty} \frac{1}{|\lambda + l|^{3-\beta}}, \quad -\frac{1}{2} \leq \lambda \leq \frac{1}{2} \tag{2.7}$$

where c is a constant given by the normalization, $\int_{-1/2}^{1/2} f(\lambda)\, d\lambda = \sigma^2$. The aggregated processes of (exactly or asymptotically) second order self-similar processes $\mathbf{X}^{(m)}$ possess a non-degenerate correlation structure, as $m \to \infty$. Finally, we note that for the self-similar processes with $H > 0.5$, there is a positive correlation among their values (page no: 84, [52]).

[1] In the context of traffic process, $Y(t)$ can be thought of as a counting process and $X_k^{(m)}$ then denotes the rate of the process at scale m.

2.3 Wavelet Analysis of Self-similar Processes

In this section, we present the application of the wavelet transform to the analysis of self-similar processes. For an excellent review of the wavelet transform, see [53].

Wornell proved [45, 54] that as long as the analyzing mother wavelet has at least one vanishing moment, the variance of the wavelet coefficients for a "self-similar" process behaves in the following manner:

$$Var(x_n^m) \;=\; \sigma^2 \; 2^{\gamma m} \tag{2.8}$$

or

$$log(Var(x_n^m)) \;=\; 2log(\sigma) + m \;\; log(2^{\gamma}) \tag{2.9}$$

where x_n^m is the wavelet coefficient at scale m. The constant γ is related to the Hurst parameter H by, $\gamma = 2H - 1$.

Thus, a plot of $log(Var(x_n^m))$ versus the scale m should be a straight line with slope $log(2^{\gamma})$ for a self-similar process. Abry *et al* [46] have shown that the time average of the square of the wavelet coefficients at a given scale ($|x_n^m|^2$) gives an estimate of the power spectral density of the analyzed process about the frequency $2^{-m}\nu_0$, where ν_0 is an arbitrary reference frequency fixed by the choice of the analyzing wavelet. More formally,

$$\hat{S}_{xx}(2^{-m}\nu_0) \;=\; \frac{1}{k_m}\sum_n |x_n^m|^2 \tag{2.10}$$

where k_m is the number of available wavelet coefficients at scale m (which in practical cases is always finite due to the finite length of the analyzed signal). Again, a log(variance) versus scale plot should turn out to be a straight line with the slope giving the Hurst parameter. This interpretation of the wavelet transform gives a convenient way of looking at the variance-scale plots as given in [46].

2.4 Inference for Self-Similar Phenomena

All the methods for testing and estimating the degree of self-similarity of a given data fall into two major categories: (i) time domain analysis and (ii) frequency domain analysis. The method based on R/S statistics, method based on variance of aggregated processes $\mathbf{X}^{(m)}$ and variance based estimation of parameter H belong to time domain methods, where as, maximum likelihood type estimates (MLE) and related methods based on *periodogram*, Whittle's approximate MLE approach for Gaussian or approximately Gaussian processes and the more recent wavelet based estimators [46] belong to frequency domain methods. Here, we discuss only those methods which are implemented in this thesis. For more details about the inference for self-similar phenomena, see [55].

2.4.1 Time Domain Methods

2.4.1.1 R/S analysis

R/S analysis or *rescaled adjusted range plot* is a graphical method of estimation of Hurst parameter. For a given set of observations $(X_k : \quad k = 1, 2, \ldots, n)$ with sample mean $\bar{X}(n)$ and sample variance $S^2(n)$, compute,

$$R(n)/S(n) = \frac{1}{S(n)}[max(0, W_1, W_2, \ldots, W_n) - min(0, W_1, W_2, \ldots, W_n)] \tag{2.11}$$

with $W_k = (X_1 + X_2 + \ldots + X_k) = k\bar{X}(n)$ for $(k \geq 1)$. Next step is to plot the $log(R(n)/S(n))$ versus $log(n)$. This results in the *rescaled adjusted range plot* (also called *pox diagram of R/S*). A least squares line is fit to the points of the R/S plot omitting the R/S analysis of the smallest values of n (as they are dominated by the short-range correlations). The slope of the regression line for these R/S samples is an estimate $(\hat{H})$ of the Hurst parameter.

2.4.1.2 Variance-Aggregation Level Plot

The *variance-time plots* are obtained by plotting $log(var(\mathbf{X}^{(m)}))$ against $\log(m)$ ("time") and by fitting a simple least squares line through the resulting points in the plane. We noted that for second-order self-similar processes, the variances of the aggregated processes $\mathbf{X}^{(m)}, \ \ m \geq 1$, decrease linearly (for large m) in log-log plots against m with slopes arbitrarily flatter than -1. A least squares line is fit to the data ignoring the the smaller values for m. Values of the estimate $\hat{\beta}$ of the asymptotic slope between -1 and 0, gives directly the estimate $(\hat{H})$ of the Hurst parameter, through the relation, $\hat{H} = 1 - \hat{\beta}/2$.

2.4.1.3 Variance Based Estimation of Parameter H

This procedure assumes that, the dataset is a realization of fractional Brownian traffic. We denote the fractional Brownian traffic observed over time interval $[0, T]$ by $A(T)$. Now consider the sums of squares,

$$S_k^2 = \sum_{i=1}^{2^k} \left(A(i2^{-k}T) - A((i-1)2^{-k}T) - \hat{m}2^{-k}T\right)^2 \tag{2.12}$$

for $k = k_0, k_1, \ldots, k_N$, where $\hat{m}$ is the corresponding mean rate. Note that this sum denotes the variance of the dataset, computed at different scales depending on the value of k. Note that bigger k corresponds to higher resolution (and a larger data set). Then, the parameter H can be estimated as

$$\hat{H}_k = \frac{1}{2}\, log_2 \left(\frac{2S_{k-1}^2}{S_k^2} \right) \tag{2.13}$$

Here, the estimator converges to the right values when k increases. However, this convergence is slow for high H since $S_k^2/2^k$ is then a poor estimator for the corresponding variance. For any stationary traffic process, the numbers $\hat{H}_k$ can be used to get a quick look at the dependency structure of the process. If the process is self-similar, the H_k's are approximately equal for all k's for which the variance estimator has reasonable accuracy.

2.4.2 Frequency Domain Methods

2.4.2.1 A Wavelet based H Estimator

A wavelet based Hurst parameter estimator has been developed by Abry *et al* [46], which is shown to be unbiased under very general conditions and is efficient under Gaussian assumptions.

Definition of the Estimator The coefficient $|d_x(j,k)|^2$ measures the amount of energy in the analyzed signal about the time instant $2^j k$ and frequency $2^{-j}\nu_0$, where ν_0 is an arbitrary reference frequency selected by the choice of the analyzing mother wavelet ψ_0. Using the notation n_j to denote the number of wavelet coefficients at scale j, if one considers the spectral estimator,

$$\hat{\Gamma}_x(2^{-j}\nu_0) = \frac{1}{n_j} \sum_k |d_x(j,k)|^2 \tag{2.14}$$

which is basically a time average of the $|d_x(j,k)|^2$ at a given scale for a wide-sense-stationary process. This spectral estimator suffers the standard convolutive bias. But for LRD signals this bias reduces naturally to a simple form, enabling an unbiased estimation of H. It is, therefore, possible to design an estimator $\hat{H}$ for the parameter H by performing a weighted least squares fit between the scales (octaves) j_1 and j_2 yielding the following explicit formula for the estimator of H:

$$\hat{H}(j_1, j_2) \equiv \frac{1}{2} \left[\frac{\sum_{i=j_1}^{j_2} S_i \sum_{k=j_1}^{j_2} k\eta_k S_k - \sum_{i=j_1}^{j_2} i\, S_i\, \sum_{k=j_1}^{j_2} \eta_k S_k}{\sum_{i=j_1}^{j_2} S_i\, \sum_{k=j_1}^{j_2} k^2 S_k - \left(\sum_{k=j_1}^{j_2} k\, S_k \right)^2} + 1 \right] \tag{2.15}$$

where

$$\eta_j = log_2 \left(\frac{1}{\eta_j} \sum_k |d_x(j,k)|^2 \right) \tag{2.16}$$

and the weight $S_j = (n \;\; ln^2 2)/2^{j+1}$ is the inverse of the theoretical asymptotic variance of η_j.

2.5 Discussions

In this chapter, we have discussed the basic relationships among discrete valued stochastic processes for which counting processes form a special case. We have presented this classification, in order to give a global picture, which allows one to understand better, the region in which he is operating, as we proceed with our studies of CIPP (a non-renewal counting process) in Chapter 4 and the non-renewal queueing theory and associated performance implications in Chapter 5. The counting process introduced in this thesis (CIPP) forms a special case of a discrete state stochastic process with dependent holding times, very much like the Poisson process forming a special case of a Markov chain. We will come back to the main diagram 2.1 when we discuss CIPP in Chapter 4. We will see in subsequent chapters that, when one moves from independent case to long-range dependent case (defining the arrival processes), the variability in the arrival processes increases. Correspondingly, the queueing performance of the queue with such arrival processes, degrades (in fact, there are empirical evidence [24] that the queueing performance of a queue with a LRD process as input is *worst*). The above is true even in the case of simple Markovian service times for the queue.

We will see in later chapters that there is a close nexus among the positive correlation, self-similarity and queueing performance. For a self-similar process $X(t)$ with $H > 0.5$, the values of $X(t)$ are positively correlated [52]. We will see in Chapter 5, that the positive correlation degrades the queueing performance.

While the attractive feature of the R/S analysis is its relative robustness against changes of the marginal distribution, the wavelet based estimator is attractive for a variety of reasons: (i) the estimator is unbiased under very general conditions and efficient under Gaussian assumptions, (ii) highly robust against the presence of deterministic trends allowing their detection and identification, (iii) a trade off between the computational overhead and the accuracy of the estimate of Hurst parameter, by properly selecting the mother wavelet (characterized by the number (N) of vanishing moments) to be used in the estimator. In the model based estimators of Hurst parameter (variance based estimator given in Section 2.4.1.3), the estimation depends on the model assumed.

Chapter 3

Overview of Broadband Teletraffic Modeling

In this chapter, we discuss the techniques and methods pertaining to broadband teletraffic modeling, apart from a brief overview of traffic models addressed in the literature. First we give a brief account of broadband traffic characteristics. Then we discuss, traffic modeling from a practical perspective, where a measurement based tool is described. We present an overview of conventional traffic models discussed in the literature. Finally, queueing performance measures which are of practical interest are described. The chapter ends with discussions.

3.1 Broadband Teletraffic Characteristics

As mentioned previously, since the broadband networks support a variety of services, the traffic generated from these services differ by a wide range of statistical characteristics and bit rates. Understanding the statistical nature of *real-world* traces of such a traffic is necessary for coming up with a valid analytical model. This analytical model would be used to characterize the queueing performance and subsequently, the performance evaluation of these networks. In this section, we will review some of the empirical observations made by teletraffic researchers pertaining to *real-world* traces of individual sources (in particular, VBR video traces), traffic specific to a service (TELNET, FTP, SMTP, NNTP, x11 and WWW traces) and the composite traffic (Bellcore Ethernet traffic traces).

3.1.1 Video Source Characteristics and Requirements

In integrated broadband networks, a major part of the traffic will be generated by multimedia sources like video conferencing, video telephone, digital TV and video-on-demand servers [1]. The video output from most of these services, is MPEG coded. The output of the MPEG video encoder can be either constant bit rate (CBR) or variable bit rate

(VBR) video bit streams. By virtue of constant bit rate, CBR does not need complicated transmission mechanism. But, CBR suffers variable quality of service [56]. On the other hand, VBR delivers constant quality video at the receiver apart from achieving significant multiplexing gain at the cost of more complex transmission. Since, VBR is basically statistical, it is obvious that we need to model only the VBR video traffic. In this section, we discuss the statistical characteristics of VBR video traffic and their QoS requirements.

VBR Video Source Characteristics: Video has extremely high bit rate compared to both voice and data sources. In [16], the authors observe that the video traffic (video telephony) exhibits exponential correlation function. Huang [57, 58], observes that the autocorrelation of video traffic is *not* exponential. Furthermore, Joseph *et al* [59] conclude that the bit rate statistics of video is not even stationary. Finally, Beran *et al* [26] note the fractal-like behaviour in VBR video traffic.

QoS Requirements: The delay requirements are very stringent for video sources. For frame rates of 25-30 frames per second, the delay requirement is within 30-40 ms (minus the time required to decode, uncompress and play out the stream at the receiving end). The requirements for uncompensated cell loss ratio for video are in the order of 10^{-7} to 10^{-10} [60].

3.1.2 WAN Traffic Characteristics

In [32], the authors argue convincingly that, Poisson processes are valid *only* for modeling the arrival of user sessions in wide area network (WAN) traffic (TELNET connections, FTP control connections and RLOGIN connections). However, they fail to accurately model other WAN arrival processes, like packet arrivals. It intuitively makes sense as these connection arrivals (TELNET, FTP and RLOGIN) are user initiated sessions and generally any user initiated sessions are well modeled by Poisson model, in contrast with the machine initiated arrivals.

Other protocols' arrivals namely, SMTP, NNTP, WWW and x11 are *not* well modeled as Poisson. The fact that SMTP and NNTP are not well modeled by Poisson, is not surprising since both types of connections are machine initiated and can be timer driven. Consecutive SMTP interarrival times show consistent *positive* correlation. It is believed that, the reason for x11 connection arrivals deviating much from Poisson distribution is that, during a single x11 *session* (corresponding to running a *xterm*, say) a user initiates multiple x11 connections, while TELNET and RLOGIN sessions are comprised of single

TCP connections. The WAN packet arrival processes appear to be modeled better using self-similar processes.

3.2 Traffic Modeling - A Practical Viewpoint

The whole of statistical modeling in general, involves a theoretical basis which attempts to explain sufficiently the hidden philosophy of a real-world process which one tries to model. Moreover, it allows the investigator to deduce the logical implications of this theoretical model. Here we present in brief, the techniques of traffic modeling and succinctly recall the concept of the measurement based tool for management and control of networks addressed in [61].

3.2.1 Traffic Modeling Methodology

In the following we give the key steps involved in the process of traffic modeling.

1. *Propose a model:* The proposed model could be based on the theoretical beliefs of the investigator, after a thorough investigation of the real-world phenomena. Here, the real-world phenomena is the teletraffic process (could be the Ethernet traffic process) and the model could be any arrival process. The validation of the model is evaluated critically by statistical methods. For instance, statistical measures can be used to determine how well the model "resembles" the real-world process.

2. *Estimation of parameters:* The next step is to estimate the parameters of the model, given the real-world data. This is achieved in two steps: estimation of statistical measures (first/higher order statistics) of the data and next, matching the moments of the analytical model and that of data, to get the parameters of the model. In traffic modeling problem, this amounts to the estimation of first/higher order statistics of interarrival sequence (or alternatively that of the process of counts).

3. *Performance evaluation:* The final step is to evaluate the performance of the proposed model in the real-world situation. In the context of traffic modeling, the real-world situation could be the output queue in a statistical multiplexer. (Usually, the output queue of a statistical multiplexer is modeled as a single server queue with deterministic services). If the model is analytically tractable, then queueing performance measures can be obtained. Otherwise, the queueing performance

measures are evaluated by performing the trace (generated from the model) driven simulations.

One of the end objective of teletraffic modeling is real-time control of networks. This is achieved by a "measurement based system" which involves three steps: real-time traffic measurement, estimation of parameters of the model (by matching some of its moments) and queueing solutions (see Figure 3.1). The queueing performance solutions obtained are used for real-time control of the networks.

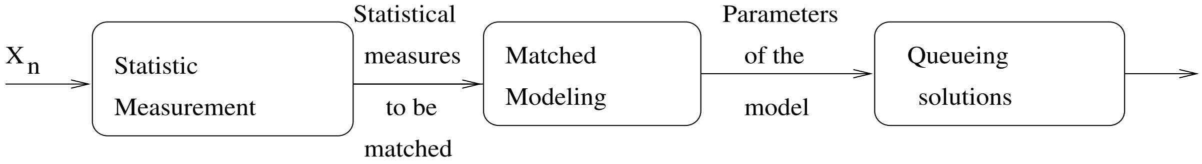


Figure 3.1: A measurement based system

3.3 Current Traffic Models

The traffic models addressed in the literature can be classified into three major categories: Independent, SRD and LRD models. The increments in independent models are independent. Examples are the Poisson process and the Bernoulli process. In SRD, the dependence is limited to a finite time and the autocorrelation decays exponentially fast. Typical SRD models are MMPP, AR(1) and DAR(1) models. In LRD processes, the dependence persists throughout the time axis. Examples: self-similar processes.

The fluid flow model belongs to SRD category, since it is different from all other models conceptually, it needs a special mention. The fluid flow model dispenses with the discrete nature of packets arriving into the network. Instead, it views the traffic as a rate process (rather, as a stream of fluid), so that the traffic count is replaced by a traffic volume [13]. The advantages of fluid model are: (1) conceptual simplicity (2) negligible computational overhead and (3) feasibility of infeasible cell traffic simulations (which involves vast CPU and memory resources) by conventional methods.

We would briefly review the Neuts process (which encompasses most of the SRD models) and some self-similar models addressed in literature. For a more elaborate list of teletraffic models and their detailed description, refer [7, 31]. For details about fluid flow models and their applicability, refer [13, 16, 62].

3.3.1 *Neuts* Process and its Special Cases

The versatile Markovian model developed by Neuts [8] has great appeal in its applicability to real-world situations especially those involving some kinds of fluctuations of arrivals by renewals. This modulating feature renders this class of processes, some correlations in the arrival process. Its formulation includes Poisson processes, phase type renewal processes, MMPP and others as special cases. It also has the appealing property that the superposition of independent Neuts traffic streams results in a traffic stream of the same class governed by a Markov process, whose state space is the cross product of the component state spaces. Moreover, it is analytically tractable [9]. The purpose of this section is to summarize the definition of the Neuts process, give its special cases and add a few words about MMPP - an important special case. A detailed treatise on Neuts process, its properties and related queues can be found in [8, 9, 11]. In order to define Neuts process, we need Phase-type distribution, which we describe now. We use the same notation as in [8].

Phase-type Distribution: Consider a (n+1) state, continuous time Markov chain with n transient states and one absorbing state. Let its infinitesimal generator be Q. Then,

$$Q = \begin{vmatrix} T & \mathbf{T}^{\circ} \\ \mathbf{0} & 0 \end{vmatrix} \tag{3.1}$$

where, T is an $n \times n$ matrix, with $T_{ii} < 0, \quad T_{ij} \geq 0$, for $i \neq j$ and such that T^{-1} exists. The column vector $\mathbf{T}^{\circ}$ has non-negative entries and satisfies $T\mathbf{e} + \mathbf{T}^{\circ} = 0$ where $\mathbf{e} = (1, \ldots, 1)^T$. A row vector (α, α_{n+1}) of initial probabilities is also given and satisfies $\alpha\mathbf{e} + \alpha_{n+1} = 1, \quad 0 \leq \alpha_{n+1} < 1$. The probability distribution $F(.)$ of the time till absorption in the state $(n+1)$ is given by,

$$F(x) = 1 - \alpha exp(Tx)\mathbf{e}, \text{ } for \text{ } x \geq 0 \tag{3.2}$$

The probability distribution function $F(.)$ is called PH-distribution and the pair (α, T) is called a *representation* of $F(.)$. Here, usually it is assumed that $\alpha_{n+1} = 0$, so that $F(.)$ does not have transition at 0.

The Markov Chain Q^*: Before defining the Neuts process, we need a Markov chain Q^*, constructed from Q. We consider the matrices T° with $T^{\circ}_{ij} = T^{\circ}_i$ and A° with $A^{\circ} = diag(\alpha_1, \ldots, \alpha_m)$. The infinitesimal generator, $Q^* = T + T^{\circ}A^{\circ}$ describes the Markov chain, obtained by resetting the original chain Q instantaneously using the same initial probabilities, whenever an absorption into state $(m+1)$ occurs. It is assumed that Q^* is *irreducible* if neccessary after deletion of superfluous states form the chain Q.

Given the Markov chain Q^*, a transition from the state i to the state j, which does not involve a renewal (i.e., no visit to the obsorbing state $(n+1)$) will be called an (i,j) - *transition*. A transition from i to j, which involves a renewal, will be called an (i,j) - *renewal transition*. Note that a renewal transition may involve (i,i) transition.

Definition of Neuts process: We now define the Neuts process as below. In Neuts process, multiple arrivals are allowed and the terminology of *group arrivals* is used to denote them. There are three types of arrival epochs,

1. During any sojourn time of the Markov process Q^* in the state i, $1 \leq i \leq m$, there are Poisson arrivals of rate λ_i and group density $\{p_i(k),\ k \geq 0\}$. We shall denote the probability generating function of $p_i(k)$ by $\phi_i(z)$ and we may assume without loss of generality that $\phi_i(0) = 0$, for $1 \leq i \leq n$.

2. At (i,j) renewal transitions, there are group arrivals with probability density $\{r_{ij}(k),\ k \geq 0\}$. The generating function of the density $\{r_{ij}(k)\}$ is denoted by $\Phi_{ij}(z)$, for $1 \leq i,\ j \leq m$. The matrix $\{\Phi_{ij}(z)\}$ is denoted by $\Phi(z)$ and we allow $\Phi(0)$ to be positive for some (i,j).

3. At i,j transitions, there are group arrivals with probability density $\{q_{ij}(k),\ k \geq 0\}$. The generating function of the density $\{q_{ij}\}$ is denoted by $\Psi_{ij}(z)$, for $i \neq j$. We shall find it convenient to define $\Psi_{ij}(z) = 1$, for $1 \leq i \leq n$. The matrix $\{\Psi_{ij}(z)\}$ is denoted by $\Psi(z)$. We do allow some of the $\Psi_{ij}(0)$, $i \neq j$, to be positive.

Special cases of Neuts process:

- **Phase-type Renewal Processes**
  An important special case of renewal models occurs when the inter-arrival times are of phase type (PH - distribution). In Neuts process if $\Phi(z) = zE$ where, $E_{ij} = 1$ and $\Psi_{ij}(z) = E$ and $\lambda = \mathbf{0}$, then what we have is phase-type renewal arrival process. In words, we have single arrivals at phase-type renewal epochs.

- **Markov Modulated Poisson Process (MMPP)**
  If $\Phi(z) = \Psi(z) = E$, and $\phi_i(z) = z$, for $1 \leq i \leq n$, where $E_{ij} = 1$ we obtain the familiar Markov modulated Poisson process (MMPP) model. In this case, at the state k of the Markov chain Q^*, arrivals occur according to a Poisson process at rate λ_k. The MMPP has been extensively used for modeling the broadband teletraffic since it qualitatively models the time-varying arrival rate and captures correlations while still remaining analytically tractable (despite its computational complexity).

- **Interrupted Poisson Process (IPP)**
  In IPP, Poisson arrivals occur on alternating, exponentially distributed intervals. Heffes [5], considered IPP for modeling packet arrivals in digital telephony. This corresponds to the particular choice,

$$T = \begin{vmatrix} -\sigma_1 & \sigma_1 \\ 0 & \sigma_2 \end{vmatrix} \tag{3.3}$$

$$T^\circ = \begin{vmatrix} 0 \\ \sigma_2 \end{vmatrix} \tag{3.4}$$

  , $\alpha = (0,1)$, $\lambda_l = \lambda$, $\phi_l(z) = z$, $\lambda_2 = 0$, $\phi_2(z)$ arbitrary,

$$\Phi(z) = \Psi(z) = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} \tag{3.5}$$

- **Markov Arrival Process**
  In Markov arrival process, arrivals occur at transition instants of Markov chain. This corresponds to the particular choice $\lambda = \mathbf{0}$, $\mathbf{\Phi(z)} = \mathbf{zE}$, and $\Psi_{ij}(z) = z$, for $i \neq j$.

3.3.2 Self-similar Models

There are two formal mathematical models that yield elegant representations of the self-similar phenomenon, namely, *fractional Gaussian noise* and the class of *fractional autoregressive integrated moving-average (ARIMA) processes.* The above processes are pure self-similar processes in the sense of Mandelbrot as given in Section 2.2.1. Since (pure) self-similar processes are *not* analytically tractable, researchers were trying to define a process which possesses the self-similarity in a very weak sense namely, asymptotically second order self-similarity (def. 2.3) and develop the corresponding queueing theory. This was successfully attempted by Tsybakov *et al* [28], wherein they define a process, which is asymptotically second-order self-similar. The authors give bounds for some queueing measures. Apart from these, there are a number of other methods reported in literature, which generate self-similar traffic [24, 32]. Here, we will not discuss such empirical models, rather we will give a short review of the two "formal" self-similar models mentioned above.

Fractional Gaussian noise $X = (X_k : \ k \geq 0)$ with parameter $H \in (0,1)$ [43] is a stationary Gaussian process with mean μ, variance σ^2 and autocorrelation function $r(k) = (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})/2,\ k > 0$. One may verify that, the fractional Gaussian noise is strictly second-order self-similar for $1/2 < H < 1$.

Fractional ARIMA (p, d, q) *processes* are generalizations of the traditional ARMA and ARIMA processes [63]. To define fractional ARIMA(p, q, d), consider the discrete time random walk $X_t = \sum_1^t a_k$, where a_k is a Gaussian white noise process. Denoting by B, the backward shift operator $BX_t = X_{t-1}$, a fractional ARIMA(p,d,q) process is formally defined by the equation

$$\phi(B)(1-B)^d\, X_t = \psi(B)a_t$$

where the fractional difference operator $(1-B)^d$ is defined by the binomial series

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \tag{3.6}$$

and ϕ and ψ are polynomials of orders p and q, respectively. These processes were introduced by Granger and Joyeux [64] and Hosking [65] who showed that fractional ARIMA (p, q, d) processes are asymptotically second-order self-similar with Hurst parameter $1/2 < H < 1$ where, $H = d + 1/2$ and $0 < d < 1/2$.

Queueing Theory for Self-similar Processes: The behaviour of a queue with self-similar traffic input is not very well understood till this date. But, only an approximate formula for the buffer occupancy distribution in the queue with fractional Brownian motion as input is known. Norros [49] has derived the following asymptotic lower bound for the complimentary probability,

$$P(Q > x) \approx exp(-cx2 - 2H) \tag{3.7}$$

with

$$c = \frac{n^{2H-1}\left(\frac{1-\rho_u}{\rho_u}\right)^{2H}}{2a}\left[\left(\frac{1-H}{H}\right)^{H} + \left(\frac{H}{1-H}\right)^{1-H}\right]^2 \tag{3.8}$$

Here, ρ_u is the utilization (ratio of arrival rate to link capacity). Duffield *et al* [66] have shown that the Norros bound is asymptotically tight. Tsybakov *et al* [28] have derived an exact expression for stationary distribution of cell delay of an asymptotically second order self-similar process. They also give some results on the merging of second order self-similar processes.

3.4 Queueing Performance Measures

The queueing performance measures of much relevance for broadband networks are derived from the following distributions namely, delay, cell loss, conditional cell loss and overflow. *Delay distribution* is the distribution of the random time from the moment a cell arrives until its service is complete. *Overflow probability* is the probability (or the proportion of time) that the number of cells in the buffer exceeds a certain threshold, in an infinite capacity queue. *Cell loss probability* is the probability that an arriving cell is lost when the buffer is full. *Conditional cell loss probability*[1] is the probability that an arriving cell is lost given that previous k cells are lost.

[2]. Since the above are probabilities, we need to talk about their first order or higher order moments for computational convenience, while estimating parameters of the proposed model. We now give the queueing performance measures mean delay, cell delay variation(cdv)[3], overflow probability for a given threshold and cell loss rate for a given buffer size.

In this thesis, for our purpose of performance modeling, we use the above queueing measures generalized to packets (variable sized information units). Thus, we use a slightly twisted terminology for the above measures namely, mean waiting time, mean number of packets in the buffer, packet delay variation and overflow probability. These performance measures will be used for for our estimation and performance modeling purposes in subsequent chapters. For the finite capacity queues, we use the packet loss probability and conditional packet loss probability as queueing performance measures.

3.5 Discussions

In this chapter, we have discussed the source characteristics, WAN traffic characteristics and traffic modeling aspects. The empirical observation of VBR video traffic has revealed that it is highly correlated. This correlation is believed to be arising out of the correlations within the frame and among adjacent frames even after encoding (which uses compression algorithms). We have also noticed that all user initiated TCP session arrivals are modeled reasonably well by the Poisson process. But, within a session, the packet/cell arrivals deviate much from the Poisson process. This empirical observation has lead researchers to consider arrival processes with correlations in their increments and develop corresponding queueing theory. The process proposed in this thesis (namely the CIPP), can be used

[1]This is slightly different from the definition of *burst loss length* as used by Chan *et al* [30]. Nevertheless, one can be obtained from the other.

[2]This is slightly different from the *burst loss length* as used by Chan *et al* [30]. Nevertheless one can be obtained from the other.

[3]Some publications use delay jitter to denote a related quantity. Delay jitter is the difference between minimum delay and worst case delay, whereas cell delay variation is the variance of delay [67, 68]. This definition of cell delay variation (or packet delay variation) is used in this thesis.

as a full fledged model in the measurement based system shown in Section 3.2, as it is amenable to analysis (as shown in Chapter 5). The applications of such a system is obvious in real-time control of broadband networks. The Neuts process is a more generalized counting process to which many other processes form a special case. Note that, in Neuts process there are renewal instants over which the process regenerates itself. This is in sharp contrast with the CIPP case. We will discuss this in Chapter 4. Since the ARMA and ARIMA processes are short range dependent, the fractional ARIMA processes are capable of simultaneously modeling the short term and long term behaviour of a time series than fractional Gaussian noise (fGn). The reason for fGn not being capable of capturing the wide range of correlation is that it has only three parameters μ, σ^2 and H and thus possess very rigid correlation structure.

Chapter 4

The CIPP Model, Properties and its Modeling Applications

In this chapter, we derive a counting process with correlated interarrivals and justify the modeling of VBR video traffic and composite Ethernet traffic. We investigate its statistical properties and other relevant properties for broadband traffic modeling.

Motivated by the empirical observation that the real-world traces show considerable correlation in interarrivals, we would like to define a counting process which inherits all nice properties of the Poisson processes but, has correlated interarrivals and is amenable to analysis. In light of the above, we derive a process (here in called as Correlated Interarrival time Poisson Process (CIPP)) with correlation in interarrivals. This is accomplished by first generating the random points on the real line (such that the intervals between consecutive points are exponentially distributed and are correlated) and then constructing a counting process over them such that, the increments of the resulting counting process are stationary. For the increments to be stationary, one has to select the interval in which origin lies, in a certain probabilistic way. This method has been used to construct the (stationary) CIPP. We will see that the CIPP is parameterized by a correlation parameter 'ρ', which represents the degree of correlation in adjacent interarrivals in addition to 'λ', the intensity of the process.

We propose CIPP for modeling both composite arrival process of packets in broadband networks as well as for individual traffic source behavior. We consider the *real-world* MPEG-1 VBR video traffic traces and the composite Bellcore Ethernet traffic for our experiments. We use the following measures namely, the index of dispersion for counts (IDC), the index of dispersion for intervals (IDI) and the autocorrelation function for our fitting procedure. We then investigate the statistical properties of CIPP, which are used in later chapters for the estimation of parameters of the process. A notable feature of CIPP is its correlational structure in which the memory of the process extends up to the arrival instant previous to the latest arrival instant. We believe that this feature is responsible for the self-similar property of CIPP. In sequel, we discuss the position of CIPP in Figure 2.1.

We give some of the important properties of CIPP which are of interest in the context of teletraffic modeling. We evaluate the *burstiness factor* [35] for CIPP. We examine the widely addressed property in broadband teletraffic literature, namely, the self-similar behavior, corresponding to CIPP by various methods. First we give the "pictorial proof" of self-similar behavior. In order to explore the relation between the positive correlation in interarrivals and the degree of self-similarity, we use the wavelet based estimator for the Hurst parameter (H) proposed by Abry *et al* to get an empirical relationship between ρ, the correlation parameter of CIPP and H. We then present the autocorrelation function of a simulated CIPP trace at various scales. We study a variance based method of investigating the degree of self-similarity namely the log(variance) versus log(aggregation level) plot. Finally, we give the wavelet analysis of CIPP process, which explores the log(variance of wavelet coefficient at a scale) versus log2(scale) plot. We study the self-similar property of superposed process of two independent CIPPs, by estimating the Hurst parameter.

Note that CIPP was basically derived by generating correlated exponential sequence (specifically, exponential auto-regressive process (EAR(1)) pioneered by Gaver and Lewis [69]) for interarrivals and constructing counting process over the sequence of arrival instants thus generated. There are other methods of generating correlated sequences addressed in the literature namely the exponential moving average (EMA(1)) process (having exponential marginals) [70] and the one given by Finch (having gamma marginals) [71]. A short review of EMA(1) and Finch model are included here to show that the correlation structure in them is weak (The weak correlation is a consequence of the moving average structure) and thus poorly models the broadband traffic. We give here the phase plot corresponding to CIPP and compare with EMA(1) and Finch model. Finally, we clarify that CIPP is *not* a special case of Neuts process.

The chapter is organized as follows: Section 4.1 gives the related literature survey and the derivation of CIPP. We justify the CIPP model for broadband teletraffic for both composite Ethernet traffic and MPEG-1 VBR video traffic in Section 4.2. Section 4.3 discusses the statistical properties of CIPP. Correlational structure of CIPP and its implications are investigated in Section 4.4. Properties of CIPP pertinent to broadband teletraffic modeling are presented in Section 4.5. Other methods of generating correlated sequences are given in Section 4.6. Finally, we clarify that CIPP is not a special case of Neuts process in Section 4.8. Section 4.9, contains discussions on the chapter.

4.1 Correlated Interarrival time Poisson Process (CIPP)

In this section, we give the literature survey of related attempts to generalize the Poisson process. Then, we derive the distribution $P_n(t)$ for CIPP [72]. Some remarks are given to highlight some important features of the CIPP model.

4.1.1 Introduction

In Poisson probability law, the number of occurrences of an event in an interval is independent of the number of occurrences of the same event in some other interval, when the intervals are disjoint. This is a consequence of the interarrival times being independent and exponentially distributed. This principle of complete randomness, providing the Poisson distribution, may be an excellent ideal, but it is not very natural or practical for all situations. For example, teletraffic researchers have concluded after extensive empirical observation that there are correlations in the interarrivals of broadband teletraffic trace (We reproduce below the exact sentence, given in page 229 of [32], ".. If we require fixed rates only over 10 minutes intervals, then SMTP and FTPDATA burst arrivals are not terribly far from Poisson, though neither is statistically consistent with Poisson arrivals, and consecutive SMTP interarrival times show *consistent positive correlation*". While the above is Ethernet trace, a similar behavior has been observed even in encoded VBR video cell streams [30]). Hence, it is of interest to come up with a process, in which this randomness is removed, i.e., interarrivals are correlated.

Numerous researchers have *compounded, modified and generalized* the Poisson process [73]. Their attempts have been to manipulate λ (say, expressing it as a linear function of 't') and to get a model which suits a specific application. While the motivation for the Generalised Poisson Distribution(GPD) as introduced by Consul and Jain [74], is to explain the extraordinary value of frequency or unequal mean and variance in the numerical data in different fields, the motivation for correlated Poisson Distribution (CPD) introduced by Drezner [75] is to explain the correlations. They do not consider explicit assumption of correlation in interarrival times. But this has been attempted by Wold [76], Cox [77], and Gaver and Lewis [69]. Wold [76] has made an attempt to generalize the Poisson process by obtaining dependent, but marginally exponentially distributed intervals between events. Similarly, Cox [77] has defined a sequence of random variables with conditionally exponential distributions. Here, we assume correlation in interarrivals explicitly, and arrive at a distribution.

4.1.2 Axiomatic Derivation of CIPP

We wish to define a counting process $N(t)$ which retains as far as possible all the nice properties of the Poisson process, but has correlated interarrivals. Typically, we would like to have interarrivals being exponentially distributed, apart from being correlated. Motivated by the analytical simplicity of the method of generating stationary correlated exponential sequence (first order exponential auto-regressive process (EAR(1))) developed by Gaver and Lewis [69], we present the following formalism for defining CIPP.

Consider arrivals occurring in time on the interval $(0, \infty]$. For $t > 0$, let $N(t)$ be the number of arrivals that have occurred in the half-open interval $(0, t]$. Consequently, $N(t)$ and $\{N(t+h) - N(t)\} \quad (h > 0)$, assume only non-negative integer values. We now state the following axioms.

Axioms:

1. Since we begin counting arrivals at time 0, we define $N(0) = 0$.

2. **Correlation Structure for Interarrivals:** Let X_n be the interarrival time between the n^{th} arrival instant (T_n) and $(n-1)^{th}$ arrival instant (T_{n-1}). Then,

$$X_{n+1} = \rho X_n + \epsilon_{n+1} \qquad 0 \le \rho < 1 \quad n = 1, 2, 3 \ldots \tag{4.1}$$

   where, $\{\epsilon_n\}_{n=2}^{\infty}$ is an iid sequence, with ϵ_n, $n > 1$, being a product of Bernoulli random variable (B) with parameter ρ and exponential random variable (V) with parameter λ. B is statistically independent of V. $\{X_n\}$ forms a stationary sequence with the exponential distribution characterized by the parameter λ. Note that, $X_1 = T_1 - T_0$ where, T_0 is the 0th arrival instant.

3. The counting process $N(t)$ is strictly stationary, that is, for any $r = 1, 2, \ldots$ the joint distribution of $\{N(h+t_{u1}) - N(t_{l1}), N(h+t_{u2}) - N(t_{l2}), \ldots, N(h+t_{ur}) - N(t_{lr})\}$ is independent of h, for any $h > 0$ and $0 < t_{l1} < t_{u1} < t_{l2} < t_{u2} < \ldots < t_{lr} < t_{ur} < \infty$

Definition: A counting process $\{N(t), t > 0\}$ satisfying Axioms 1 through 3 is called Correlated Interarrival time Poisson Process (CIPP) with parameters λ and ρ.
Remark 1: The axiom 2 is *more specific* and the following properties of the counting process $N(t)$ directly follow from it.

1. In any interval (no matter how small) there is a non-zero probability that an arrival will occur. Mathematically, $0 < Pr\{N(t) > 0\} < 1$ for any $t > 0$.

2. For any $t \geq 0$,
$$\lim_{h \to 0} \frac{Pr\{N(t+h) - N(t) \geq 2\}}{Pr\{N(t+h) - N(t) = 1\}} = 0 \tag{4.2}$$
This means that in sufficiently small intervals, at most one arrival can occur, i.e., it is not possible for arrivals to occur simultaneously. The process is *orderly*, in the sense of Daley *et al* [78].

Remark 2: The main difference between the axiomatic definition of a Poisson process and CIPP, is the addition of Axiom 2 and the relaxation of the independence assumption in Axiom 3.

Let $P_n(t) \triangleq Pr\{N(t) = n\}$, be the probability that the number of arrivals in the interval $(0, t]$ is n.

Lemma 1 *CIPP with parameter λ and ρ has the distribution,*
$$P_n(t) = \begin{array}{l} (1+\rho-(1-\rho)\alpha_n) \\ \times \sum_{j=0}^{n} \frac{\alpha_j \prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\rho(\alpha_j+1)-\alpha_n+1)(\alpha_j-\alpha_n+1)(\alpha_j-\rho\alpha_n)\prod_{k=0,k\neq j}^{n}[\alpha_j-\alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \quad 0 \leq \rho < 1 \end{array} \tag{4.3}$$
for $\quad n = 0, 1, 2, \ldots\ , \ \lambda > 0, \ \ t \geq 0,$ $\quad$ *and* $\quad \alpha_j = \sum_{i=0}^{j} \rho^i$

Proof: See Appendix A for proof.

It is worth noting that the mean of the process, $\mathbf{E}\{N(t)\}$, can be shown to be λt [79].

Remark 3: The prime reason for assuming (4.1) for the interarrival sequence, is to mimic the self-similar behavior in increments for the corresponding counting process. Note that, to mimic a self-similar process with $0.5 < H < 1$ it is necessary that the increments should have positive correlations [80]. The reason for this particular correlation is that, (4.1) introduces strong positive correlation than other models based on moving average structure like EMA(1) proposed by Lawrance [70] and the one proposed by Finch [71] (see Section 4.6). We will see in later sections that, as ρ increases, the estimated local Hurst parameter $(0.5 < H < 1)$ increases. The second reason is that the correlation structure (4.1) allows one to derive a closed form expression for $P_n(t)$ with reasonable simplicity, which is very difficult if not impossible with the other aforementioned models. The third reason is the analytical simplicity of the correlation structure (4.1) which allows one to develop the corresponding queueing theory. Finally, the sequence $\{X_n\}$ is obtained as an additive random linear combination of random variables, and is thus easy to simulate on the computer.

Remark 4: The distribution (4.3) is count stationary[1] in the sense of Lawrance [70]. An interval stationary process (again in the sense of Lawrance) can be shown to be as

[1]same as the stationarity given in Axiom 3.

given below (See Appendix B for derivation). This process is constructed by choosing the origin at an *arbitrary event* (rather than choosing origin at an *arbitrary instant*).

$$P_n^I(t) = \sum_{j=0}^{n} \frac{\prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \alpha_n + 1)(\alpha_j - \rho\alpha_n)\prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \quad 0 \leq \rho < 1 \tag{4.4}$$

Remark 5: For both the cases (4.3 and 4.4), as the limit $\rho \to 0$, we obtain the Poisson distribution. Thus, the Poisson distribution is approached continuously from the CIPP distribution.

Remark 6: We believe that the models that capture the correlation in the interarrival sequence is more appropriate in the context of modeling broadband teletraffic than the models that capture the correlation in the counts sequence (as done in [18, 33] for VBR modeling) or capture correlation by the modulating Markov chain as in SRD models. Correlations in interarrivals *necessarily* imply correlations in the number of counts; however, the converse of this statement is *not* necessarily true. A good example would be a renewal process. Moreover, for an interarrival sequence forming first order Markov chain, the corresponding counts sequence (the sequence of number of arrivals in the fixed length, consecutive and disjoint intervals) possesses, in general, much stronger correlation structure (than the first order Markov chain). In this context, we would like to mention that, the discrete version of (4.1), namely the discrete auto-regressive (DAR(1)) is used in [18, 33] to model the *frame size* of VBR traffic. Here, the counts rather than interarrivals form a Markov chain.

Remark 7: Note that, the number of parameters is minimal in CIPP as compared to the conventional short-range dependent (SRD) models like MMPP, DBMAP, ON-OFF processes etc. For example in MMPP, the number of parameters is N^2 for N state MMPP ($N^2 - N$ for infinitesimal generator and N for the intensity matrix Λ) while in CIPP it is always 2 namely λ and ρ.

Remark 8: For many reasons, using an excessive number of parameters is undesirable. In particular, it increases the uncertainty of the statistical inference and the parameters are difficult to interpret. In our case, CIPP has only two parameters ρ and λ and as we mentioned, both of them have a nice physical interpretation. λ denotes the arrival rate, and ρ quantifies the degree of self-similarity. While λ can be estimated from the mean of the process, ρ may be estimated from higher order statistics.

Remark 9: There is always a convenient way of associating the parameter ρ of CIPP with the local Hurst parameter (a measure of degree of self-similarity) of the corresponding CIPP trace (see Section 4.5.3). On the contrary, in other SRD models, too many parameters decide the degree of self-similar behavior (to a range of time scales and hence pseudo-self-similarity) and one does not know which of them have crucial role in exhibiting self-similar behavior. The above feature of CIPP makes it more suited for practical purposes, say, when one needs to fit CIPP to a video sequence of given H parameter.

Remark 10: We derived the CIPP, by first generating arrival instants on the positive real line such that the interarrival time forms the correlated exponential sequence and then constructing a counting process over it. The whole exercise can be repeated in discrete time domain, by considering correlated geometric sequence generated by DAR(1) model [19]. But, the result does not seem to yield a closed form expression and will not be quoted any further in this thesis.

4.2 Validation of CIPP as a Teletraffic Model

In this section, we justify the CIPP model for modeling broadband teletraffic. In particular, we show that CIPP models well the *real-world* MPEG-1 VBR video traces and also the composite *Bellcore* Ethernet traffic. Specifically, we carry out the "goodness of fit" test of CIPP to the trace data based on various second order measures and compare it favorably with the other models addressed in literature.

4.2.1 Fitting the CIPP Model to Real Traffic Data

In order to show that CIPP models well the data, we have to "fit" the CIPP model to the data. This requires the parameters of CIPP be estimated such that this "best fit" CIPP in some sense behaves closer to the data (with respect to some statistical measures). This fitting procedure usually involves the estimation of first/higher order statistics of data and matching this with the corresponding moments of CIPP to get the parameters of CIPP. (Note that, here in CIPP case, one may use the elegant method of estimating the parameters (namely ρ and λ) as given in [69]. But, here, we would like to mention that, we are *more* interested in capturing the LRD of the real-world data. Hence, we use correlation measure in some form or other for our fitting procedure). In light of the above, we choose to estimate the parameters of the CIPP model with a measure based on any of the following second order properties [81], namely, (1) autocorrelation of counts sequence (2) index of dispersion for intervals (IDI) and (3) index of dispersion for counts (IDC). Out of the above three, the CIPP has analytical expression only for IDI. For the other two measures, one has to estimate by simulations.

We use the following notation. X_i is the interarrival time defined as the length of time between the instant when ith packet is transmitted and the instant when $(i-1)$th packet is transmitted. Note that under the above definition of X_i, the transmission time of the previous packet is included in the interarrival time. Thus, for networks that allow the variable packet sizes, this definition introduces a source of dependency between intervals and packet lengths.

4.2.1.1 Autocorrelation of Counts Sequence $\{u_T(i)\}$.

The time dependent statistics are crucial in the case of video traffic, because correlations in the video streams can affect the performance of the statistical multiplexer (in a typical broadband network). In particular, the positive correlation in the traffic process (which is the case with CIPP streams) degrades the performance of statistical multiplexer [80].

Let $\{u_T(i)\}_{i=1}^{\infty}$ denote the number of arrivals in i^{th} slot on the positive real axis where each time slot is of equal length, T time units. The sequence $\{u_T(i)\}_{i=1}^{\infty}$ is called as *counts sequence*[2]. Then the correlation coefficient for counts sequence is defined by,

$$\zeta_T(k) = \frac{E\{u_T(i)u_T(i+k)\} - E^2\{u_T(i)\}}{var(u_T(i))}$$

Note that for Poisson process $\zeta_T(k) = 0$ for all $k \neq 0$.

4.2.1.2 Index of Dispersion for Intervals (IDI)

We use the variance of the sum of n random variables normalized by the factor $nE^2(X)$, as a measure of the variability of packet arrival processes [81]. The sequence of values

$$J_n \triangleq \frac{var(X_{i+1} + \ldots + X_{i+n})}{nE^2(X)}$$

with $n = 1, 2, \ldots,$ is called *index of dispersion for intervals (IDI)*. It is shown in the Appendix C that for CIPP J_n is given by,

$$J_n = \frac{1}{n(1-\rho)^2}\left[n(1-\rho^2) + 2\rho^{n+1} - 2\rho\right] \tag{4.5}$$

Then,

$$J \triangleq \lim_{n\to\infty} J_n = \frac{1+\rho}{1-\rho} \tag{4.6}$$

Since the CIPP has an analytical expression for J_n, the modeling procedure becomes simple. We have to just estimate ρ from the data, using (4.5).

[2]Many approaches can be used to describe packet arrival process: through the *counting process* $N(t)$ or through the sequence of arrival instants $\{T_i\}_{i=0}^{\infty}$ or through the *counts sequence* $u_T(i)$. Here, we would like to mention that the first two descriptions are well defined and the *counts sequence* $u_T(i)$ can be derived from them for a given T. The counts sequence deserves primary importance throughout this thesis.

4.2.1.3 Index of Dispersion for Counts (IDC)

For packet counts, we can define a function similar to the index of dispersion for intervals [81]. The *index of dispersion for counts (IDC)* at time t is the variance of the number of arrivals in an interval of length t divided by the mean number of arrivals in t:

$$I_t = \frac{\mathrm{var}(N(t))}{E(N(t))} \tag{4.7}$$

where $N(t)$ indicates the number of arrivals in an interval of length t. The IDC has been defined in this manner so that for a Poisson process, the value of the IDC is 1, for all t.

In estimating the IDC of the measured arrival processes, we will only consider discrete, equally spaced time instants, $\tau_i \quad (i \geq 0)$. Denoting the number of arrivals in ith interval of duration $T = \tau_i - \tau_{i-1}$, by $u(i)$, we have,

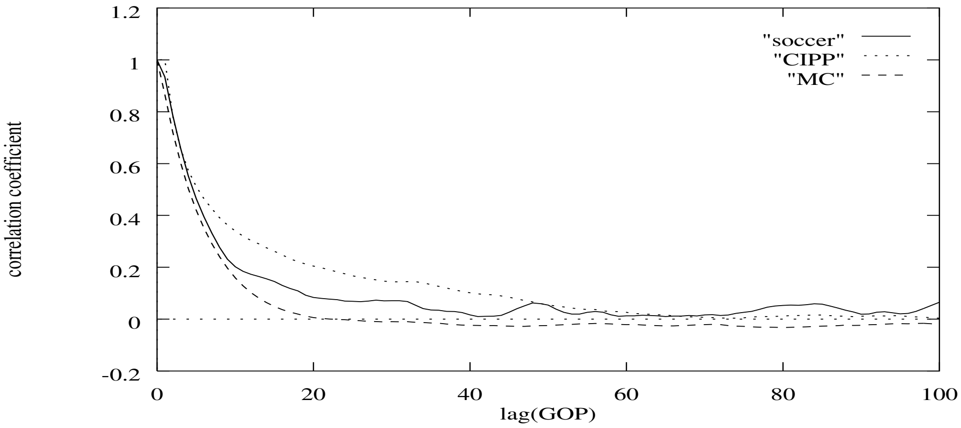
$$\begin{aligned} I_n(T) &= \frac{var\left(\sum_{i=1}^{n} u_T(i)\right)}{\mathrm{E}\left(\sum_{i=1}^{n} u_T(i)\right)} \\ &= \frac{var(u_T(i))}{E(u_T(i))}\left[1 + 2\sum_{j=1}^{n-1}(1-\frac{j}{n})\zeta_T(j)\right] \end{aligned} \tag{4.8}$$

where $\zeta_T(j)$ is the autocorrelation coefficient of $u(i)$ at lag j. Note that, in general, $I_n(T)$ will not be a constant for renewal processes where counts in the disjoint intervals are correlated (exception is the Poisson process). Note also that the IDC is a function of the length of time unit T, over which the counts sequence is constructed from the original arrival process (characterized by the arrival instant sequence). While we can estimate $\mathrm{var}(N_t)/\mathrm{E}(N_t)$, the representation on the RHS of (4.8) is valid only if the data is stationary.

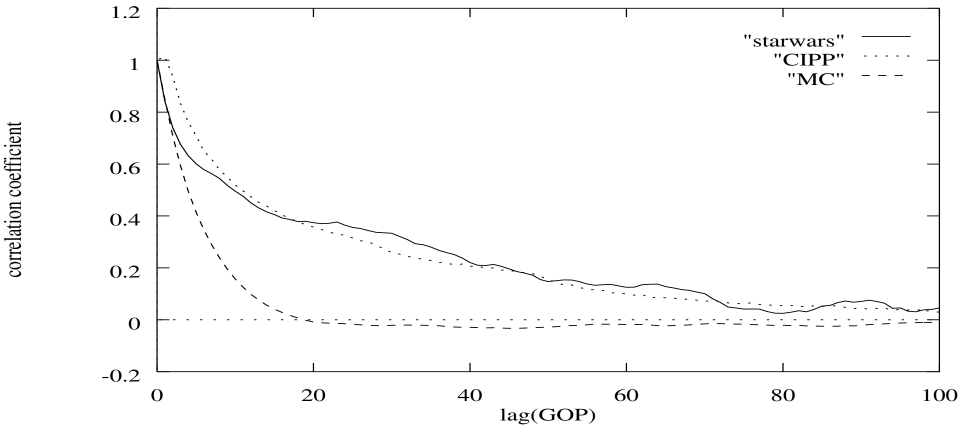
4.2.2 Experimental Results

4.2.2.1 Autocorrelation of Counts $\{u_T(i)\}$.

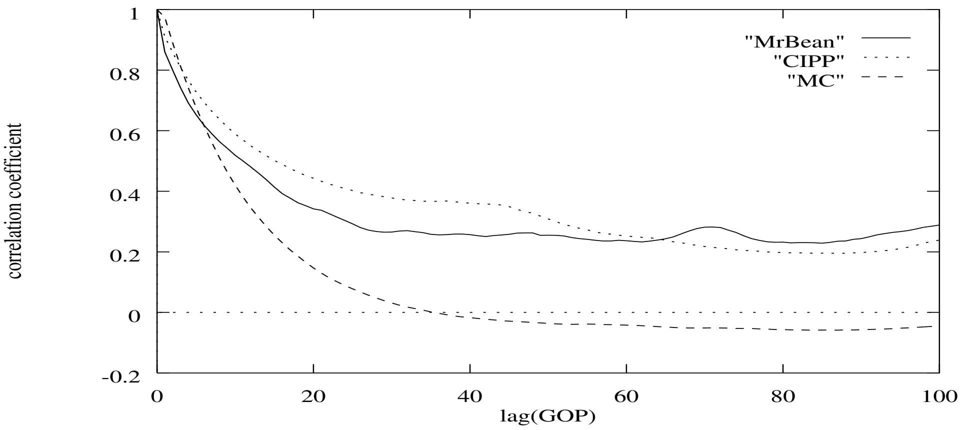
We now attempt to fit a CIPP model to MPEG-1 encoded video trace. GOP sizes (sum of frame sizes of one GOP) in the frame size trace from MPEG-1 encoded video sequence, can be thought of as the number of (cell) arrivals (after segmentation of frame size trace) $\{u_{T_{GOP}}(i)\}$ in a GOP time slot (T_{GOP}). We fit a CIPP by using the following algorithm.



(a)



(b)



(c)

Figure 4.1: Experimental check on the "goodness of fit" of the CIPP as compared to MC [1] model for various video sequences [2] with autocorrelation as a measure: (a) *soccer* video sequence (b) *starwars* video sequence (c) *MrBean* video sequence

---

Fitting procedure

1. Segment the bits in each frame into cells[3]. We then have the counts sequence $u_{T_f}(i)$ of cells.

2. Estimate the correlation coefficient of this counts sequence of cells. This is the correlation coefficient $\zeta^{data}_{T_{GOP}}(.)$ corresponding to data, to which we intend to match the correlation coefficient of the CIPP model.

3. **Algorithm to estimate parameters of CIPP:**
   Set the rate λ_{opt} of CIPP equal to $\mathbf{E}(u_{T_{GOP}}(i))/T_{GOP}$.

   Algorithm to estimate ρ_{opt} of CIPP:
   begin
       min = 99999999.9
       step = 0.0001
       for i = 1:9999
           $\rho = step * i$
           estimate the correlation coefficient $\zeta^{CIPP}_{T_{GOP}}(.)$ for CIPP by reasonably large realizations.
           compute the mean square error (P) between $\zeta^{CIPP}_{T_{GOP}}(.)$ and $\zeta^{data}_{T_{GOP}}(.)$
           if$(min > P)$
               min = P
               $\rho_{opt} = \rho$
               $\zeta^{fit}_{T_{GOP}}(.) = \zeta^{CIPP}_{T_{GOP}}(.)$
           endif
       end
   end

The correlation coefficient $\zeta^{fit}_{T_{GOP}}(.)$ is plotted along with the correlation coefficient corresponding to data $\zeta^{data}_{T_{GOP}}(.)$. For comparison, we also plot the result by fitting a Markov chain (MC) [1] to the GOP size process (Our procedure is very similar to [1]). The number of states M in the Markov chain is taken to be G_{max}/σ_G, where G_{max} denotes

[3]This is just to avoid the computational overflow while estimating the correlation coefficient of the data

the size of largest GOP and σ_G the standard deviation of the GOP sizes. Thus, the size of quantization interval is σ_G. Note that it is conventional to consider the autocorrelation function (or equivalently correlation coefficient) of the GOP sizes since, it is difficult to get a clear picture of the long-range correlations of video traffic stream, from frame level correlations.

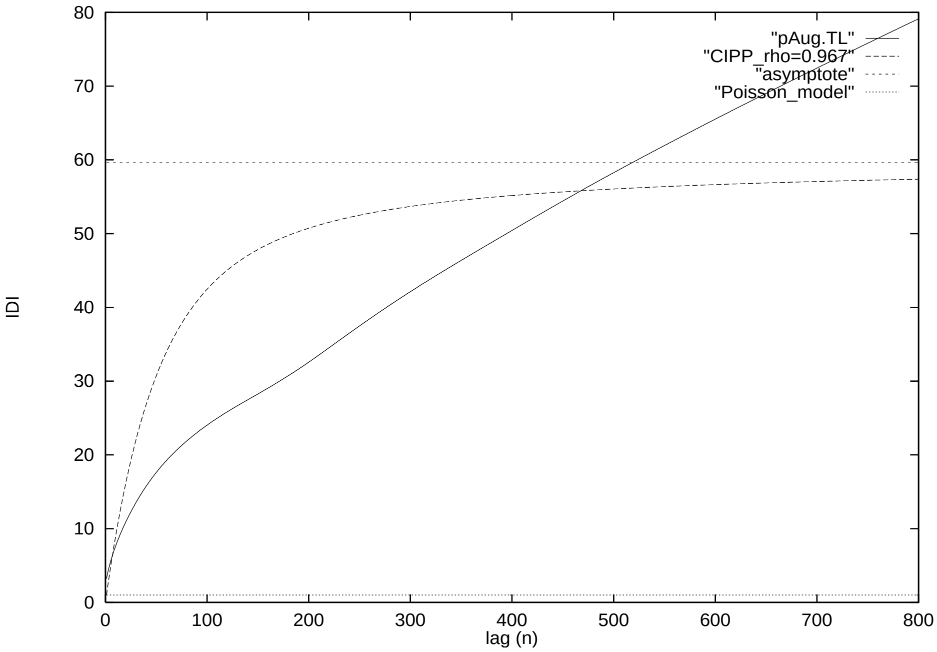
We consider the frame size trace of MPEG-1 encoded video sequences with various Hurst (H) parameters (ranging from 0.77 to 0.96) namely, "soccer" (H = 0.77), "starwars" (H = 0.85) and "Mr.Beans" (H = 0.96) [2] for our experimental purposes. We carry out the autocorrelation fit for each of these sequences. Our results are illustrated in Figure 4.1. One can see that for "soccer" (H = 0.77), although both models show almost the same approximation quality, the MC model is more optimistic. However, for higher H parameter video sequences, like "starwars" and "MrBean", the CIPP performs substantially better than the MC model. Note that the rho of the 'best fit' CIPPs to the sequences "soccer", "starwars" and "MrBean" are $0.992, 0.993$ and 0.9932 respectively.
Discussions: Note that the correlation coefficient corresponding to MC model, decays very fast (with respect to lag) which is a feature of Markov models. In contrast, the counts sequence, $\{u_T(i)\}_1^\infty$ in CIPP case, may have more complex correlation structure, (may not even be a first order Markov chain) despite the fact that the interarrivals form a first order Markov chain. We strongly believe that this is the reason for non-exponential decay of correlation coefficient of counts sequence corresponding to CIPP (and we will see in a later section that, this also accounts for the self-similar behavior to a range of time scales.)

4.2.2.2 Index of Dispersion for Intervals (IDI)

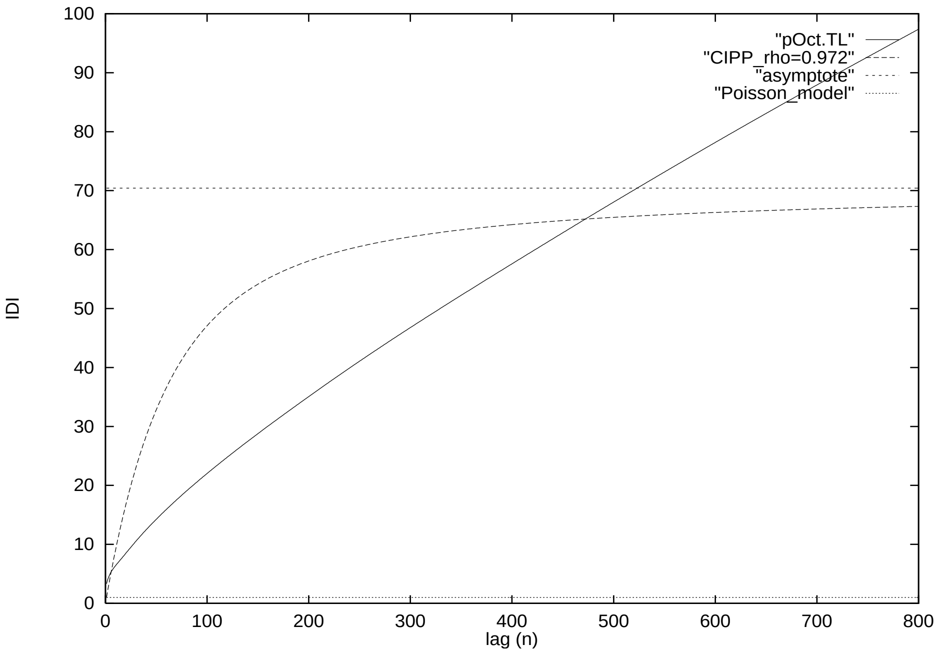
Figure 4.2 shows IDI plot for Bellcore traffic datasets [82] `pAug.TL` (Figure 4.2 (a)) and `pOct.TL` (Figure 4.2 (b)) and their corresponding (approximating) CIPP models with $\rho = 0.967$ and $\rho = 0.972$ respectively. It is seen that IDI rises sharply (initially corresponding to small lags) and eventually saturates at a point. Note this saturating point is given by (4.6). Note that the CIPP does substantially better than that of the Poisson process.

4.2.2.3 Index of Dispersion for Counts (IDC)

Figure 4.3 (a) shows the IDC plot for *race* video sequence and the corresponding CIPP match. CIPP again shows reasonable approximation quality. The corresponding plot for Bellcore dataset is shown in Figure 4.3 (b). Here, the approximation is not very

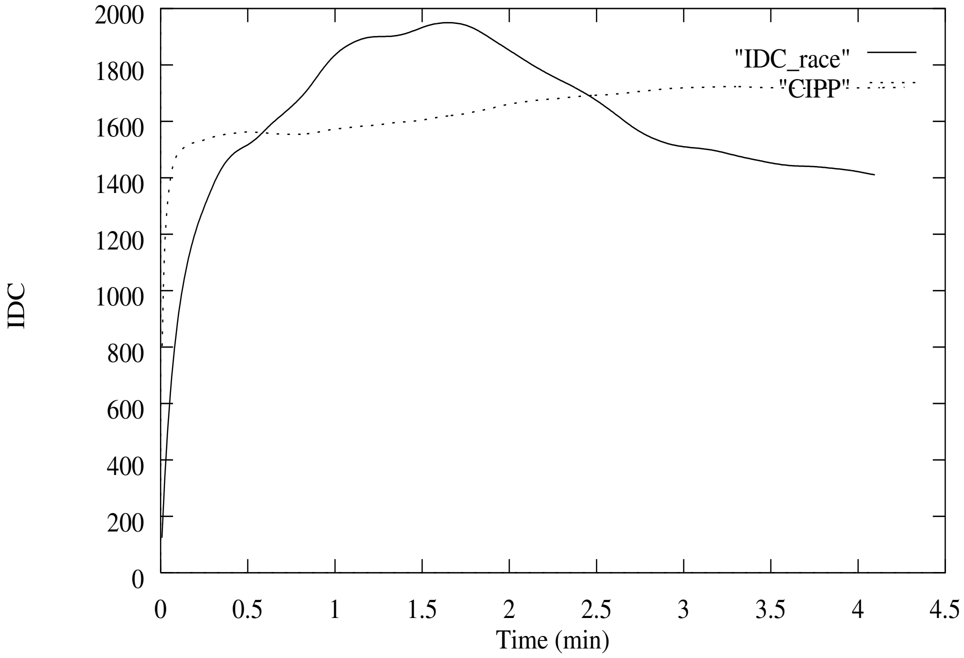


(a)

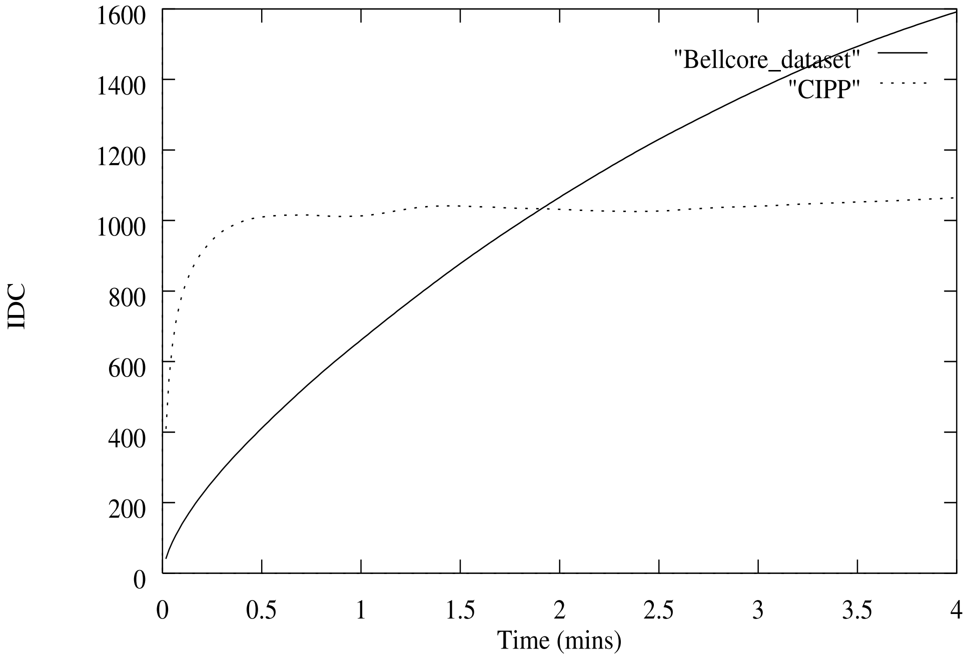


(b)

Figure 4.2: IDI of Bellcore datasets and their corresponding (approximating) CIPP models (a) `pAug.TL` (b) `pOct.TL`



(a)



(b)

(a) IDC of *race* video sequence (b) IDC of *Bellcore* dataset.

Figure 4.3: IDC of *real-world* traffic traces and their respective approximating CIPP models: (a) *race* video sequence and (b) Bellcore Ethernet dataset $pAug.TL$

good, possibly because CIPP is a pseudo self-similar process, to be discussed in the next section.

4.3 Statistical Properties of CIPP

For fitting the CIPP model to real-world arrival processes and the performance modeling of real-world queueing situations using CIPP queues, it would be advantageous to have explicit expressions for the moments of CIPP in terms of the parameters of CIPP. To that end, we investigate some statistical properties of CIPP, in this section. These results will be used in subsequent chapters.

4.3.1 Mean

By virtue of (3.2) in [79], the mean of the CIPP process is λt. Mathematically, $\mathbf{E}\{N(t)\} = \lambda t$.

4.3.2 Correlation in Interarrivals

The correlation coefficient for interarrival time sequence is defined by,

$$\eta(k) \triangleq \frac{E\{X_i X_{i+k}\} - E^2\{X_i\}}{\sigma_{X_i}\sigma_{X_{i+k}}}$$

where, σ_{X_i} is the standard deviation of the ith interarrival time. Exploiting the stationarity of the random sequence $\{X_i\}$, one can show that the correlation coefficient of the interarrival time sequence $\{X_i\}$ in CIPP is a geometric function with parameter ρ (see Appendix D for derivation).

$$\eta(k) = \rho^k \tag{4.9}$$

4.4 Correlational Property of CIPP

In this section, we discuss about the position of CIPP in Figure 2.1. We discuss the correlational property of the CIPP process and how it leads to large scale correlations (and thus self-similar behavior over a range of time scales).

4.4.1 Revisiting Chapter 2

We reproduce the Figure 2.1 here, for our convenience. The Figure 4.4 shows CIPP, in addition to the processes given in Figure 2.1. (MMCIPP will be introduced in Section 6.5).

The figure shows that CIPP forms a subset of a more general discrete state stochastic processes whose holding times are dependent. Note that, in the case of CIPP, holding times are interarrival times, since CIPP is a counting (birth) process. One may see that there is an overlap between CIPP and renewal process, the region where the Poisson process lies. In next section, we will see how, the dependent interarrival times in CIPP leads to self-similar behavior (over a range of time scales) with ($H > 1/2$).

4.4.2 Correlational Structure of CIPP leading to Self-similar Behaviour

Consider a sample path of CIPP illustrated in Figure 4.5. Let t be the instant of observation. Let $L_n(t)$ be the time required for the occurrence of nth event after t. It is well known that $L_n(t)$ is exponentially distributed (See Figure 4.5) for memoryless processes like, Poisson processes. Since the presence of after-effects (or memory) is an important correlational property, a process can be specified by a complete pre-history up to the instant 't' under consideration [83]. Varying degrees of non-Markovian nature may reduce the degree of indispensability of the complete pre-history. For instance, if the process is the renewal process, the pre-history need extend only up to the latest event. For this reason, Khintchine calls a renewal process, "a process with limited after-effect". This is illustrated in Figure 4.5. For a non-renewal process, it may be necessary to specify the history even beyond the latest event. In CIPP case, the pre-history extends up to the occurrence of the event, previous to the latest event. We strongly feel that this feature is responsible for the self-similar behavior over a range of time scales.

Moreover, the counts sequence, $\{u_T(i)\}_1^\infty$ in CIPP case, might have more complex autocorrelation structure, (since, the counts sequence $u_T(i)$ may not be a first order Markov chain) despite the fact that, the interarrivals are forming a first order Markov chain. This is an another manifestation of the above correlation structure, which we believe that, is responsible for self-similar behavior over a range of time scales.

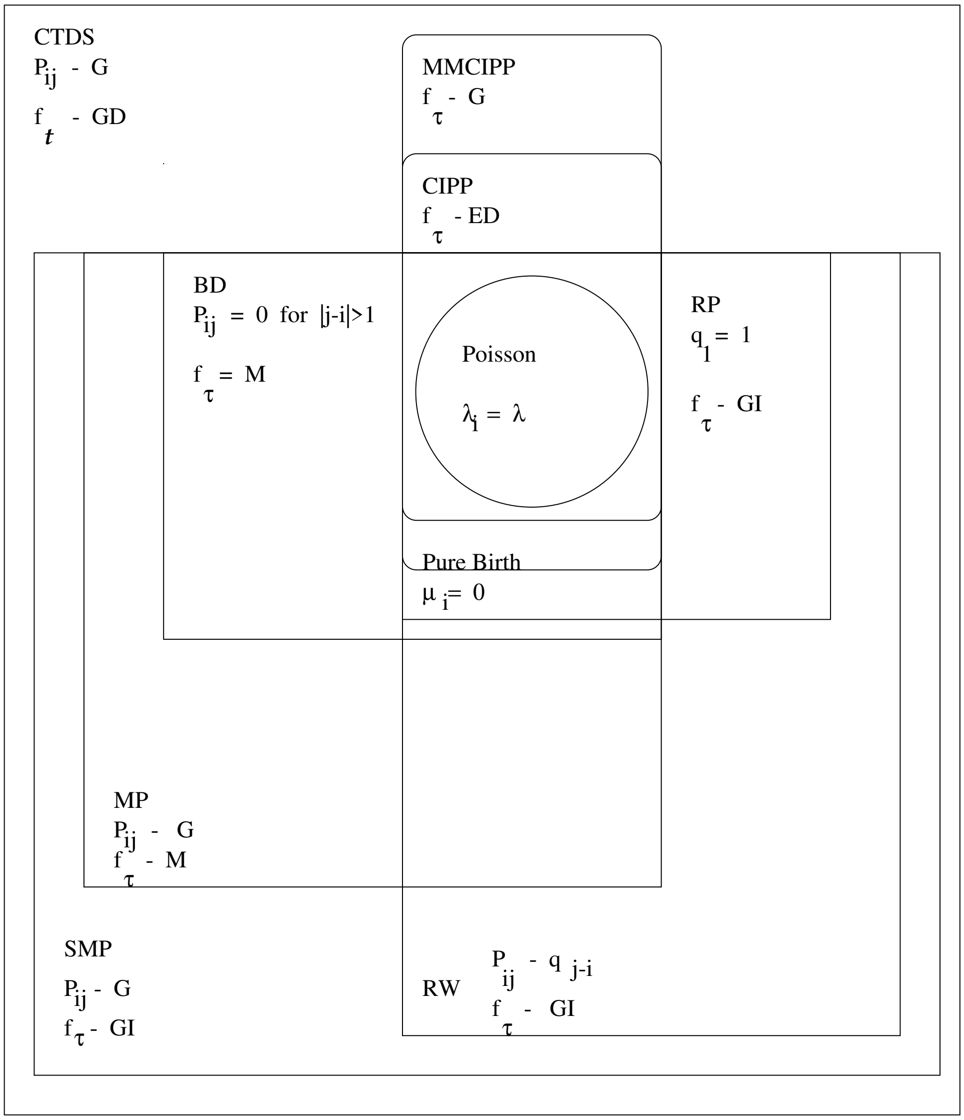


Figure 4.4: Position of CIPP in the map illustrating interrelationships between various stochastic processes of interest to a queueing theorist: **Legend:-** CTDS - Continuous Time Discrete State Stochastic Processes. SMP - Semi-Markov Processes. MP - Markov Processes. BD - Birth-Death Processes. RW - Random Walk. RP - Renewal Processes. CIPP - Correlated Interarrival Poisson Processes. MMCIPP - Markov Modulated CIPP. M - Memoryless distribution. GI - General and Independent. GD - General and Dependent. ED - Exponential and Dependent. P_{ij} - Probability of transition from state i to state j. f_τ - holding time distribution.

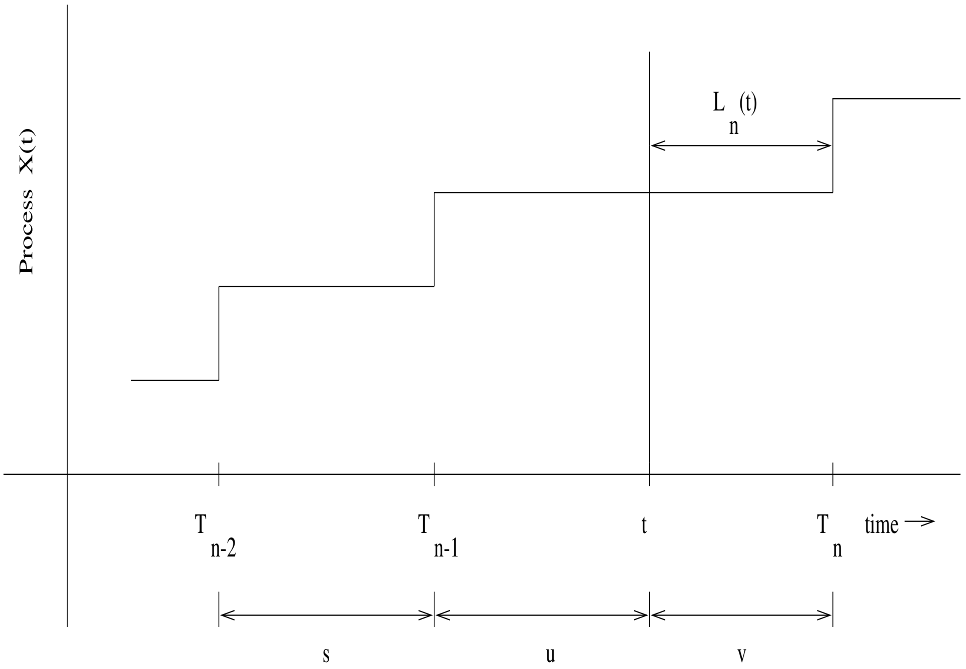


Figure 4.5: Illustration of correlational structure in CIPP leading to self-similar behavior

4.5 Properties of CIPP pertaining to Traffic Modeling

In this section, we will describe the properties of CIPP relevant for teletraffic modeling, namely burstiness measure and self-similarity. The self-similar nature of superposed process of independent CIPP streams is explored.

4.5.1 Burstiness Measure

We consider the burstiness measure as defined in [35]. We repeat the definition here for convenience,

$$C_p(z) = \sum_{k=0}^{\infty} C_k z^k / (\lambda^{-1})^2, \qquad |z| < 1 \tag{4.10}$$

where λ^{-1} is the mean interarrival time. $C_k = \mathbf{E}\{(X_1 - \lambda^{-1})(X_{k+1} - \lambda^{-1})\}$ denotes the correlation with lag k for interarrival times and $X_1, X_2, \ldots$ is a sequence of interarrival times. $C_p(z)$ includes the complete information for the second moment of interarrival times. For Poisson process, $C_p(z) = 1$ for all z. For a renewal arrival process, $C_p(z) =$ (squared coefficient of variation in the renewal process) for all z. For an MMPP, the burstiness measure has been discussed in [35]. For CIPP, it can be shown that the correlation (C_k) with lag 'k' of interarrival times is given by $C_k = \rho^k, \quad |\rho| < 1$. As

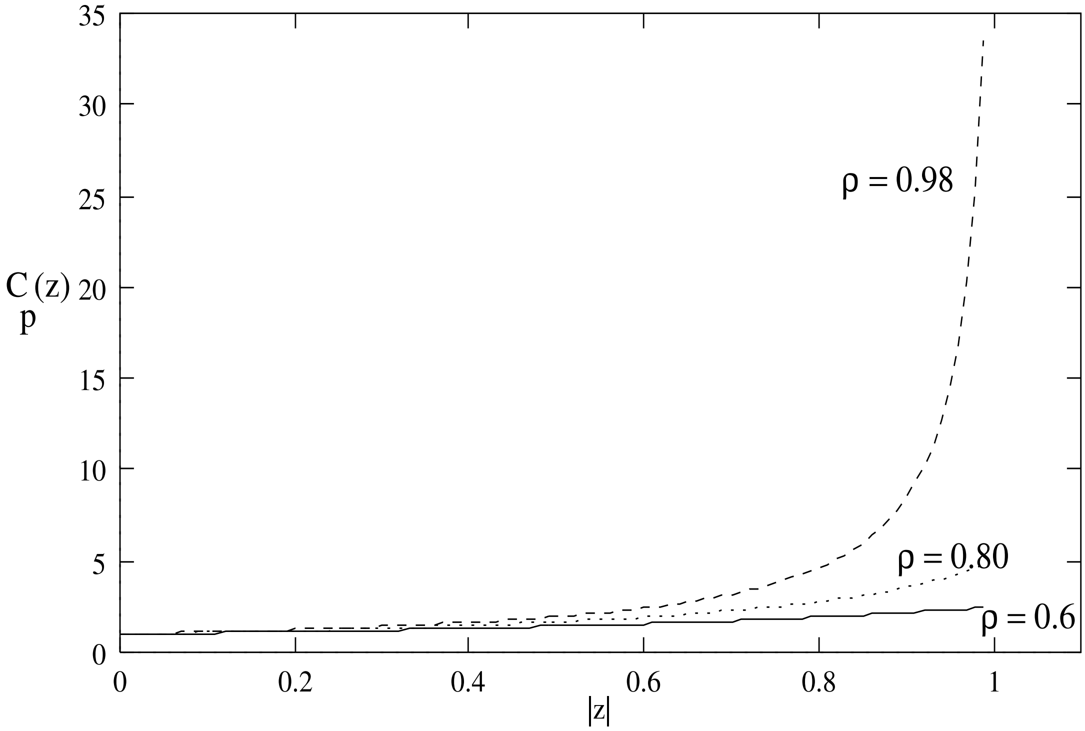


Figure 4.6: Burstiness factor of CIPP: $C_p(z)$ versus $|z|$ for CIPP

a result, $C_p(z) = \frac{1}{1-\rho z}$. Figure 4.6 shows $C_p(z)$ versus z. One can see that CIPP does display burstiness for large ρ.

4.5.2 Superposition of Independent CIPP Streams

In this subsection, we study the self-similar property of superposed process resulting from merging of two independent CIPP traffic streams. We estimate the Hurst parameter of superposed process and of the component processes. We consider two independent CIPPs with arbitrary ρ with same λ. We estimate Hurst parameter of the component processes and also estimate the Hurst parameter of the merged process. In Figure 4.7, we give our results for the estimated Hurst parameter of the superposed process versus the estimated Hurst parameters of the component processes. One can see from the Figure, that the Hurst parameter (H), of the superposed CIPP lies somewhere in between the Hurst parameter (H) of the component processes.

4.5.3 Self-similar Property of CIPP

To compare the behavior of CIPP with the Bellcore dataset, we generate samples which are equal in length and with the same mean value as the various traces analyzed in [24]

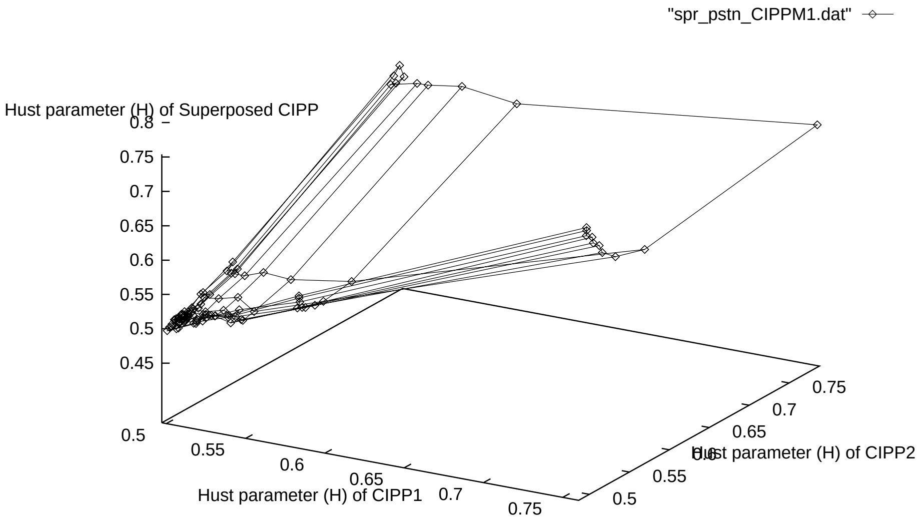
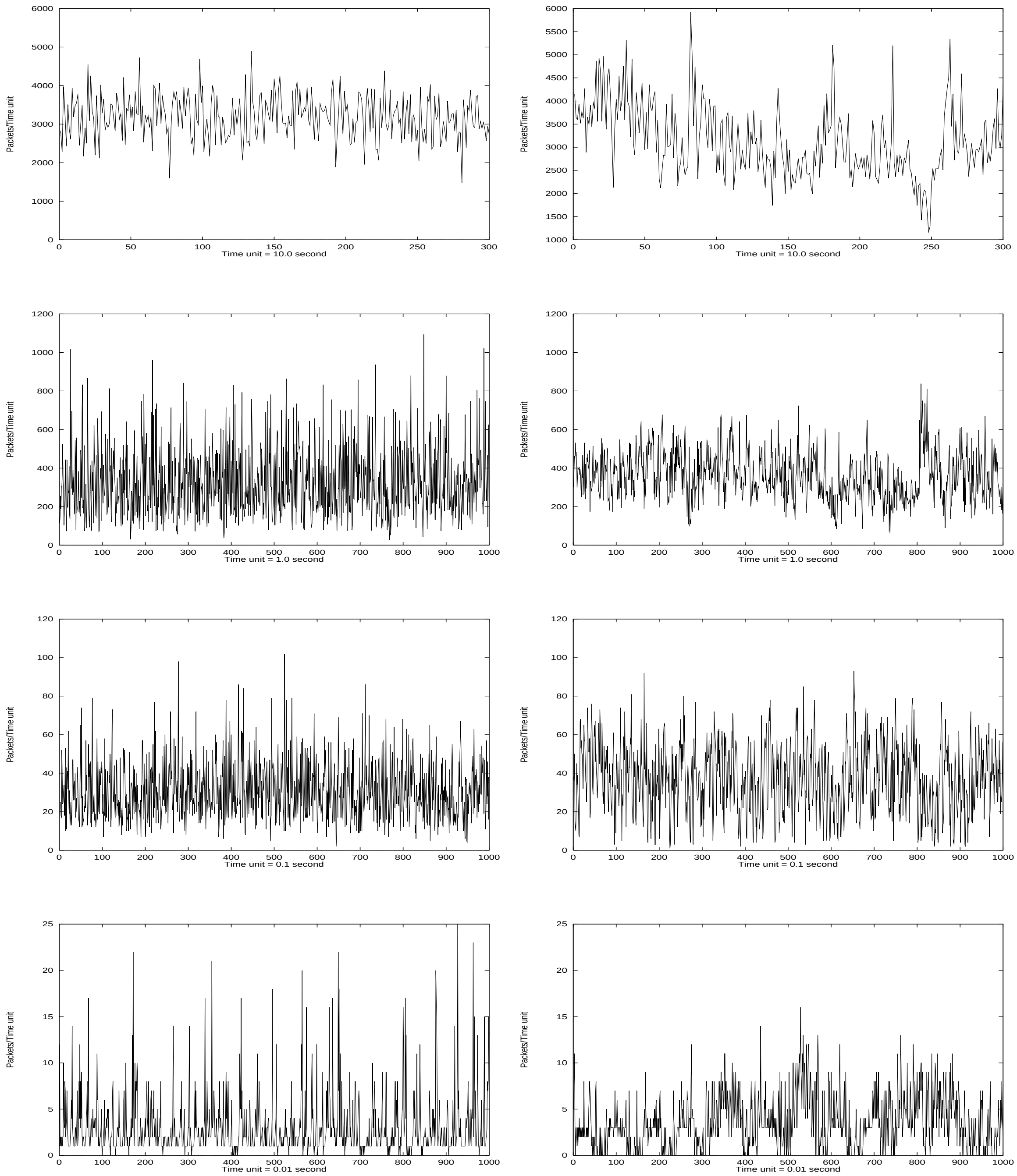


Figure 4.7: The self-similar behavior of superposed process of two independent CIPP streams

and [84]. For the synthetic traces we generate, we use $\rho = 0.976$ and $\lambda = 318.18$ arrivals per sec. The λ selected here is equal to the mean packet arrival rate in August Bellcore dataset (`pAug.TL`) and so also with ρ. To get an idea of self-similarity of CIPP, we resort to a pictorial representation analogous to those given in [24] and [84]. We select, at random, sections of the simulated traffic and the August Bellcore dataset and plot them side by side at the same resolution level. This is shown in Figure 4.8. The simulated CIPP process exhibits a similar bursty behavior as the Bellcore dataset over all plotted scales. Note that the length of the Bellcore August dataset limits the number of points we can plot at the coarsest scale.

Next, we generate CIPP with different ρ and estimate the Hurst (H) parameter of the process, by the wavelet estimator proposed by Abry *et al* [46]. In Figure 4.9, we plot the estimated Hurst parameter for CIPP versus the correlation parameter. Note that the estimated Hurst parameter H is 0.5, corresponding to $\rho = 0$, which is true for the Poisson process. The plot shows that we can model teletraffic with higher H by using a CIPP with a larger ρ.

Next, the autocorrelation function at various scales is estimated. Figure 4.10 shows autocorrelation at various scales. One can see, as T (block size or slot length) increases, the autocorrelation tends to approach that of the Poisson case (that is, zero correlation for non-zero lag). This is a typical feature of pseudo-self-similar processes (see [1] pp.



Simulated CIPP trace Bellcore August Dataset

Figure 4.8: Demonstration of self-similar property of CIPP: A "Visual" justification.

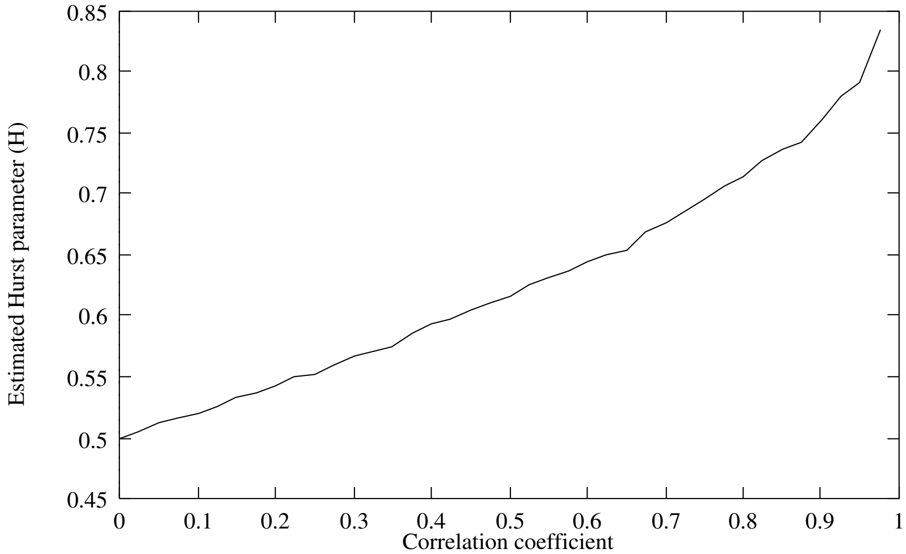


Figure 4.9: Estimated Hurst parameter H versus the correlation parameter ρ of CIPP

338)

The next plot (Figure 4.13) is the log (variance) versus log (aggregation level) plot for CIPP and *Bellcore* dataset. The lower-most curve is a reference curve with slope 1. The behavior of CIPP is strikingly similar to the *Bellcore* dataset, with the variances for both processes decaying at the same rate, with slope less than slope 1.

We finally present the wavelet analysis of CIPP. Wornell [45] has proved that as long as the analyzing mother wavelet has at least one vanishing moment, the *log of variance* of the wavelet coefficients for a self-similar process increases linearly with the scale, the slope being the Hurst parameter. This interpretation of wavelet transform gives a convenient way of looking at the variance-scale plots which has been given in [84] and is used here. Figure 4.11 shows such a plot for various processes and CIPP. One may see that for pure self-similar processes, the plot results in a linear curve. For pseudo self-similar processes the curve increases linearly and saturates at a point, where the slope is zero. Note that for CIPP the curve is linear to a much wider scale than ON-OFF processes. Also, observe that the linearity increases as ρ increases.

We now compare CIPP with Bellcore August dataset, using wavelet analysis. We have analyzed the two signals using the Daubechies-4 [85] wavelet. The plots of log of variance of wavelet coefficient versus log2(scale) for the simulated CIPP and Bellcore August dataset shows a strong resemblance in the properties of the two processes as seen in Figure 4.12.

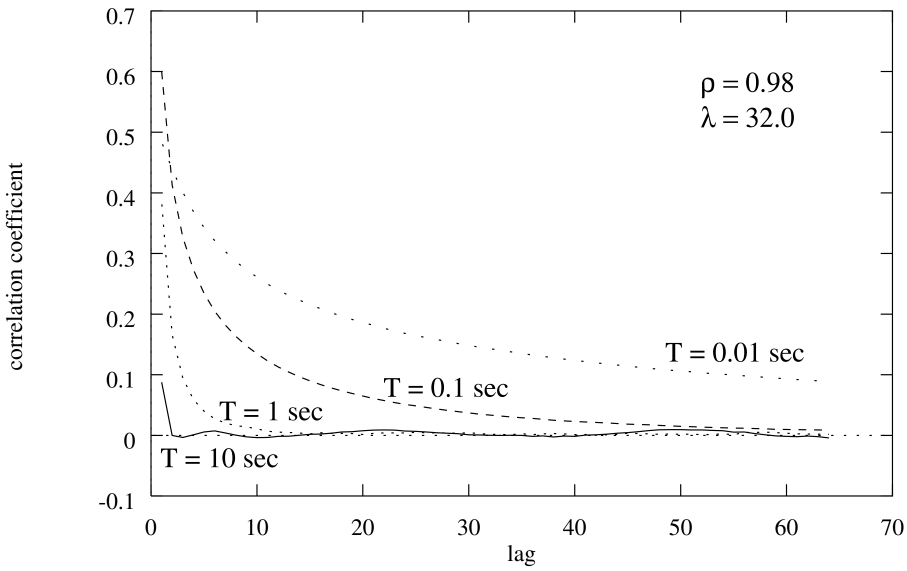


Figure 4.10: Autocorrelation of CIPP at various scales

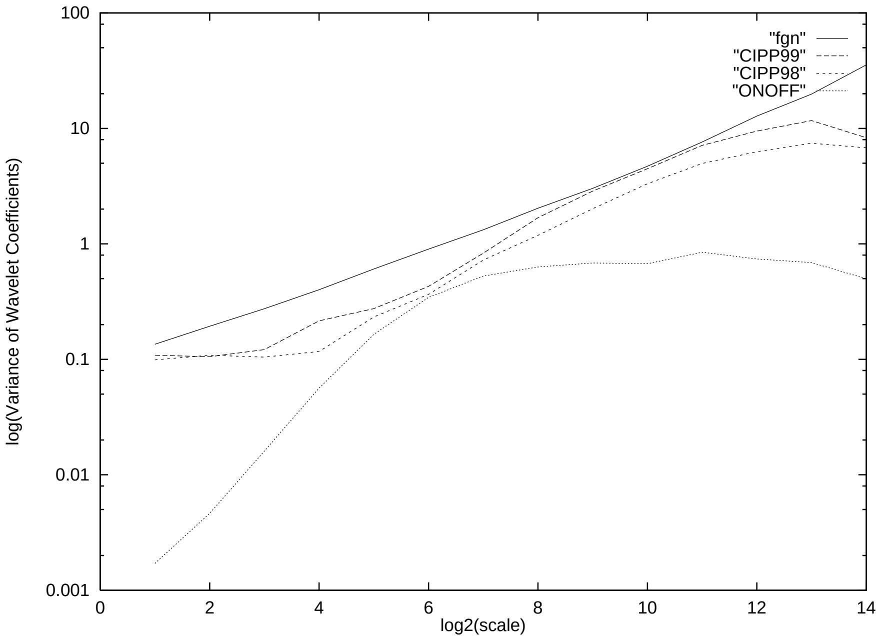


Figure 4.11: Variance scale plots for CIPP and other processes

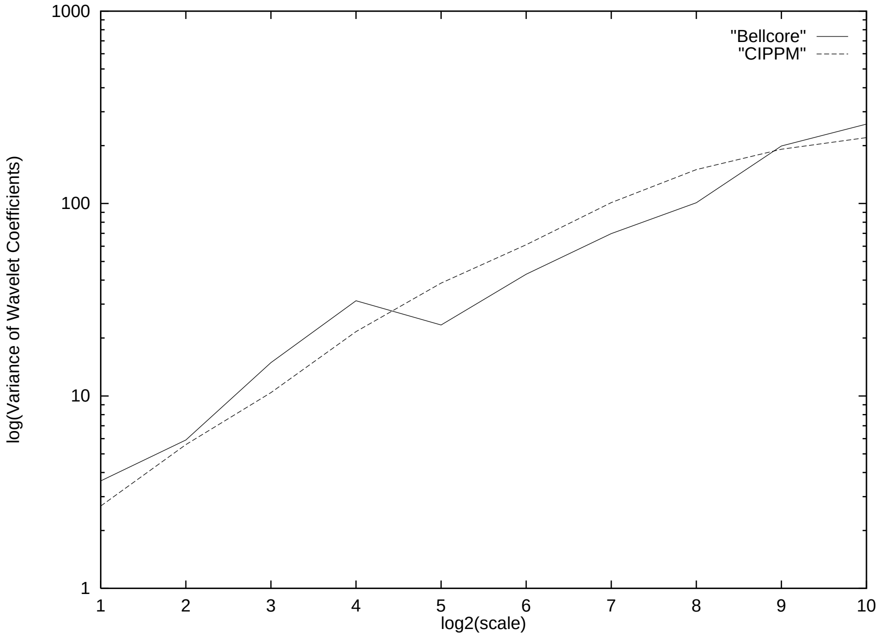


Figure 4.12: Variance scale plots for Bellcore dataset and the approximating CIPP process

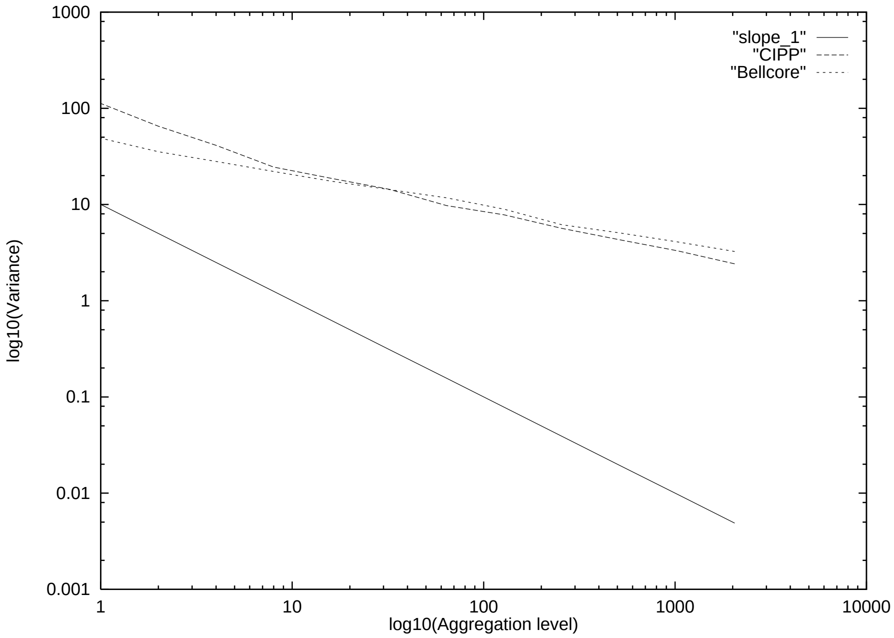


Figure 4.13: The variance and aggregation level plot for Bellcore dataset and the approximating CIPP process. The lower most curve is a reference curve of slope -1

4.6 Other Correlated Sequences

In this section, we discuss two models which generate stationary, correlated, positive, random sequence, the former having exponential marginals, the later being gamma distributed. Here, these are included in order to explain, why the EAR(1) process developed in [69] is more suitable for our modeling purpose than the EMA(1) model developed by Lawrance or the modified Finch model, despite the later having queueing theory results.

4.6.1 Exponential Moving Average EMA(1)

The exponential moving average (EMA(1)) sequence was introduced by Lawrance *et al* [70]. The sequence $\{X_i\}$ is generated by random linear combinations (of order one) of an *iid* exponential sequence $\{\epsilon_i\}$ as given below.

$$X_i = \begin{cases} \beta\epsilon_i & \text{with prob } \beta \\ \beta\epsilon_i + \epsilon_{i-1} & \text{with prob } 1-\beta \end{cases}$$

where $0 \leq \beta \leq 1$ and $i = 1, 2, \ldots$. It has been shown that $\{X_i\}$ form a sequence of exponential random variables and it will be seen from above, that the adjacent members of the sequence are correlated. But, in this model, the autocorrelation is zero beyond lag 1. (Note that this is a feature of moving average models [63].) We study the phase plot of EMA(1) along with the phase plots for other processes in the next section.

4.6.2 Modified Finch Process (MF)

We consider the following modified version of the model developed by Finch in [71].

$$X_i = (1-\alpha)\epsilon_i + \alpha\epsilon_{i-1} \tag{4.11}$$

for $i \geq 1$ and $0 \leq \alpha \leq 1$. The model is again very similar to Lawrance *et al* [70] except that the coefficients are constants here. The advantage of this model is that queueing theory exists [86]. (This modification is necessary to make the mean of interarrival sequence to be independent of α. This allows one to see the effect of correlation on the queueing performance in a more transparent manner).

Limitations of the above models:-

The main problem with the above models is that they are basically moving average models. Moving average models are *not* Markovian [63]. In other words, the correlation is weak, that is, the autocorrelation for the nth order MV process has a *cut-off* at lag n). This limits their usefulness as a model for broadband teletraffic. Another disadvantage, specific to MF process is that, the interarrivals are gamma distributed.

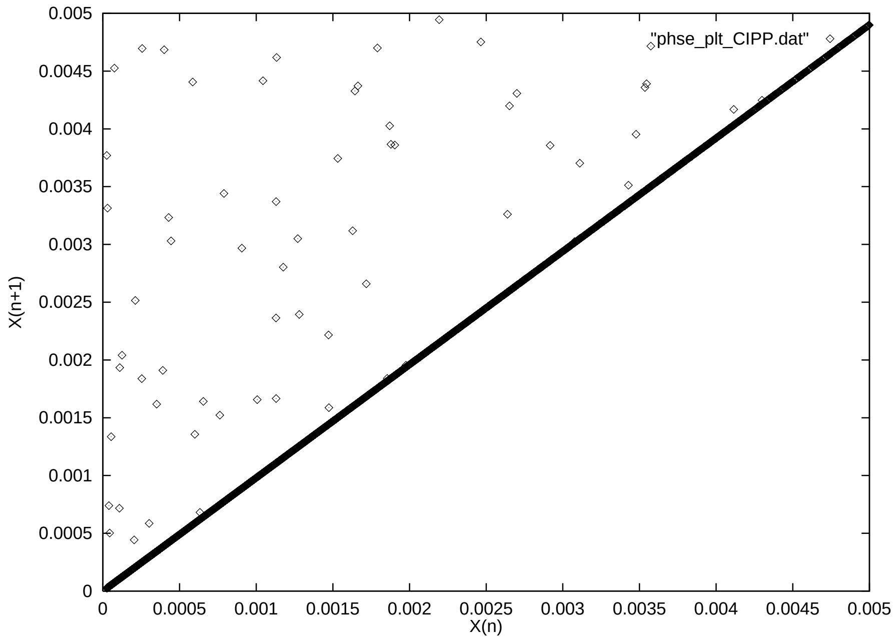


Figure 4.14: Phase plot of CIPP

4.7 Phase Plot

The phase plot [29] of time series of interarrival times $\{X_i\}_{i=1}^{\infty}$ is the plot of X_{i+1} versus X_i. Phase plot brings out the stochastic features of the process under study at the interarrival time level. Here, it is included as it brings out some key stochastic features of the arrival process (whether there are any bursts or positive correlation in interarrival times). Thus, it is used to bring out the correlation structure in CIPP and compare with the corresponding plots for Poisson process, EMA(1) and MF processes. The mean arrival rate of all the processes considered above, is selected as the mean packet arrival rate of the Ethernet Bellcore traffic (`pAug.TL`) namely 318.3 arrivals per second (in order to compare with the phase plot of `pAug.TL` in Chapter 6). We give the phase plot for CIPP with $\rho = 0.976$ in Figure 4.14. The corresponding plots for Poisson, EMA(1) and MF process are shown in Figures 4.15, 4.16 and 4.17 respectively. We have taken $\beta = 0.5$ for EMA(1) process and $\alpha = 0.5$ for MF process to see the effect of maximum correlation. (Note that when $\beta = 0.0$ or 1 ($\alpha = 0.0$ or 1), the EMA(1) process (MF process) reduces to Poisson process.)

From the figures, the plot corresponding to Poisson process, has random points throughout the range. This should not be surprising, as this behavior is a manifestation of the independence of interarrival times. It is for this reason that the Poisson process is sometimes referred to as white noise process [78]. For MF process, the phase

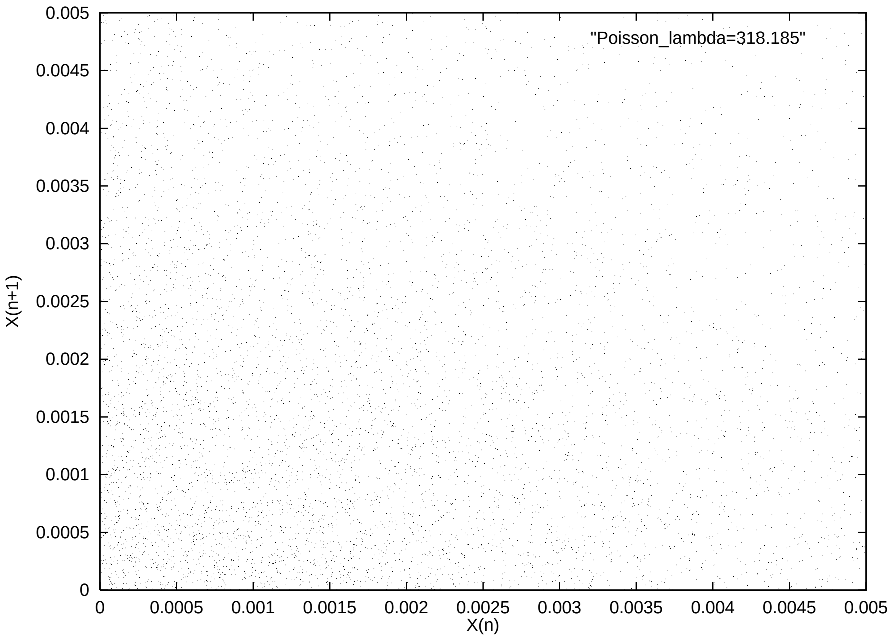


Figure 4.15: Phase plot of Poisson process

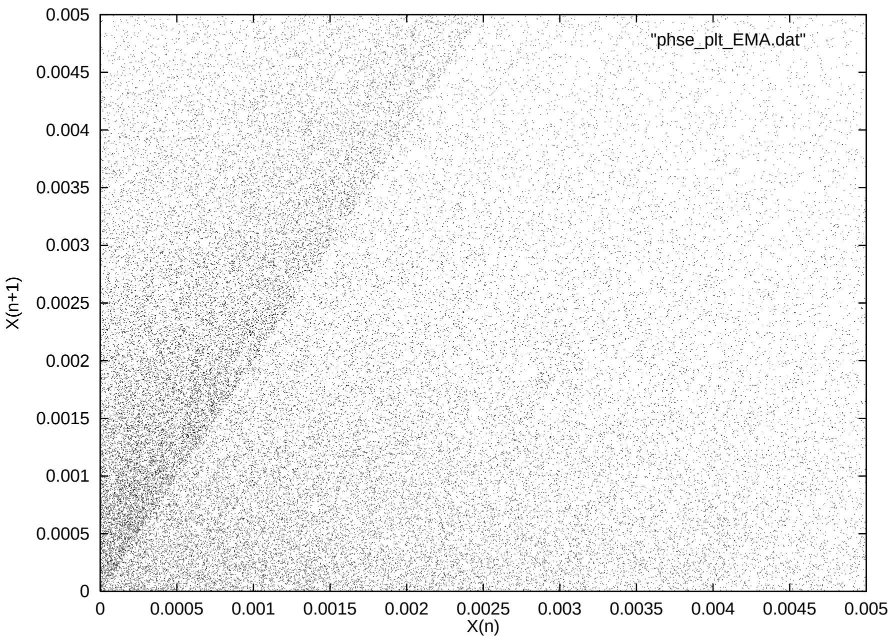


Figure 4.16: Phase plot of EMA(1) process

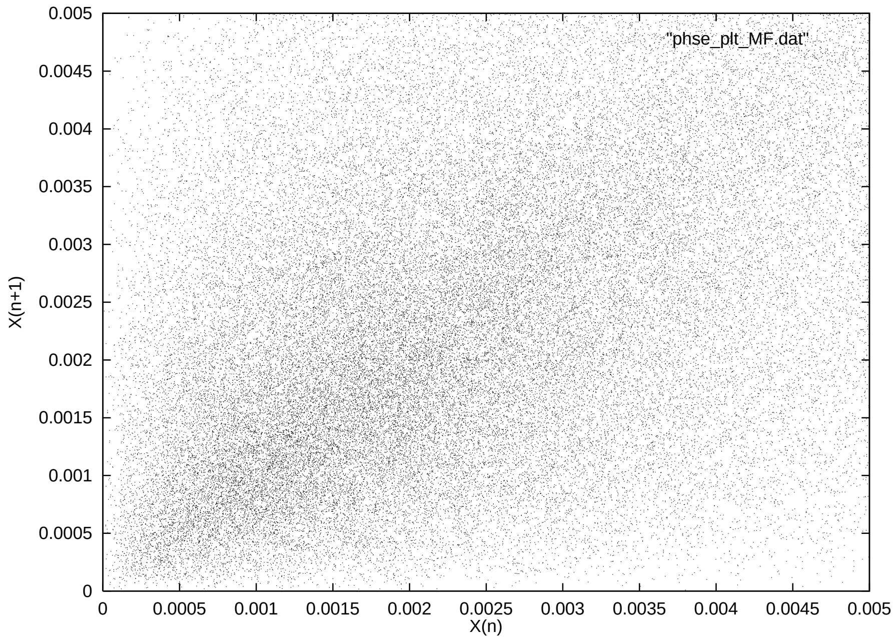


Figure 4.17: Phase plot of MF process

plot shows points concentrated along the diagonal, rather than distributed throughout the range, for high ρ. This is the consequence of correlation in interarrivals, but is very weak. The phase plot corresponding to EMA(1) process exhibits slightly a different behavior. Note that a diagonal line clearly demarcates the triangular patch of dense points, (on the top left) from relatively less dense points. This line denotes that there is a slight positive correlation in the interarrival sequence and it is very weak. (We observe that the line has the slope of $1/\alpha$). This behavior is a manifestation of weak correlation, which is a typical feature of moving average structure. (Note, EMA(1) is also a moving average process, but with random coefficients).

Now, we give the phase plot of CIPP. Figure 4.14 gives the phase plot corresponding to a single realization of CIPP with $\rho = 0.976$ and mean arrival rate same as that in Bellcore `pAug.TL` Ethernet dataset. From figure, one may see that for the CIPP, all the points are *above* the line of slope ρ. This line is the locus of the points for which $X_{n+1} \approx \rho X_n$. The larger the value of ρ, the more prominent is the line. The above feature is the consequence of interarrivals forming a special case of AR(1) model as mentioned in 4.1, wherein X_{n+1} is lower bounded by ρX_n. We will see in Chapter 6, that, the phase plot of composite Ethernet traffic exhibits complex features of diagonal lines (diagonal lines with positive slope implies positive correlation). While this feature of diagonal lines of positive slope is captured by CIPP (which the above models were not been able to), the feature that points lie every where in the phase plot is poorly modeled by CIPP with high ρ. This is the limitation of CIPP process and we will come back to this point, when we discuss about the phase plot of real-world traffic traces in Section 6.4.

4.8 CIPP and the *Neuts* Process

Here, we will see that CIPP is *not* a special case of Neuts process [8]. For details about Neuts process, the reader is referred to Section 3.3.1. In the Neuts process [8], the memory of the process is limited to the latest phase renewal instant and the process completely *forgets* the past after that point. Note, to exhibit self-similarity or long-range dependence, the process should have infinite memory. In CIPP, there is *no* renewal instant at all. The memory of the CIPP process at time t, is limited to the arrival instant just prior to the latest arrival instant (and the latest interarrival is again dependent on its previous interarrival and so on). The memory is "passed" on to the next interarrival which affects the future event in some sense. Thus, CIPP possesses large scale correlations. It is precisely this feature that is responsible for CIPP to exhibit self-similarity over a range of time scales. Quantitatively, this extent of memory is controlled by the correlation parameter ρ.

In Neuts process, within a phase renewal interval, if we allow an arrival *only* during (i, j) transitions (the process so constructed is called *Markov arrival process*), one may find that the interarrivals are correlated. (Note that, in the case of MMPP also, interarrivals can be correlated, and such events have very low probability, by *assumption*). But, this is due to the difference in marginal distribution of *holding times* of the modulating Markov chain. When there is no difference in the marginal distribution of *holding times* at different states, it reduces to a renewal process (within the phase renewal interval), irrespective of probability transition matrix. In contrast, the interarrivals form a stationary sequence in CIPP and are correlated.

4.9 Discussions

Since the discovery of self-similar nature of Ethernet traffic, teletraffic researchers have put considerable effort in coming up with a model which is self-similar on the one end and is analytically tractable on the other end. Ideally one would like to have a stochastic process pertinent to modeling with following properties.

1. A counting process which is easy to simulate on the computer.

2. The increment (rate) process corresponding to this process should exhibit self-similar behavior to a wide range of time scales.

3. It should be analytically tractable, meaning, the queueing theory should exist for a queue with such a process as input.

One way to achieve the above objectives is to consider models which introduces correlation in counts as in DAR or in the rate process as in AR models or by introducing correlation by the modulating Markov chain. Another way is to define a counting process, in which the dependency (or memory) goes back as far as possible, through *correlations in interarrival times*, still retaining some nice properties which allows one to develop queueing theory for the same. In this chapter, we saw that, such an exercise was undertaken for the case of interarrival times forming the first order Markov chain. In sequel, we have introduced CIPP with correlated intervals for modeling broadband teletraffic.

We have derived the distribution ($P_n(t)$) for CIPP. It has been shown by simulations that this arrival process models the real-time traffic source quite well, (in particular the video source) in addition to the composite Ethernet traffic. The motivation for the model is two fold: first, the real-world Ethernet traces show a strong positive correlation, second, a strong positive correlation in interarrivals leads to the corresponding counts sequence exhibiting more self-similar behavior over a range of time scales. This range of time scales depends on the choice of parameter ρ of the model.

We saw that both the MC and CIPP models, approximates the frame size trace of video sequences with low H parameters, with same accuracy. We saw that, for modeling the VBR video sequence, both the MC and CIPP models, approximate with same accuracy for video sequences of low H parameters. However, for high H parameter video sequences like "starwars" and "MrBean", the CIPP performs substantially better than the MC model, exhibiting a non-exponential decay of autocorrelation function. We believe that, this is the consequence of large scale correlations exhibited by CIPP with high ρ.

Instead of having first order Markov process for interarrival sequence, one may have in general, nth order Markov process. In that case, the following issues are to be sorted out: the stationarity of the interarrival sequence, ease of generation in the computer and the corresponding queueing theory. We expect that the process thus generated, would exhibit a higher degree of self-similar behavior. It is to be noted that Finch *et al* has developed queueing theory corresponding to a process whose interarrivals forming nth order moving average process. But, it lacks an important property that we are looking for, namely, the strong (positive) correlation. Note here that the interarrivals form moving average sequence and the correlation introduced is weak (and thus does *not* exhibit the self-similar behavior).

One of the attractive features of CIPP is that there exists an empirical relationship between the correlation parameter ρ of CIPP and the estimated Hurst parameter H. This has a lot of implications on the real-time control and management of networks, in particular reference to the fact that the network traffic is self-similar. For instance, CIPP is feasible as an efficient tool in studying the performance evaluation of various network elements in broadband networks with respect to the degree of self-similarity of the network traffic. The later feature is demonstrated in Chapter 7 by considering the problem of interference effect on MPEG Video over ATM networks by network traffic.

Chapter 5

Queues with CIPP input

This chapter concentrates mainly on the derivation of stationary probabilities for the CIPP/M/1 queue and the corresponding queueing performance measures, which will be used in later chapters for performance modeling purposes.

In most cases, where very complex models are encountered, simulation techniques are successfully employed and sometimes preferred to analysis due to the intractability of such models. However there is a certain need for analytical results when possible to get deeper insight, reduce run-times, handle very rare events and optimize system performance. Towards this end, we consider single server FCFS queue with CIPP input and exponential service times (CIPP/M/1 queue). We derive the expressions for the stationary distributions for buffer occupancy[1] and the waiting time. Our approach is similar to that in Tin [87]. The method exploits the embedded Markov chain at arrival instants to obtain the fundamental recursive equations in joint probability involving queue length. Then, the generating function method is used to transform the fundamental equations in a form amenable for further analysis. Finally, the complex analysis methods are used to obtain explicit expression for stationary probabilities.

We derive the queueing performance measures of practical interest, namely, mean packets in the system, mean waiting time for packets, packet delay variation and overflow probability corresponding to the CIPP/M/1 queue. In order to verify the above theoretical results, we compute the analytical queueing performance measures and estimate these measures by simulating the CIPP/M/1 queue with similar parameter values for comparison.

We also consider the CIPP queues with deterministic (constant) service intervals (CIPP/D/1 queue), in view of the fact that the service time[2] of a fixed length packet in a *real-world* broadband switch is constant. We use the approach of Benes, to obtain the

[1] Buffer occupancy here denotes the number of customers in the *system* (both the queue and the server).

[2] service time includes switching and transmission time.

approximate expression for buffer occupancy distribution. The (approximate) analytical expression for the buffer occupancy distribution poses a numerical problem for large values of ρ. Hence, we estimate the expressions for stationary distribution, by suitable simulations.

We show that our theoretical results corresponding to CIPP/M/1 queue are consistent with Loynes [88], who has given a frame work for the stability of a queue with more general arrival and service times (with arbitrary correlation structure). The queueing theory implications of the above theoretical results are explored. In particular, the relation between the correlation in interarrivals and the queueing performance is discussed.

CIPP queues with finite capacity does not seem to be analytically tractable. Hence, we resort to simulation based study of CIPP/M/1/K queue. The performance measures investigated are packet loss probability and conditional packet loss probability. We also study the behavior of the buffer occupancy distribution. We finally illustrate with a numerical example, the way in which the (positive) correlation affects the queueing performance and its implications in buffer length design.

This chapter is organized as follows: Section 5.1 presents an introduction to non-renewal queueing theory. In Section 5.2, the stationary queue length distribution and the stationary waiting time distribution in the CIPP/M/1 queue are derived. In Section 5.3, the queueing performance measures (corresponding to CIPP/M/1 queue) of interest to teletraffic modeling are derived. Using the Benes approach, the buffer occupancy distribution for CIPP/D/1 queue is derived in Section 5.4. Simulation study of CIPP/M/1/K queue is undertaken in Section 5.5. An example illustrating implication of correlations in interarrivals on the buffer length design is given in Section 5.7 while Section 5.8 comprises of discussions.

5.1 Introduction to Non-renewal Queueing Theory

Before we attempt to derive the stationary probabilities for the CIPP/M/1 queue, we give a brief overview of the non-renewal queueing theory. The non-renewal queueing theory deals with the study of queues with arrival and service processes such that, they have correlations within themselves and may also have cross correlations. In a single server FCFS queue, the waiting times W_n of successive customers[3], are related by the

[3]In this section and next section, we use the term 'customers' to denote the arrival entities, while developing queueing theory, but use 'packets' while computing the queueing performance measures of practical importance and applying the theory to the *real-world* situations in the rest of the thesis.

equation,

$$W_{n+1} = [W_n + V_n]^+ \tag{5.1}$$

where

$$V_n \triangleq S_n - X_n \tag{5.2}$$

where again, S_n and X_n are the service time for the n^{th} customer, and the interarrival time between $(n-1)^{th}$ and n^{th} customer respectively. The particular case of $\{V_n\}$ being an *iid* sequence has been analyzed by Lindley [89] and the theory has been further developed by Smith [90]. The general solution given by Smith is complex and is obtainable from the Wiener-Hopf technique or the equally specialised technique due to Spitzer [91]. In a more general case, when V_n is a strictly stationary process (thus relaxing the independence assumption as done in previous works) it has been shown by Loynes [88] that under a simple condition, the existence of an unique stationary distribution is ensured. Again, Loynes [92] shows that, (when the above condition is satisfied) the stationary distribution may sometimes be found and he obtains some related qualitative results. In [20], a single server queue has been considered by Addie *et al* with V_n being a stationary, ergodic, Gaussian discrete time stochastic process wherein the approximations for the tail distributions are obtained.

The case of $\{X_n\}$ forming a finite-state Markov chain, has been considered by Runnenburg [93],[94]. He has used this to study the waiting time process of a single server queue with discrete-valued Markov-dependent interarrival times and negative exponential services. The benign structure of moving average input model for interarrivals allowed Finch *et al* [86] to obtain *exact* expressions for the stationary distribution for the buffer occupancy distribution. But, unfortunately their method is applicable only to queues with moving average input (of any finite order [71]). Although Franken *et al* [95] and Borovkov [96] considered these queues in a much more general framework, their results are more on limit theorems, their existence and other rigorous mathematical results rather than exact structure of distributions. Tin [87] considers a single server queueing system with Markov-dependent interarrival times. The queue size and waiting time processes are investigated. In another direction, Szekli *et al* [97], have studied the effect of dependencies in the arrival process to a single server queue on mean queue lengths and mean waiting times in the context of a queue with Markov renewal arrival process with general independent services (MR/GI/1 queue). Our method is similar to Tin's method. The method exploits the special structure of the particular (embedded) Markov chain of the model considered.

5.2 CIPP/M/1 Queue

In this section, we consider a single server FCFS queueing system with Correlated Interarrival time Poisson Process (CIPP) as arrival process. The service times are exponentially distributed. The queue size and the waiting-time processes are investigated. Both the transient and the limiting results are given.

5.2.1 Definitions and Notations

Let the instants of arrivals be $T_0, T_1, T_2, \ldots$, and define $X_n = T_n - T_{n-1}$ for $n \geq 1$. We set $T_0 = 0$. We assume that the service times are independent and identically distributed random variables with common distribution function $1 - e^{-\mu x}, \;\; x \geq 0$. With no loss of generality, we take, $T_0 = 0$.

For CIPP with parameter ρ and λ, the marginal distribution of interarrival time is exponential. Hence, the density and its Laplace-Steiltjes Transform (LST) are given by,

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{5.3}$$

and

$$G(s,x) = \frac{\lambda}{\lambda + s} \tag{5.4}$$

respectively. The conditional density of $X_{n+1} = t$ for a given $X_n = x$ by,

$$f(t/x) = \rho\delta(t - \rho x) + (1-\rho)\lambda e^{-\lambda(t-\rho x)} U(t - \rho x) \tag{5.5}$$

and the corresponding conditional LST is given by

$$G(s,x) = \left(\rho + \frac{(1-\rho)\lambda}{\lambda + s}\right) e^{-\rho x s} \tag{5.6}$$

and with serial correlation ρ such that $0 \leq \rho < 1$.

Let $Q(t)$ denote the number of customers waiting or being served at the instant t and let $Q_n \triangleq Q(T_n)$. Then, it is seen that $\{Q_n, X_n\}$ is a bivariate Markov chain.

In Sections 5.2.2 and 5.2.3, we investigate transient and limiting behavior of the process $\{Q_n, T_n\}$ respectively. The method is similar to that used by Tin [87] (see also Pearce [98] and Finch and Pearce [86] for a moving-average interarrival process). In Section 5.2.4, the waiting time distribution is derived. In Section 5.3, we give some experimental verification of CIPP/M/1 queue.

5.2.2 Transient Behavior of the Queue Size

Let us define,

$$P_{ij}^{(n)}(t) \triangleq Pr\{Q_n = j, X_n \le t | Q_0 = i\}. \quad (5.7)$$

and the probability, $k_j(t,x)$ of j departures during an interarrival interval $X_{n+1} \le t$, given that $X_n = x$ is given by

$$\begin{aligned} k_j(t,x) &= \int_0^t \frac{e^{-\mu y}(\mu y)^j}{j!} f(y|x)dy && (5.8)\\ k_j^*(s,x) &= \int_0^\infty e^{-st} d_t k_j(t,x) && (5.9)\end{aligned}$$

Throughout this section, we use the following notation,

$$\begin{aligned} P_{ij}^*(s;n) &= \int_0^\infty e^{-st} d_t(P_{ij}^{(n)}(t)) && (5.10)\\ P_i^*(s,z;n) &= \sum_{j=0}^\infty P_{ij}^*(s;n)z^j && (5.11)\\ P_i^*(s,z,w) &= \sum_{n=0}^\infty P_i^*(s,z;n)w^n && (5.12)\\ &= \sum_{j=0}^\infty P_{ij}^*(s,w)z^j, \quad (|z| \le 1, |w| < 1, s \ge 0) && (5.13)\\ C_m^*(s;n) &= \int_0^\infty \sum_{l=0}^\infty k_{l+m+1}^*(s,x) d_x(P_{il}^{(n)}(x)) && (5.14)\\ C(s,z,w) &= \sum_{n=0}^\infty \sum_{m=0}^\infty (1-z^{-m}) C_m^*(s;n) w^n, \quad (|z| \ge 1, |w| < 1, s \ge 0) && (5.15)\end{aligned}$$

Fundamental Equations: If a customer arriving at T_{n+1} finds $j \ge 1$ individuals in the system, the customers arriving at T_n must have found at least $(j-1)$ customers already present. If there were $j-1+l \quad (l \ge 0)$ customers present there would need to be l departures in (T_n, T_{n+1}) for the queue length at $T_{n+1} - 0$ to be j. A similar argument holds for $j = 0$. Thus, we have the following recurrence relations,

$$\begin{aligned} P_{ij}^{n+1}(t) &= \int_0^\infty \sum_{l=0}^\infty k_l(t,x) d_x(P_{i,j+l-1}^{(n)}(x)), \quad (j \ge 1) && (5.16)\\ P_{i0}^{n+1}(t) &= \int_0^\infty \sum_{l=0}^\infty \sum_{m=l+1}^\infty k_m(t,x) d_x(P_{il}^{(n)}(x)) && (5.17)\end{aligned}$$

Hence we have,

$$\begin{aligned} P_{ij}^*(s;n+1) &= \int_0^\infty \sum_{l=0}^\infty k_l^*(s,x) d_x(P_{i,j+l-1}^{(n)}(x)), \quad (j \ge 1) && (5.18)\\ P_{i0}^*(s;n+1) &= \int_0^\infty \sum_{l=0}^\infty \sum_{m=l+1}^\infty k_m^*(s,x) d_x(P_{il}^{(n)}(x)) && (5.19)\end{aligned}$$

It is shown in Appendix E.1, that by substituting (5.9) into (5.18) and (5.19) and then taking generating functions with respect to j, we finally obtain after using (5.8), (5.10) and (5.11),

$$P_i^*(s,z;n+1) = zB(s,z)P_i^*(A(s,z),z;n) + \sum_{m=0}^{\infty}(1-z^{-m})C_m^*(s;n) \tag{5.20}$$

where,

$$\begin{aligned} A(s,z) &= \rho(s+\mu-\mu z^{-1}) && (5.21)\\ B(s,z) &= \rho + \frac{(1-\rho)\lambda}{\lambda+\mu+s-\mu z^{-1}} && (5.22)\end{aligned}$$

Once again taking the generating functions with respect to n we have, from (5.20), (see Appendix E.2 for derivation),

$$P_i^*(s,z;w) = wzB(s,z)P_i^*(A(s,z),z;w) + z^i + wC(s,z,w) \tag{5.23}$$

$A(s,z) = s$ has exactly one positive root for s. Denoting the root by ξ and putting $s = \xi$ in (5.23), we have,

$$P_i^*(\xi,z,w) = \frac{z^i + C(\xi,z,w)}{1-wzB(\xi,w)} \tag{5.24}$$

We note that for CIPP, $\xi = \frac{\rho\mu(1-z^{-1})}{1-\rho}$ and is positive. Now define,

$$\Phi(s) = P_i^*(s,z,w) - P_i^*(\xi,z,w) \tag{5.25}$$

Equation (5.23) then reduces to,

$$\Phi(s) = h(s,\Phi(A(s,z))) \tag{5.26}$$

where

$$\begin{aligned} h(s,y) &= wzB(s,z)[y+H(s)] && (5.27)\\ H(s) &= [wzB(s,z)]^{-1}\left[z^i + wC(s,z,w) - \frac{1-wzB(s,z)}{1-wzB(\xi,z)}[z^i+wC(\xi,z,w)]\right] && (5.28)\\ H(\xi) &= 0 && (5.29)\end{aligned}$$

We now solve the functional equation (5.26) in s (for fixed w and z) by making use of a theorem due to Read [99] (see Appendix F for Read's theorem). It can be seen that the following conditions are satisfied:

1. $A(s,z)$ is regular at $s=\xi$,

2. $0 < A'(\xi, z) < 1$,

3. the equation $\eta = h(\xi, \eta)$ has a solution $\eta = 0$.

4. $h(s, y)$ is regular.

5. since $|1 + (1-\alpha)(\lambda - \lambda z^{-1} + \xi)| > 1$, $|\frac{dh}{dy}|_{s=\xi, y=0} = |wzB(\xi, z)| < \frac{1}{|A'(\xi,z)|}$.

Hence, the solution of the (5.26) is given by,

$$\Phi(s) = \lim_{n\to\infty} \Phi_n(s) \tag{5.30}$$

where $\Phi_n(s) = h(s, \Phi_n(A(s,z)))$ and $\Phi_0(s)$ is an arbitrary function taking the value 0 at ξ. This solution is regular at ξ with $\Phi(\xi) = 0$. Taking $\Phi_0(s) = 0$, we obtain,

$$\Phi(s) = \sum_{j=0}^{\infty} H(f^j) \prod_{i=0}^{j} [wzB(f^i, z)] \tag{5.31}$$

$$f^{i+1} = A(f^i, z) \tag{5.32}$$

with $f^0 = s$.
By substituting $H(s)$ from (5.28) and using (5.25), we finally obtain,

$$P_i^*(s, z, w) = H_1(s, z, w) + H_2(s, z, w), \tag{5.33}$$

where

$$H_1(s,z,w) = z^i \left[1 + \frac{wzB(s,z)}{1 - wzB(\xi,z)} + \sum_{j=1}^{\infty} (wz)^j \left[1 - \frac{1 - wzB(f^j,z)}{1 - wzB(\xi,z)}\right] \prod_{m=0}^{j-1} B(f^m, z)\right]$$

$$H_2(s,z,w) = \begin{cases} wC(s,z,w) + \frac{wzB(s,z)}{1-wzB(\xi,z)} wC(\xi,z,w) \\ \\ + \sum_{j=1}^{\infty} (wz)^j \left[wC(f^j,z,w) - \frac{1-wzB(f^j,z)}{1-wzB(\xi,z)} wC(\xi,z,w)\right] \prod_{m=0}^{j-1} B(f^m,z) \end{cases}$$

We now consider a function $F(s, z, w)$ defined by,

$$F(s,z,w) = \begin{cases} [1 - zT(w)]P_i^*(s,z,w), & (|z| \le 1, |w| < 1, s \ge 0) \\ (1 - zT(w))Q_i(s,z,w), & (|z| \ge 1, |w| < 1, s \ge 0) \end{cases} \tag{5.34}$$

where, $Q_i(s, z, w)$ denotes the right-hand side of (5.33) and $T(w)$ is the root of

$$z = wB(\xi, z^{-1}) \tag{5.35}$$

inside the unit circle. The quantity $T(1)$ is of interest in the following sections. It is straightforward to show that,

$$T = T(1) \tag{5.36}$$

$$= \frac{(1-\rho) + \rho(\mu/\lambda)}{\mu/\lambda} \tag{5.37}$$

Since $P_i^*(s,z,w)$ is the generating function of a probability distribution $P_i^*(s,z,w)$, $F(s,z,w)$ must be a regular function of z for $|z|<1$. Also, as the only zero of $1-wzB(\xi,z)$ outside the unit circle is that of $(1-zT(w))$, $F(s,z,w)$ must be regular for $|z|>1$. $F(s,z,w)$ is also continuous on the boundary. Hence, by analytic continuation, $F(s,z,w)$ is a regular function of z for all finite z.

Since $C(s,z,w)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}(1-z)^{-m}C_m^*(s;n)w^n$, it follows by Abel's theorem (see Appendix F for statement of Abel's theorem) on the continuity of power series that $C(s,z,w)$ converges to $\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}C_m^*(s;n)w^n$ and $C(f^i,z,w)$ converges to $\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}C_m^*(a_j(s);n)w^n$ as $z\to\infty$ where $a_j(s)=\lim_{z\to\infty}f^j$. By using an iterative technique, which is similar to that used by Phatarford [100], we obtain, for (5.21),(5.22), (5.32),

$$\begin{aligned} a &= \lim_{j\to\infty} a_j(s) \\ &= \frac{\mu\rho}{1-\rho} \end{aligned} \tag{5.38}$$

and $a_j(a)=a$ for all $j\geq 0$. Thus, $\lim_{z\to\infty}(F(s,z,w)/z^{i+1})$ exists and vanishes at $s=a$. Hence, $F(s,z,w)$ must be of the form

$$F(s,z,w)=\sum_{m=0}^{i+1} z^m I_m(s,w) \tag{5.39}$$

where, $I_m(s,w)$ $(m=0,1,2,\ldots,i+1)$ are functions of s and w alone, satisfying

$$I_{i+1}(a,w)=0 \tag{5.40}$$

$$P_i^*(s,z,w)=[1-zT(w)]^{-1}\sum_{m=0}^{i+1} z^m I_m(s,w) \quad (|z|\leq 1, |w|<1, s\geq 0), \tag{5.41}$$

giving

$$P_{ij}^*(s,w)=\begin{cases} \sum_{m=0}^{j}[T(w)]^{j-m}I_m(s,w), & (j\leq i) \\ \sum_{m=0}^{i+1}[T(w)]^{j-m}I_m(s,w), & (j\geq i+1) \end{cases} \tag{5.42}$$

It now remains to determine the functions I_m $\forall m$ so that the complete $P_{ij}^*(s,w)$ and hence $P_{ij}^*(s,w)$, can be obtained. However, because of tedious algebra we consider here only the case where the initial queue size, Q_0 is zero.

From (5.9),(5.18) and (5.19) we have,

$$P_{0j}^*(s;n+1)=\begin{cases} \sum_{l=0}^{\infty}\frac{(-\lambda)^l}{l!}\frac{\partial^l}{\partial y^l}\left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)}\right]^r \right. & \times \\ \left. P_{0,j+l-1}^*\left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)};n\right)\right]_{y=\lambda} & (j\geq 1) \end{cases} \tag{5.43}$$

By substituting $P^*_{0j}(s,w)$ from (5.42) in (5.43) and using Taylor's series theorem, we finally obtain, (see Appendix E.3 for detailed derivation),

$$T(w)U(s,w) = wB(s,T(w)^{-1})U(A(s,T(w)^{-1}),w) \tag{5.44}$$

where,

$$U(s,w) = T(w)I_0(s,w) + I_1(s,w) \tag{5.45}$$

In order to solve this functional equation, we first note that, $(T(w)/w)[B(g^i,T(w)^{-1})]^{-1} \to 1$ as $i \to \infty$, where g^i is the i^{th} iterate of $A(s,T(w)^{-1})$ and $g^i \to \xi$ geometrically fast. Further, noting that $|\partial B(s,T(w)^{-1})/\partial s|_{s=\xi} < \infty$, it can be proved that,

$$\lim_{n\to\infty} V_n(s) = \lim_{n\to\infty} \prod_{i=0}^{n-1} (T(w)/w)[B(g^i,T(w)^{-1})]^{-1} \tag{5.46}$$

tends to a limit, $V(s)$, say. Then, the general solution of the functional equation in this case is given by (Kuczma [101], p.48, Theorem 2.2, reproduced in Appendix F)

$$U(s,w) = \frac{D}{V(s)} \tag{5.47}$$

where D is independent of s. Although we were not able to obtain the closed form expression for $V(s)|_{s=0,w=1}$, we show that the sequence $V_n(s)|_{s=0,w=1}$ converges. In Appendix E.4, we give the expression for $V_n(s)|_{s=0,w=1}$.

$$V_n(s)|_{s=0,w=1} = \frac{(\rho+\frac{\lambda}{\mu}(1-\rho))(1-\rho+\rho\frac{\lambda}{\mu})}{(1-\rho)(\rho+\frac{\lambda}{\mu})+\rho^2\frac{\lambda}{\mu}} \prod_{i=1}^{n-1} \frac{(\rho+(1-\rho)\frac{\lambda}{\mu})(1-(1-\frac{\lambda}{\mu})\rho^{i+1})}{(1-\rho)\frac{\lambda}{\mu}+\rho(1-(1-\frac{\lambda}{\mu})\rho^{i+1})} \tag{5.48}$$

Note that the element in the product term approaches 1, as $i \to \infty$ and hence the sequence $V_n(s)|_{s=0,w=1}$ converges.

We note in Appendix E.5, that,

$$P^*_i(s,1,w) \;\; = \;\; \sum_{n=0}^{\infty} \frac{\lambda+\rho^n s_n}{\lambda+s_n} w^n \tag{5.49}$$

which is the Laplace-Steiltjes transform of the distribution of T_n given $T_0 = 0$, we have from (5.41), for $i = 0$,

$$I_0(s,w) + I_1(s,w) = [1-T(w)] \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n s_n}{\lambda+s_n}\right) w^n \tag{5.50}$$

By solving (5.44) and (5.50) for I's and using (5.40), we obtain D so that I's are fully determined. By substituting these I's in (5.42), we finally obtain, (see Appendix E.6 for derivation)

$$\begin{aligned} P_{0j}(w) &= P^*_{0j}(0,w) & (5.51)\\ &= \begin{cases} (1-w)^{-1} - T(w)\frac{M(a,w)}{M(0,w)}\sum_{n=0}^{\infty}\left(\frac{\lambda+\rho^n a}{\lambda+a}\right)w^n, & j=0 \\ (1-T(w))[T(w)]^j\frac{M(a,w)}{M(0,w)}\sum_{n=0}^{\infty}\left(\frac{\lambda+\rho^n a}{\lambda+a}\right)w^n, & j>0 \end{cases} & (5.52) \end{aligned}$$

where,

$$M(s,w) = \begin{cases} \frac{T(w)}{w\left(\rho+\frac{\lambda(1-\rho)}{\lambda+s+\mu(1-T(w))}\right)} \times \\ \lim_{n\to\infty}\prod_{i=1}^{n-1}\left(\frac{T(w)}{w+\left(\rho+\frac{\lambda(1-\rho)}{\lambda+\rho^i s+\mu(1-T(w))\frac{(1-\rho^{i+1})}{1-\rho}}\right)}\right) \end{cases}$$

5.2.3 Limiting Behavior of the Queue Size

We have from (5.6) and (5.49), that for $|\rho|<1$,

$$\lim_{w\to 1}(1-w)\sum_{n=0}^{\infty}\left(\frac{\lambda+\rho^n s}{\lambda+s}\right)w^n = \left(\frac{\lambda}{\lambda+a}\right)^r \tag{5.53}$$

It can be seen that from (5.35), $T(1)$ lies in the unit circle iff $\lambda/\mu<1$. Thus, we have from (5.52) that $\lim_{w\to1}(1-w)P_{0j}(w)$ exists if and only if $\lambda/\mu<1$ and $|\rho|<1$. In this situation, we obtain, from (5.52),

$$\begin{aligned} P_j &= \lim_{w\to 1}(1-w)P_{0j}(w) & (5.54)\\ &= \begin{cases} 1-\frac{TM(a,1)}{M(0,1)}\left(\frac{\lambda}{\lambda+a}\right), & j=0 \\ \frac{(1-T)T^jM(a,1)}{M(0,1)}\left(\frac{\lambda}{\lambda+a}\right), & j>0 \end{cases} & (5.55) \end{aligned}$$

where, $T=T(1)$ is given by (5.37), which is <1 for $\mu/\lambda>1$ and $a=\frac{\mu\rho}{1-\rho}$. $M(a,1)$ and $M(0,1)$ are given by,

$$M(0,1) = \begin{cases} \frac{(\rho+\frac{\lambda}{\mu}(1-\rho))(1-\rho+\rho\frac{\lambda}{\mu})}{(1-\rho)(\rho+\frac{\lambda}{\mu})+\rho^2\frac{\lambda}{\mu}} \times \\ \lim_{n\to\infty}\prod_{i=1}^{n-1}\frac{(\rho+(1-\rho)\frac{\lambda}{\mu})(1-(1-\frac{\lambda}{\mu})\rho^{i+1})}{(1-\rho)\frac{\lambda}{\mu}+\rho(1-(1-\frac{\lambda}{\mu})\rho^{i+1})} \end{cases} \tag{5.56}$$

$$M(a,1) = \begin{cases} \frac{((1-\rho)\frac{\lambda}{\mu}+\rho)(1-\rho(1-\rho)(1-\frac{\lambda}{\mu}))}{\left[(1-\rho)^2\frac{\lambda}{\mu}+\rho(1-\rho(1-\rho)(1-\frac{\lambda}{\mu}))\right]} \times \\ \lim_{n\to\infty}\prod_{i=1}^{n-1}\frac{\left(\rho+(1-\rho)\frac{\lambda}{\mu}\right)\left[(1-\rho)(1-(1-\frac{\lambda}{\mu})\rho^{i+1})+\rho^{i+1}\right]}{(1-\rho)^2\frac{\lambda}{\mu}+\rho\left[(1-\rho)(1-(1-\frac{\lambda}{\mu})\rho^{i+1})+\rho^{i+1}\right]} \end{cases} \tag{5.57}$$

Note that P_j is the stationary probability of j number of customers seen by an arriving customer.

5.2.4 Waiting-time Distribution

In this section, we deduce the form of the waiting-time distribution from that of the queue-size distribution. Let W_n be the waiting time of the nth customer; then $W_n = \sum_{i=1}^{Q_n} v_i$, where $v_1, v_2, \ldots, v_n$, are independent and identically distributed as $1 - e^{-\lambda x}$. Thus, if we let $\Omega_n(s)$ be the Laplace-Steiltjes transform of $Pr\{W_n \leq x/Q_0 = 0\}$, then we have

$$\Omega_n(s) = \sum_{j=0}^{\infty} \left(\frac{\mu}{\mu+s}\right)^j P_{0j}^{(n)} \tag{5.58}$$

where $P_{0j}^{(n)} = Pr\{Q_n = j|Q_0 = 0\}$. Hence, by taking generating functions with respect to n and using (5.55), we obtain,

$$\begin{aligned} \sum_{n=0}^{\infty} \Omega_n(s) w^n &= \sum_{j=0}^{\infty} \left(\frac{\mu}{\mu+s}\right)^j P_{0j}(w) \\ &= \begin{cases} (1-w)^{-1} - \frac{T(w)M(a,w)}{M(0,w)} \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n a}{\lambda+a}\right) w^n + \\ \frac{(1-T(w))M(a,w)\mu T(w)}{M(0,w)(s+\mu(1-T(w))} \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n a}{\lambda+a}\right) w^n \end{cases} \end{aligned} \tag{5.59}$$

If $\lambda/\mu < 1$ we obtain, from (5.53) and (5.59),

$$\lim_{n\to\infty} \Omega_n(s) = \lim_{w\to 1}(1-w) \sum_{n=0}^{\infty} \Omega_n(s) w^n \tag{5.60}$$

$$= 1 - \frac{TM(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda+a}\right) + \frac{(1-T)M(a,1)\mu T}{M(0,1)(s+\mu(1-T))} \left(\frac{\lambda}{\lambda+a}\right) \tag{5.61}$$

By inversion and denoting the limiting waiting-time distribution by $Pr\{W \leq x\}$, we finally obtain,

$$Pr\{W \leq x\} = 1 - \frac{TM(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda+a}\right) exp(-\mu x(1-T)) \tag{5.62}$$

Remark: It is worth noting that some of known results for the $M/M/1$ system can be obtained from the results presented here, by setting $\rho = 0$. To see this, set $\rho = 0$ in (5.21) and (5.38). This gives $\xi = 0$ and $a = 0$. Hence, from (5.55) and (5.62), we have

$$P_j = (1-T)T^j \quad j \geq 0 \tag{5.63}$$

$$Pr\{W \leq x\} = 1 - Texp\{-\mu(1-T)x\} \tag{5.64}$$

where $T = \lambda/\mu$.

5.3 Experimental Verification of Theoretical Results

In order to verify the theoretical results corresponding to CIPP/M/1 queue, we need the statistical measures over which we can make the comparisons between analytical and estimated results (from simulations). For that purpose, we derive the queueing performance measures of practical importance (which were introduced in Section 3.4), for the case of CIPP/M/1 queue from the expressions for stationary distributions derived in the previous section. We first simulate the CIPP/M/1 queue for some parameter values of CIPP and service rates. We then compare the estimated values of the queueing performance measures from simulations with the corresponding *analytical* queueing performance measures computed for the same set of parameters.

5.3.1 Derivation of Queueing Performance Measures

We derive the expressions for the queueing performance measures corresponding to CIPP/M/1 queue from the stationary distributions derived in the previous section. We define

$$\beta \triangleq \frac{M(a,1)}{M(0,1)} \frac{\lambda}{\lambda + a} \tag{5.65}$$

where, $M(0,1)$, $M(a,1)$ and a are given by (5.56), (5.57) and (5.38) respectively. T is given by (5.37).

5.3.1.1 Mean Number of Packets in the System

From (5.55), we have

$$\mathbf{E}\{Q\} = \frac{TM(a,1)}{(1-T)M(0,1)} \left(\frac{\lambda}{\lambda + a}\right). \tag{5.66}$$

5.3.1.2 Mean Waiting Time

From (5.62), we have

$$\mathbf{E}\{W\} = \frac{TM(a,1)}{\mu(1-T)M(0,1)} \left(\frac{\lambda}{\lambda + a}\right). \tag{5.67}$$

5.3.1.3 Packet Delay Variation

In a broadband network, for a given connection, packets follow a fixed route and their sequence is preserved. However, they may encounter different traffic condition in shared buffers and therefore experience different packet transfer delays [67]. Hence, we use the term *packet delay variation (pdv)* (denoted by τ) to denote the variance of the packet transfer delay (Y) on the same connection. Note that $Y = W + S$ where, W is the stationary waiting time and S is the stationary service time. It is shown in Appendix G, that Y has the following density

$$f_Y(y) = \mu\left((T-\beta)e^{-(1+\frac{\beta T}{1-T})}e^{-\mu y} + \beta(1-T)e^{-(1-T(1-\beta))}e^{-\mu y(1-T)}\right) \tag{5.68}$$

where, β is as given above. Then, it is straight forward to derive the expression for *pdv* (see Appendix G for derivation),

$$\tau = \frac{2}{\mu^2}\left[1-\beta+\frac{\beta}{(1-T)^2}\right] - \left[\frac{1}{\mu}\left(1+\frac{\beta T}{(1-T)}\right)\right]^2 \tag{5.69}$$

5.3.1.4 Overflow Probability

Overflow probability is the probability that the number of packets in the buffer exceeds a threshold in an infinite buffer queue. Sometimes, in literature [67], overflow probability is used as an approximation to loss probability in corresponding finite capacity queue. For CIPP/M/1 case, it is straight forward to show that the overflow probability for the threshold of K packets[4] for $K > 0$, is given by,

$$Pr\{Q \geq K+1\} = \frac{M(a,1)}{M(0,1)}\left(\frac{\lambda}{\lambda+a}\right)T^{K+1} \tag{5.70}$$

where Q denotes the number of packets in the system (queue and server).

5.3.2 Experimental Results

For the theoretical computations of queueing measures which involve the computation of $M(a,1)$ and $M(0,1)$, we take 'n' on the RHS of $M(a,1)$ and $M(0,1)$ such that the percentage error (between n^{th} and $(n+1)^{th}$ term) is less than 2%. We consider the parameters $\lambda = 1.2\ \ (sec)^{-1}$ and $\mu = 2.0\ \ (sec)^{-1}$ for our experiments.

5.3.2.1 Mean Number of Packets in the System

We compute the theoretical mean number of packets as given by (5.66). On the other hand, we simulate the CIPP/M/1 queue corresponding to the above parameters. Figure

[4] The unit "packet" refers to the mean packet length (in bits). This terminology is used in the rest of the thesis.

5.1 depicts our results. From the figure, it is seen that the simulation results agree well with the theoretical results for the mean number of packets in the system. Note also that the mean number of packets degrades as the ρ increases. This means that the queueing performance measures degrades as ρ increases. We will soon see that, this is the case with all other queueing performance measures we considered.

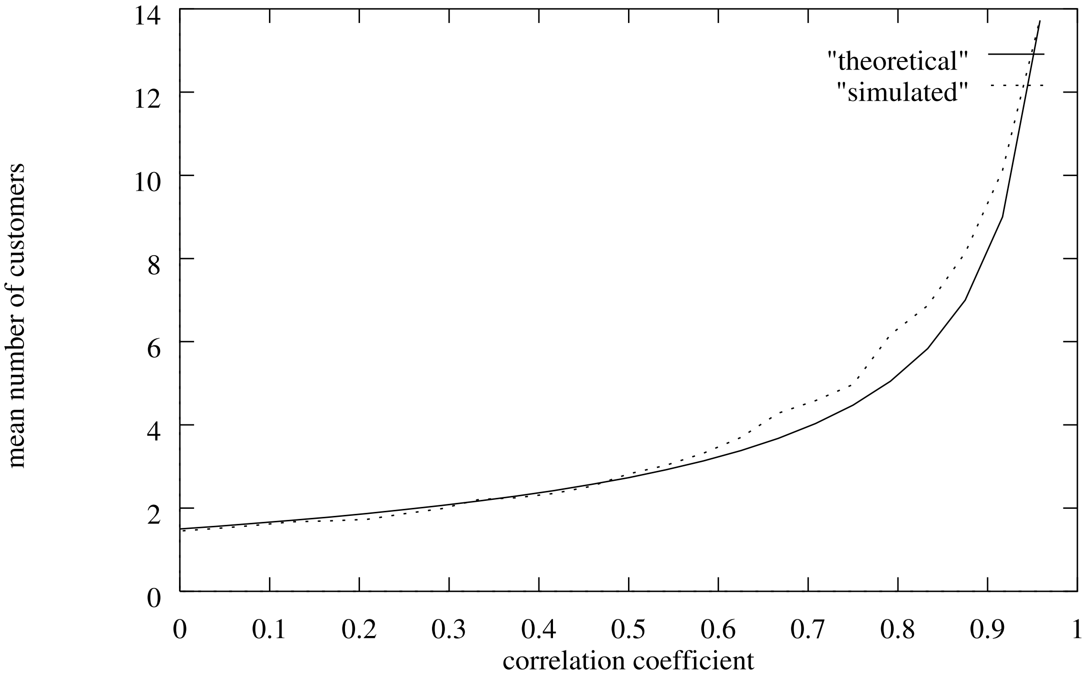


Figure 5.1: Experimental verification of theoretical results for CIPP/M/1 queue with respect to the measure, the mean number of packets in the system.

5.3.2.2 Mean Waiting Time

In order to validate the theory, we have computed the theoretical mean waiting time as given by (5.67) and simulated the CIPP/M/1 queue with same parameters. Our results are illustrated in Figure 5.2. One can see from the figure that the simulation results agree well with the theoretical results for the mean waiting time. As we saw for the mean number of customers previously, mean waiting time also increases as ρ increases.

5.3.2.3 Packet Delay Variation

Figure 5.3 shows the *pdv* versus correlation parameter ρ. One may see from figure that the simulation results agree well with the theoretical result given by (5.69). Again, the behavior of *pdv* follows the behavior as the above measures.

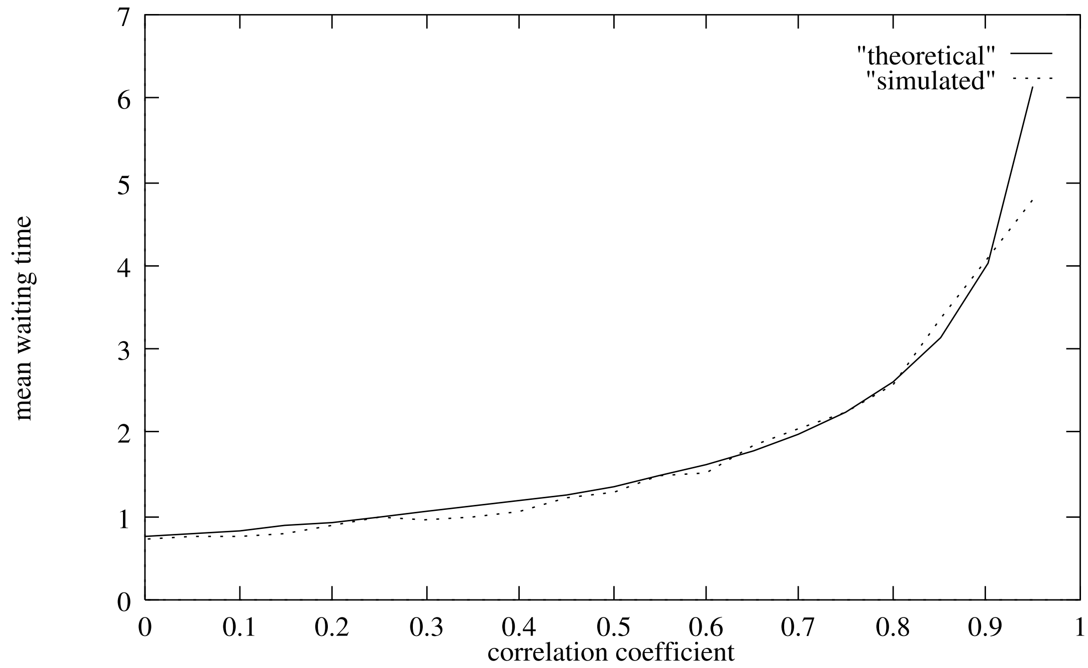


Figure 5.2: Experimental verification of theoretical results for CIPP/M/1 queue with respect to the measure, the mean waiting time for packets.

5.3.2.4 Overflow Probability

For the purpose of simulations, we have taken a threshold of 32 packets. Figure 5.4 shows our results for overflow probability estimated from simulations and from analytical expression given by (5.70). It is seen form the figure that our simulation results agree well with the theoretical results. Again there is a degradation in the queueing performance as ρ increases.

5.4 CIPP/D/1 Queue

In this section, we discuss the single server, FCFS, infinite capacity queue with CIPP input and constant service intervals. A theorem due to Benes [102], expressing the virtual waiting time V_t (the time that a fictitious customer has to wait until he enters service, given that he had entered the system at time t) in a G/G/1 queue is used below to compute the mean delay in the CIPP/D/1 queue. In this section, we use the term *cell* to denote the 53 byte ATM packet.

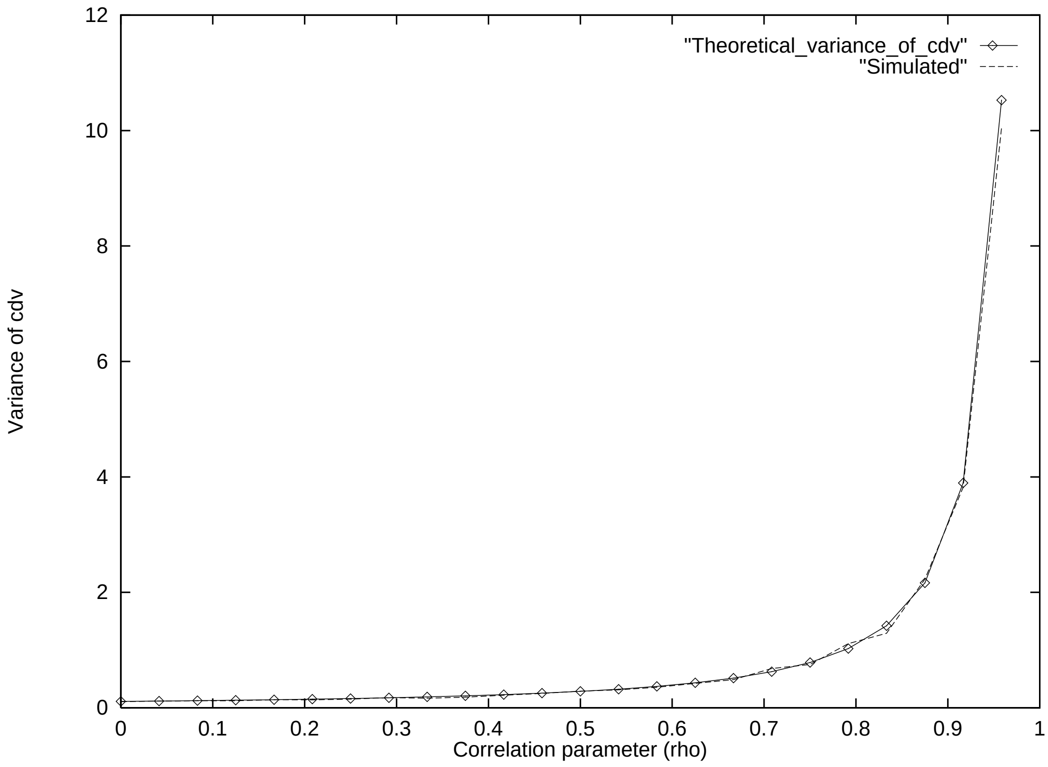


Figure 5.3: Experimental verification of theoretical results for CIPP/M/1 queue with respect to the measure, the packet delay variation *pdv*

5.4.1 Benes Approach

The results presented in this subsection hold for a discrete-time, infinite capacity queueing model with service times equal to one slot. Moreover, the work (the amount of service in seconds which an arriving customer requires) arrives discontinuously in quanta of equal size (in our case, cells). Let $\nu(t) =$ number of cell arrivals in $[-t, 0)$. To ensure stability (note that, we have infinite capacity queue), we require that,

$$\lim_{t\to\infty} \frac{\nu(t)}{t} < 1 \quad \text{a.}s \tag{5.71}$$

Let $\xi(t)$ denote the excess work arriving in $[-t,\ 0)$. Then, it follows that $\xi(t) = [\nu(t) - t]^+,\ \ t \geq 0$. Now, V_0 is given by the Reich's formula (see page 374, [1]), $V_0 = \sup_{t\geq 0} \xi(t)$. Denoting the complimentary distribution function of V_0, by $Q(x)$, we have

$$Q(x) \triangleq Pr\{V_0 > x\}. \tag{5.72}$$

If the process $\xi(t)$ is piecewise continuous and differentiable, we can deduce the Benes formula for constant service time queue,

$$Q(x) = \sum_{n>x} Pr\{\nu(n-x) = n, V_{(n-x)} = 0\}. \tag{5.73}$$

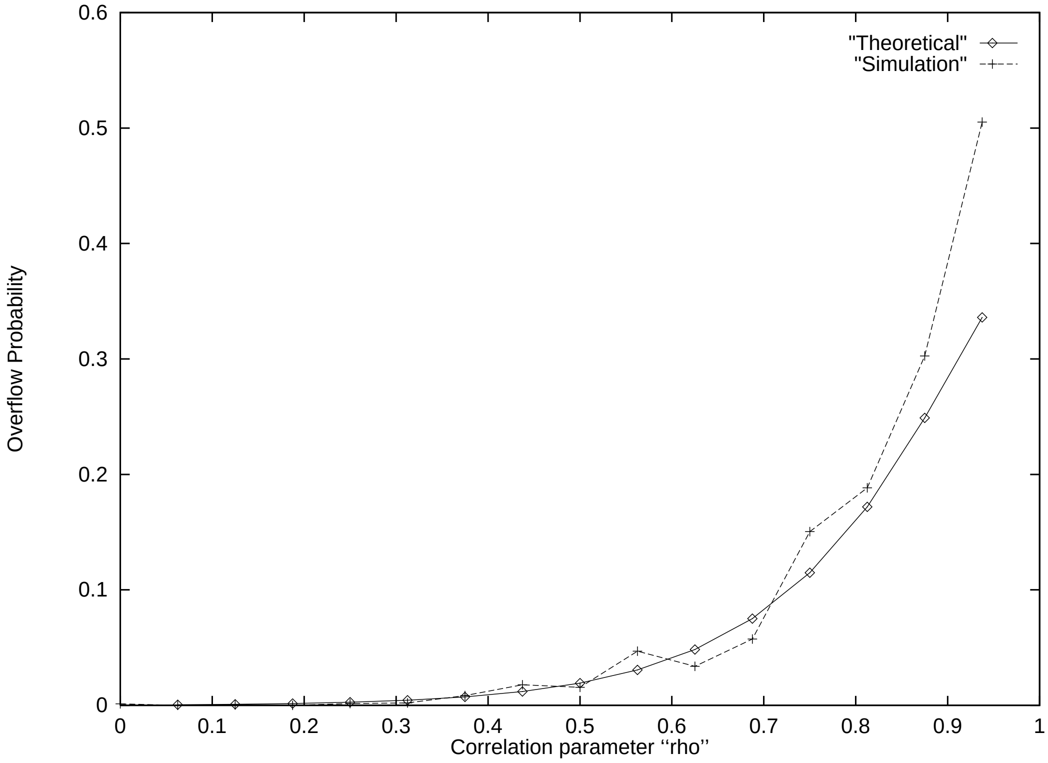


Figure 5.4: Experimental verification of theoretical results for CIPP/M/1 queue with respect to the measure, the overflow probability for a threshold of 32 packets.

The joint probability given above, is difficult to calculate exactly because of the difficulty of taking the condition $V_{(n-x)} = 0$ into account. But, it is possible to derive a good approximation, as noted in [1]. We give here only the result,

$$Q(x) \approx \sum_{n>x} \left[Pr\{\nu(n-x) = n\} - \rho \;\; Pr\{\nu(n-x) = n | \text{one arrival at } -(n-x)\}\right] \quad (5.74)$$

Note that the probability distribution in the first term (inside the summation in the above equation) is count stationary, whereas the distribution in the second term is interval stationary[5]. Hence, for CIPP, (5.74) takes the following form,

$$Q(x) \approx \sum_{n>x} \{P_n(n-x) - \gamma P_n^c(n-x)\} \quad (5.75)$$

where,

$$P_n^c(t) = \left\{ \sum_{j=0}^{n} \frac{\prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \alpha_n + 1)(\alpha_j - \rho\alpha_n)\prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \qquad 0 \le \rho < 1 \right. \quad (5.76)$$

and,

$$P_n(t) = \begin{cases} (1 + \rho - (1-\rho)\alpha_n) \quad \times \\ \sum_{j=0}^{n} \frac{\alpha_j \prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\rho(\alpha_j+1) - \alpha_n + 1)(\alpha_j - \alpha_n + 1)(\alpha_j - \rho\alpha_n)\prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \quad 0 \le \rho < 1 \end{cases} \quad (5.77)$$

[5]This is apparant, after shifting the origin to $-(n-x)$.

where, γ is the load[6]. Both (5.76) and (5.77) are valid for $n = 0, 1, 2, \ldots\ , \ \lambda > 0, \ \ t \geq 0, \ $ and $\alpha_j = \sum_{i=0}^{j} \rho^i$. However, (5.75) poses a numerical problem due to the appearance of large alternating, very nearly canceling terms in both the expressions for $P_n^c(t)$ and $P_n(t)$ given by (5.76) and (5.77) respectively for large values of n and ρ. Hence, we resort to simulations for the estimation of $P_n(t)$ and $P_n^c(t)$.

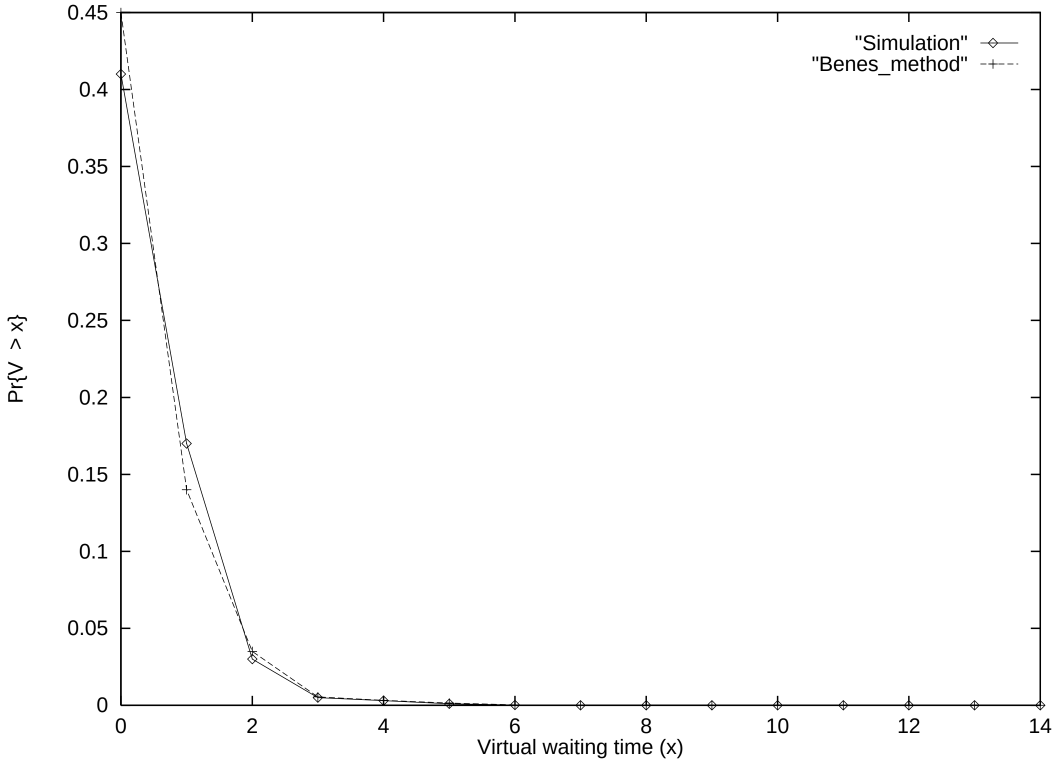


Figure 5.5: Complimentary distribution for virtual waiting time in a CIPP/D/1 queue: analytical result given by (5.75)) and simulated result

Once $Q(x)$ is computed, the stationary distribution for cell delay W_d (the number of slots an arbitrary cell resides in the buffer, including its transmission slot) can be computed by inter-relationships between the steady-state probability mass function of the buffer occupancy Y, cell delay W_d and complimentary distribution $Q(x)$ [1]. We now present those relationships.

In the constant service (and discrete) time queueing system, the buffer occupancy at the beginning of the k^{th} slot is represented by Y_k. Then, the evolution of the queue can be described by the system equation,

$$Y_{k+1} = (Y_k - 1)^+ + b_k \tag{5.78}$$

[6]the ratio of the arrival rate to the service rate. While the term 'load' is well entrenched in queueing theory literature, the term link utilization factor (LUF) is used in the traffic modeling literature to denote the same quantity.

where b_k denotes the number of cells entering the buffer during the slot k. Here, it is implicitly assumed that the the maximum number of cells transmitted per slot is 1. Then, one can show that (refer page 373, [1]),

$$Pr\{W_d = n\} = Pr\{Y = n\}/\gamma \ \text{ for } \ n \geq 1 \tag{5.79}$$

independent of the statistical details of the cell arrival process, except for the utilization,

$$\gamma = \lim_{k \to \infty} \mathbf{E}\{b_k\}/c \tag{5.80}$$

Note that, by definition, $Pr\{W_d = 0\} = 0$. Observe that the following relation exists between the complimentary distribution $Q(.)$ and the $Pr\{Y = n\}$,

$$Pr\{Y = n\} \quad = \quad (Q(n-1) - Q(n)) \tag{5.81}$$

5.4.2 Simulation Results

We considered $\rho = 0.9$ and $\lambda = 7.5$ arrivals per second. Figure 5.5 shows the complimentary distribution $Q(x)$ for CIPP/D/1 queue computed using (5.75) and the complimentary distribution estimated from the simulation of a CIPP/D/1 queue. It is seen from the figure that the simulation results match with the theoretical results even though the expression we have is an approximate one. Using (5.74), (5.79) and (5.81) we have computed the theoretical mean waiting time, by truncating the sum (such that the error is less than 8 %) in the definition of expectation. To validate the theoretical results, we simulated the CIPP/D/1 queue. We have taken link capacity as 150.95 cells/second (equivalent to 64 Kbps) and $\rho = 0.98$. Figure 5.6 shows the theoretical mean waiting time and the simulated mean waiting time for CIPP/D/1 queue.

The percentage errors[7] computed are as given below: a minimum percentage error of 0.4385 occurs at a load of 0.73 and a maximum percentage of 48.48 occurs at 0.33. One may see from the figure and the error calculations that our theoretical results agree reasonably closely with the simulated result, even though the analytical expression is an approximation.

5.5 Implications of Queueing Theory Results

Two important issues in the context of the theory for CIPP/M/1 queue catch immediate attention: first, the implications of these results and its relation to the work of Loynes [88],

[7] See Section 6.1.1 for formal definition of percentage error.

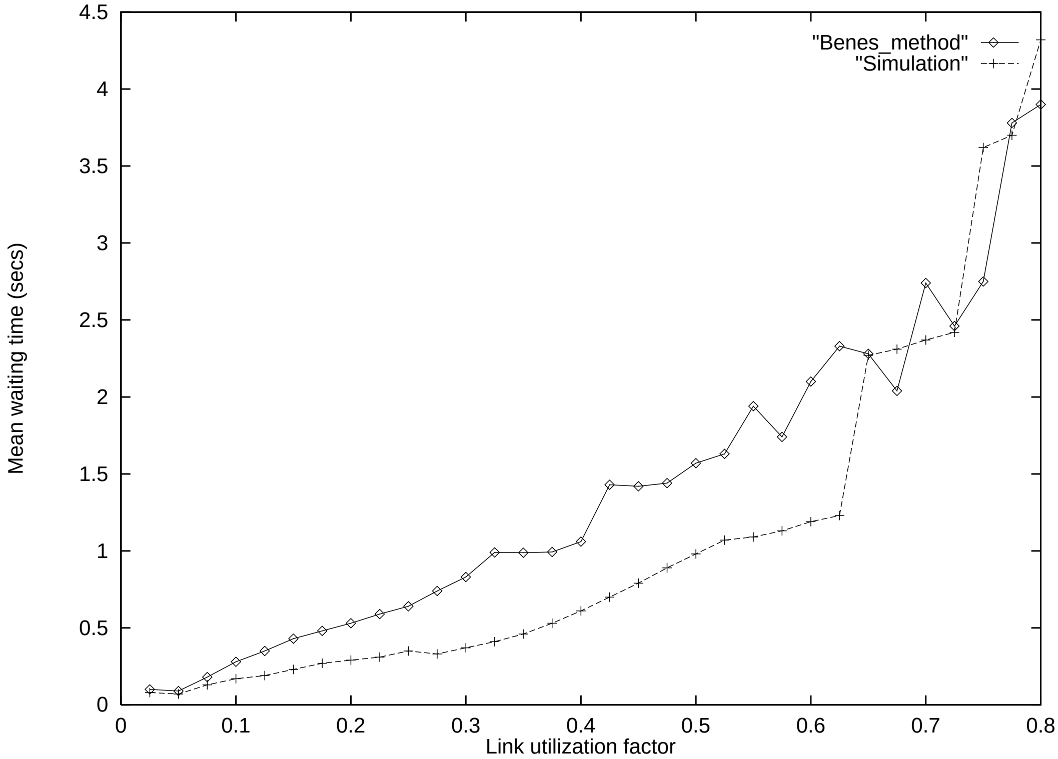


Figure 5.6: Experimental verification of theoretical results for the mean waiting time in a CIPP/D/1 queue

second, the impact of correlation in interarrivals on the queueing performance measures. This section is devoted to explain these issues in detail. In addition, we discuss the computational advantage of CIPP/M/1 queue.

5.5.1 Implications of Loyne's Result

Loynes in [88], has considered more general queues with interarrival and service times forming stationary sequences (with arbitrary dependence among themselves and arbitrary cross correlation). He has proved that, for the stability of the queue, the ratio of arrival rate to service rate should be less than 1. The correlation structure among the interarrival sequence has nothing to do with the stability of the queue. In the CIPP/M/1 case also, one can see from the (5.55) and (5.62) that the correlation parameter ρ has no role in deciding the stability of the queue. It is only the ratio of λ to μ that decides the stability of the queue.

5.5.2 Impact of Interarrival Correlations on Queueing Performance

One of the potential issues in the queue with correlated arrivals is the impact of correlation in interarrivals on the queueing performance. Livny *et al* [80] and Patuwo *et al* [103] have independently investigated the effect of correlations on queueing performance *empirically*. After carrying out extensive simulations, they have concluded that, the positive correlation in interarrivals *always* leads to degradation in queueing performance (in their case, mean waiting time is the queueing performance measure). The effect of negative correlations, however, is not clear; it might either degrade or improve the queueing performance. We saw that the increments have positive correlation in CIPP. Moreover, the analytical results given in Sections 5.2.3 and 5.2.4 and their verification by simulations given in Section 5.3, demonstrate that the queueing performance degrades with respect to the correlation parameter ρ of CIPP. Thus we have analytical results corroborating Livny's empirical observation. Note that, earlier workers [4] also had analytical result that the positive correlation leads to degradation [80] in queueing performance. But, the correlation considered there is in the counts sequence (or in the rate process) rather than in the *interarrival* sequence.

5.5.3 Computational Advantage of CIPP/M/1 Queue

The matrix geometric method developed by Neuts, suffers computational overhead, due to the recursive computation to find the minimal nonnegative solution of some nonlinear matrix equations. Moreover, the low (linear) convergence rates of the iterative algorithms employed makes, the use of this method impractical for large dimensionality. In contrast, in CIPP/M/1 queue, the stationary probability measures of practical interest have closed form expressions over the whole range of parameters of CIPP. Also, CIPP possesses the other advantages as mentioned in the Section 4.1.2.

5.6 Finite Capacity Queues with CIPP Input

Although infinite capacity queues with CIPP input is analytically tractable, we may not be able to apply the corresponding results to real-world systems, since all real systems have limited storage capabilities. But, the analysis of finite capacity queues with CIPP input, seems intractable. Hence, we resort to simulation based performance evaluation study. We consider exponential service times (hence, we have CIPP/M/1/K queue eventually). In this section, we concentrate on the queueing performance measures for the finite capacity queues namely, packet loss probability and *conditional packet loss probability*. We also study empirically the buffer occupancy distribution for CIPP/M/1/K queue.

5.6.1 Buffer Occupancy Distribution

An intriguing feature of the buffer occupancy distributions of a finite capacity queue with highly correlated arrivals is that, the distribution shows peaks at both the extremes (empty state and fully occupied state). This means that, the buffer spends most of its time either almost empty or almost full. (This is in sharp contrast with the conventional M/M/1/K queue, where the distribution is geometrically decaying with respect to the buffer occupancy). This behavior has also been noted by Lazar *et al* [38] on the MAGNET II network and by Skelly *et al* [39] on video conferencing traces.

Here, we show that the above behavior is exhibited by finite capacity CIPP queue also. For the purpose of simulations, we consider the CIPP with $\rho = 0.97$. We take $K = 12$ and a load of 0.8. Our result for buffer occupancy distribution is shown in Figure 5.7. Also shown are the theoretical results corresponding to M/M/1/K queue with same load as for CIPP/M/1/K queue. Note that the buffer occupancy distribution for the CIPP/M/1/K queue peaks at empty state and fully occupied state. In contrast, observe that the buffer occupancy distribution for the M/M/1/K queue is *geometric*.

5.6.2 Packet Loss Probability

We study the queueing measure namely, the packet loss probability of CIPP/M/1/K for varying correlation parameter ρ of the input CIPP process. We take the buffer size of 64 packets ($K = 64$). We take $\lambda = 318.27 \quad (sec)^{-1}$ (which is the arrival rate of Bellcore `pAug.TL` dataset) and a load of 0.9. Figure 5.8 shows our results for packet loss probability corresponding to CIPP/M/1/K queue versus the correlation parameter ρ of the CIPP process. One may see that, the packet loss ratio increases as ρ increases. This means that, the queueing performance degrades as ρ increases. (Recall that, for the infinite capacity queue case, we noted in Section 5.3.1 that the queueing performance degrades as ρ increases.)

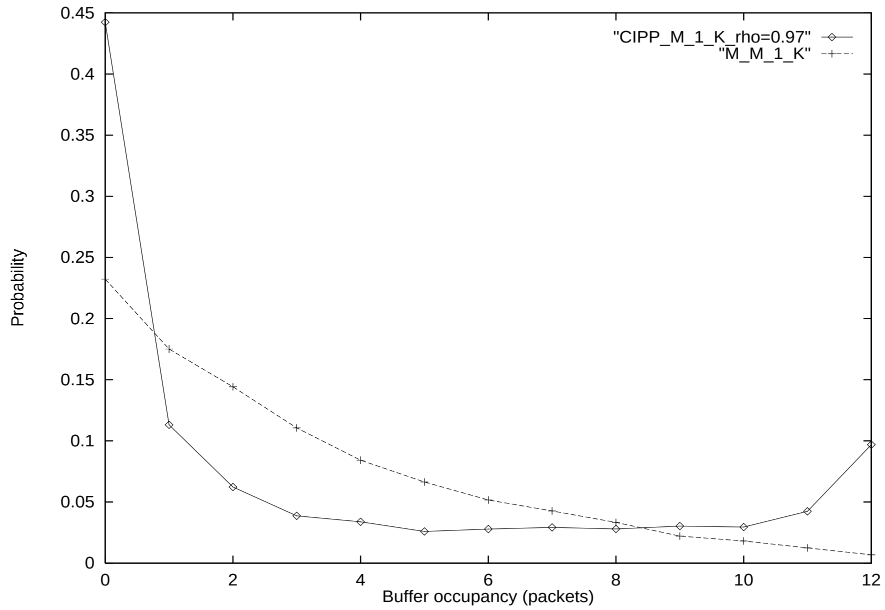


Figure 5.7: Buffer occupancy distribution in a CIPP/M/1/K queue and the corresponding behavior in an M/M/1/K queue

5.6.3 Conditional Packet Loss Probability (CPLP)

In many cases, the packet loss probability is not a sufficient performance measure as an indicator of (real-world) packet loss. Specifically, for the guaranteed services like video services in ATM networks, the probability of consecutive packet (cell) loss is of prime importance. Simulation [104] and analytical studies [105] show that even if the overall packet loss of the superposition is small, individual sources may experience the loss of a string of consecutive packets (cells), and hence suffer from an important quality of service degradation. This consecutive packet loss in composite traffic has two origins: first, there may be temporal overload situation due to the fact that the number of active users exceeds the capacity of the system and second, due to the periodical nature of the arrival process. Hence, a call losing a packet is likely to lose a string of packets. For the present discussion, the later situation is of interest, namely, the periodicity of a single source results in the consecutive packet (cell) loss. This statistical characteristics of highly correlated (almost periodic) arrivals is best captured by CIPP (with high ρ).

We explore the conditional packet loss probability for CIPP/M/1/K queue with $K = 64$ for different ρ's in order to see the effect of correlations in interarrivals on the performance measure of conditional packet loss probability. Here again, we take $\lambda = 318.27\ (sec)^{-1}$ (same as the packet arrival rate in the Bellcore `pAug.TL` dataset) and a load of 0.98. Figure 5.9 shows the conditional packet loss probability for the CIPP/M/1/K

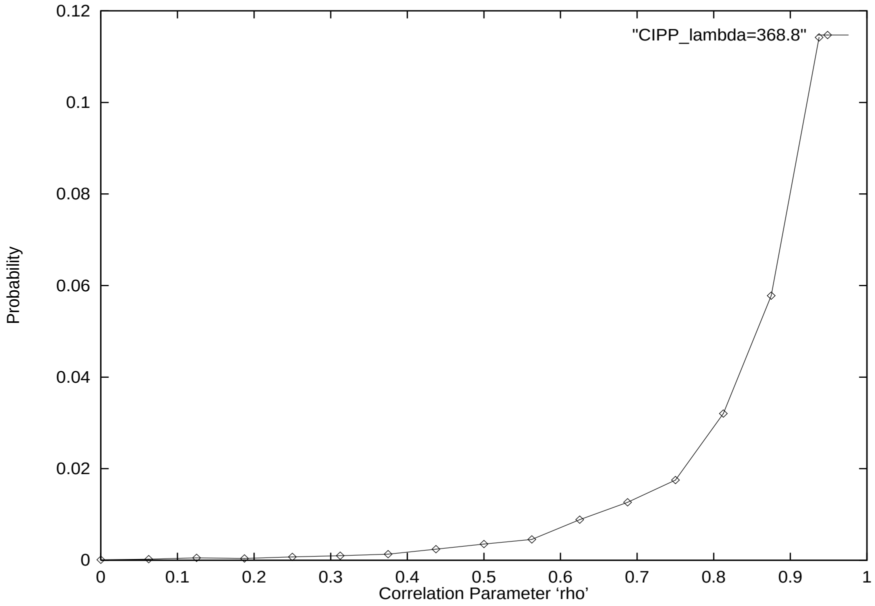


Figure 5.8: Packet loss probability versus correlation parameter 'ρ'of CIPP in a CIPP/M/1/K queue with $K = 64$ packets

queue with the CIPP characterized by the correlation parameter $\rho = 0.3, 0.6, 0.9$ and 0.98. Also shown is the corresponding results for the M/M/1/K queue case. It can be seen from the figure that, for the M/M/1/K queue case, the probability of k *consecutive* packet losses are decaying *very fast* with respect to k. But, in the CIPP/M/1/K queue case, this decay is slow, and this feature is highly pronounced for high ρ.

5.7 Some Examples

In this section, we will see some numerical examples illustrating the implications of correlations in interarrivals on the queueing performance and on some key design issues.

5.7.1 Buffer Design

A data source is connected to a data sink by a 128 kbps line. The source generates random packet traffic with a mean of 12 packets/sec. The packet lengths are exponentially distributed with a mean of 5335 bits. Now, the system designer has to determine the buffer length at the source, given the maximum loss rate of packets equal to 2 in 1000.

We model the arrival process as the Poisson process and the above situation can be treated as an M/M/1/K queue with buffer capacity (K) selected as follows. Denoting

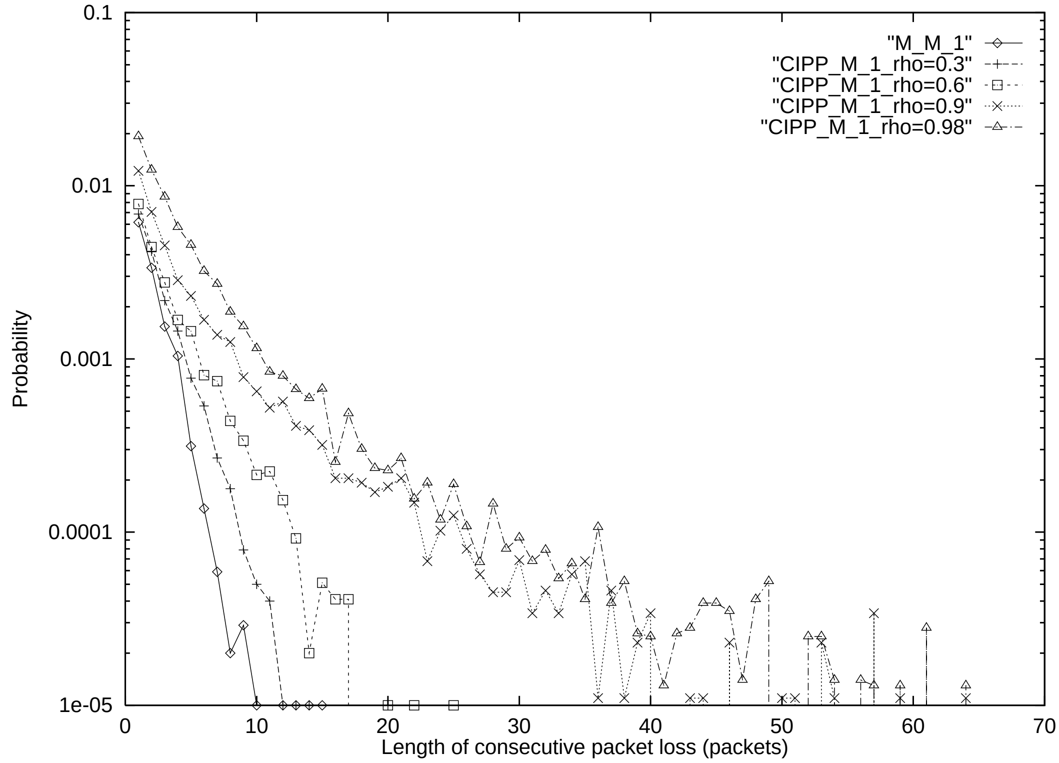


Figure 5.9: Conditional packet loss probability in a CIPP/M/1/K queue with the buffer capacity of $K = 24$ packets.

the buffer occupancy distribution by P_k, for the M/M/1/K system, we have,

$$P_K = \frac{(1-\gamma)\gamma^K}{1-\gamma^{K+1}} \tag{5.82}$$

where γ is the ratio of arrival rate to service rate. This is equal to the probability of loss of an arrival. Numerically, we solve for K, this yields the minimum buffer size as $K = 8$ packets. This implies that, the memory required is 5 KB.

Now, we model the packet arrivals by the CIPP with relatively high ρ. Here, we make an approximation. The loss probability in a finite queue is approximated by the overflow probability in an infinite queue. This is usually a reasonable approximation for a wide range of arrival processes [67]. Hence, we make use of the (5.70). We find that corresponding to $\rho = 0.7$ we have $K = 31$ packets, which means that we need a memory of size 20.670 KB. On the other hand, if we have $\rho = 0.98$ then, we have a whopping $K = 230$ packets! (This means that we need a memory of size 1.227 MB, 245 times the buffer space required corresponding to the M/M/1/K queue case). This just tells that how disastrous can be the design result if we use M/M/1/K model, while the *real-world* traffic exhibits high correlation.

In fact, it has been discovered that the Ethernet traffic (LAN, MAN and WAN) exhibits high correlations and is statistically self-similar [24]. We note that the queueing performance with such a traffic input, is shown *experimentally* to be worst [1], [28] (meaning that the mean waiting time or the mean number of packets in the queueing system with such input processes are highest as compared to the corresponding quantities in an M/M/1/K queue (or any other queue) for the same mean arrival rate). In that case, the M/M/1/K model seriously underestimates the queueing performance measures. On the other hand, CIPP captures this self-similar behavior and gives a reasonably *realistic* estimate of the queueing performance measures. Recall that we noted in Section 4.5.3 that, there is a monotonically increasing relationship between the correlation parameter ρ of CIPP and the estimated local Hurst parameter H of the CIPP. Through this relationship one may quantify the degree of self-similarity of the real-world process. This feature of CIPP makes it best suited for dimensioning purposes in broadband networks apart from the real-time control and management of such networks where the network traffic is self-similar.

5.7.2 Dedicated Versus Shared Channels

A data source is connected to a data sink by a 64 kbps line. There are eight parallel sessions using the line. Each session generates Poisson traffic with a mean of 2 packets/sec. The packet lengths are exponentially distributed with a mean of 2000 bits. The system designer must choose between giving each session a dedicated 8 Kpbs piece of bandwidth (via TDM or FDM) or having all packets compete for a single 64 Kpbs shared channel. Which alternative gives a better response time?

For TDM or FDM, each 8-kbps channel operates as an independent queueing system with $\lambda = 2$ packets/sec and $\mu = 4$ packets/sec. We know for the M/M/1 queue, the total average time spent (T_{av}) in the queue is given by $T_{av} = \frac{1}{\mu - \lambda}$ (Little's result). Hence, $T_{av} = 500$ msec. Now, the analysis of a single 64-kbps system with, $\lambda = 16$ and $\mu = 32$, gives T_{av} is 66.7 msec. This makes sense intuitively also. Splitting up a single channel into k fixed size pieces makes the response time k times worse. The reason is that it frequently happens that several of the smaller channels are idle, while other ones are overloaded. The lost bandwidth can never be regained. Again we assume arrivals follow the CIPP distribution. The conclusion is same as above, but with a difference: as ρ increases, T_{av} for both the situation increases.

5.8 Discussions

In this chapter, we have derived the analytical expressions for stationary probabilities in a CIPP/M/1 queue. The method used here for the derivation of stationary probabilities is specific to the model considered. In particular, the method exploits the particular Markovian structure implicit in the model. We have verified the above theoretical results (corresponding to CIPP/M/1 queue) by simulations with respect to the conventionally used queueing performance measures.

We noted that our queueing theory results are consistent with the Loyne's results which state that, for the stability of the queue, it is the the ratio of arrival rate to service rate, that matters and *not* the correlation structure among the interarrival sequence. We observed that our *analytical* results corroborated with the results of Livny *et al* [80] and Patuwo *et al* [103] who *empirically* established that positive correlation in the arrival process leads to degradation in the queueing performance. We noted that, since CIPP/M/1 queue has closed form expressions for stationary probabilities, computationally it is much more efficient than the queues with conventional SRD models (say an MMPP, wherein even the moments of the distributions of interest have to be computed recursively).

We used the approach of Benes, to obtain the approximate expression for buffer occupancy distribution. Since, this (approximate) analytical expression for the buffer occupancy distribution posed a numerical problem for large values of ρ, we had to estimate the distribution of the CIPP, by suitable simulations. This leads to the estimation of density, which is computationally intensive. For this reason, we will not use the results of CIPP/D/1 queue for performance modeling purposes in later chapters. Nevertheless, we showed here that the simulated mean delay agrees to an extent with theoretical mean delay in a CIPP/D/1 queue.

Then, we considered the simulation based performance evaluation study of CIPP/M/1/K queue. We observed that the packet loss probability increases as ρ increases. Also this observation reflects in the other queueing measures considered also. Notably, in the case of the conditional packet loss probability, there is a slow decay of probability of loss of 'bunch' as 'bunch' size increases, when the arrivals are CIPP with high ρ. We observed that the buffer occupancy distribution for CIPP/M/1/K queue shows peaks at both the extremes (empty state and fully occupied state), for high ρ and at high utilizations. This was found to be consistent with results by earlier workers. Also, we observed that the buffer occupancy distribution is dependent *only* on the load and not on the individual values of arrival rate or service rate. Note that, this is the case with M/M/1/K queue. To conclude, the correlation in interarrivals seems to degrade the queueing performance in the *finite* capacity CIPP queue (CIPP/M/1/K) *also*. (Earlier we noted that, from the analytical expressions for the queueing performance measures of the CIPP/M/1 queue, the queueing performance degrades with the correlation parameter ρ).

We saw in Section 4.5.3 that, there exists an empirical relationship between the correlation parameter (ρ) and the estimated Hurst parameter (H) of the CIPP trace. Thus, the method given in this chapter, allows one to understand the mutual relationships among positive correlation, the high degree of self-similarity (by virtue of the above relationship) and degradation of the queueing performance. We finally saw a numerical example in which the buffer size required, differed by a factor of 245 when arrival process is highly correlated (CIPP with $\rho = 0.98$) instead of the conventional Poissonian assumption for the arrival process. (This illustrates the worst case effect of correlation on the buffer length design). The theory developed in this chapter, will be used in the subsequent chapters for application of the CIPP process for modeling the queueing situation as encountered by composite Ethernet traffic (Chapter 6) and for modeling the queueing situation as encountered by VBR video traffic (Chapter 7) in the output queue of a statistical multiplexer.

Chapter 6

Performance Modeling of Composite Traffic by CIPP Queues

In previous chapters, we observed that the network traffic exhibits self-similarity or long-range dependence (LRD) or fractal behavior. Apart from the statistical significance of such phenomena, this self-similar behavior (or LRD) has considerable impact on queueing performance [29]. Hence, it is imperative to come up with an analytical model for *self-similar* counting process[1] which should *also* be analytically tractable. In light of the above, in Chapter 4, we proposed the CIPP model and noted that, it did exhibit self-similarity to a range of time scales. In Chapter 5, we developed the queueing theory for the same. In the present chapter, we will apply those results, for studying the queueing situation as encountered by composite Ethernet traffic in a statistical multiplexer.

First, we use the queueing performance measures derived in Section 5.3 for the performance modeling of statistical multiplexer with composite traffic input and exponential service by CIPP/M/1 queue. We use the queueing performance measures, namely, packet loss probability and conditional packet loss probability for performance modeling of finite capacity statistical multiplexer[2] with composite traffic input and exponential service, using the CIPP/M/1/K queue. In addition, we compare the buffer occupancy distribution of the finite capacity queue with composite Ethernet traffic and the "best fit" CIPP/M/1/K queue with K arbitrary.

We refer back to the limitation of CIPP model discussed in Section 4.7 and define a new process (which is a modification of CIPP) to overcome this limitation. This new process, herein called Markov Modulated CIPP (MMCIPP) possesses correlation in interarrivals in addition to correlation introduced due to the modulating Markov chain. We show by various methods that this process is self-similar over a wider scale than that

[1] By self-similar counting process, we mean a counting process whose corresponding rate process $(u_T(i)/T)$ behaves in exactly the same way at all time scales (T).

[2] By "finite capacity statistical multiplexer", we mean, the buffer capacity of the output queue of a statistical multiplexer is finite. A mere "statistical multiplexer", would imply infinite buffer capacity. The case of finite capacity is more realistic for obvious reasons.

of MMPP. A rigorous mathematical treatment of MMCIPP is outside the scope of this thesis.

The following is the organization of this chapter. The performance modeling of statistical multiplexer with composite Ethernet traffic using CIPP/M/1 queue is undertaken in Section 6.2. Section 6.3 gives the performance modeling of finite capacity statistical multiplexer with composite Ethernet traffic using CIPP/M/1/K queue. Phase plot of Bellcore August (`pAug.TL`) dataset is compared with the phase plot for CIPP in Section 6.4. Section 6.5 describes a modification of CIPP, namely MMCIPP. In Section 6.5.2, we show that MMCIPP is more self-similar than CIPP or MMPP. Some discussions on the chapter are given in Section 6.6.

6.1 Simulation Methodology

Before discussing the performance modeling of statistical multiplexer with real-world Ethernet traffic by CIPP queues, we present here the simulation methodology for using the Bellcore Ethernet traffic traces for our modeling purposes.

For our experiments, we consider only the arrival instants of packets of the Ethernet *Bellcore* dataset [82] (and ignore the second column which gives the corresponding packet sizes). We note in passing that the multiplexer is characterized by a constant switching time while the transmission time[3] of the packet (over a transmission link of fixed bandwidth) being proportional to the length of the packet (in bits). Hence, it is reasonable to think of the statistical multiplexer with packet traffic as a queue with a traffic model (here CIPP) of suitable parameters and service times being deterministic (and proportional to length of the packet).

In this chapter, since we intend to use the theoretical results corresponding to CIPP/M/1 queue developed in Section 5.2, we set the service times as exponentially distributed random variables. Accordingly, the packet stream is then fed to an infinite buffer single server FCFS queue with exponential service. (In the present context, this amounts to the assumption of exponential distribution for ratio of packet length to the link capacity. This is reasonable if the time to serve a bit is sufficiently small and the mean packet size is sufficiently large). Then, the queueing measures mentioned in Section 3.4 are estimated. We thus get the experimental results corresponding to data. To get the *theoretical* performance results, we fit the CIPP to composite traffic trace data by

[3]Here, we assume that the switching time is negligible as compared to the transmission time of the packet.

autocorrelation matching and compute the queueing performance measures by using the expressions given in Section 5.3, corresponding to this "best" fit CIPP.

The simulation methodology for performance modeling with finite capacity CIPP queues is same as above except that the computation of theoretical performance measures for infinite capacity queues is replaced by the estimation of queueing performance measures (namely, packet loss probability and conditional packet loss probability) from simulations. We also study the buffer occupancy behavior of finite capacity CIPP queue. We repeat the experiment with various link utilization factors (LUF) (the ratio of arrival rate to service rate[4]) to get the plot of queueing performance measures versus LUF for both the finite and infinite capacity cases.

6.1.1 Error Quantification

In this chapter (and in next chapter), since we undertake the performance modeling study, we would like to quantify the error between the performance measures corresponding to the "best fit" CIPP/M/1 and that of the queue with real-world data. For this purpose, we use the percentage error, which is defined below. The percentage error $P(i)$ between the data $d(i)$ and theoretical curve $g(i)$ for $0 \leq i \leq N$ is given by,

$$P(i) = \frac{|d(i) - g(i)|}{|d(i)|} \times 100 \tag{6.1}$$

The maximum (P_{max}) and the minimum (P_{min}) percentage errors, are given by,

$$P_{min} = \overset{min}{i}\ P(i) \quad \text{occurring at} \quad I_{min} = \overset{argmin}{i}\ P(i) \tag{6.2}$$

and

$$P_{max} = \overset{max}{i}\ P(i) \quad \text{occurring at} \quad I_{max} = \overset{argmax}{i}\ P(i) \tag{6.3}$$

respectively.

6.2 Performance Modeling of Statistical Multiplexer using CIPP/M/1 Queue

There are two main reasons which justify the study of infinite capacity queueing systems (although all real systems have finite capacity): first, infinite models have been

[4]here, the service rate is equal to the band width of the output link. Note, here the arrival rate is fixed (and is set to the arrival rate of the given trace).

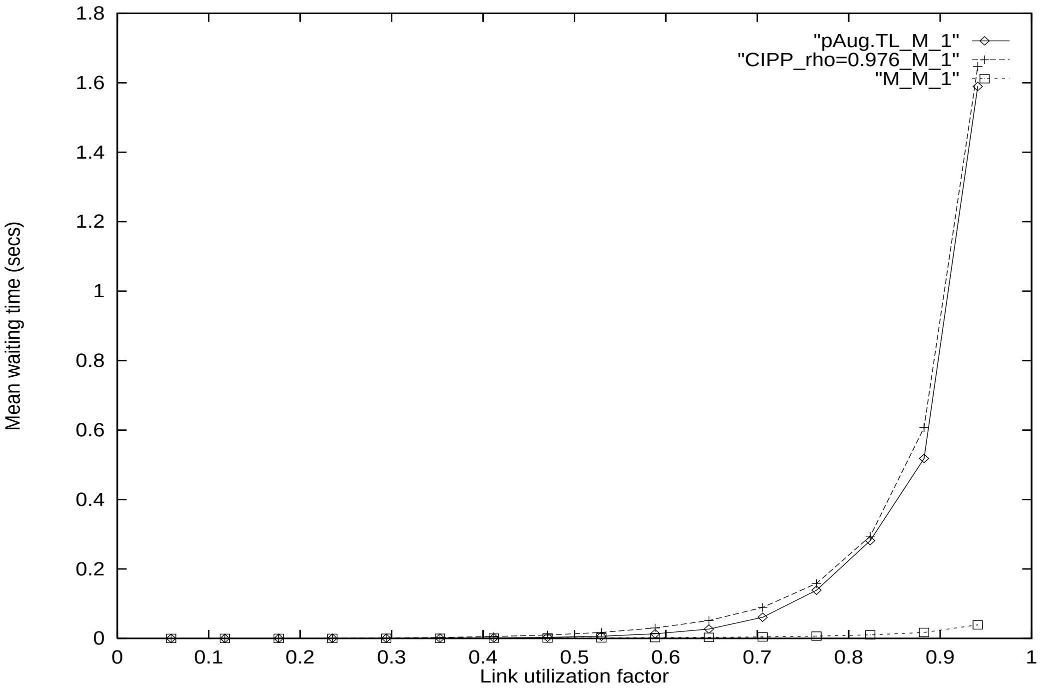
Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
`pAug.TL`	6.1	92.75	0.71	97.54	0.94	3.63	0.94	46.95	0.7	0.976
`pOct.TL`	6.1	94.35	0.76	98.24	0.94	4.32	0.88	38.44	0.76	0.98

Table 6.1: Percentage error between the mean waiting time of packets in queues with `pAug.TL` (Figure 6.1 (a)) input and the approximating CIPP/M/1 queue. The corresponding error for `pOct.TL` (Figure 6.1 (b)) is also given. In both the cases corresponding error for the approximating M/M/1 queue is shown for comparison.

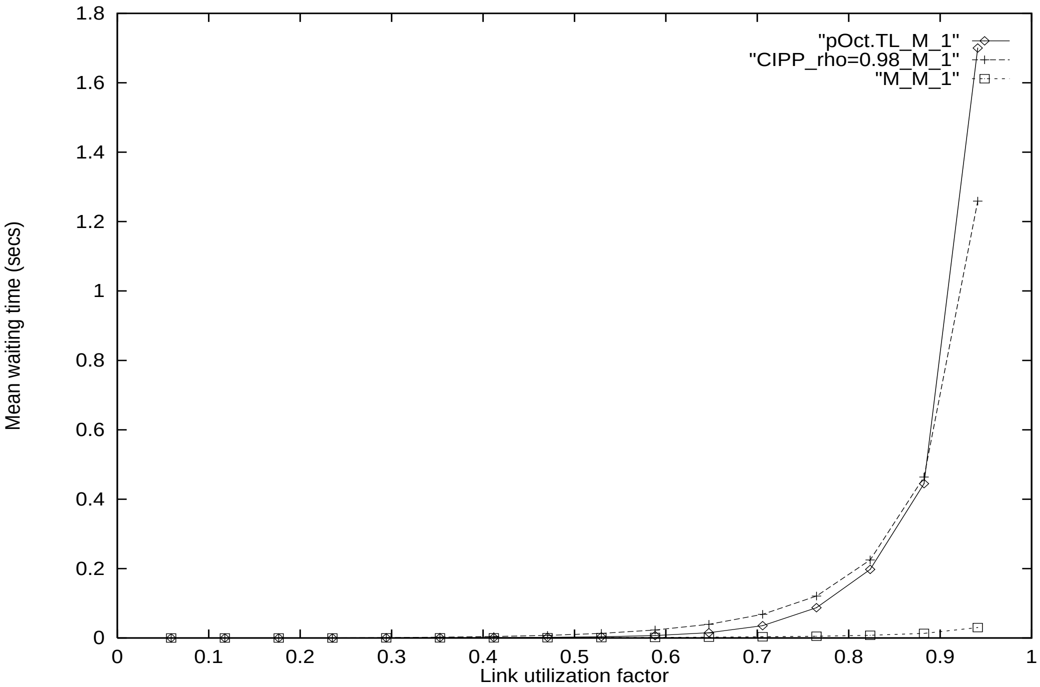
used in the analysis of asymptotic queue length behavior which is closely linked with effective bandwidth computations for call admission control and second, the overflow probability (in infinite capacity queue) is used usually as an approximation for packet loss probability in finite capacity queueing systems (when the finite capacity queue is analytically intractable). In this section, we consider the performance modeling of statistical multiplexer with the composite Ethernet traffic input, using CIPP/M/1 queue. We extensively use the analytical expressions for performance measures obtained in Section 5.3, for our *theoretical* performance modeling study. We consider the Bellcore Ethernet traffic datasets namely, `pAug.TL` and `pOct.TL`, for our simulation purpose.

6.2.1 Mean Waiting Time

We first consider the performance modeling using CIPP/M/1 queue, with respect to the performance measure, the mean waiting time of a packet. Figure 6.1 shows the mean waiting time performance for `pAug.TL` and `pOct.TL` and their corresponding (approximating) CIPP/M/1 and M/M/1 queues. The Table 6.1 gives the maximum and minimum percentage error between the theoretical performance measure corresponding to CIPP/M/1 queue and the estimated performance measure from simulations of the queue with the above Bellcore datasets. From table and the figure, it can be seen that CIPP/M/1 queue approximates well the queueing situation of statistical multiplexer with Ethernet Bellcore traffic input. Our *theoretical* performance modeling results establishes that the M/M/1 queue seriously under-estimates the performance of a queue with real-world Ethernet traffic as input. It has been observed by Paxson *et al* [32] that the Poisson process poorly models the Ethernet traffic. The under-estimation of performance measures, by M/M/1 queue is an another manifestation of the above observation.



(a)



(b)

Figure 6.1: Mean waiting time of packets in a queue driven by Bellcore dataset and the corresponding quantity in the approximating CIPP/M/1 and M/M/1 queues (a) `pAug.TL` (b) `pOct.TL`

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
`pAug.TL`	6.2 (a)	92.74	0.71	97.54	0.94	3.33	0.94	47.14	0.71	0.976
`pOct.TL`	6.2 (b)	94.36	0.76	98.24	0.94	4.23	0.88	38.23	0.76	0.98

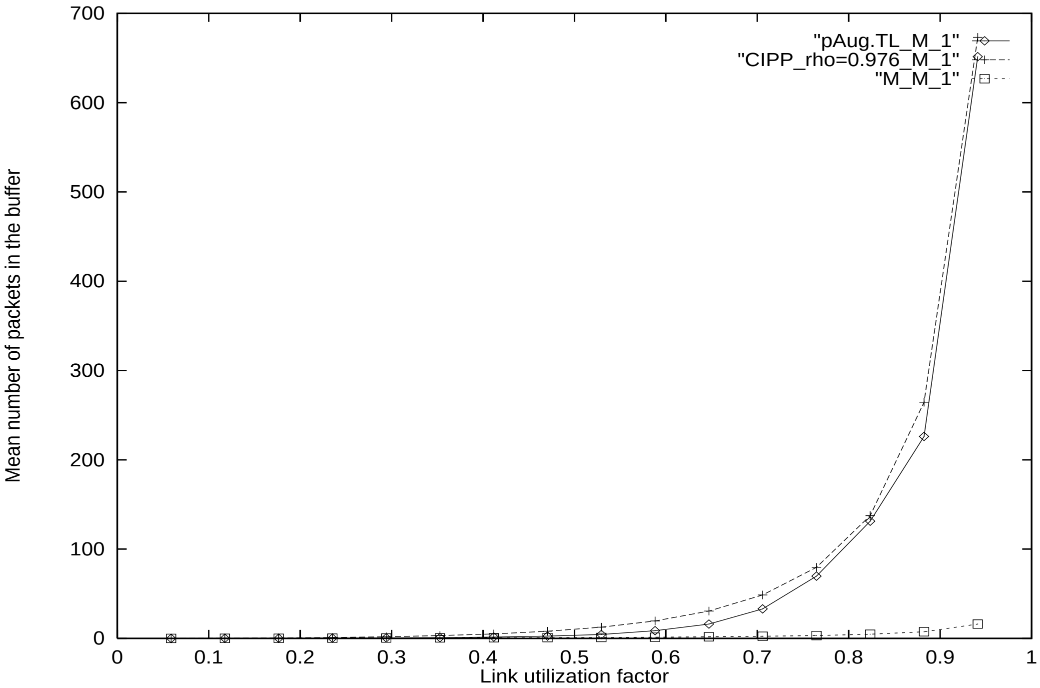
Table 6.2: Percentage error between the mean number of packets in the system with `pAug.TL` (Figure 6.2 (a)) input and the approximating CIPP/M/1 queue. The corresponding error for `pOct.TL` (Figure 6.2 (b)) is also given. In both the cases corresponding error for the approximating M/M/1 queue is shown for comparison.

6.2.2 Mean Number of Packets in the System

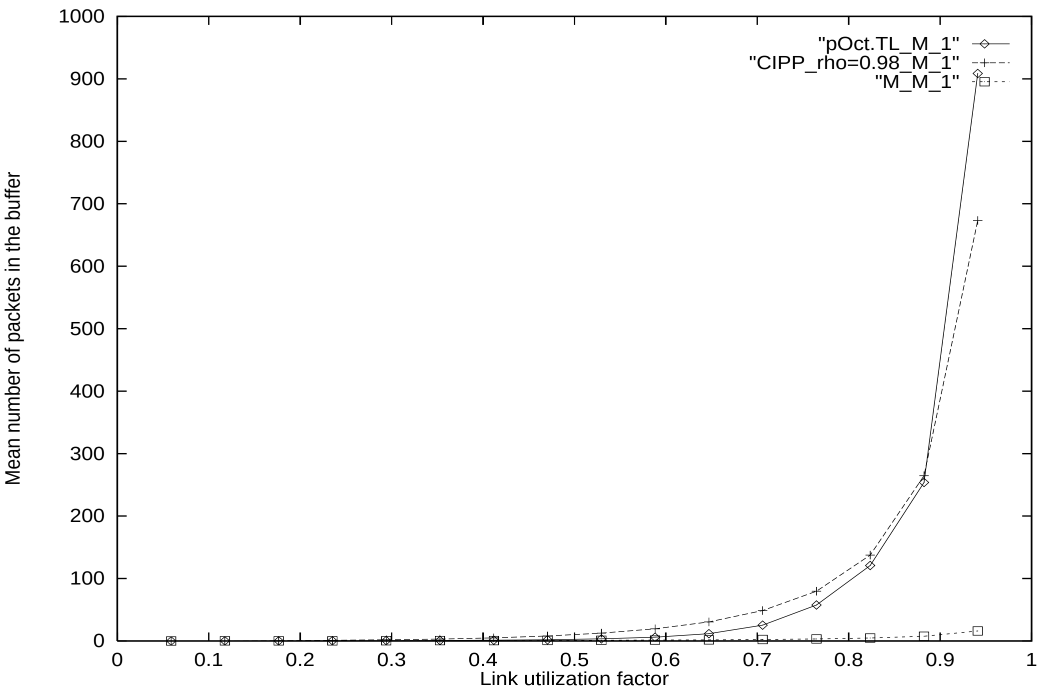
Here, we undertake the performance modeling with the mean number of packets in the system as a performance measure. For that purpose, we use the procedure given in Section 6.1. Figure 6.2 shows the mean number of packets for a queue with `pAug.TL` (`pOct.TL`) and (approximating) CIPP/M/1 queue. Table 6.2 shows the error between the performance measure of 'best fit' CIPP/M/1 queue and the performance measure corresponding to the queue with Bellcore traffic dataset as input. One can see from the figure and the table that the CIPP/M/1 queue models (both qualitatively and quantitatively) quite well, the queueing situation as encountered by the Ethernet traffic in a statistical multiplexer, with respect to the measure of mean number of packets. The corresponding results for M/M/1 queue are also presented. Here again, the our results establish (beyond the observation of Paxson *et al* [32] that the Poisson process models the broadband traffic poorly) that, the M/M/1 queue seriously under-estimates the performance measures for the queue with real-world composite traffic input.

6.2.3 Packet Delay Variation (pdv)

In this subsection, we repeat the above experiment for the measure of packet delay variation for the Bellcore datasets. Our results are shown in Figure 6.3 for both `pAug.TL` and `pOct.TL`. Table 6.3 gives the error statistics. It is seen from the figure and the table that CIPP/M/1 queue models reasonably well the queueing situation as encountered by the Ethernet traffic in a statistical multiplexer, with respect to the queueing performance measure of *pdv*. The results corresponding to M/M/1 queue for the measure of *pdv* exhibit a similar behavior as above and the results are readily interpretable.

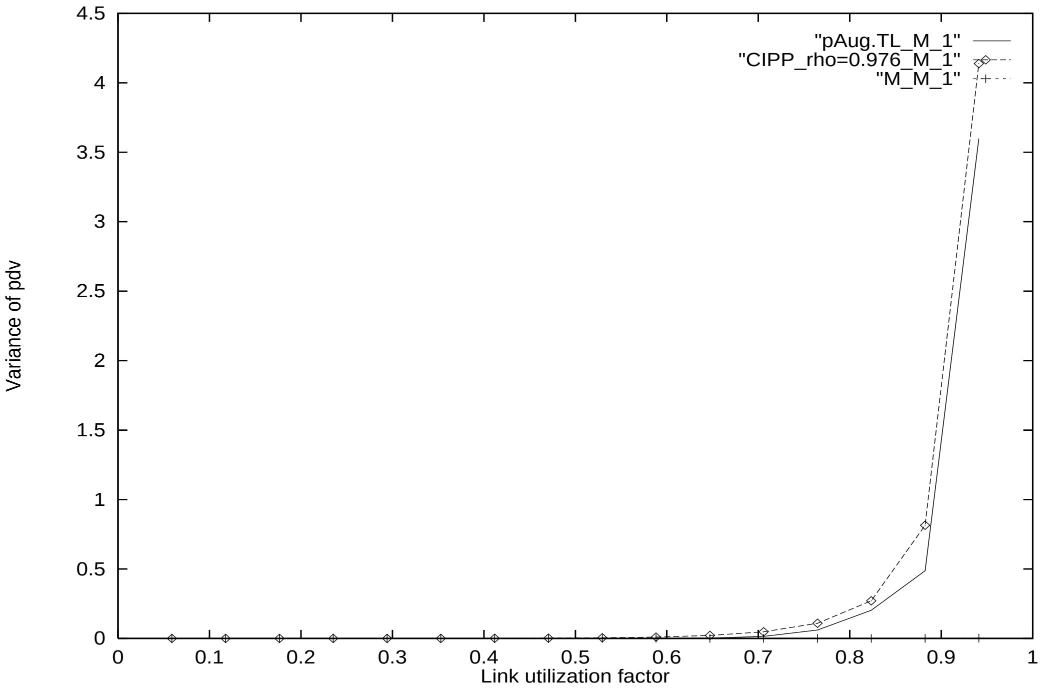


(a)

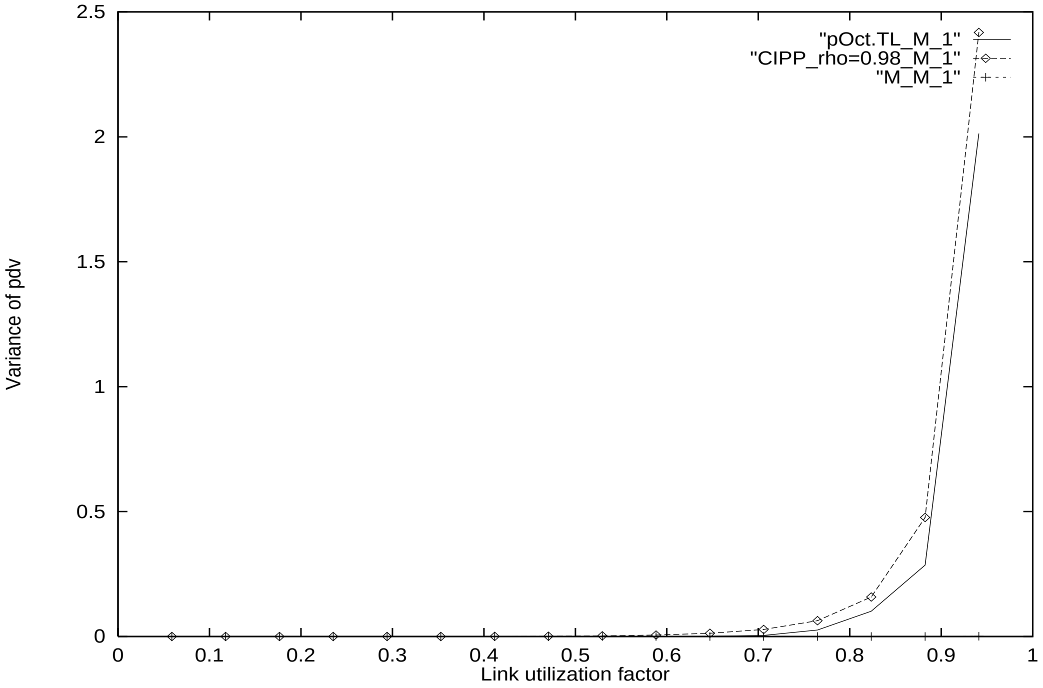


(b)

Figure 6.2: Mean number of packets in a queueing system driven by Bellcore dataset and the corresponding quantity in the approximating CIPP/M/1 and M/M/1 queues (a) `pAug.TL` (b) `pOct.TL`



(a)



(b)

Figure 6.3: *pdv* of packets in a queue driven by Bellcore dataset and the corresponding quantity in the approximating CIPP/M/1 and M/M/1 queues (a) `pAug.TL` (b) `pOct.TL`

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
`pAug.TL`	6.3 (a)	99.92	0.88	99.95	0.94	15.0	0.94	67.32	0.88	0.976
`pOct.TL`	6.3 (b)	60.0	0.72	93.42	0.94	20.1	0.94	66.7	0.88	0.98

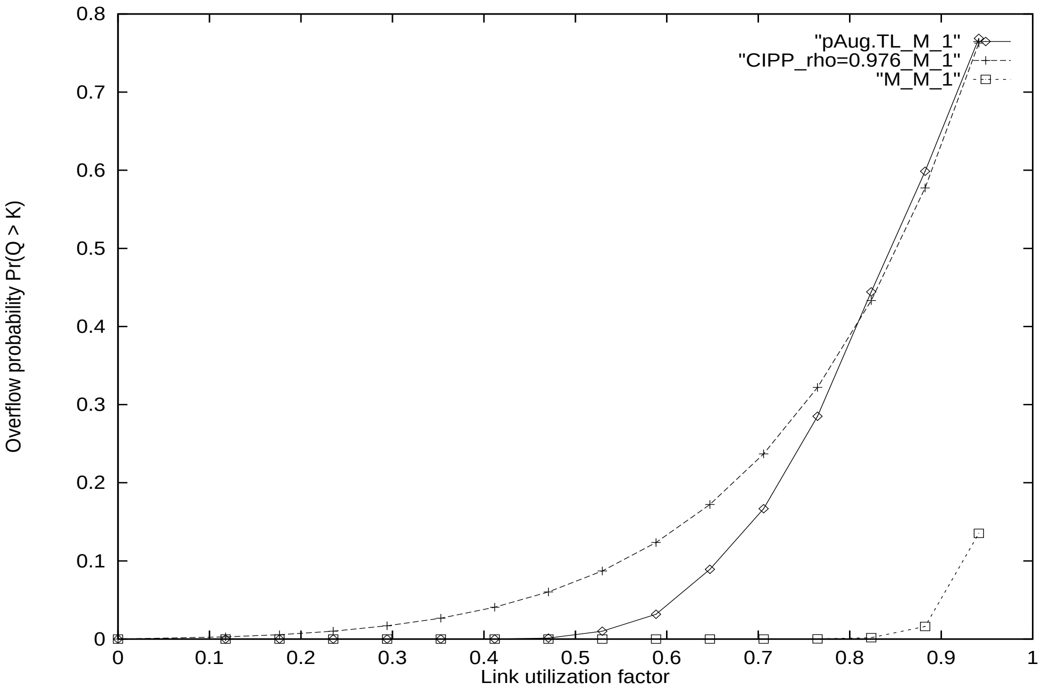
Table 6.3: Percentage error between the packet delay variation in queues with `pAug.TL` (Figure 6.3 (a)) input and the approximating CIPP/M/1 queue. The corresponding error for `pOct.TL` (Figure 6.3 (b)) is also given. In both the cases corresponding error for the approximating M/M/1 queue is shown for comparison.

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
`pAug.TL`	6.4 (a)	82.41	0.94	99.99	0.59	0.78	0.48	42.07	0.36	0.976
`pOct.TL`	6.4 (b)	84.05	0.94	100.0	0.53	7.76	0.65	33.79	0.76	0.98

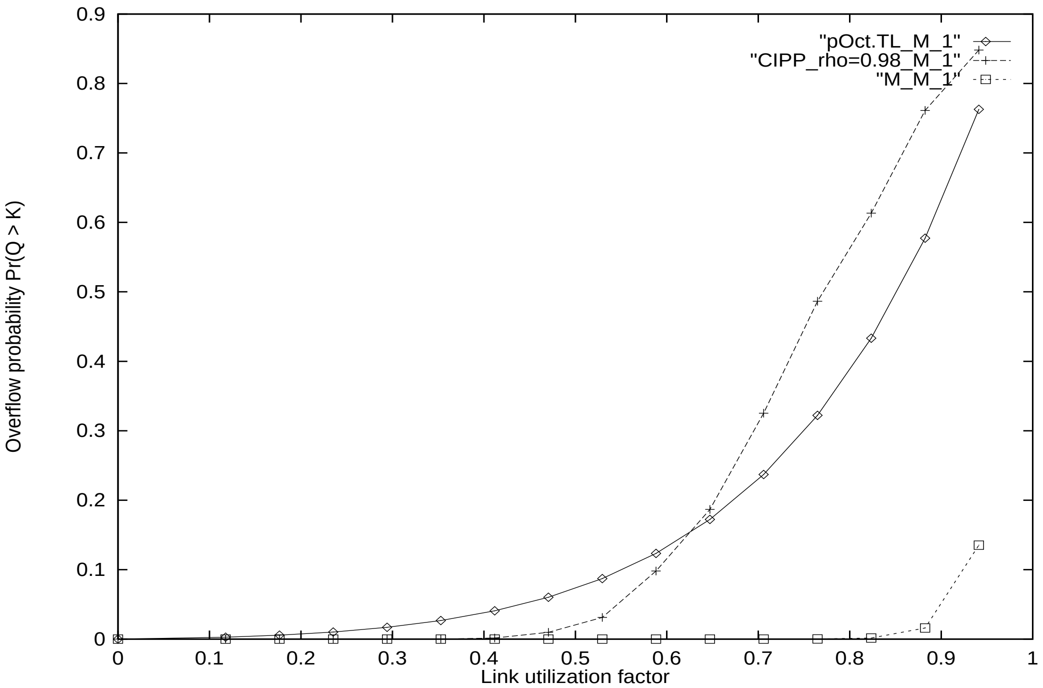
Table 6.4: Percentage error between the overflow probability in queues with `pAug.TL` (Figure 6.4 (a)) input and the approximating CIPP/M/1 queue. The corresponding error for `pOct.TL` (Figure 6.4 (b)) is also given. In both the cases corresponding error for the approximating M/M/1 queue is shown for comparison.

6.2.4 Overflow Probability

Here, we concentrate on the queueing performance measure of overflow probability for our performance modeling study. We consider a threshold of 32 packets. Figure 6.4 shows the overflow probability for `pAug.TL` and `pOct.TL`. Also shown are the respective overflow probabilities of their corresponding 'best fit' CIPP/M/1 and M/M/1 queues. The results for error statistics are listed in the Table 6.4. From the figure and table, note that the CIPP again approximates well the data. The results corresponding to M/M/1 queue follow the same behavior as above and are readily interpretable.



(a)



(b)

Figure 6.4: Overflow probability in a queue driven by Bellcore dataset and the corresponding quantity in the approximating CIPP/M/1 and M/M/1 queues (a) `pAug.TL` (b) `pOct.TL`. Here, the threshold buffer size is $K = 32$ packets

6.3 Performance Modeling of a Finite Capacity Statistical Multiplexer with CIPP/M/1/K Queue

In this section, we consider the performance modeling of statistical multiplexer (with *finite* capacity buffer) with composite Ethernet traffic input, using CIPP/M/1/K model. We use the packet loss probability and conditional packet loss probability as performance measures for our experiment (see Section 5.6). Finally, we compare the buffer occupancy distribution in CIPP/M/1/K queue with that in a composite Ethernet traffic queue. We consider the Bellcore Ethernet traffic datasets namely, `pAug.TL` and `pOct.TL`, for our simulation purpose.

6.3.1 Packet Loss Probability

We first consider the performance measure of packet loss probability for our performance modeling study. We consider a single server queue with buffer capacity of 32 packets and Bellcore Ethernet dataset as input and repeat the procedure as given in Section 6.1 for the finite capacity case. Repeating the experiment for various link utilization factors, we plot the packet loss probability in Figure 6.5. From this figure, one can see that the CIPP makes a good approximation to the performance of finite buffer statistical multiplexer with real-world Ethernet traffic. Result corresponding to M/M/1/K queue is seen to seriously underestimate the packet loss probability.

6.3.2 Conditional Packet Loss Probability (CPLP)

As noted in Section 5.6.3, the consecutive packet losses in composite traffic may arise in the following situations: first, there may be temporal overload situation due to the fact that the sum of current rate of active users exceeds the capacity of the system (may be due to the bursts at peak rate of some of the users) and second, due to the periodic nature of a highly correlated (component) arrival process. In the present context, the former situation is of interest to us. There are two causes for the rate to exceed the capacity of the system: first, there might be simultaneous bursts of guaranteed services (like CBR and VBR) and second, more number of available bit rate (ABR) connections accepted and they all burst at a time. This behavior of consecutive packet losses is best captured by CPLP.

CPLP is important in ATM networks in the context of rate based feedback mechanisms used in ABR services. The enhanced rate control mechanism based on explicit

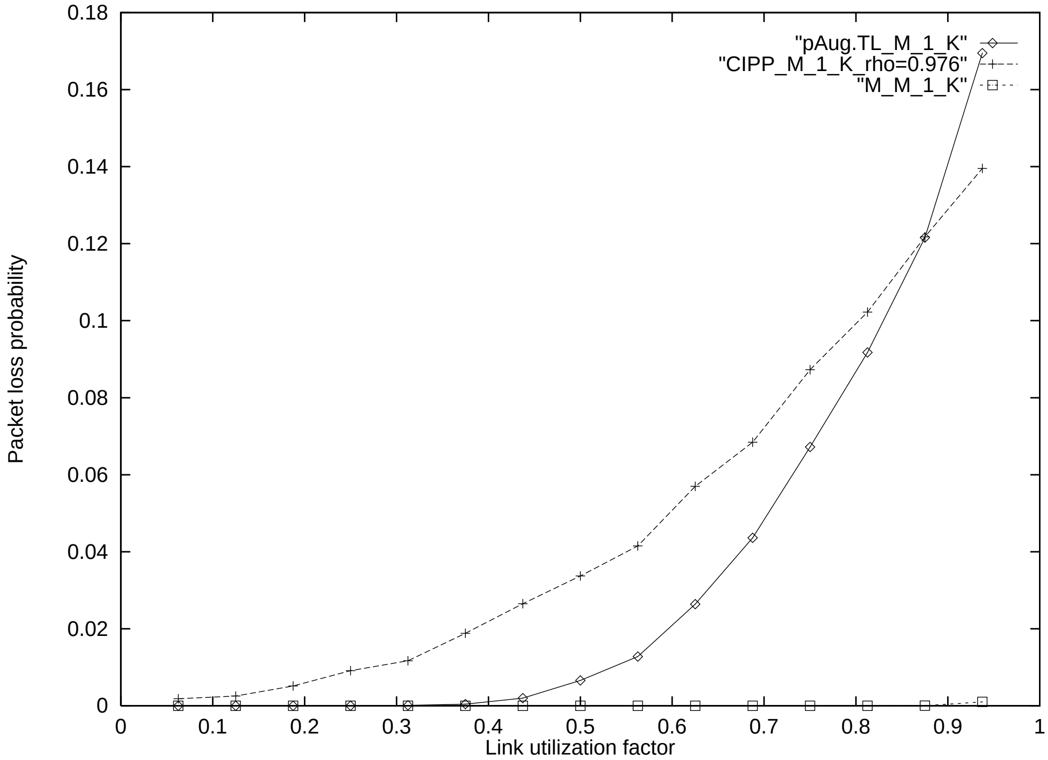


Figure 6.5: Packet loss probability in a queue with `pAug.TL` input and that of the approximating CIPP/M/1/K and M/M/1/K queues

rate feedback control allows the ABR source to adapt it's sending rate according to the availability of bandwidth resources [106]. The ABR sources reduce their sending rate, the moment their cells come across a congested node by these explicit rate feedback control mechanism. When the network experiences a sudden burst loss (consecutive cell loss) of cells in the composite traffic, all those ABR sources which experience losses, eventually reduce their rate to such an extent that the channel is under-utilized. Then gradually they recover resulting in the saw tooth behavior for their (available) channel utilization. This phenomena largely depends on the available band width and the statistical behavior of ABR sources. The above behavior is best captured by CPLP.

6.3.2.1 Simulation Results

Here, we present the performance modeling of a finite capacity statistical multiplexer with the Bellcore Ethernet traffic input using CIPP/M/1/K queue with respect to the measure of conditional packet loss probability. We have taken $K = 64$. We consider the high traffic condition (a load of 0.96) to study the worst case situation. Figure 6.6 shows the conditional packet loss probability for `pAug.TL` queue and that of the approximating

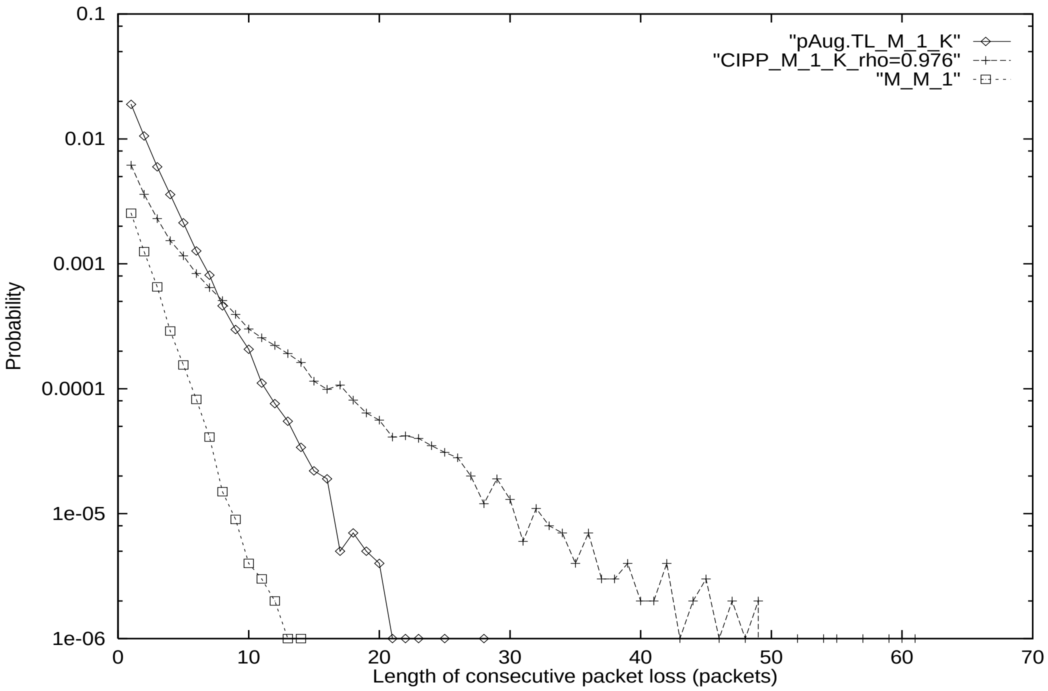
CIPP/M/1 queue. (The corresponding results for `pOct.TL` is also given). Also shown are the corresponding results for M/M/1 queue case. One can see from the figure that, the CPLP corresponding to M/M/1/K queue decays very fast, while the decay (with respect to k) is slow for the CIPP/M/1/K case (in particular, for high ρ). It may be noted that, the M/M/1 queue poorly models (probability decays much faster) the data. On the contrary, CIPP (for high ρ) decays slowly and approximates well the data, except for high k, where it overestimates the CPLP. The slow decay of conditional packet loss probability is the feature of highly correlated arrivals [17]. For CIPP with high ρ, this trend of 'slower' decay is highly pronounced.

6.3.3 Buffer Occupancy Distribution

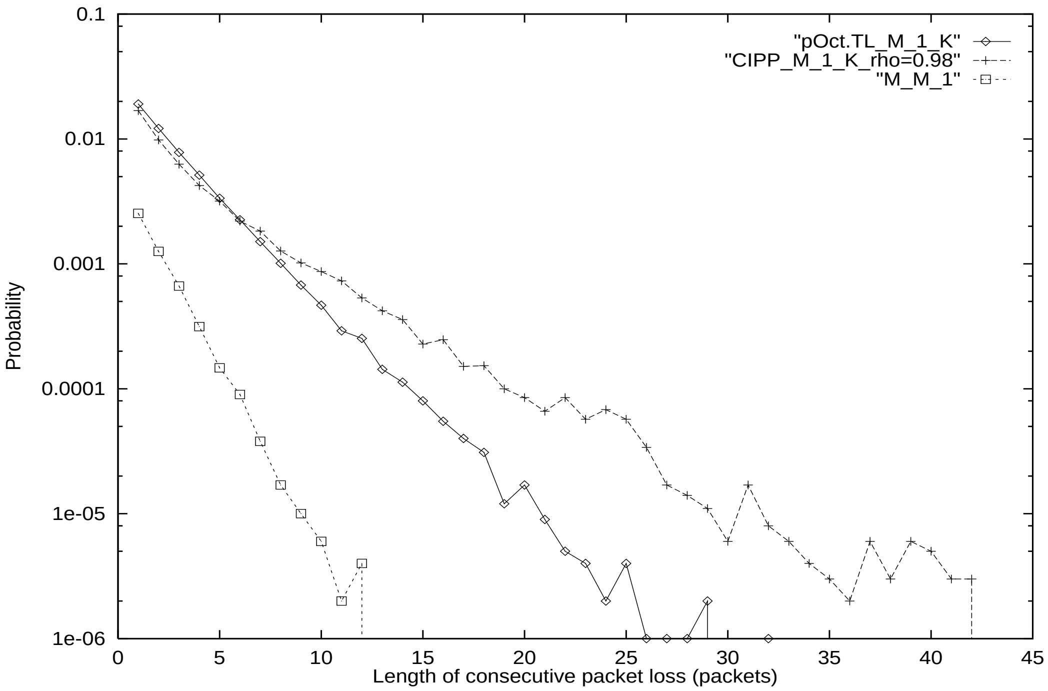
We noted in Section 5.6.1, that the buffer occupancy distribution of a CIPP queue with finite capacity for high ρ shows peaks at empty state and fully occupied state, particularly at high utilizations. It has been shown empirically, by earlier researchers that this is in general true for highly correlated arrivals. Here, we show that our simulation results also corroborate their empirical observation for Bellcore Ethernet traffic datasets. Moreover, we show that CIPP/M/1/K approximates reasonably well, the buffer occupancy behavior of a finite capacity queue with *real-world* Bellcore Ethernet traffic as input.

We consider queues with different buffer capacities namely, 64, 96 and 128 packets, all at high link utilization factors of 0.96 for our experiment. We compute the buffer occupancy distribution for Bellcore dataset, the approximating CIPP and Poisson process (CIPP with $\rho = 0$). In Figure 6.7, we have plotted our results for the Bellcore datasets, `pAug.TL` and `pOct.TL` side by side. In each of the plots, the corresponding results for the CIPP/M/1/K and M/M/1/K queues are also given.

In the first place, note that the distribution peaks at empty state and fully occupied state both for CIPP/M/1/K queue and the queue with Bellcore traffic datasets as we claimed. One can see from the figure that the CIPP/M/1/K queue captures well the behavior of a queue with *real-world* Ethernet traffic for the finite buffer case. Result corresponding to M/M/1/K queue, shown in all the plots, is an another manifestation of the observation made by Paxson *et al* [32].

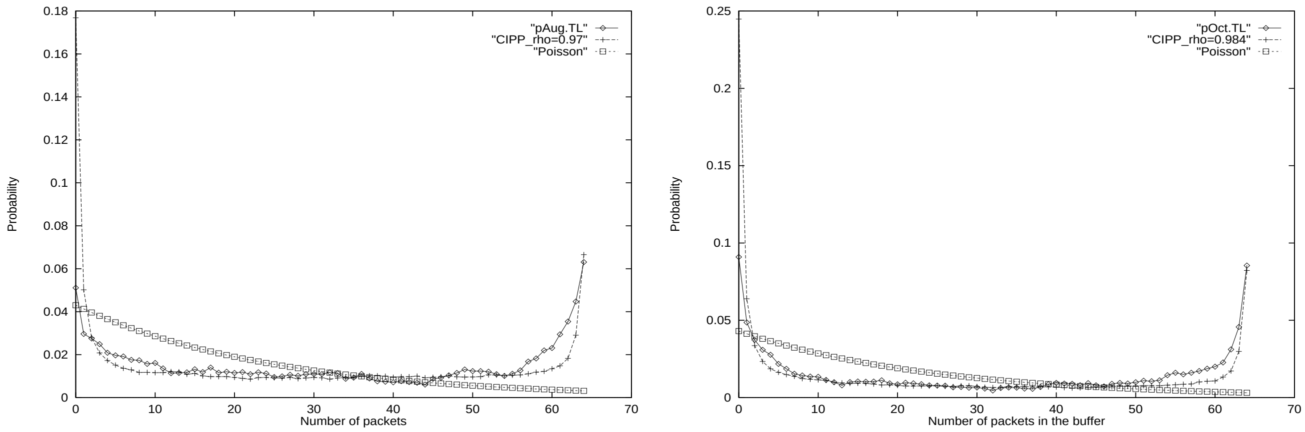


(a)

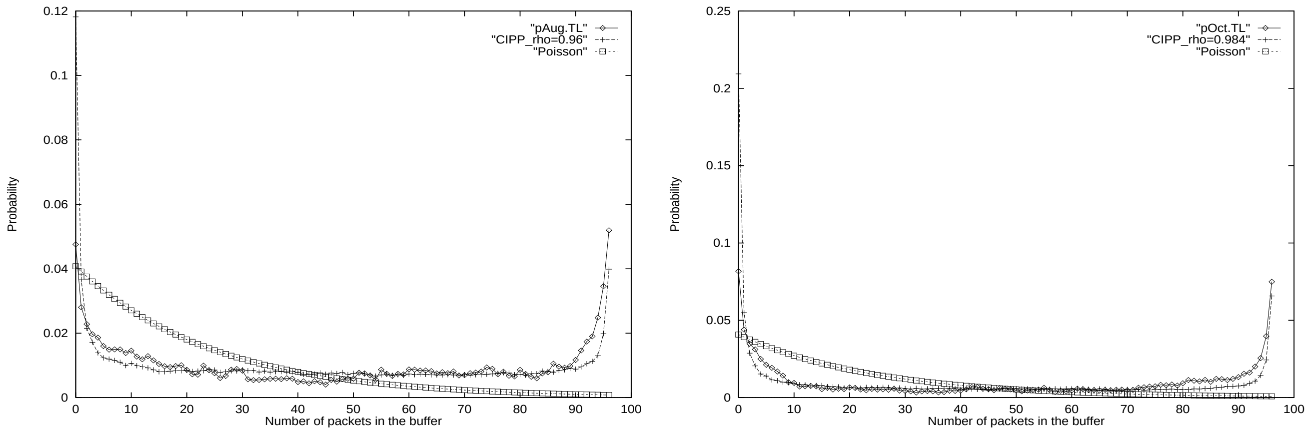


(b)

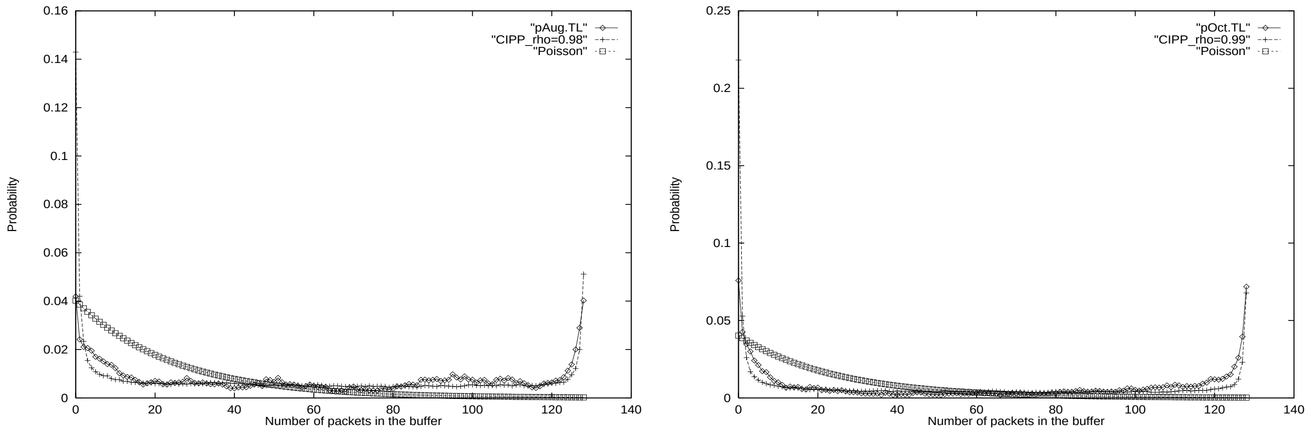
Figure 6.6: Conditional packet loss probability of a queue driven by Bellcore datasets and that of their approximating CIPP/M/1/K and M/M/1/K queues: (a) `pAug.TL` (b) `pOct.TL`



Buffer capacity 64 packets



Buffer capacity 96 packets



Buffer capacity 128 packets

(a) Bellcore August Dataset (`pAug.TL`) (b) Bellcore October Dataset (`pOct.TL`)

Figure 6.7: Buffer occupancy distribution in a finite capacity queue driven by Bellcore Ethernet datasets and that of their approximating CIPP/M/1/K and M/M/1/K queues: (a) `pAug.TL` (b) `pOct.TL`

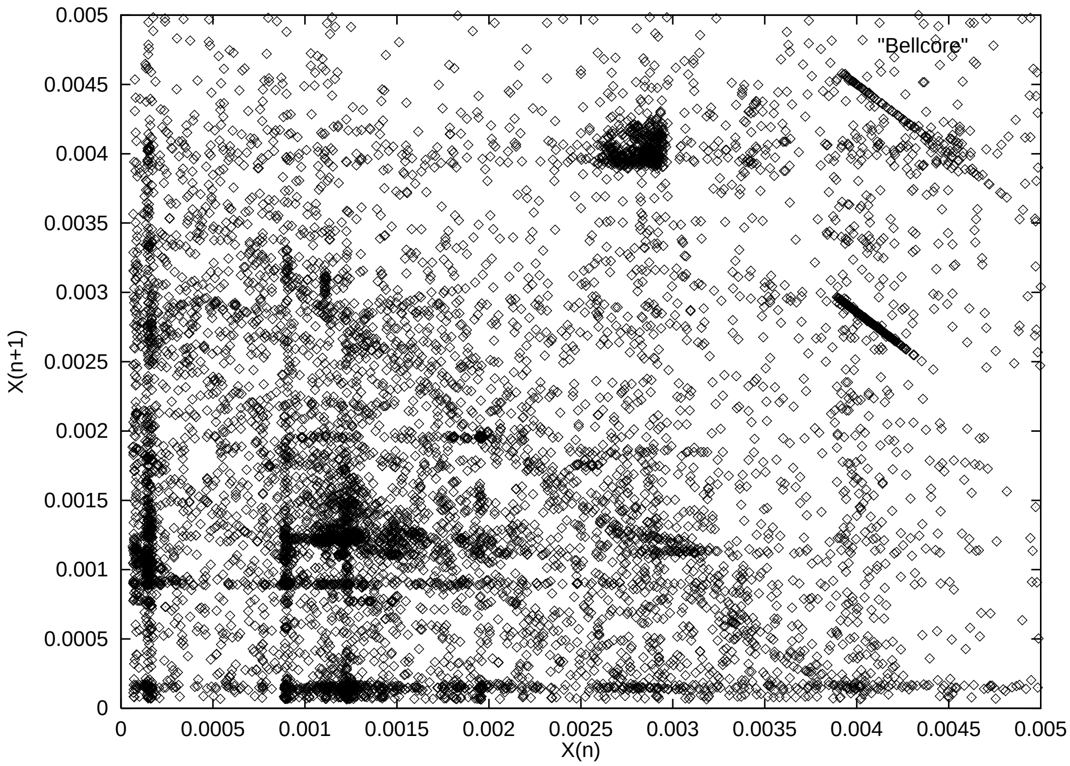


Figure 6.8: Phase plot of interarrival times X_{n+1} versus X_n (in seconds) for 100000 packets of `pAug.TL` traffic trace

6.4 Phase Plot of Bellcore Traffic Dataset

As introduced in Section 4.7, the phase plot is the plot of $(n+1)^{th}$ interarrival time versus n^{th} interarrival time. In this section, we intend to explore the stochastic behavior of *real-world* Ethernet traffic dataset. We start by giving the phase plot for Bellcore Ethernet dataset. Examining its phase plot would facilitate us to understand the stochastic behavior of a *real-world* traffic in the interarrival time level. This would provide us an intuitive idea to propose a model for characterizing the real-world data.

The phase plot of Bellcore August dataset is given in Figure 6.8. If the interarrival times were truly independent, Figure 6.8 would have mostly vertical and horizontal lines and more complicated structures like the diagonal lines would have been absent. Given, that there is some degree of quantization in the interarrival times, the horizontal lines correspond to the first packet in the burst, while the vertical lines correspond to the last packet in a burst. The diagonal lines are probably caused by a superposition of two or more sources which generate deterministically spaced interarrival times when they are 'ON'.

Going back to Section 4.7, we see that, for the phase plot for CIPP (see Figure 4.14), all the points are above the line with slope ρ, whereas, in the corresponding plot for Bellcore traffic dataset, points are distributed all over the region. This is a limitation of CIPP, despite the fact that CIPP model exhibits diagonal lines in its phase plot (which corresponds to positive correlations). In what follows, we will define a modification of CIPP process, which overcomes this limitation.

6.5 Markov Modulated CIPP (MMCIPP)

Given the discovery of self-similarity or LRD in measured packet traffic and its demonstrated implications on queueing performance, stochastic modeling of long-range phenomena incurs prime importance. Traditional approaches include mimicking LRD by short-range dependent models. (We noted earlier that SRD models possess correlations in the *rate process* or *counts sequence*). In earlier chapters, we have considered the CIPP process in which interarrivals are correlated. In addition to the interarrival correlation, if we introduce correlation by modulating the CIPP by a Markov chain, then we intuitively expect that the resulting process will exhibit self-similar behavior on a much wider scales. We call such a process as Markov Modulated CIPP (MMCIPP). Note that, it is a generalization of Markov Modulated Poisson Process (MMPP) considered in the literature. In this section, we show that MMCIPP, overcomes the limitation of CIPP as mentioned in Section 4.7. We also show that this process, exhibits self-similar behavior over much wider range of time scales than those exhibited by either MMPP or CIPP. A thorough mathematical characterization of MMCIPP and its queueing analysis are outside the scope of this thesis. Nevertheless, we give the following notation for our convenience.

MMCIPP is parameterized by the N-state continuous-time Markov chain with the infinitesimal generator Q_0, the rate matrix $\mathbf{\Lambda}$ and the correlation vector $\mathbf{\Gamma}$.

$$\mathbf{Q}_0 = \begin{bmatrix} -\sigma_1 & \sigma_{12} & \sigma_{13} & \dots & \sigma_N \\ \sigma_{21} & -\sigma_2 & \sigma_{23} & \dots & \sigma_N \\ \sigma_{31} & \sigma_{32} & -\sigma_3 & \dots & \sigma_{3N} \\ \vdots & & & \ddots & \vdots \\ \sigma_{N1} & & & \dots & \sigma_N \end{bmatrix} \tag{6.4}$$

$$\sigma_i = \sum_{j=1, j \neq i}^{N} \sigma_{ij} \tag{6.5}$$

$$\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \dots, \lambda_N) \tag{6.6}$$

$$\mathbf{\Gamma} = (\rho_1, \rho_2, \dots, \rho_N)^T \tag{6.7}$$

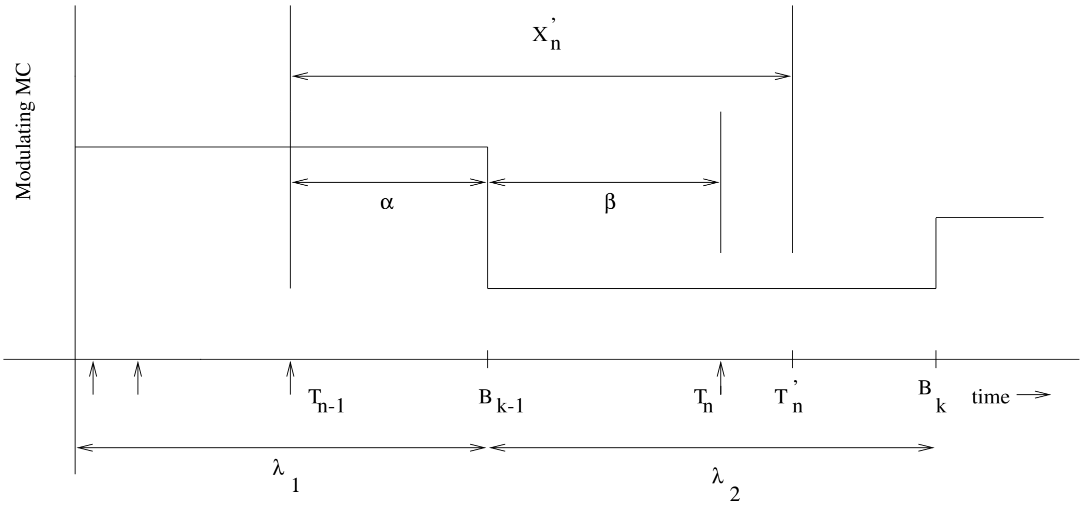


Figure 6.9: Illustration of correlation mapping during state changes in MMCIPP, by which dependency is "passed on" to the first arrival in the next sojourn time of the modulating Markov chain.

6.5.1 Correlational Structure of MMCIPP

In MMPP, within a sojourn time (holding time) of the modulating Markov chain, the arrivals are Poisson and these arrivals are independent of the past state, except that the (Poisson) rate follows the Markov chain. The dependency in counts (in disjoint intervals) arises *only* due to the (state of the) modulating Markov chain. On the contrary, in MMCIPP, the correlation in counts arises due to two factors: (i) correlation in interarrivals and (ii) correlation introduced due to the modulating Markov chain. Note that in MMCIPP, within a sojourn time of the modulating Markov chain, the arrivals are CIPP. By definition, we assume that during the state change of the modulating Markov chain, the first arrival in the next sojourn interval depends on the last arrival in the last sojourn interval as given below (see Figure 6.9). Had there been no state change, the interarrival time between $(n-1)$th and nth arrival would have been X_n' (and nth arrival instant would have been T_n'). Since there is a change in the mean arrival rate from λ_1 to λ_2 at B_{K-1} with $\lambda_1 > \lambda_2$, the interarrival X_n should be $\alpha + \beta$ where, $\beta = (\lambda_2/\lambda_1)(X'_n - \alpha)$.

An MMPP can be a renewal process under certain conditions [107]. For example, interrupted Poisson process (IPP) (which is a special case of MMPP) is a renewal process [108]. On the contrary, in MMCIPP, within a sojourn interval of the modulating Markov chain, the arrivals are CIPP. Hence, for non-zero $\mathbf{\Gamma}$, MMCIPP can not be a renewal process.

6.5.2 Self-similar Property of MMCIPP

In this subsection, we explore the self-similar property of MMCIPP. As a test for self-similarity, we use the following methods.

1. Various methods of estimating Hurst parameters, namely, time domain methods like, variance-time curve and wavelet based methods like wavelet estimator proposed by Abry *et al* [46].
2. Study of a queue with the shuffled versions of synthetic trace data generated by MMCIPP as done in [29].
3. Phase plot of a synthetic trace of MMCIPP [29].
4. Study of the superposition of independent streams of MMCIPP.

6.5.2.1 Hurst Parameter Estimation of MMCIPP

In order to investigate the self-similar nature of the MMCIPP, we generate synthetic traffic by MMCIPP for a given set of parameters and estimate the Hurst parameter. Here we use both time domain and wavelet domain methods for estimation.

R/S Analysis: First, we use the R/S analysis or *rescaled adjusted range plot* discussed in detail in Section 2.4.1. For our experiments we consider 3 - state MMCIPP with $\rho_i = 0.9 \quad i = 1, 2, 3$. We select the parameters of MMCIPP such that the mean arrival rate is 7.98 arrivals/sec. Figure 6.10 shows the R/S plot for 3-state MMCIPP. One can see from figure, that the best fit line has a slope of around 0.8, giving the Hurst parameter.

Variance-Aggregation Level Plot: Here, we compute the variance of the trace at various scales as given in Section 2.4.1. Results are shown in Figure 6.11. Note that the curve corresponding to MMCIPP decays slowly and its slope is distinctly different from the reference curve of slope -1. (Note that, values of the estimate $\hat{\beta}$ of the asymptotic slope, gives the Hurst parameter estimate, $\hat{H}$ through the relation, $\hat{H} = 1 - \hat{\beta}/2$.) The slope is estimated to be 0.52 which leads to $\hat{H} = 0.74$.

Variance Based Estimation of H Parameter: This method allows one to see to what range of scales, a given dataset is self-similar. The plot is obtained by computing the point estimates of Hurst parameter, versus sample size assuming that the data set is a realization of fractional Brownian motion versus sample size. The point estimates are obtained by the method given in Section 2.4.1. Note that, if the process is self-similar, the

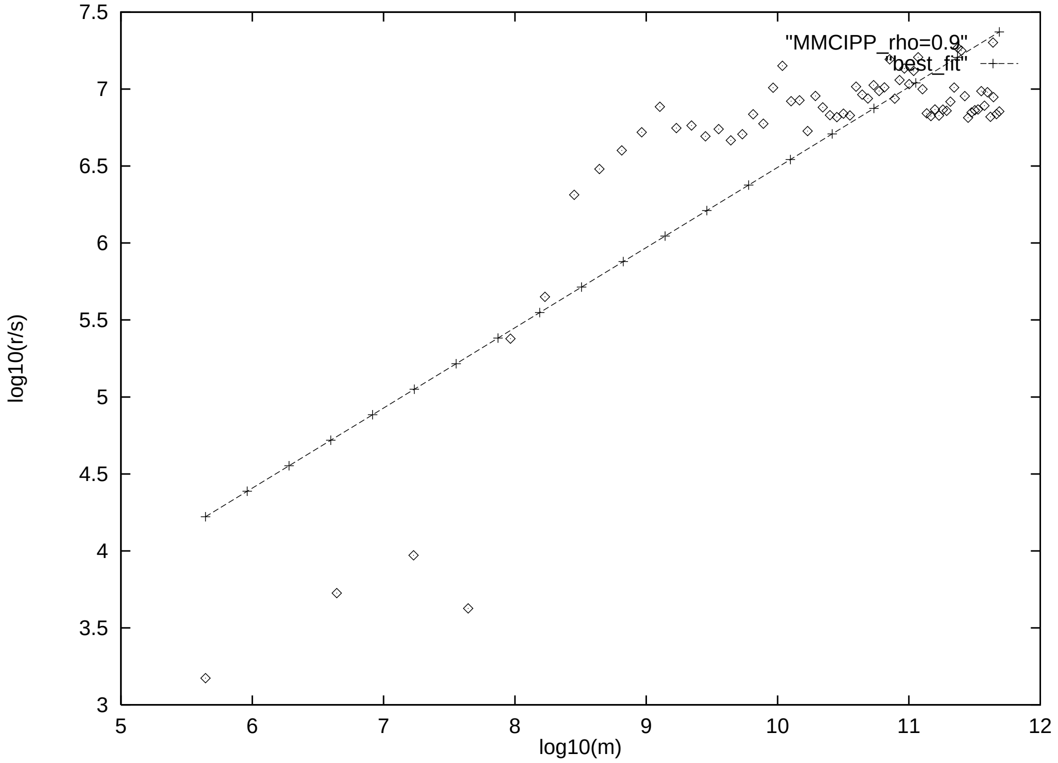


Figure 6.10: Graphical proof of LRD in MMCIPP: R/S analysis

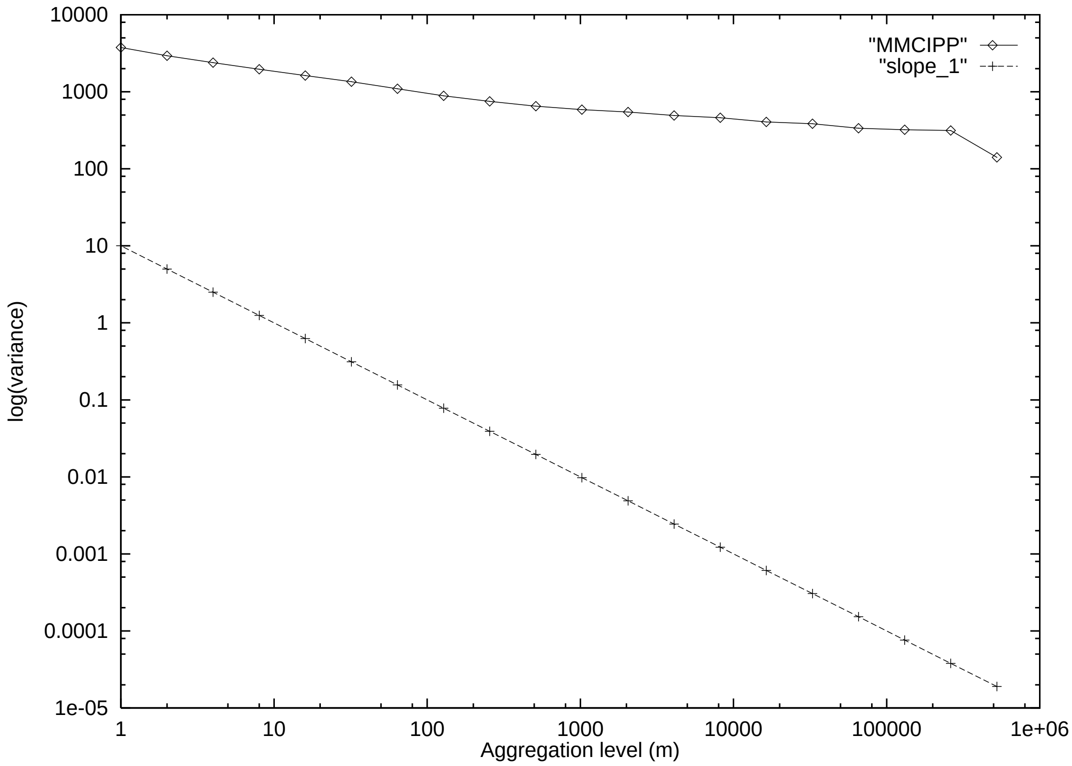


Figure 6.11: Graphical proof of LRD in MMCIPP: Variance-time plot

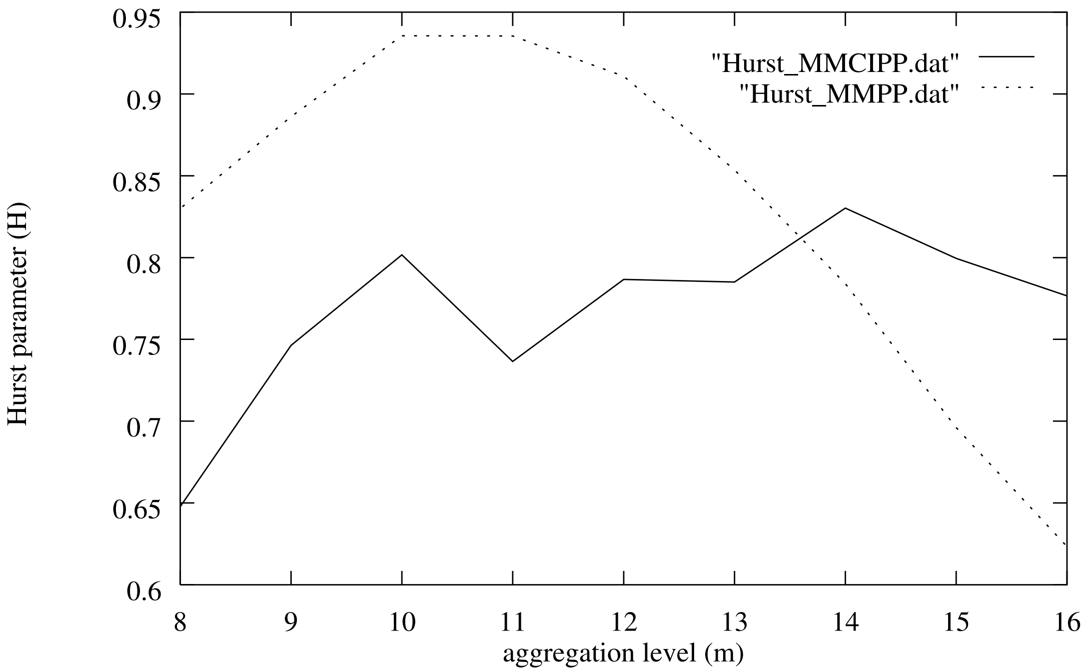


Figure 6.12: Graphical proof of LRD in MMCIPP: Variance-time

$\hat{H}$ will be approximately same at all scales (for which variance estimator has reasonable accuracy.)

Our results for MMCIPP and MMPP for the same mean arrival rate are shown in Figure 6.12. It is seen from the figure that, the $\hat{H}$ estimate corresponding to MMCIPP is almost constant for various scales over the range. For this set of parameters of MMCIPP, the estimate is approximately, 0.75. In contrast, the curve corresponding to MMPP, reaches a peak and decays immediately.

Variance of Wavelet Coefficients versus Wavelet Scale Plot: We finally establish the self-similar nature of MMCIPP using wavelet analysis. For that purpose we use the result due to Wornell [45], for the estimation of Hurst parameter of any given dataset. (See Section 2.3 for the Wornell's result). We plot the log(variance of wavelet coefficients) versus the log2(scale). Our results are shown in Figure 6.13. It is seen from the figure that the curve corresponding to MMCIPP is linear. Note that Wornell's result stipulates that for a process to be self-similar, the above plot should be linear. Observe that the curve corresponding to MMPP is non-linear.

6.5.2.2 Queues with Shuffled Traces

In [29], the authors show that the LRD (or self-similar nature) in the input traffic is responsible for its *worst* queueing performance, by conducting various experiments with

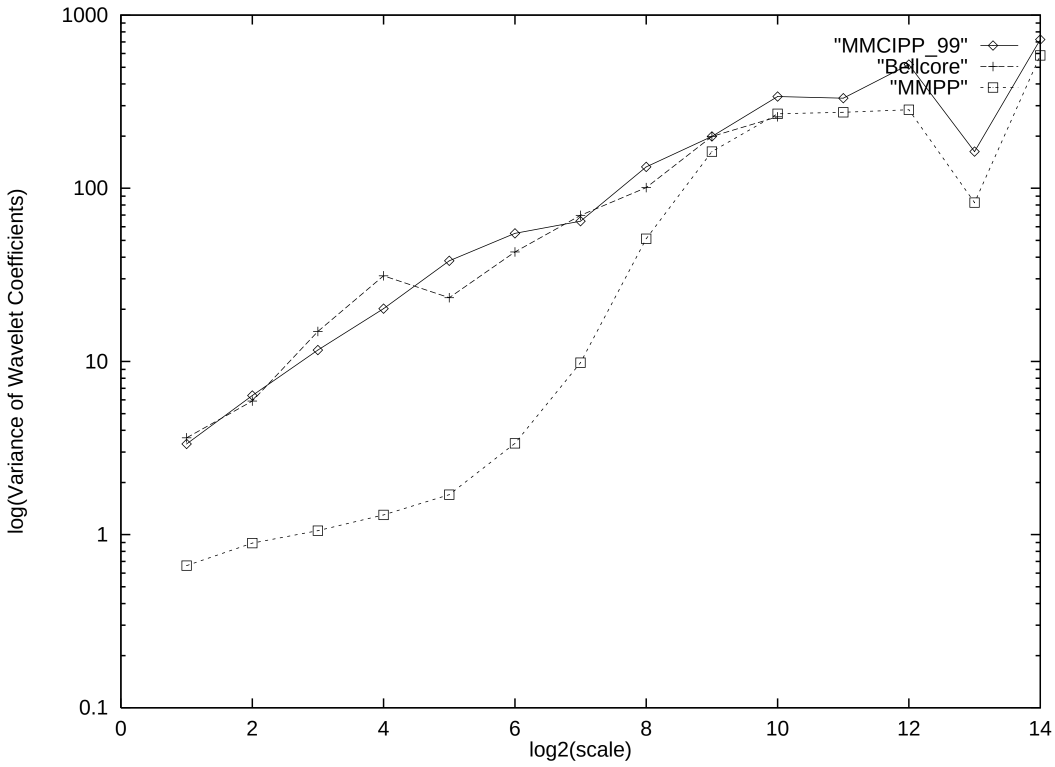


Figure 6.13: Graphical proof of LRD in MMCIPP: log(variance of wavelet coefficients) versus log(scale) plot

shuffled versions of the original *real-world* Ethernet trace. In particular, they consider shuffling, that makes the packet interarrival times independent, retain the short-range correlations, or exhibit the same long-range dependency as in the original trace. Finally, they show that the performance of the queue with original traffic is matched by the performance of the queue with the shuffled version of the original traffic which retains the LRD feature. Here, we use their methodology just as an another method to demonstrate the self-similar nature of MMCIPP.

Performance of a Queue with Shuffled Traces of MMCIPP: We divide the Ethernet data (sequence of arrival instants) into blocks of size m; with N interarrival times (hence there are N/m such blocks). We then perform an "external shuffle", i.e., the order of the blocks is shuffled. When $m = 1$, we have shuffling which results in retaining the marginal distribution of interarrival times. To get the shuffling which retains the one-step correlations, we pick the next interarrival time X_{n+1} from the possible set of interarrival times (as indicated by the phase plot) for a given interarrival X_n. This shuffling preserves the marginal distribution and one step correlations. When $m > 1$ in the external shuffling, we have shuffling which retains the short-range correlations. In

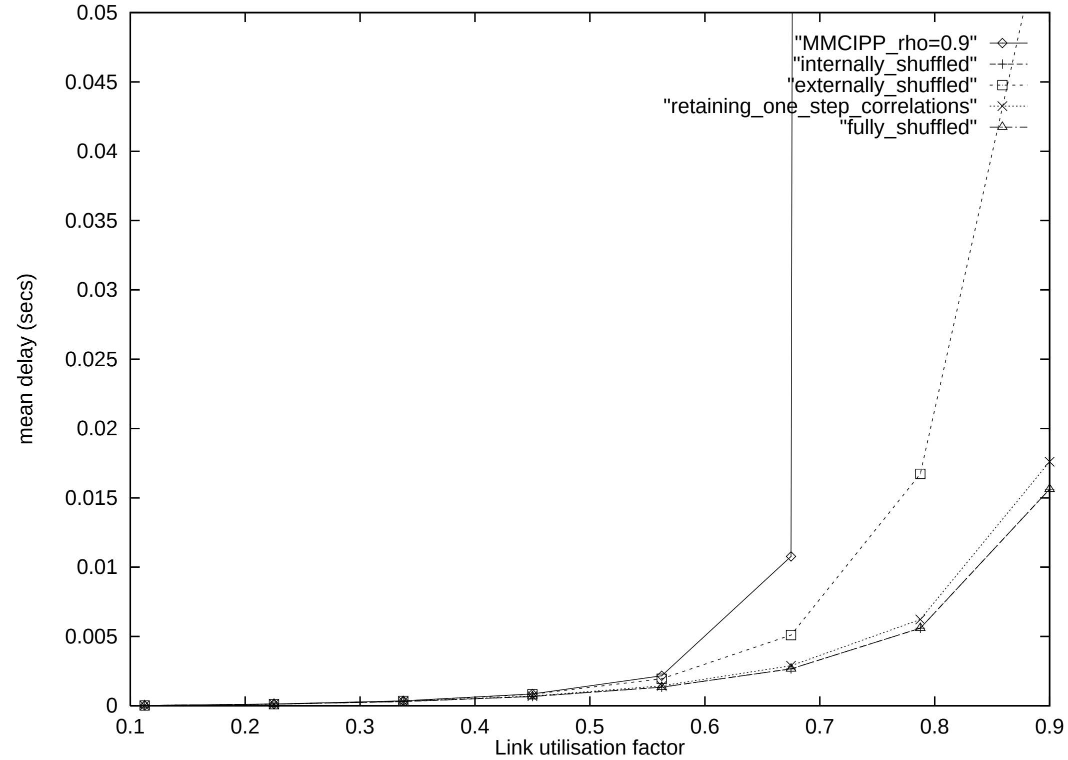


Figure 6.14: Proof of LRD in MMCIPP: performance of queues with shuffled versions of MMCIPP trace

"internal shuffling", the order of the blocks are unchanged, but the sequence within each block is randomized. For $m > 1$ in the internal shuffling retains LRD.

Figure 6.14 shows the performance of a queue with various random shuffled versions of the synthetic MMCIPP trace. The lowermost curve corresponds to shuffling (of block size $m = 1$) which retains the marginal distributions. The solid line (uppermost curve) corresponds to internal shuffling of block size $m > 1$ (which retains the LRD feature). The curves in between from lower most to the uppermost solid curve correspond to one step correlations and short-range correlations respectively.

The solid curve distinctly differs from the curves which correspond to the shuffled versions, which retain the short-range, one step or independence in interarrivals. Moreover, observe that the curves corresponding to shuffled version retaining LRD and that of the original trace merge. This supports our claim that MMCIPP exhibits LRD.

6.5.2.3 Phase Plot of MMCIPP

In this subsection, we show that, the phase plot of MMCIPP exhibits similar features (namely, vertical, horizontal and diagonal lines), as exhibited by the phase plot of Bellcore

Ethernet trace. We noted that in Section 6.4 that with some degree of quantization in the interarrival times, points which constitute the horizontal lines in a phase plot correspond to the first packet in a burst, while the vertical lines correspond to the last packet in a burst. The diagonal lines represent a high correlation (nearly periodic) during a burst. More precisely, the diagonal lines with positive slope correspond to positive correlation in interarrivals. In the following, we give intuitive reasons why the MMCIPP can very well capture the burst arrivals. The arrivals within a burst are nearly periodic (high correlation). Outside the "burst regime", the arrivals are highly random (low correlation). As an illustration, here we attempt to model an arrival process (with bursty arrivals) by 3-state MMCIPP, to see the intuitive implications. We model the "burst regime" to a state of MMCIPP (call this as state 1) where arrivals are CIPP with high ρ say $\rho_1 = 0.98$. For state 2, we consider $\rho_2 = 0.6$. For state 3, we choose the $\rho_3 = 0.0$ (which corresponds to the Poisson process). Now, whenever the state of the modulating Markov chain changes, either from or into the state 1, we have either vertical or horizontal lines respectively in the phase plot. The arrivals during state 1, account for the diagonal line of positive slope equal to ρ_1 of state 1 and the interarrivals corresponding to arrivals during this state are above this line. The arrivals during state 2, accounts for the diagonal line of slope equal to ρ_2 of state 2, All interarrivals corresponding to the arrivals during this state are above the line of slope ρ_1. Since, state 3 is characterized by Poisson arrivals, we have points in the phase plot all over the domain. Clearly, MMPP can't exhibit the diagonal lines since arrivals are Poisson at whatever the state be. Conclusively, we observe that the phase plot of MMCIPP is markedly different from that of MMPP (See Figures 6.15 and 6.16).

6.5.2.4 Superposition of Independent Streams

It is well known in the literature that, the superposition of a large number of SRD models will *smooth* out the resultant traffic. But, in the case of LRD models, the superposed process will retain it's LRD nature.

Superposition Process of Independent MMCIPP Streams: In this subsection, we study the self-similar property of superposed process resulting from merging of two independent MMCIPP traffic streams. We estimate the Hurst parameter of superposed process and of the component processes. We consider two independent MMCIPPs with arbitrary Γ with same Λ. We estimate Hurst parameter of the component processes and also estimate the Hurst parameter of the merged process. In Figure 6.17, we give our results for the estimated Hurst parameter of the superposed process versus the estimated Hurst parameters of the component processes. One can see from the figure, that the

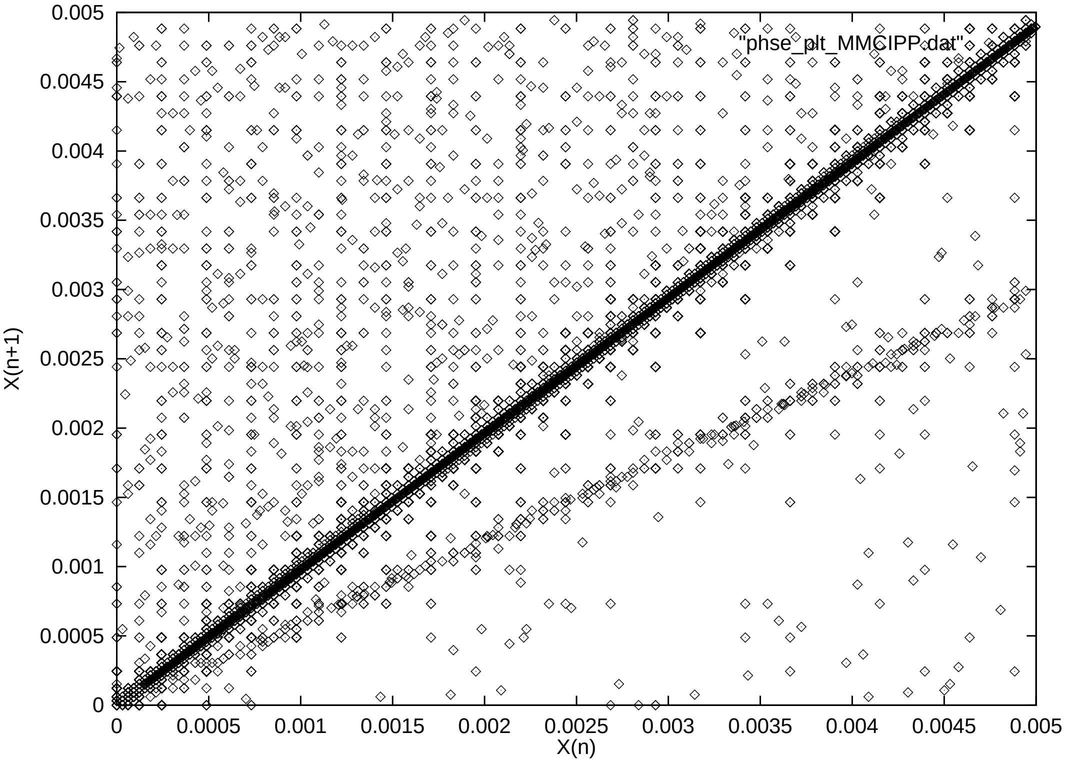


Figure 6.15: Phase plot of MMCIPP traffic trace

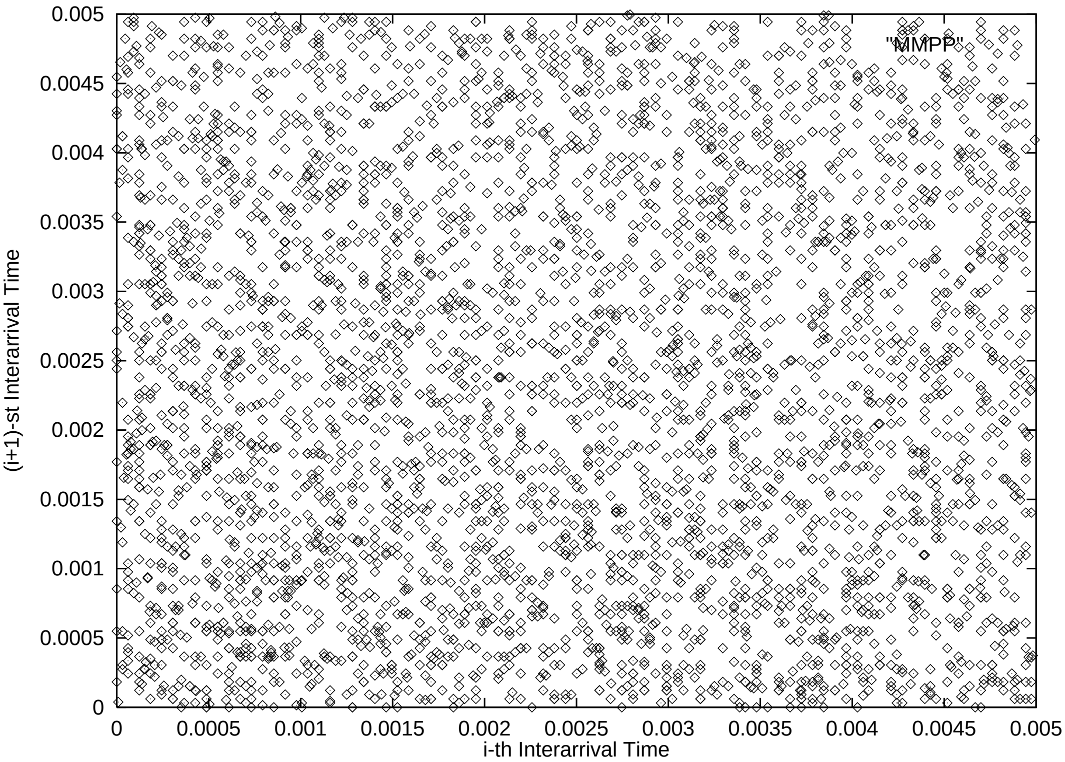


Figure 6.16: Phase plot of MMPP traffic trace

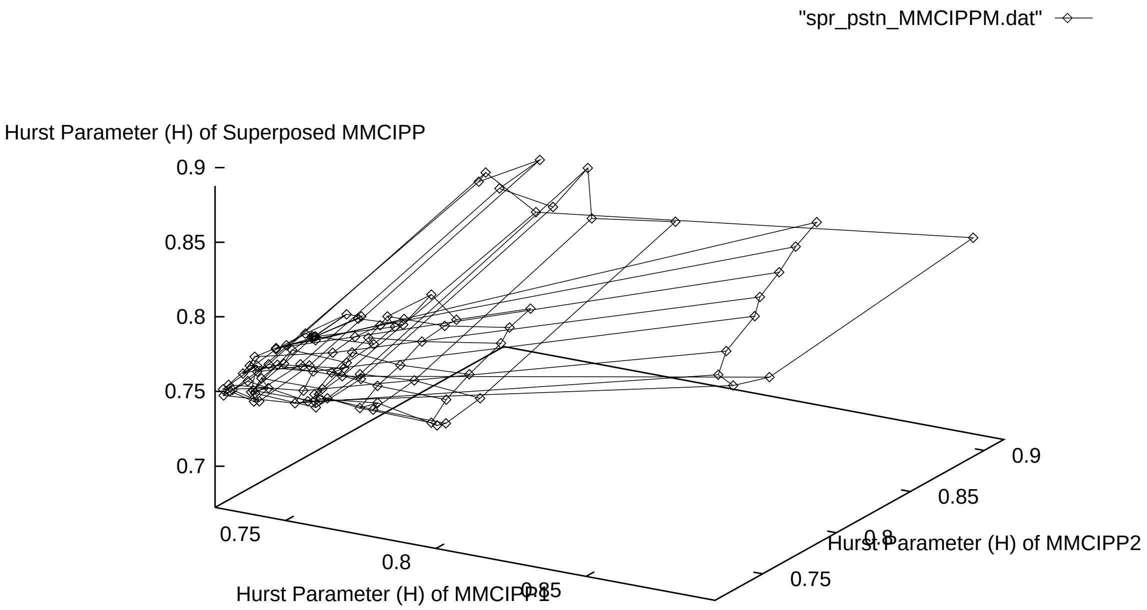


Figure 6.17: Superposition of two independent MMCIPP processes

Hurst parameter H, of the superposed MMCIPP lies somewhere in between the Hurst parameter H of the component processes.

6.5.3 Autocorrelation of MMCIPP Processes

Next, we estimated autocorrelation function of a synthetic MMCIPP trace at various scales from reasonably large number of realizations. Figure 6.18 shows autocorrelation at various scales. One can see, as T (block size) increases, the autocorrelation function decays faster.

6.6 Discussions

As mentioned in Section 3.2, after validating a model for the real-world data, one has to evaluate the performance of the model in the real-world situation. While the first part had been carried out in Chapter 4, the second part is undertaken in this chapter. We have applied the queueing theory results of CIPP/M/1 queue for performance modeling of the statistical multiplexer with composite Ethernet traffic and with exponential service times. Note that the modeling is on the arrival process component of the queue.

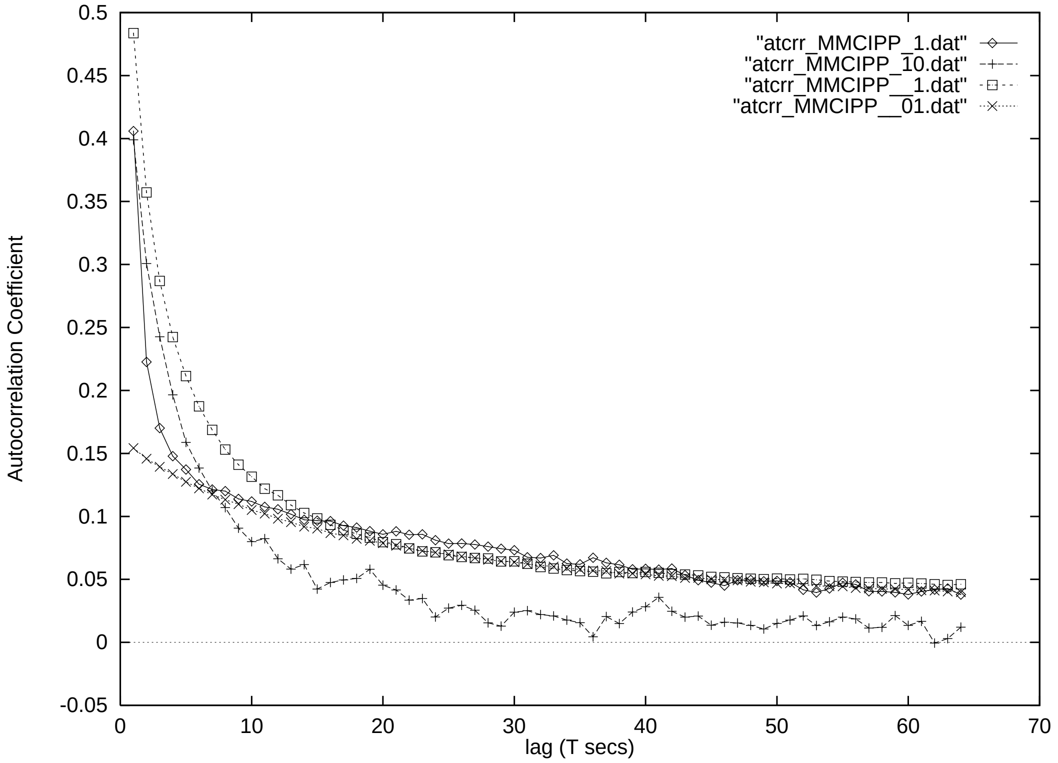


Figure 6.18: Autocorrelation of MMCIPP at various scales

The experimental results showed that the *theoretical* performance of CIPP/M/1 queues matches quite well with the performance of the queue with real-world traffic input. Also observe that in our performance modeling study, we not only consider the first order moments, but also the overflow probability and the higher order moment like *pdv*. The serious underestimation of performance measures of the *real-world* Ethernet traffic queue by M/M/1/K queue is an another manifestation of the observation made by Paxson *et al* [32]. Owing to the fact that, all real systems have finite storage, we also considered the performance modeling of statistical multiplexer with composite broadband traffic input using CIPP/M/1/K queue. The queueing performance measures considered are packet loss probability and conditional packet loss probability. The experimental results show that the behavior of finite capacity CIPP queue also matches closely with that of the real-world Ethernet traffic queue. We noted that the buffer occupancy behavior in M/M/1/K queue is distinctly different from the corresponding CIPP/M/1/K queue and the *real-world* traffic queue.

We considered a modification of CIPP, namely, the Markov Modulated CIPP (MM-CIPP). The motivation for defining this process is to overcome the limitation of CIPP mentioned in Section 4.7. MMCIPP has following three attractive features: (i) a model which considers *both* correlation in interarrivals and correlation introduced due to modulating Markov chain, (ii) the phase plot of MMCIPP captures all the features in the phase plot of Bellcore Ethernet traffic and (iii) is capable of capturing the burst arrivals. We showed that, the MMCIPP indeed exhibits self-similar behavior on a much wider range of scales than that of MMPP, by various methods. Until now, we discussed about the performance modeling of statistical multiplexer with Ethernet traffic which is basically composite in nature. In the next chapter, we concentrate on the performance modeling of statistical multiplexer with VBR traffic using CIPP queues.

Chapter 7

Performance Modeling of Statistical Multiplexer with VBR Traffic Input by CIPP Queues

Digital video communications (like video phone, video conferencing, television distribution etc.,) are the major class of services provided by broadband networks. The multiplexing gain of a statistical multiplexer is restricted by quality of service (QoS) guarantees which are basically bounds on queueing performance measures introduced in Section 3.4. Accordingly, the performance modeling of variable bit rate (VBR) traffic receives much attention in light of modeling and analysis of traffic generated by video communications in broadband networks. In this context, there are two issues, which are worth mentioning here: first, to come up with a process, which models well the real-world data (statistical modeling part) and second, to evaluate how closely the behavior of a queue with *this* model as input, approximates the queueing behavior of a queue with real-world traffic data as input (performance modeling part). While we have discussed the first part in Chapter 4, we will concentrate on the performance modeling of a statistical multiplexer with VBR video traffic input, using CIPP queues in this chapter.

In the literature on VBR modeling, it has been emphasized that, the correlation in the frame size trace of MPEG encoded bit streams is enough to capture the correlation structure. Two classes of traffic models have been investigated for modeling the bit rate of a VBR video source, namely, short-range dependent (SRD) models [16, 18, 62, 109] and long-range dependent (LRD) (or self-similar) models [26, 27]. The SRD models, broadly fall into two categories: single source models [16, 18, 62, 110] and multiplexed source models [16, 109]. For low motion video telephony and medium-motion video conferencing, we have the auto-regressive (AR) model [16, 62], the discrete auto-regressive (DAR(1)) model [18], and the discrete state continuous time Markov model [16, 62]. For full motion MPEG video source, a three class AR model with time-varying coefficients has been developed in [110]. Xu *et al* proposed a first order gamma auto-regressive

(GAR(1)) model, for VBR traffic. In [39], a model based on arrival rate histograms of VBR traffic has been proposed and a quasi-static approximation made for buffer occupancy distributions and packet loss probability. Most of the models proposed for VBR traffic are analytically *intractable* [18, 39, 109, 110].

The above models are intended to capture the correlation in the counts sequence (packets per frame) either by an ARMA model as in [18] or by the modulating Markov chain as in SRD models. In models based on self-similarity or LRD, correlations are in rate process.

In a counting process in which the interarrivals forming first order Markov sequence, the corresponding counts sequence ($u_T(i)$, in the same notation as used in Chapter 4) may not form a first order Markov chain; rather it may have complicated correlation structure. This results in the large scale correlations in the corresponding counts sequence. (This is in sharp contrast with the conventional models in which the counts sequence forms the first order Markov chain. Moreover, they lack analytical tractability). This is desirable for VBR modeling as VBR traffic shown to exhibit LRD or self-similarity. Thus, we contend here that, models having correlations in the interarrivals is more appropriate in the context of broadband teletraffic modeling, than models having correlations in the counts sequence. Moreover, we note that LRD models lack analytical tractability, except that bounds are obtained for some queueing measures [28, 49]. Those SRD models which are analytically tractable [16, 62] predict the queueing performance measures over optimistically.

In this chapter, we concentrate on the performance modeling of statistical multiplexer with VBR video traffic using CIPP/M/1 queue. For that purpose, we use the theory developed in Chapter 5. We consider exponential services here, to make use of the analytical results given in Section 5.2. In order to take into consideration, the finite capacity of the output buffer of a real-world statistical multiplexer, we characterize the behavior of a statistical multiplexer with real-world MPEG-1 VBR video traffic input using the CIPP/M/1/K queue. Since, CIPP/M/1/K queue is not analytically tractable, we resort to simulation based performance modeling study. We consider the fluid-flow approximation of CIPP queues for modeling the statistical multiplexer with VBR traffic. We compare the performance of CIPP fluid-flow queue with the conventional models addressed in literature for performance modeling of VBR video traffic.

This chapter is organized as follows. In Section 7.2, we consider the performance modeling of statistical multiplexer with VBR video traffic as input using CIPP/M/1 queue. In Section 7.3, we explore the performance modeling of realistic finite buffered statistical multiplexer with VBR traffic input using CIPP/M/1/K queue. We present the packet loss simulation results in Section 7.4, wherein we use the *fluid flow* approach for our simulations. This method is used to compare the performance modeling capabilities of CIPP with the Markov chain (MC) model in modeling the statistical multiplexer with VBR video traffic input. Some discussions are given in Section 7.5.

7.1 Simulation Methodology

In this section, we discuss the simulation methodology for performance modeling and related techniques which will be used in subsequent sections to handle the *real-world* MPEG-1 VBR video traffic data for our simulation purposes.

There is a problem (discussed in detail in next section) in simulating the several multiplexed video sources with same frame periods. In literature, smoothing is used to get around this problem. But, conventional smoothing gives only fixed size packets. Here, since we wish to use the queueing theory results developed in Section 5.2 for our performance modeling study, the service times are to be exponentially distributed. This calls for a method of smoothing which gives the variable sized packets (owing to the deterministic service nature of a broadband switch). In light of the above, we propose a variation of the deterministic smoothing which gives geometrically distributed packet sizes. This is explained in next section.

The smoothed packet stream is then fed to an infinite buffer single server queue with exponential service (here, exponential service times are approximated by the ratio of geometrical packet lengths (which we generated earlier while smoothing) to the link capacity). Then, the queueing measures mentioned in Section 5.3 are *estimated* from simulations. We thus get the experimental results corresponding to data. To get the *theoretical* performance results, we fit the CIPP to video trace data by autocorrelation matching and compute the queueing performance measures (by using the expressions given in Section 5.3), corresponding to this “best” fit CIPP.

Since, the finite capacity CIPP queue is analytically intractable, we rely on simulations for the estimation of its queueing performance measures namely, packet loss probability and conditional packet loss probability. Nevertheless, the simulation methodology for performance modeling using the CIPP/M/1/K queue is similar to the above. We repeat the experiment with various link utilization factors to get the plot of queueing performance measures versus link utilization factor for both the finite and infinite capacity cases.

7.1.1 Smoothing of VBR Video Traces

As mentioned earlier, there is a problem in simulating the several multiplexed video sources with same frame periods. If the multiplexed video sources are synchronized, they all begin their burst at the same instant for every frame, overloading the multiplexer's buffer. However, if the traffic is smoothed (for simulations) the sensitivity to synchronization is eliminated. Mathematically, smoothing gives an equivalent arrival process (equivalent in a statistical sense) specified by interarrival sequence $\{T_i\}_{i=1}^{\infty}$ from the given sequence of number of arrivals $(u_{T_f}(i))$, in a fixed length (T_f) of adjacent intervals. Two kinds of smoothing, namely, the Poisson smoothing and the deterministic smoothing are reported in literature [39]. In Poisson smoothing, within a frame interval (T_f), the Poisson process generates the packet stream, with the mean rate of the Poisson process equal to the equivalent packet rate (number of packets divided by T_f) in the corresponding frame interval of the *real-world* VBR video traffic trace. The above procedure is repeated for each frame of the video trace to get the sequence of packet arrival instants equivalent (in a statistical sense) to the original real-world traffic trace (which is in frame size in bits). Note that the interarrival times between packets (within a T_f) are exponentially distributed in Poisson smoothing. In deterministic smoothing, in every frame of the real-world traffic, all the packets in the VBR trace are accommodated at equally spaced intervals. Thus, packet interarrivals are deterministic. Deterministic smoothing is used in next chapter.

In this chapter, we use a variation of deterministic smoothing, which generates variable sized packets. In particular, the method generates geometrically distributed packet sizes. In the following, we explain the deterministic smoothing with variable packet sizes.

7.1.1.1 Deterministic Smoothing with Variable Packet Sizes

Given a *real-world* VBR video traffic trace, the bit stream in the ith frame of size $u_{T_f}(i)$) is segmented into $N_{T_f}(i)$ packets, such that $N_{T_f}(i)$ is the maximum number of packets whose cumulative size does not exceed $u_{T_f}(i)$, while the packet sizes are geometrically distributed (see (7.1)). Then, all the packets in the ith frame interval $(N_{T_f}(i))$ are accommodated with equal intervals within the frame interval. The ratio of this geometrical packet sizes to the link capacity is used as service intervals, when we simulate queues with smoothed video traffic. We use the above procedure as an approximation for exponential service[1]. The purpose of the whole procedure is two-fold, one being feasibility of

[1] This is reasonable if the time to serve a bit is sufficiently small and the mean packet size is sufficiently large.

theoretical performance modeling and the another being realistic compulsions.

1. We wanted to have exponential distribution for service intervals for the following reason: the *theoretical* performance modeling using CIPP/M/1 queue is possible only if you have exponential distribution for service intervals.

2. We note that a *real-world* broadband switch serves a packet, proportional to its length. (See Section 6.1 for the behavior of *real-world* statistical multiplexer and related assumptions). In light of the first point mentioned above, we consider geometrical distribution for packet lengths. Now, the service times are set to the ratio of geometrical packet lengths to the link capacity. We beleive that, this is a reasonable approximation for exponential services.

The whole process of deterministic smoothing with variable sized packets is given below with more mathematical clarity.

Problem: Given the frame size sequence $\{u_{T_f}(i)\}_{i=1}^{N}$ (denoting the frame size of the ith frame interval in bits), generate the sequence of packet arrival instants $\{T_j\}_{j=1}^{n}$ with packet sizes $S(j)$ for $1 \leq S(j) \leq n$ where N is the total number of frame intervals in the frame size trace and n is the total number of arrival of (variable sized) packets after smoothing.

Methodology:

Step 1: During the ith frame interval, generate the geometrically distributed packet sizes $V_i(j)$ for any $N_{T_f}(i)$ such that,

$$N_{T_f}(i) = \overset{argmax}{k} \ \{k | \sum_{j=1}^{k} V_i(j) < u_{T_f}(i)\} \tag{7.1}$$

Here, $N_{T_f}(i)$ denotes the number of packets generated in the ith frame interval T_f.

Step 2: Generate $\{T_k\}$ such that,

$$T_k \;=\; (i-1)T_f + \frac{N_{T_f}(i)}{T_f} j \quad 1 \leq j \leq N_{T_f}(i) \tag{7.2}$$

where, $k = j + \sum_{l=1}^{i-1} N_{T_f}(l)$ for $2 \leq i \leq N$ and $k = j$ for $i = 1$.

Step 3: Packet sizes are given by,

$$S(j + \sum_{l=1}^{i-1} N_{T_f}(l)) = V_i(j) \quad 1 \leq j \leq N_{T_f}(i) \tag{7.3}$$

for $2 \leq i \leq N$ and

$$S(j) = V_i(j) \quad 1 \leq j \leq N_{T_f}(i) \tag{7.4}$$

for $i = 1$.

7.1.2 Error Quantification

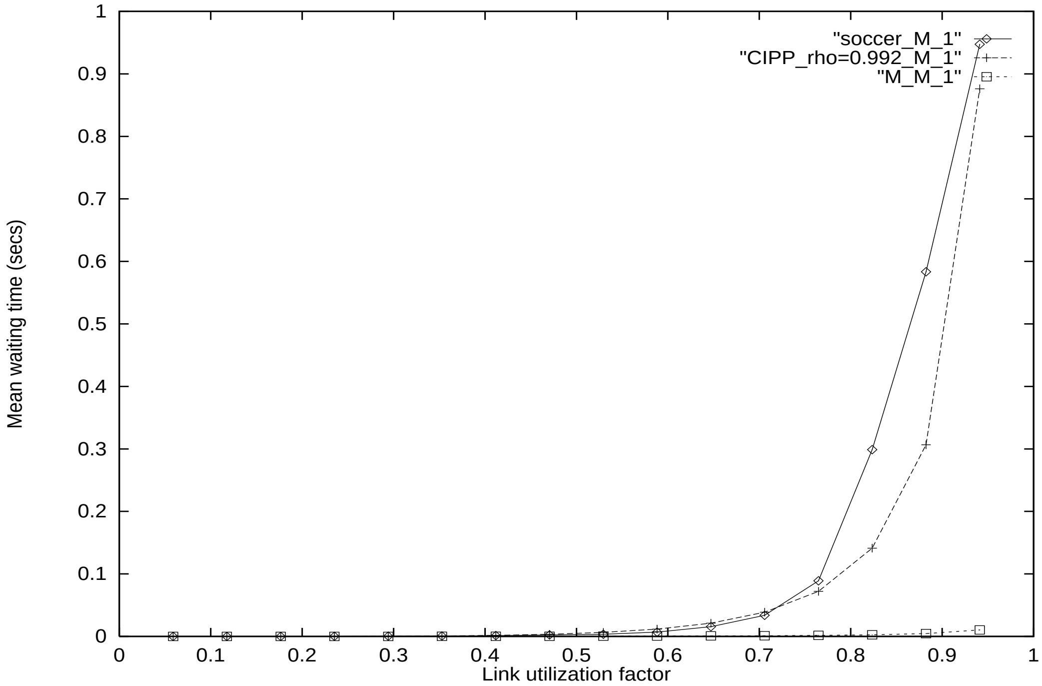
To quantify the error between the curve corresponding to real-world data and the theoretical fit, we use the percentage error as given in Section 6.1.1.

7.2 Performance Modeling of Statistical Multiplexer with VBR Input Using CIPP/M/1 Queue

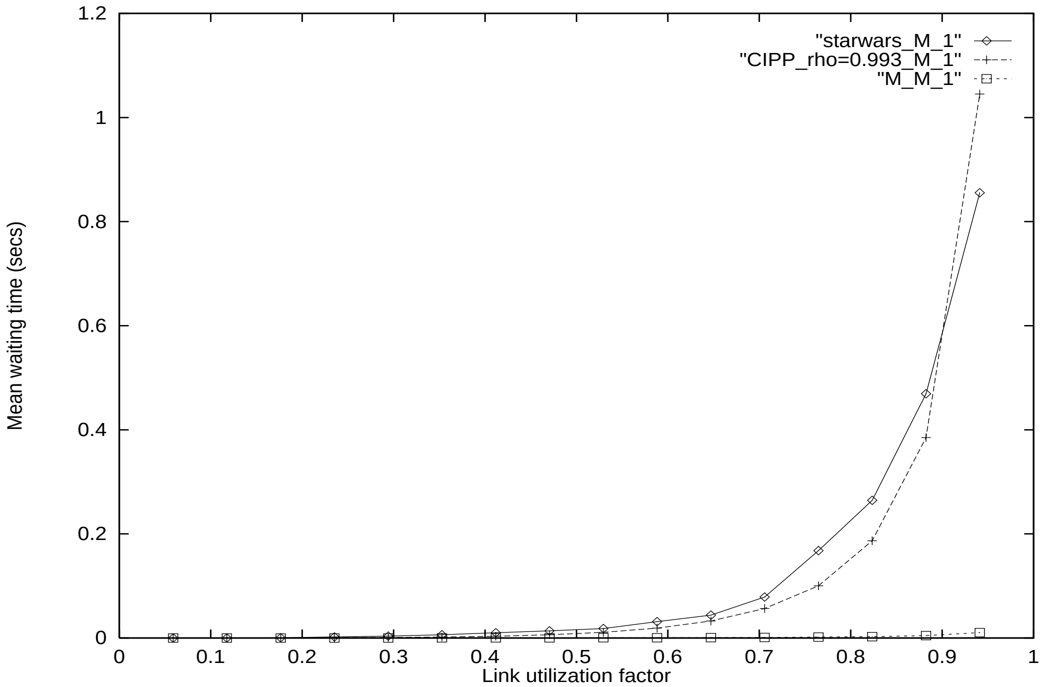
In this section, we consider the performance modeling of statistical multiplexer with *real-world* MPEG-1 VBR video traffic using CIPP/M/1 queue. Queueing performance measures (see Section 3.4) considered here are mean number of packets, mean waiting time, packet delay variation (*pdv*) and overflow probability which are given by (5.66), (5.67), (5.69) and (5.70) respectively. We consider the *soccer* and *starwars* MPEG-1 encoded VBR video sequences [2] for our experiments.

7.2.1 Mean Waiting Time

We first consider the mean waiting time for the performance modeling using CIPP/M/1 queue. Our results for the mean waiting time corresponding to *soccer* and *starwars* video sequences are shown in Figure 7.1. Also shown in the figure is the corresponding quantity for the 'best fit' CIPP/M/1 and M/M/1 queues. The error performance of the CIPP/M/1 and M/M/1 queues are presented in Table 7.1. The table gives both the maximum and minimum percentage error and the corresponding values for the link utilization factor. From the table and the figure, it can be seen that CIPP/M/1 queue approximates well the behavior of the queue with VBR video traffic with respect to the measure of mean waiting time. Note that, it has been observed by Paxson *et al* [32] that the Poisson process poorly models the traffic. Moreover, the M/M/1 queue seriously underestimates the queueing performance measures [29], which results in overly optimistic performance predictions. Our results here, corroborate their work. Moreover, our *theoretical* results quantifies the performance degradation from M/M/1 queue, given the variability (captured by ρ of CIPP) in the *real-world* traffic.



(a)



(b)

Figure 7.1: Mean waiting time of packets in a queue with VBR video sequence input and that of the approximating CIPP/M/1 and M/M/1 queues for (a) *soccer* (b) *starwars*

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
soccer	7.1 (a)	96.94	0.36	99.31	0.45	7.51	0.48	52.72	0.42	0.992
starwars	7.1 (b)	98.15	0.33	99.03	0.45	17.97	0.45	40.24	0.39	0.993

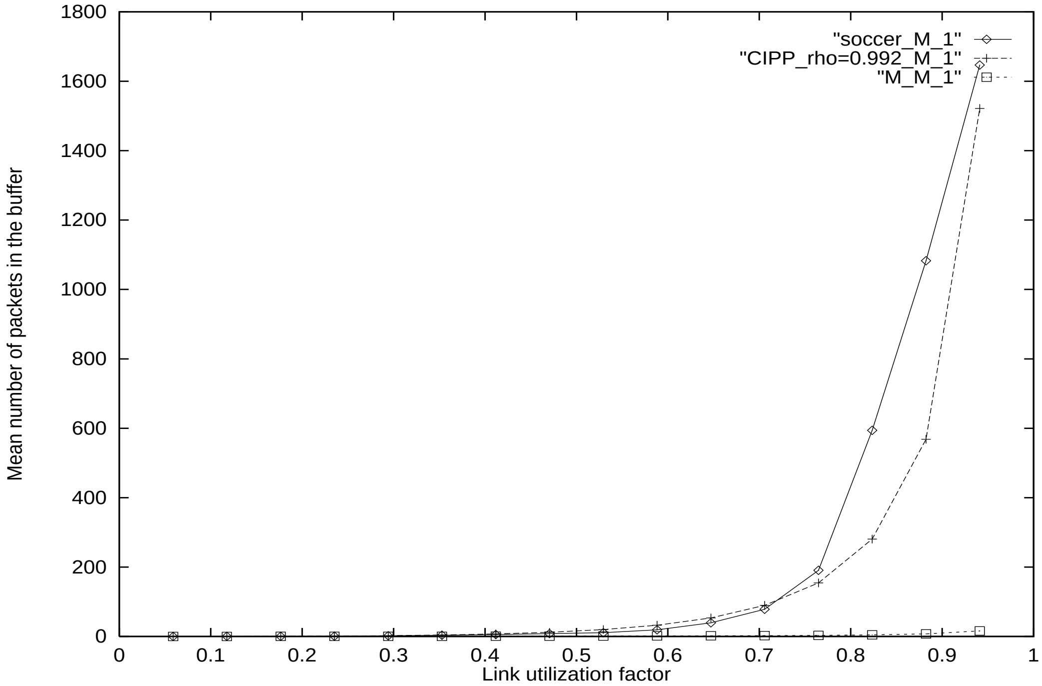
Table 7.1: Percentage error between the mean waiting time of packets in queues with *soccer* (Figure 7.1 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 7.1 (b)). In both the plots corresponding error for the approximating M/M/1 queue is shown for comparison.

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
soccer	7.2 (a)	96.93	0.36	99.31	0.45	7.59	0.48	52.79	0.42	0.992
starwars	7.2 (b)	95.52	0.33	97.68	0.24	18.08	0.24	40.23	0.4	0.993

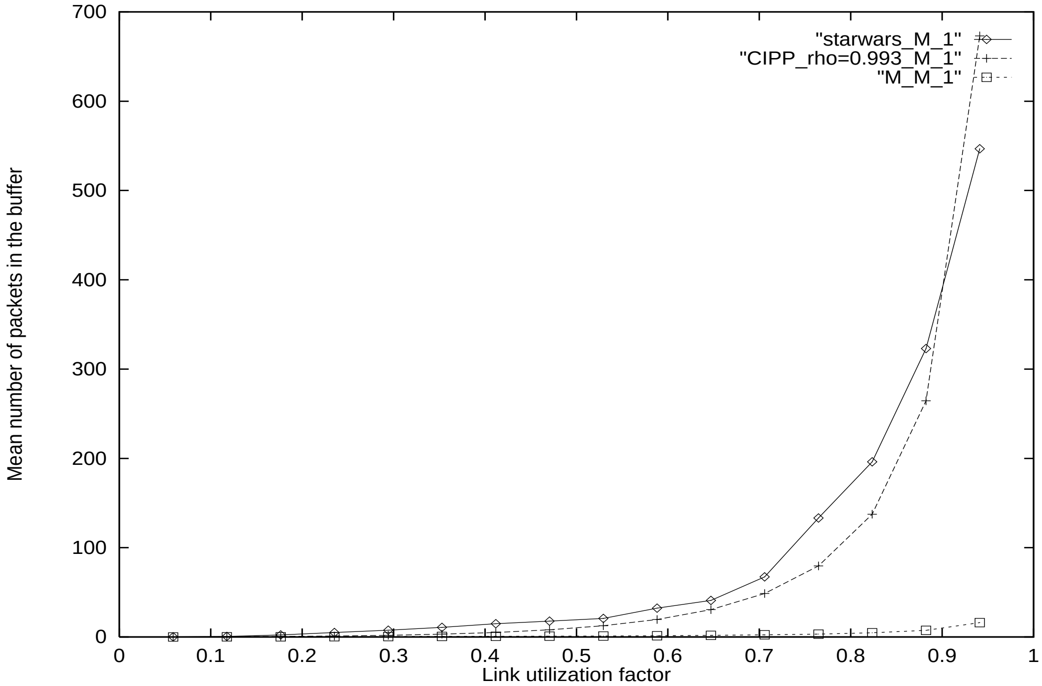
Table 7.2: Percentage error between the mean number of packets in queues with *soccer* (Figure 7.2 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 7.2 (b)). In both the plots the corresponding error for the approximating M/M/1 queue is shown for comparison.

7.2.2 Mean Number of Packets

We consider the measure of mean number of packets in the system for our performance modeling study. Figure 7.2 illustrates our results corresponding to *soccer* and *starwars* video sequences. The results corresponding to the approximating CIPP/M/1 and M/M/1 queue are also shown. The Table 7.2 gives the maximum and minimum percentage error between the theoretical performance measure, the mean number of packets corresponding to CIPP/M/1 queue and the estimated performance measure from simulations of the queue with the above VBR video sequences. Also, the error performance of the 'best fit' M/M/1 queue is indicated in the table. From the table and the figure, it can be seen that CIPP/M/1 queue performs quite well, in modeling the queue with real-world VBR traffic. Here again, the results corresponding to M/M/1 queue corroborate the work by Erramilli *et al* [29].



(a)



(b)

Figure 7.2: Mean number of packets in a queueing system with VBR video sequence input and that of the approximating CIPP/M/1 and M/M/1 queues (a) *soccer* (b) *starwars*.

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
soccer	7.3 (a)	99.9	0.48	99.95	0.45	27.87	0.42	37.95	0.45	0.992
starwars	7.3 (b)	99.95	0.48	99.99	0.36	0.99	0.48	88.08	0.33	0.993

Table 7.3: Percentage error between the packet delay variation in queues with *soccer* (Figure 7.3 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 7.3 (b)). In both the plots corresponding error for the approximating M/M/1 queue is shown for comparison

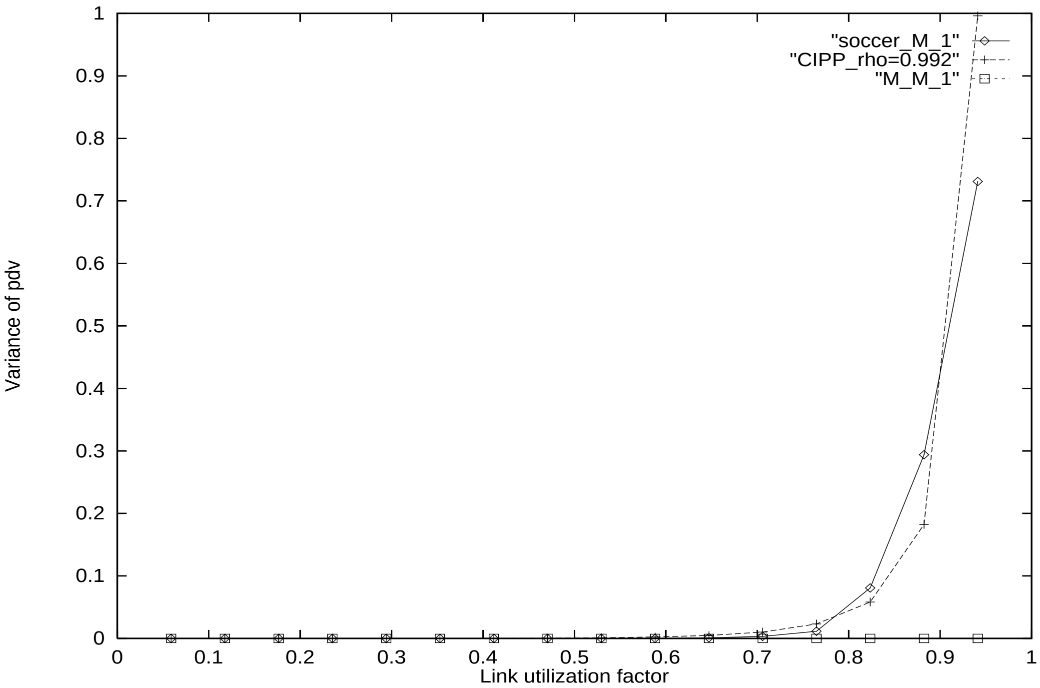
.

7.2.3 Packet Delay Variation (pdv)

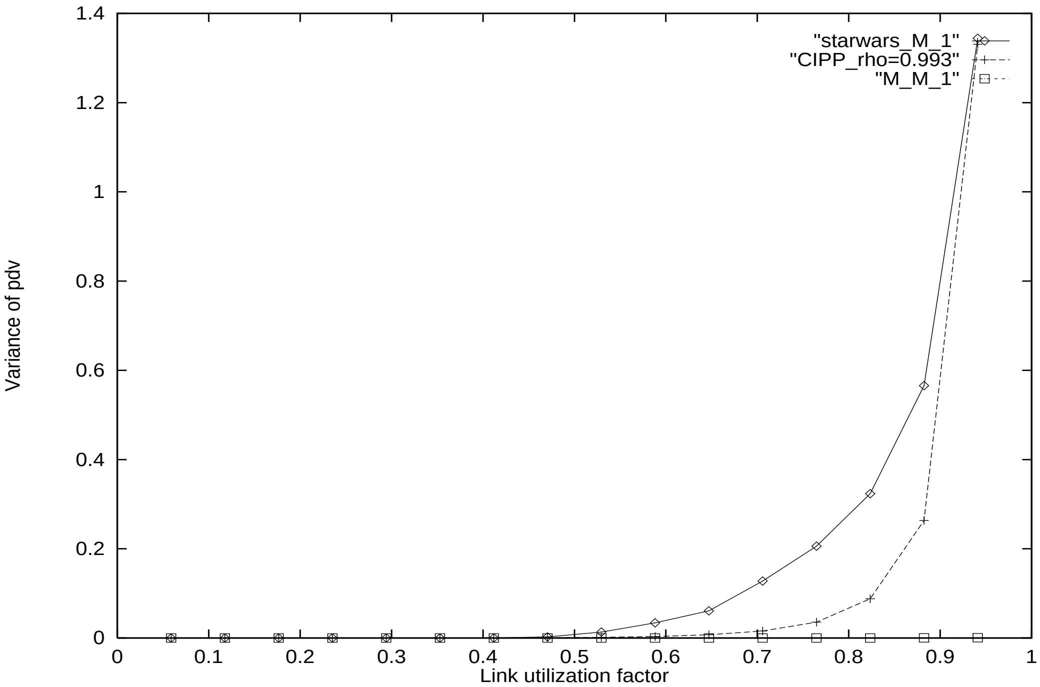
Here, our queueing performance measure will be packet delay variation (*pdv*). Our results corresponding to the video sequences *soccer* and *starwars* are depicted in Figure 7.3. The Table 7.3 gives the maximum and minimum percentage error between the theoretical performance measure, the *pdv* corresponding to CIPP/M/1 queue and the estimated performance measure from simulations of the queue with the above VBR video sequences. From the table and the figure, it can be seen that CIPP/M/1 queue performs quite well. Our results corresponding to M/M/1 queue is optimistic.

7.2.4 Overflow Probability

We now consider the queueing measure, the overflow probability. We take the threshold of 32 packets ($K = 32$) for our experiments. Figure 7.4 illustrates our results for overflow probability corresponding to *soccer* and *starwars* MPEG-1 video sequences. Also, the results corresponding to 'best fit' CIPP/M/1 and M/M/1 queues are also presented. The Table 7.4 gives the maximum and minimum percentage error between the theoretical overflow probability corresponding to CIPP/M/1 queue and the estimated overflow probability from simulations of the queue with the above VBR video sequences. Error statistics corresponding to M/M/1 queue is also indicated in the table for comparison. From the table and the figure, it can be seen that the performance of the 'best fit' CIPP/M/1 queue matches well with the behavior of the queue with VBR video traffic. Again, our results corresponding to M/M/1 queue agree well with the results of the work by Erramilli *et al* [29].

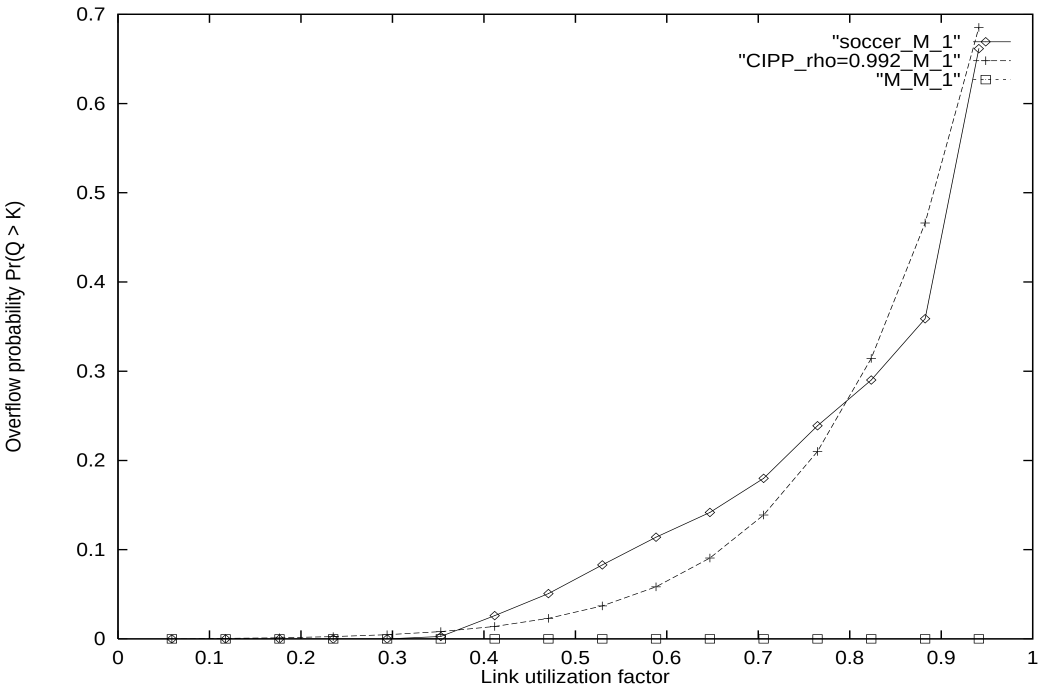


(a)

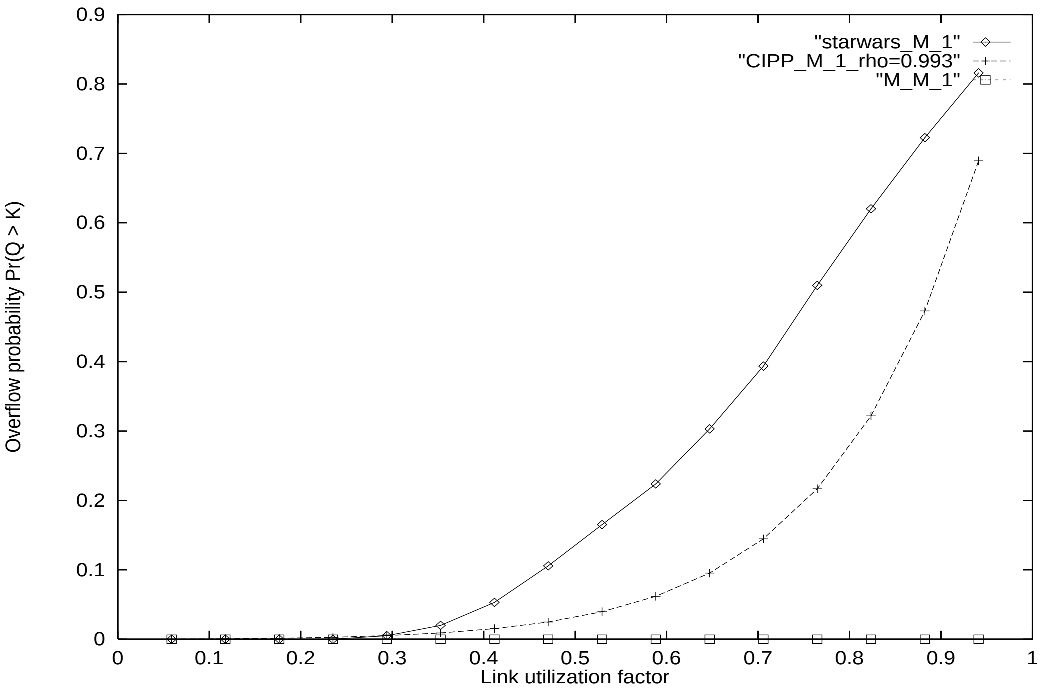


(b)

Figure 7.3: The *pdv* of a queue with VBR video sequence input and that of the approximating CIPP/M/1 and M/M/1 queues (a) *soccer* (b) *starwars*.



(a)



(b)

Figure 7.4: Overflow probability of a queue with VBR video sequence input and that of the approximating CIPP/M/1 queue (a) *soccer* (b) *starwars*. Threshold considered here is 32 packets.

Trace data	Fig. no.	M/M/1				CIPP/M/1				
		min	occurs at LUF	max	occurs at LUF	min	occurs at LUF	max	occurs at LUF	ρ
soccer	7.4 (a)	95.58	0.76	100.0	0.24	3.6	0.3	35.99	0.33	0.992
starwars	7.4 (b)	99.99	0.94	100.0	0.88	15.53	0.94	34.56	0.88	0.993

Table 7.4: Percentage error between the overflow probability in queues with *soccer* (Figure 7.4 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 7.4 (b)). In both the plots corresponding error for approximating M/M/1 is shown for comparison. A threshold of 32 packets is considered here

.

7.3 Performance modeling of VBR video traffic using CIPP/M/1/K queue

In this section, we undertake the performance modeling of finite capacity statistical multiplexer with VBR video traffic using CIPP/M/1/K queue. We use the queueing performance measures packet loss probability and conditional packet loss probability for our performance modeling study. Since the CIPP/M/1/K queue is analytically intractable, we use the simulation approach. (Refer the Section 7.1 for simulation methodology for queues with finite capacity). We first demonstrate by simulations that, in finite capacity queue with MPEG-1 VBR video input, the buffer occupancy distribution exhibits peaks at empty state and fully occupied state. This is consistent with the results observed by Lazar *et al* [38] on the MAGNET II network and by Skelly *et al* [39] on video conferencing traces. In this section, we consider the *race, dino* and *MrBean* MPEG-1 encoded VBR video sequences [2] for our experiments.

7.3.1 Buffer Occupancy Distribution

Figure 7.5 illustrates our simulation results for the buffer occupancy distribution for a queue with *race* and *MrBean* MPEG-1 video sequences and their corresponding 'best fit' CIPP/M/1/K and M/M/1/K queues. We have considered $K = 64$ packets. Note that the distribution peaks at the empty state and fully occupied state. Earlier we observed that this is the case for CIPP traffic (Chapter 5) and composite Ethernet traffic (Chapter 6). From Figure 7.5, it is also seen that CIPP models reasonably well the queueing situation with real-world MPEG-1 VBR video traffic, even in the case of finite capacity.

One can see that the distribution for both the CIPP/M/1/K and queue with *real-world* VBR video traffic are drastically different from the M/M/1/K queue case. This is an another manifestation of the observation (by Paxson [32]) that Poisson process, models the broadband teletraffic poorly.

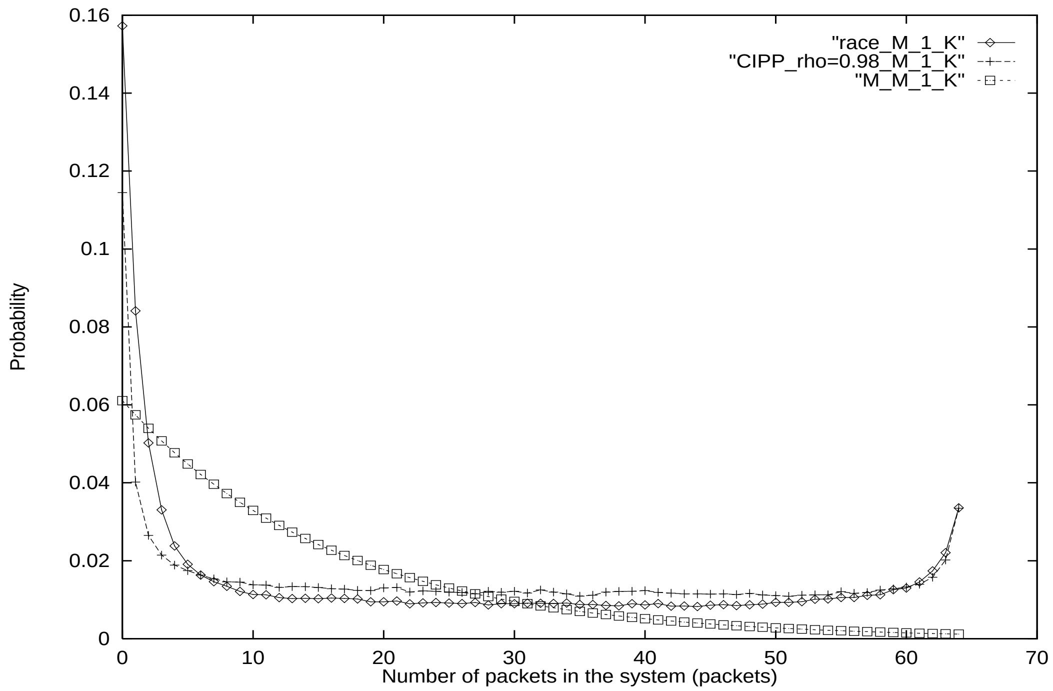
7.3.2 Packet Loss Probability

In this section, we concentrate on the performance measure of packet loss probability. We consider a buffer capacity of 96 packets for our simulation purposes. Figure 7.6 depicts our results for the video sequences *race* and *MrBean*. From the figure, one can see that, the CIPP makes a good approximation to the performance of finite buffer statistical multiplexer with real-world MPEG-1 video traffic.

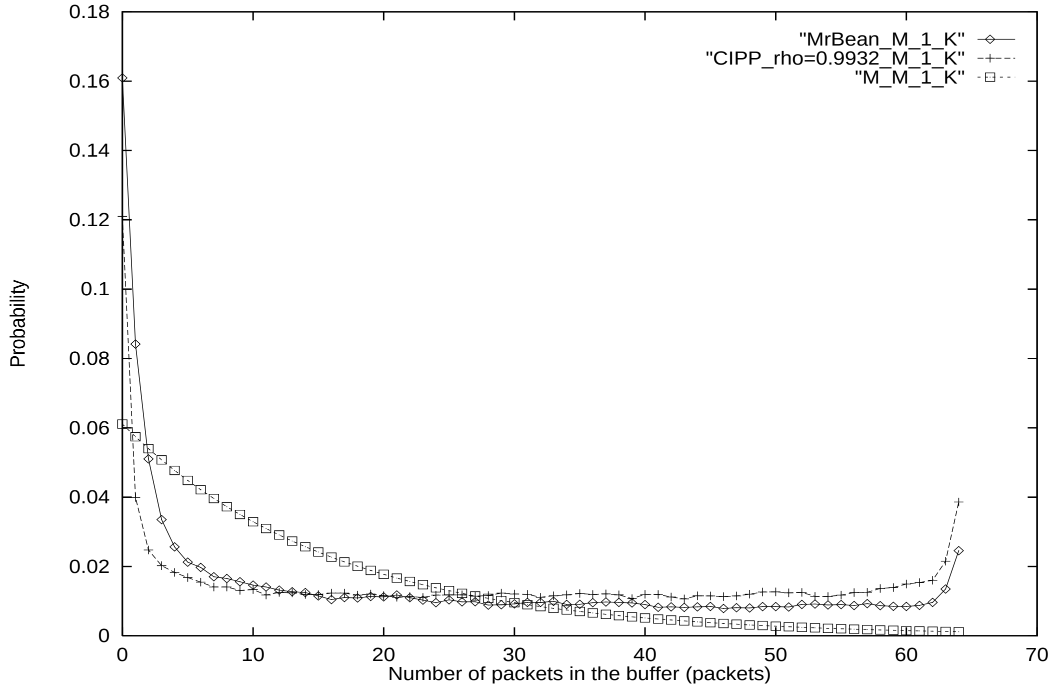
7.3.3 Conditional Packet Loss Probability (CPLP)

Often, the packet loss probability is not an appropriate indicator of performance of packet loss. In particular, for video sources using interframe and intraframe redundancy to achieve compression, if the sufficient number of consecutive packets are lost in a single video connection, it could lead to the loss of the whole block of image data at the receiver while decoding and hence suffer QoS degradation. Simulation [104] and analytical studies [105] show that even if the overall packet loss of the superposition is small, individual sources may experience the loss of a string of consecutive packets, sufficient enough to lead to the above phenomena. This consecutive packet loss has two origins: first, there may be temporal overload situation due to the fact that the number of active users exceeds the capacity of the system and second, due to the periodical nature of the arrival process, a call losing a packet is likely to lose a string of packets. Here, in the case of individual VBR source, the later situation is of much relevance. Moreover, we note that, CPLP is of prime importance, in particular, in the guaranteed services like video services wherein, a consecutive packet loss of compressed stream may lead to a loss of block of data after decoding. We will see that CIPP with high ρ is able to capture this behavior of consecutive packet loss.

Figure 7.7 illustrates the conditional packet loss probability for the queue with *real-world* MPEG-1 VBR video sequences namely, *race*, *dino* and *MrBean* as input and their corresponding 'best fit' CIPP/M/1/K queue. Also shown are the corresponding results for the M/M/1/K queue case. One can see from the figure that, the CPLP corresponding to M/M/1/K queue decays very fast, while the decay (with respect to length of consecutive

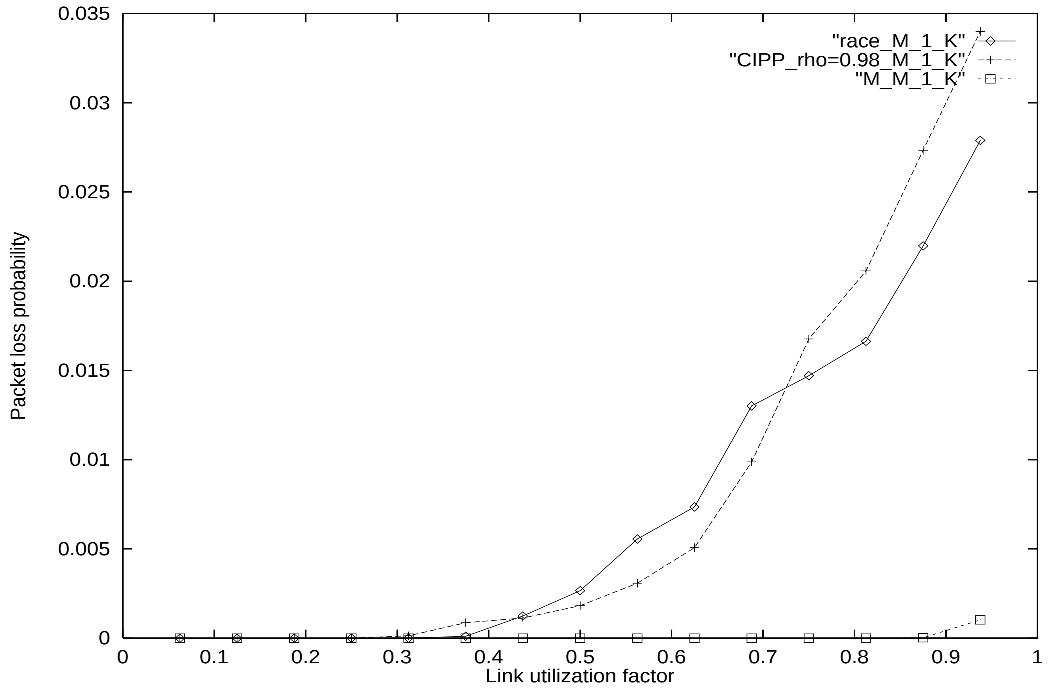


(a)

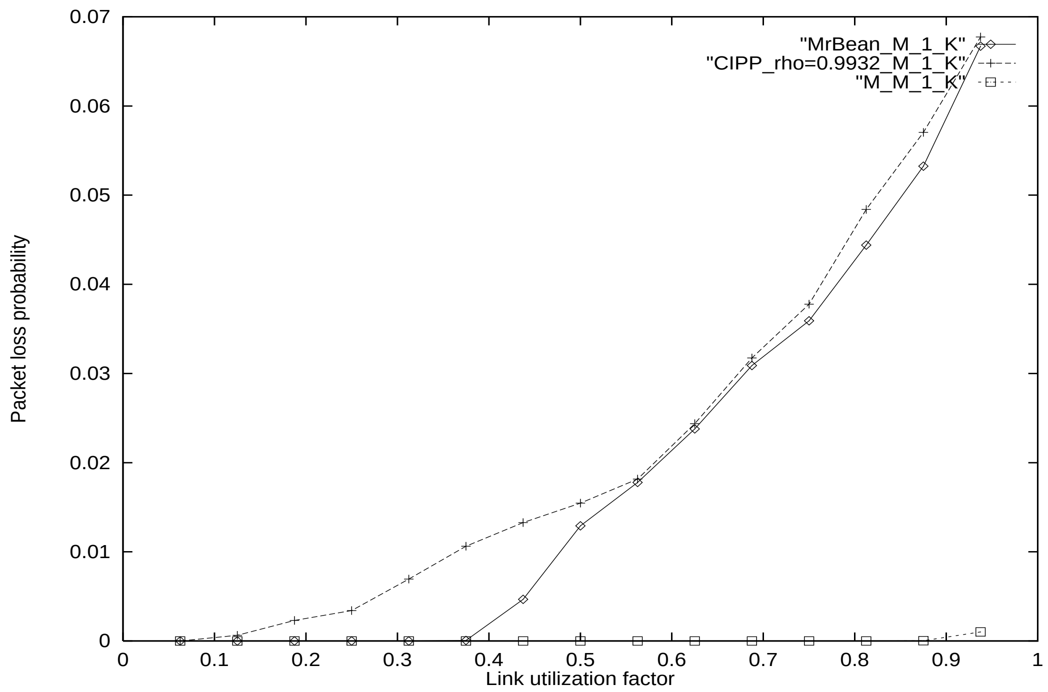


(b)

Figure 7.5: Buffer occupancy distribution of a queue with VBR video sequence input and that of the approximating CIPP/M/1/K and M/M/1/K queues (a) *race* (b) *MrBean*. Buffer capacity considered here is 64 packets.

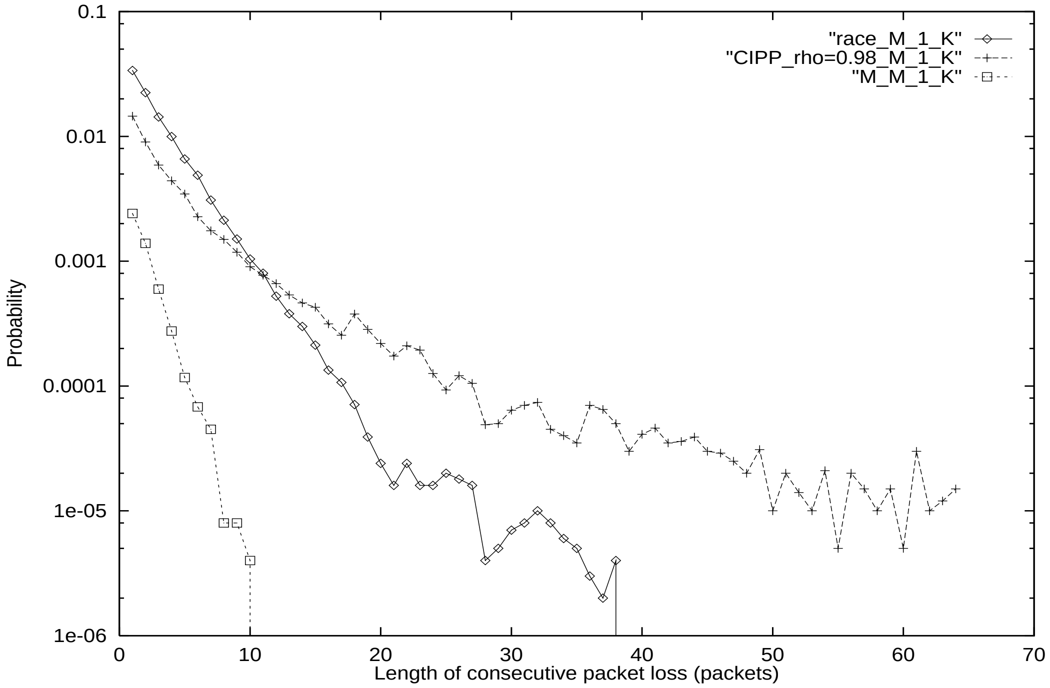


(a)

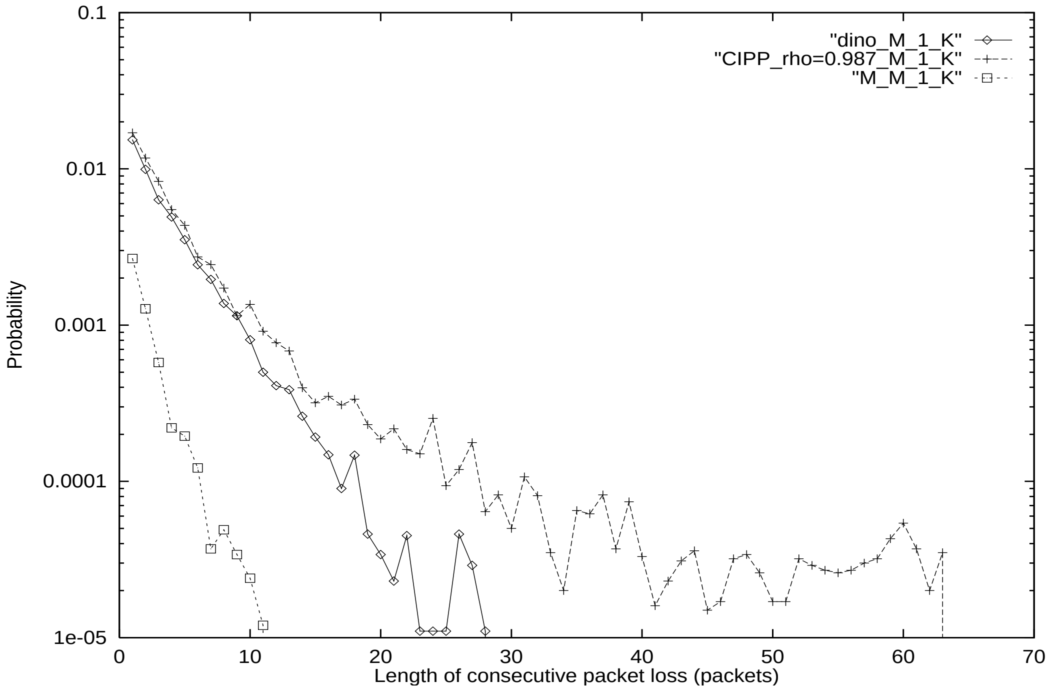


(b)

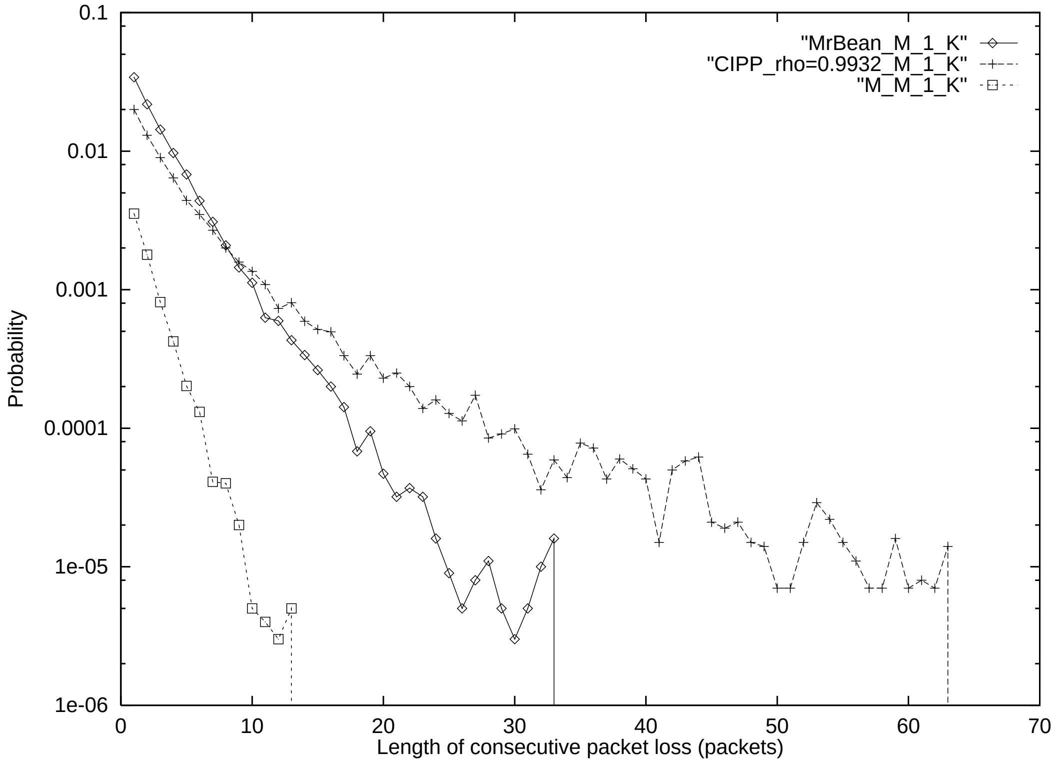
Figure 7.6: Loss probability of a queue with MPEG-1 VBR movie sequences and that of the approximating CIPP/M/1/K and M/M/1/K queues for a buffer size of 96 packets (a) *race* (b) *MrBean*


"race_M_1_K"
"CIPP_rho=0.98_M_1_K"
"M_M_1_K"
Probability
0.1
0.01
0.001
0.0001
1e-05
1e-06
0
10
20
30
40
50
60
70
Length of consecutive packet loss (packets)

(a)


"dino_M_1_K"
"CIPP_rho=0.987_M_1_K"
"M_M_1_K"
Probability
0.1
0.01
0.001
0.0001
1e-05
0
10
20
30
40
50
60
70
Length of consecutive packet loss (packets)

(b)



(c)

Figure 7.7: Conditional packet loss for a queue with the MPEG-1 VBR video sequences and the approximating CIPP/M/1/K and M/M/1 queues (a) *race* (b) *dino* and (c) *Mr-Bean*

packet loss (k)) is slow for the CIPP/M/1/K case. For CIPP with high ρ, this trend of 'slower' decay is highly pronounced. This is the consequence of high correlation in the arrival stream.

7.4 Fluid Flow Approximation of Queue with CIPP Arrivals

In this section, we use the fluid-flow approximation of CIPP queues for performance modeling of VBR traffic. The main advantage of this method is that it allows one to use the real-world traces *as they are* thus, avoiding the need for smoothing. Moreover, fluid-flow approach has another attractive feature, namely, negligible computational overhead. Throughout this section, we will use the term 'cell' to denote fixed size packet.

Here, we present the cell loss simulation results that are used to compare the present model with the conventional models. We have simulated on frame level assuming a constant video transmission rate for each frame. The frame level results are essentially

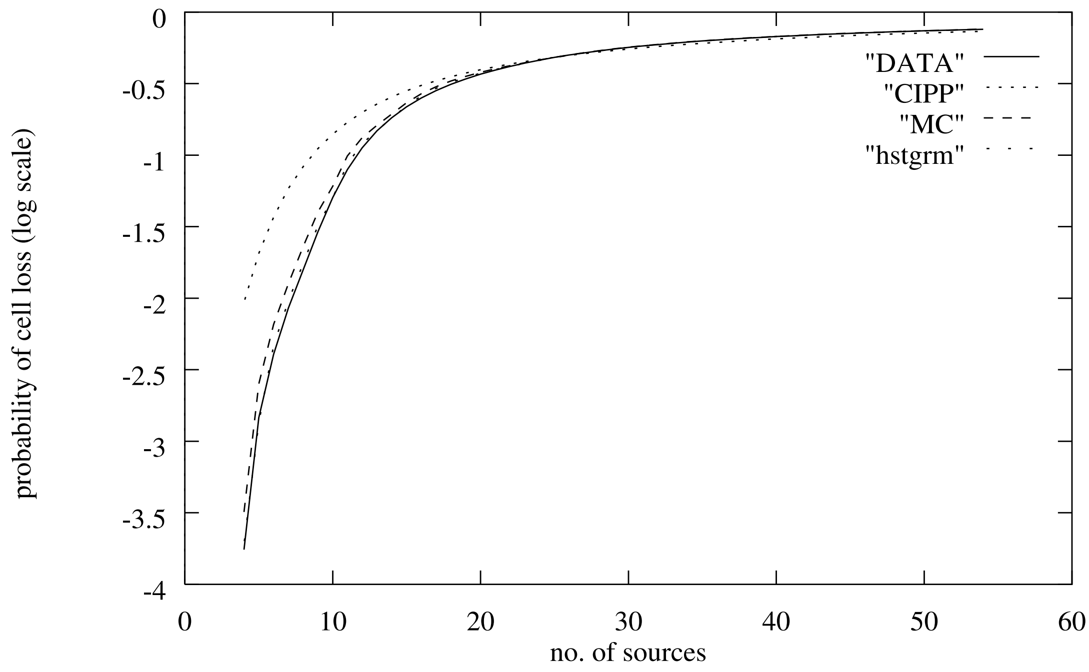


Figure 7.8: Cell loss probability for a buffer of 100 cells

equivalent to the cell level results, if we assume that each video source spreads the cells belonging to a frame over the whole frame duration.

The multiplexer has a link rate of 10 Mbps. The buffer sizes of 100, 1000, 10000 and 100000 cells have been considered. The multiplexer load is determined by the number of video input sources, ranging from 5 (a load of about 0.4) to 40 (a load of about 3). We have scaled down the link rate to 10 Mbps because we want to obtain realistic cell loss results in spite of the fact that the empirical data set has a bit rate of only 5 to 10 % of high quality full-screen MPEG video sequences.

Figure 7.8 shows the cell loss results for the histogram, MC, CIPP models [1] and *race* data set for a buffer size of 100 cells. The CIPP models better than the histogram and MC models. For a buffer size of 1000 cells (see Figure 7.9), the situation remains the same. Increasing the size to 10000 cells (see Figure 7.10) shows that the CIPP and MC models show the best approximation quality, whereas the estimates of the histogram are becoming more and more optimistic. For very large buffer size of 100000 cells (Figure 7.11), the CIPP model underestimates the cell loss probability, however, for buffer sizes of up to 10000 cells, CIPP shows marginal overestimate. This is desirable since, for network dimensioning purpose, it is more convenient to use a model which marginally overestimates the traffic.

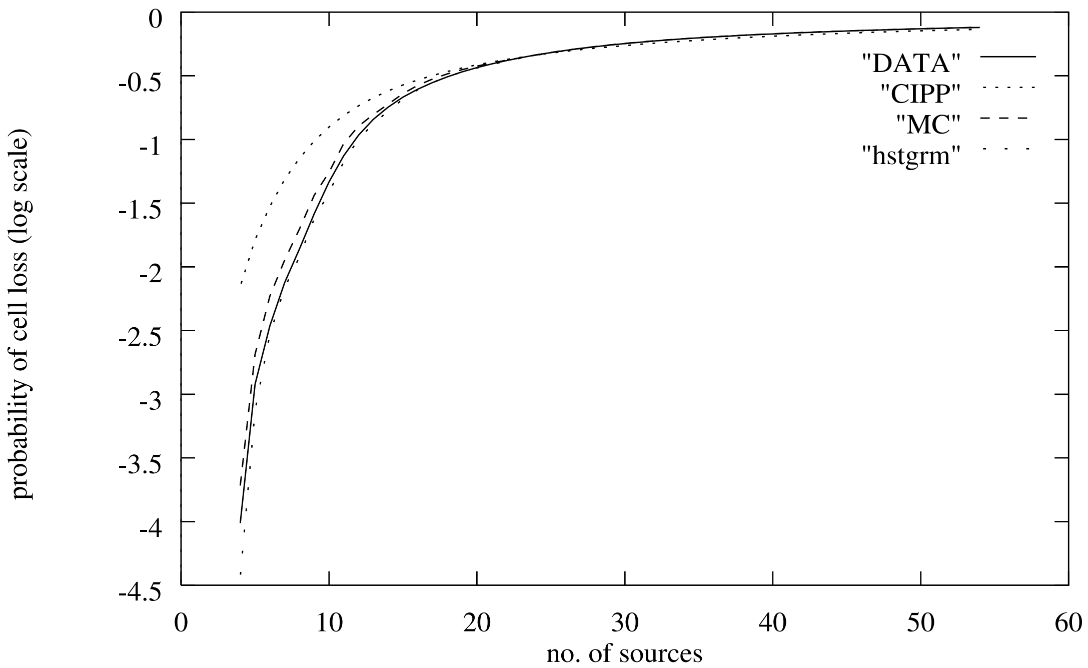


Figure 7.9: Cell loss probability for a buffer of 1000 cells

7.5 Discussions

In this chapter, we have used the theoretical results of Section 5.2 to characterize the queueing situation encountered by the VBR video traffic in a statistical multiplexer. Observe that, not only the first order statistics of queueing performance corresponding to data namely, mean number of packets and mean waiting time for packets, are well modeled by CIPP/M/1 queue, but the overflow probability and the higher order statistics namely the packet delay variation (*pdv*) as well. In summary, we note the following.

1. For the VBR video sequences of large H, the approximation by CIPP/M/1 queue is poor (that is, the performance modeling is poor). Note that, in our results, the largest value of maximum percentage error occurs for the measure of mean waiting time of packet corresponding to *soccer* sequence and is 52.73. The corresponding figure for *starwars* sequence is 88. Note that the H parameter for *soccer* sequence is 0.77, whereas for *starwars* it is 0.85.

2. In most of the performance curves given above, there is a knee point over which the curve rises sharply. The load corresponding to this 'knee', keeps decreasing for sequences with high H parameter.

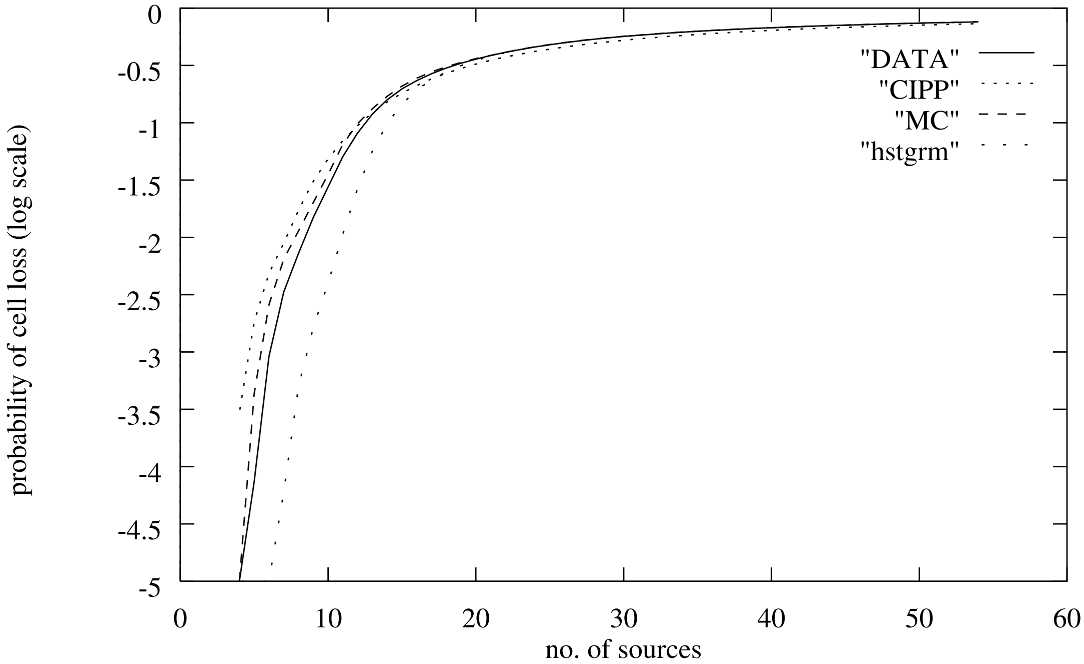


Figure 7.10: Cell loss probability for a buffer of 10000 cells

The above observations correspond to infinite capacity systems. Since, all real systems are finite buffered, we also considered the performance modeling by CIPP/M/1/K queue, to get more realistic results. A simulation study is carried out to compare the buffer occupancy distribution of the finite capacity queue with Bellcore Ethernet traffic and the "best fit" CIPP/M/1/K queue with K arbitrary. We have then undertaken a simulation based performance modeling study of statistical multiplexer with VBR video traffic using the measures, packet loss probability and conditional packet loss probability. The following are the salient features of observations from our simulation based performance modeling study for the finite capacity case.

1. Buffer occupancy distribution of *real-world* MPEG-1 VBR traces exhibit peaks at the empty state and fully occupied state, in particular, at high utilizations. This behavior is consistent with that of the earlier workers [38], [39].

2. For low values of k (denoting the length of consecutive packet loss), the CPLP for VBR trace data was found to be higher than the value predicted by CIPP, whereas, for high value of k, the CPLP for CIPP was found to be high. This is due to the fact that CIPP arrivals with high ρ is nearly periodic.

3. The results corresponding to M/M/1/K queue show that it is an another manifestation of the observation made by Paxson *et al* [32].

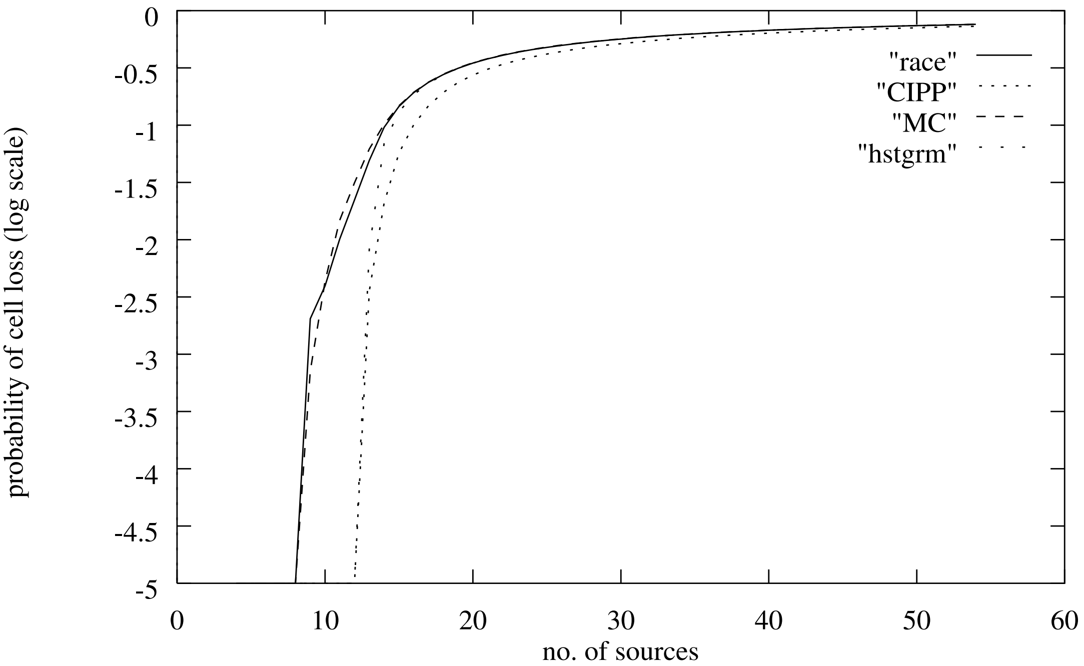


Figure 7.11: Cell loss probability for a buffer of 100000 cells

In summary, the results demonstrate that, CIPP/M/1/K queue approximates well the queueing situation encountered by the real-world VBR video traffic in a finite buffered statistical multiplexer. The main contribution in this chapter is that, we have demonstrated that considering models which capture the interarrival correlation from the real-world data is more appropriate is not only in the modeling context, but also in the performance modeling aspect.

Fluid flow approximation of CIPP queue is also shown to approximate well the queueing situation encountered due to VBR traffic in ATM networks. The simulation results show that for low buffer sizes, CIPP outperforms the MC and the histogram models. For larger buffer sizes, CIPP marginally overestimates the cell loss probability. This is desirable since, for network dimensioning purpose, it is more convenient to use a model which marginally overestimates the traffic.

Chapter 8

Interference Effects on MPEG Video over ATM Networks by Network Traffic Modeled by CIPP

In the previous chapters, we have introduced the CIPP, developed the theory for CIPP/M/1 queue, applied it for studying the performance modeling of statistical multiplexer with composite traffic input and VBR video traffic as input. In this chapter we will give an example where the empirical relationship of the correlation parameter ρ of CIPP to the Hurst parameter (as given in Section 4.5.3) is exploited to study the effect of variability of the network traffic on various network performance measures. More specifically, we present a simulation based performance evaluation study of MPEG-2 video when subjected to interference effects by a cross traffic process (background network traffic) modeled using the CIPP model.

The MPEG-2 standard [111] assumes that the end-to-end delay (of cells belonging to a video connection) from an MPEG-2 system's encoder to decoder is constant to ensure that the encoder and decoder clocks operate at the same frequency and the decoder buffer does not overflow nor underflow. But, an ATM network introduces jitter while transporting the corresponding cells[1] from source to destination, which may cause problems in decoder operations. The main source of the jitter is the statistical nature of the background traffic which is basically, multiplexed version of other highly variable individual component traffic. More precisely, the jitter suffered by cells belonging to video traffic largely depends on the degree of self-similarity of the background network traffic for the deterministic service systems. So is also, the losses of cells (belonging to the video connection under study) due to finite capacity of the ATM multiplexer. This is the kind of interference we will be addressing in this chapter. Here, the background network traffic model is modeled by CIPP. Apart from the interference study, we will also discuss the interactions among the ATM multiplexer buffer size, the decoder buffer size, the correlation parameter ρ

[1]Throughout this chapter we will freely use the term *cell* to denote the 53 byte ATM cell.

of the CIPP (characterizing the self-similar nature of the background traffic) and the various performance measures including packet loss ratio.

This chapter is organized as follows. Section 8.1 gives the overview of our experimental set up. In Section 8.2, source side modeling is presented. The mode of transport of MPEG TS packets over ATM networks is explained in Section 8.3. Section 8.4 discusses the merging of cells from the VBR video source under study with the cells represented by network traffic (modeled by CIPP). In Section 8.5, the decoder system is defined which is implemented for our simulation study. Section 8.6 presents the simulation results. Section 8.7 contains the discussions on the chapter.

8.1 Introduction

8.1.1 Related Work

To the best of our knowledge, until now, most studies, explore the decoder buffer behavior without considering the ATM network jitter [112, 113, 114], except in [115] and [116]. In [116], the network is modeled as a black box which introduces jitter into the cells, where distribution for jitter was approximated by truncated Gaussian function. In this chapter, we model the background composite network traffic by CIPP, in order to take into account the self-similar nature of the background network traffic. (Note, that it has been shown in Section 4.2 that the CIPP models well the broadband teletraffic). The advantage of using the CIPP for modeling the network traffic is that, there exists an empirical relationship between ρ, a parameter of CIPP and the estimated local Hurst parameter of the corresponding CIPP trace (see Section 4.5.3). Another attractive feature of CIPP is that it is easy to simulate on the computer.

8.1.2 Methodology

We model the network as a single server FCFS queue with the superposition of a *real world* VBR video traffic and the background network traffic generated by CIPP. By doing so, we inherently introduce jitter into the cells belonging to video traffic under study. We explore the interference effect of self-similarity of the background network traffic on the performance of VBR traffic.

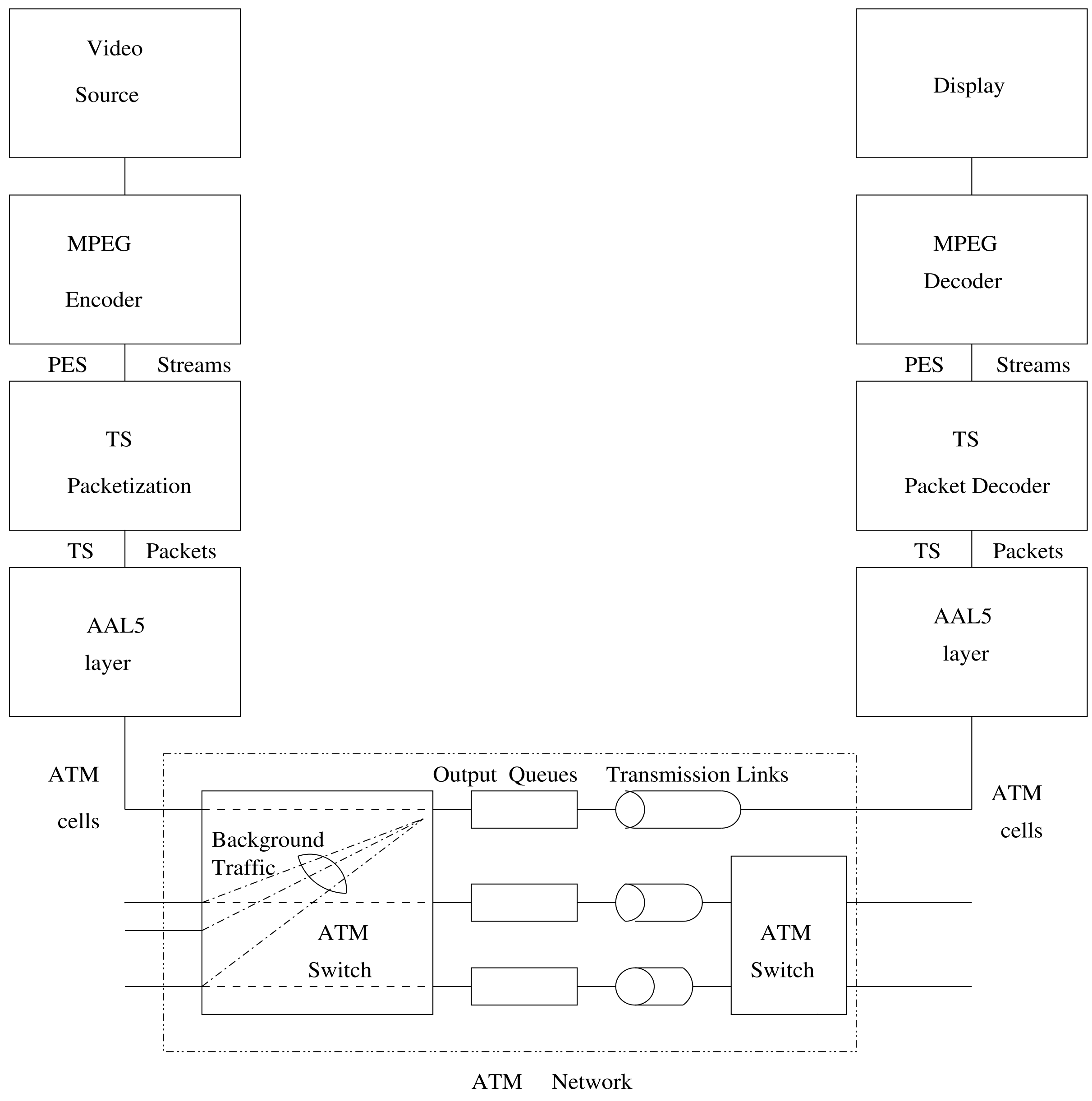


Figure 8.1: Block schematic of end-to-end simulation of MPEG transport streams over ATM networks under CIPP network traffic.

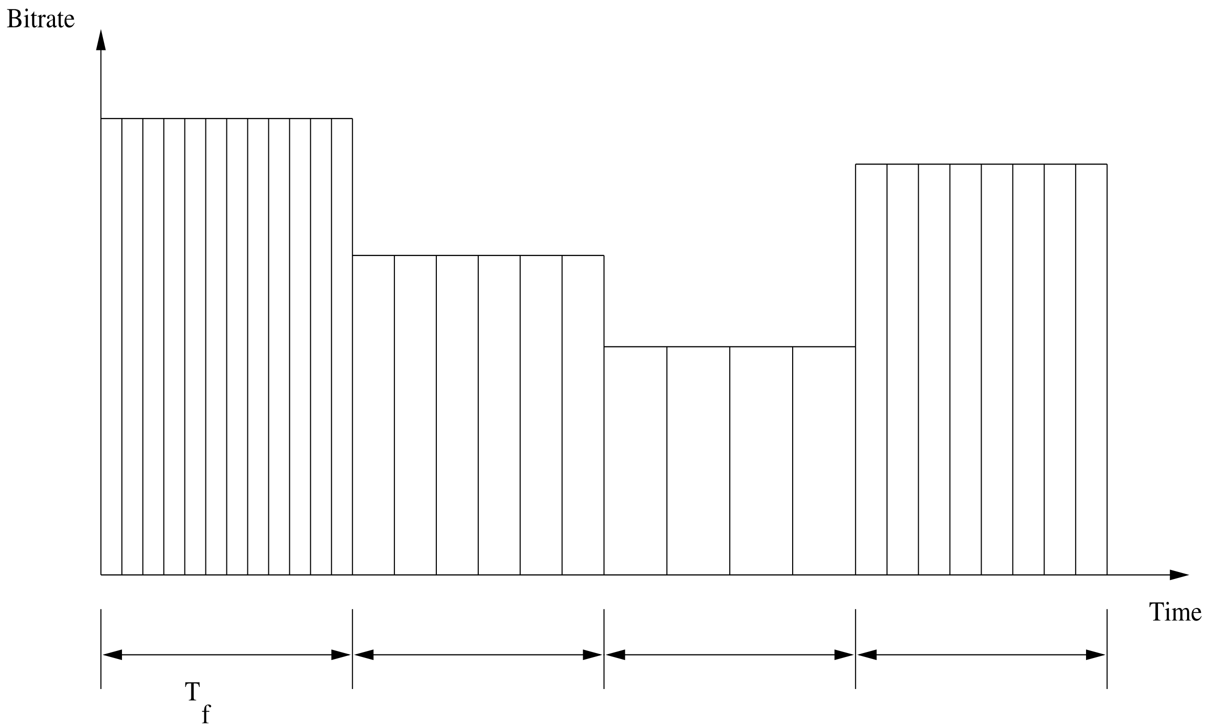


Figure 8.2: Generation of TS packets from video PES: deterministic smoothing

8.1.3 System Description

The block schematic for our experiment is shown in Figure 8.1. We first generate MPEG transport stream (TS) packets by (deterministic) smoothing [116] the *real world* MPEG traffic (see Figure 8.2). The TS packets are then encapsulated into ATM cells using the ATM adaptation layer 5 (AAL5) (see Figure 8.3). We simulate their transport over ATM networks by merging these ATM cells with the background ATM cells generated based on the CIPP model and feeding them into single server FCFS queue. Then, we segregate the cells belonging to VBR traffic under study, at the output. Finally, these cells are reassembled into TS packets at the destination, which are delivered to the upper layer TS packet decoder. We simulate the behavior of the decoder buffer under real-time requirements. We show the interactions among the ATM multiplexer buffer size, the decoder buffer size, the correlation parameter ρ of the CIPP (characterizing the self-similar nature of the background traffic) and the various performance measures.

8.2 Source-side Modeling of MPEG Transport Stream

The MPEG system first packetizes the elementary streams (compressed video sequences) to produce packetized elementary streams (PES) packets. PES packets are further combined with system information to form TS packets. A TS packet has a fixed length of 188 bytes. We use the MPEG *terminator* VBR movie sequence for our simulation study. The coding parameters used are $M = 3$ (M is the space between anchor frames) and $N = 12$ (N is the space between I frames), which yield the following fixed frame pattern within each group of pictures (GOP): IBBPBBPBBPBB. Given the video trace, we first segment into TS packets and deterministic smoothing is performed. This is illustrated in Figure 8.2.

8.3 Transporting MPEG TS Packets over ATM Networks

We consider the transport of MPEG TS packets over ATM networks using AAL5 PDU. In this scheme, MPEG TS packets are mapped into AAL5 PDUs, each containing 2 TS packets. Each PDU is then converted into eight ATM cells. At the destination, eight ATM cells are reassembled into an AAL5 PDU, which is then decomposed into two adjacent TS packets. These TS packets are then delivered to the upper layer TS packet decoder. This is illustrated in Figure 8.3.

8.4 ATM Multiplexer with Background Network Traffic Process

After the cells are formed at the source, they are merged with CIPP traffic (representing the background network traffic) which is independently generated. This merged traffic is fed to the FCFS queue which models the output queue of an ATM multiplexer. The cells belonging to VBR traffic under study is segregated at output. In this study, we set the rate of CIPP generator as 80 % of mean arrival rate of the VBR traffic under study. We study the effect of variability (or self-similarity) of the network traffic (by varying the correlation parameter ρ of the modeling CIPP) on the performance of the decoder buffer and ATM multiplexer buffer. We also study the interactions among decoder buffer behavior and multiplexer at various values of the correlation parameter ρ.

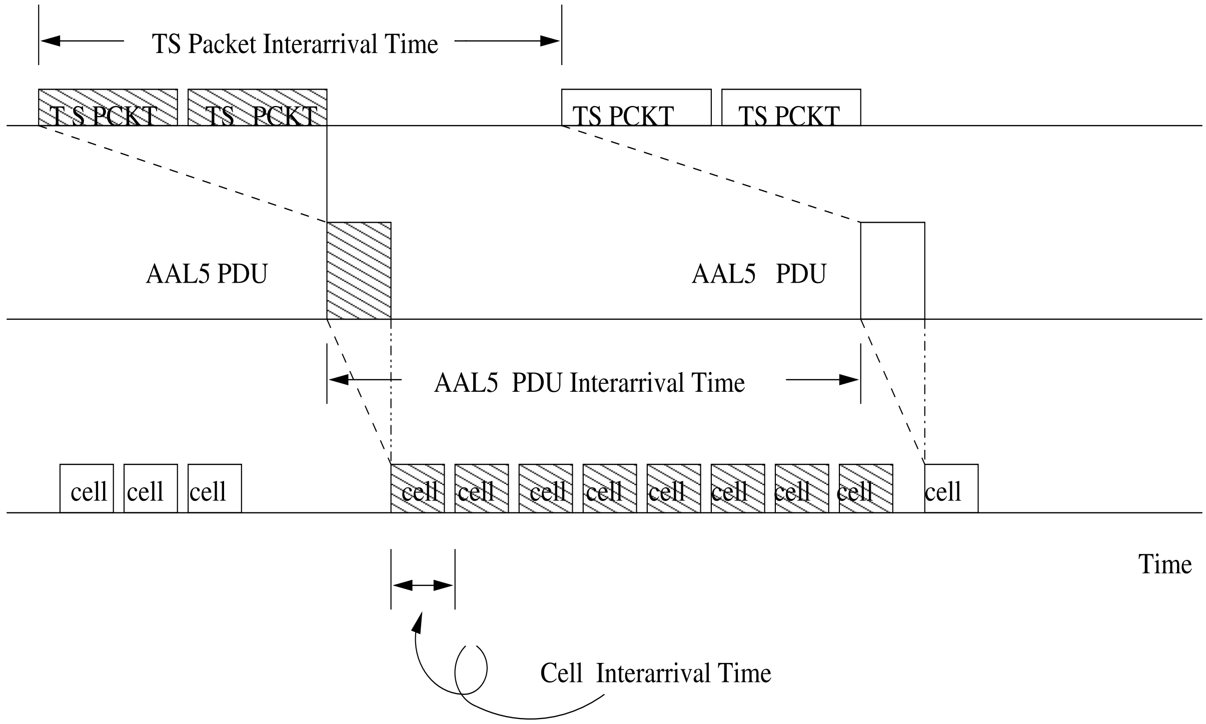


Figure 8.3: Generation of ATM cells from TS packets

8.5 Decoder Buffer System Definition

At the destination, the MPEG video TS packets enter the MPEG TS system target decoder (T-STD) decoder buffer after they are reassembled from ATM cells. We model the decoder buffer operation based on the T-STD [111]. For detailed information about the operation of T-STD, the reader is referred to [111]. In our study, we consider a simplified version of the T-STD in which decoding is performed at regular intervals (T_f secs), where $1/T_f$ is the frame rate. Packet (TS) losses occur when the buffer is full. In our simulations, at every frame interval of T_f secs, we check whether the TS packets for one video frame have arrived completely. If all the TS packets have arrived, they are removed from the decoder buffer instantly. Otherwise, we wait for one more frame interval (T_f sec) to see whether the data corresponding to this frame have arrived completely. After this extra frame interval, the arrived data for this frame are removed from the decoder buffer even though they may be incomplete. The late TS packets are counted as lost TS packets in our *delay loss rate* (defined below) calculations. We believe that even with this simplified decoder, the salient features of the MPEG decoding mechanism still remain and the overall performance would not change in any significant way. Also, simulation results based on this simplified model give us the worst case scenario and provide an upper bound on the ATM multiplexer and the decoder buffer requirements. The following are the performance measures, we consider for our study: (i) *delay loss rate* - probability of loss of TS packets due to delay at the decoder buffer, (ii) *network loss rate* - probability of loss of (equivalent) TS packets due to cell losses in the ATM multiplexer (iii) *overflow loss rate* - probability of loss of TS packets due to decoder overflow and (iv) *cell delay variation*[2] (*cdv*) while video traffic cells transiting ATM buffer.

8.6 Simulation Results

Let us denote the multiplexer buffer size and the decoder buffer size by M and D respectively. Note that M is in cells and D is in TS packets. Our study consists of three experiments. The first one aims at studying the interference effect of the network traffic on the performance of the VBR video traffic under study. The second and third, explore the interactions among the ATM multiplexer buffer size (M), the decoder buffer size (D), the degree of self-similarity of background traffic process in the network and the various performance measures.

First, we fix the buffer capacity of the ATM multiplexer (M) and the decoder buffer size (D). The experiment is carried over ρ. The results are given in Figures 8.4 and 8.5. We would also like to study the interactions between the ATM multiplexer buffer size (M) and the decoder buffer size (D) for fixed ρ. Hence, for the second set of results, we fix the ρ of CIPP and the buffer capacity of ATM multiplexer (M). We study the effect of the decoder buffer size (D) on the performance measures. We repeat the experiment for various combinations of ρ and M. The results are depicted in Figures 8.6 and 8.7. Our third experiment is designed to explore the effect of the ATM multiplexer buffer size M on the performance measures, parameterized by ρ and D. The results are presented in Figure 8.8.

8.6.1 Interference Effect of Background Network Traffic Process on VBR traffic

As one may see from Figure 8.4 (a), delay loss rate increases as ρ increases initially slowly, but increases sharply as ρ approaches 1. This is due to the fact that as the background

[2]same as packet delay variation as introduced in Section 5.3.1.3 except that the 'packet' here is a fixed size, 53 byte ATM cell.

network traffic becomes more variable (or more self-similar, captured by CIPP[3] of high ρ), it has destructive effect on the VBR video stream under study. Note that the curves corresponding to $D = 32$ and $D = 64$ merge. This intuitively makes sense, since the delay losses are independent of D. One may also note that, losses are high for the parameterization $D = 32$ and $M = 32$. Figure 8.4 (b) gives our result for overflow loss rate. Again queueing performance degrades as ρ increases. Overflow loss rate rises sharply as ρ approaches 1. Observe that the overflow losses are high for the parameterization $D = 32$ and $M = 32$. Network loss rate and cell delay variation (*cdv*) (Figure 8.5 (a) and Figure 8.5 (b) respectively) show the same behavior as delay loss rate. Results are readily interpretable. Note that, the curves corresponding to parameterization of $D = 32$ and $D = 64$ merge similar to the situation as we observed in Figure 8.4 (a).

8.6.2 Decoder Buffer Size D

Intuitively, one expects that the delay loss rate is independent of the buffer size D. This is confirmed by our experimental results given in Figure 8.6 (a). It is seen from figure that delay losses are high for $\rho = 0.9$. Figure 8.6 (b) gives our result for overflow losses. As D increases, overflow loss rate decreases. Losses are high for the parameterization $M = 32$ and $\rho = 0.9$. Careful study of Figure 8.7 (a) reveals the fact that the network losses are independent of D. Here, also one may see that higher ρ corresponds to more high. Cell delay variation follows the same behavior, which is depicted in Figure 8.7 (b)

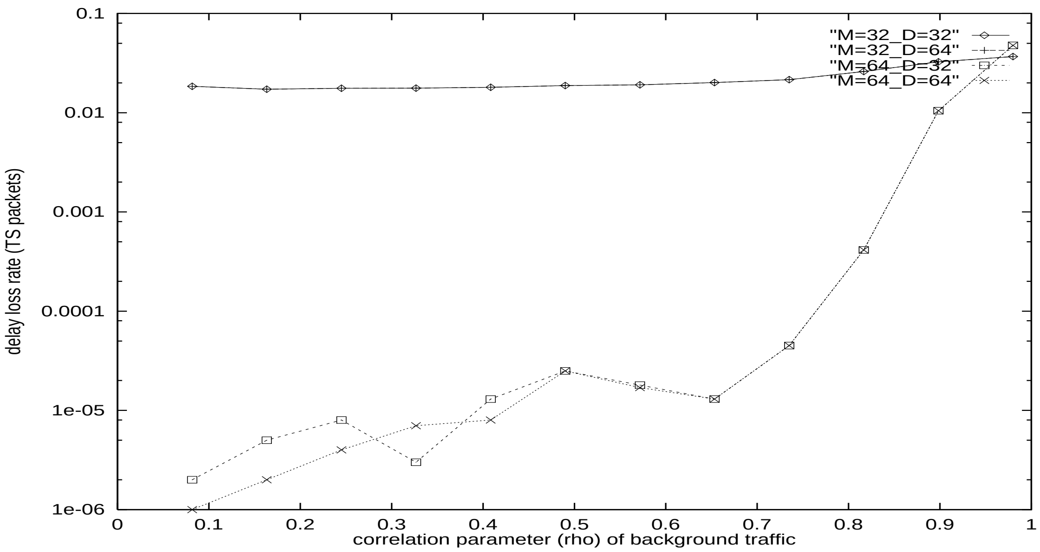
8.6.3 ATM Multiplexer Buffer Size M

Figure 8.8 (a) depicts the fact that the network losses decreases as M increases. Here, also one may see that higher ρ corresponds to higher losses. Cell delay variation follows the same behavior, which is depicted in Figure 8.8 (b). Note that the curves corresponding to $D = 32$ and $D = 64$ merge.

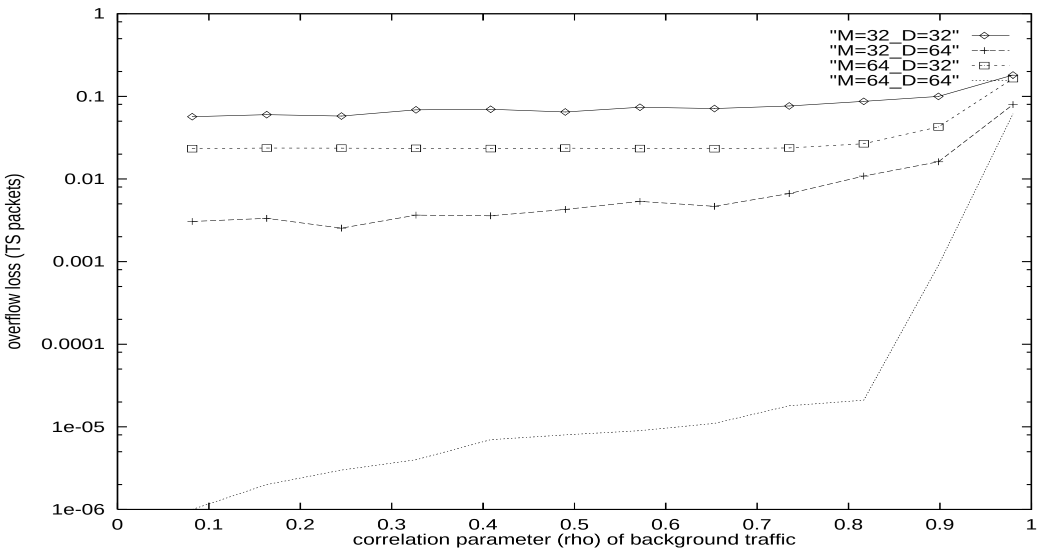
8.7 Discussions

In this chapter, we have given one example to illustrate how CIPP can be used to study the interference effect of the background composite network traffic, on the performance of VBR video traffic. We found through experiments that as the variability of background

[3] Recall, the monotonic non-linear empirical relationship between ρ and the estimated Hurst parameter H we learnt in Section 4.5.3.

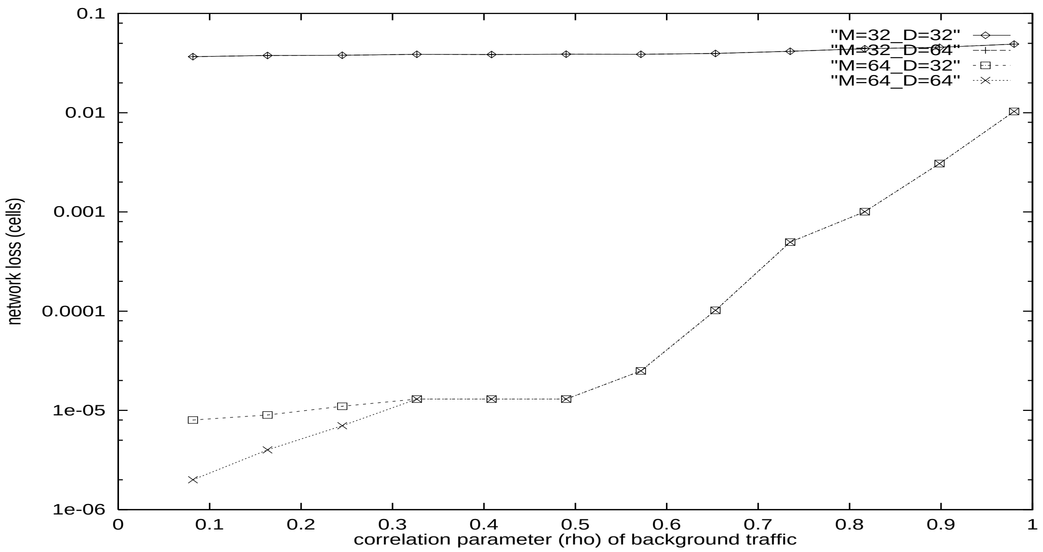


(a)

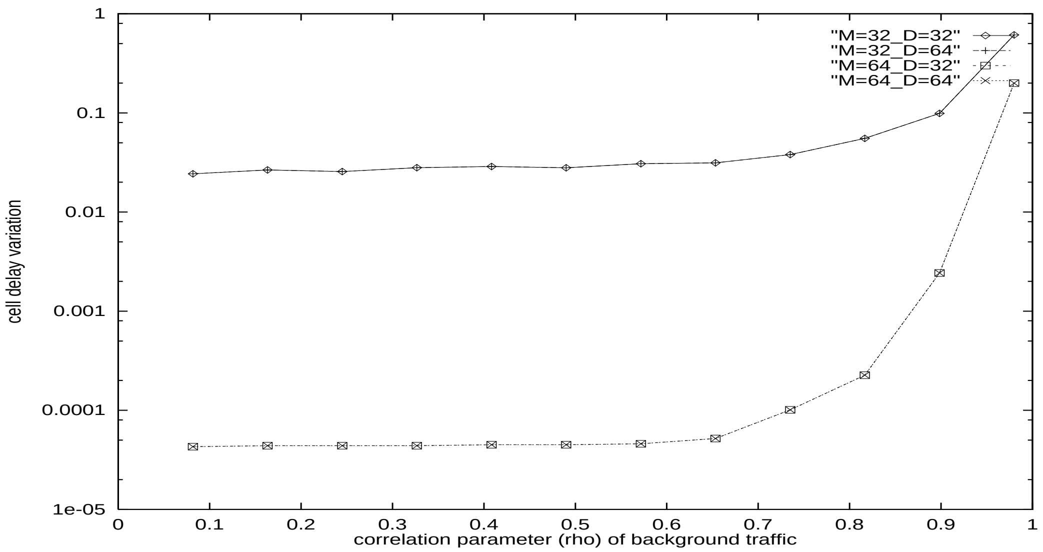


(b)

Figure 8.4: Decoder performance measures versus ρ for *terminator* video sequence (a) Delay losses (b) Overflow losses.

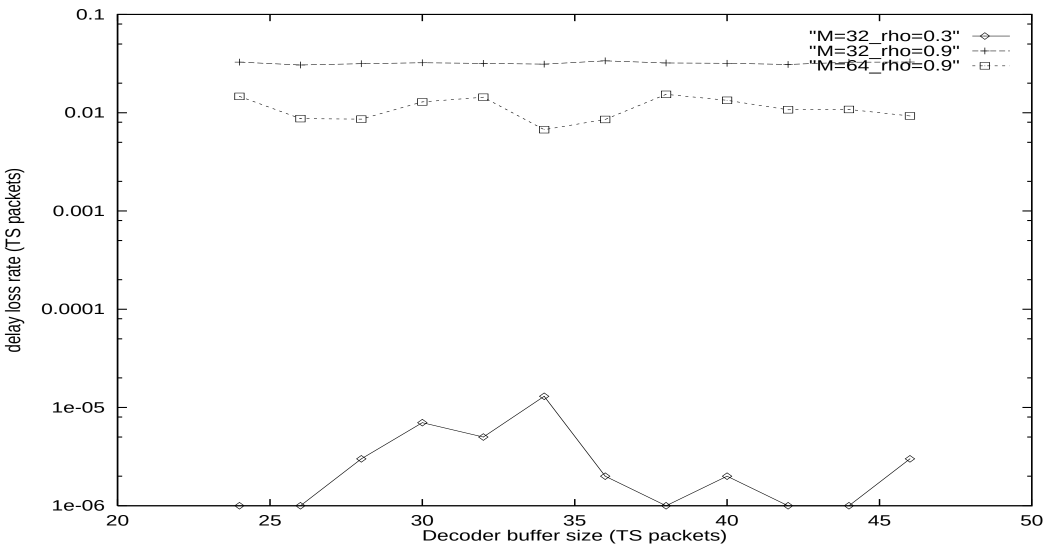


(a)

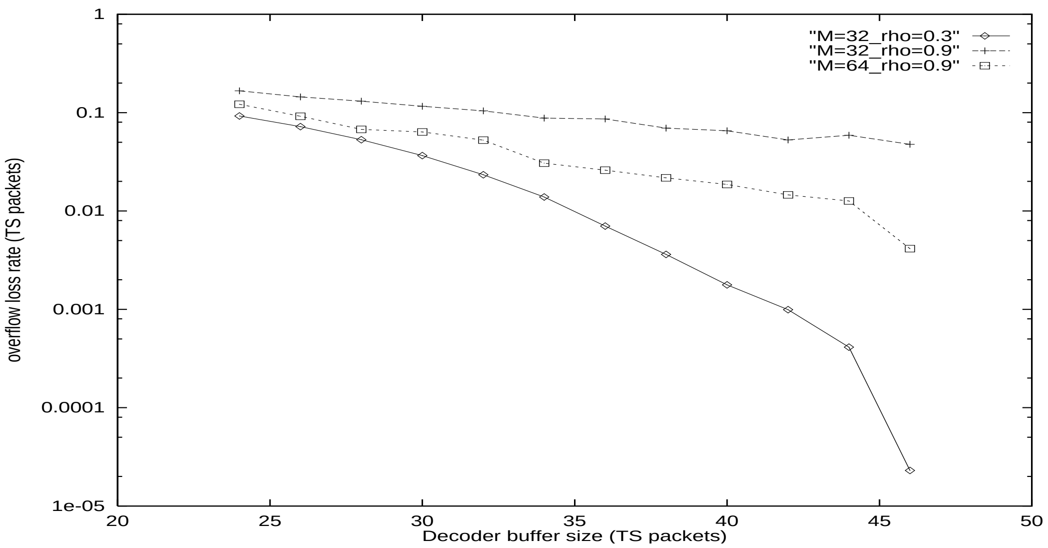


(b)

Figure 8.5: Multiplexer performance measures versus ρ for *terminator* video sequence (a) Network losses (b) cdv.

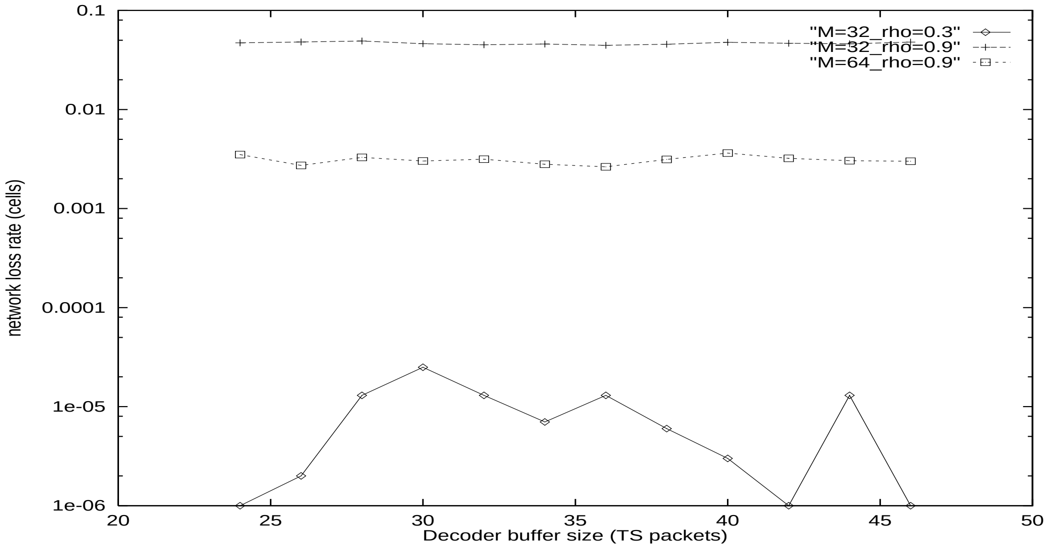


(a)

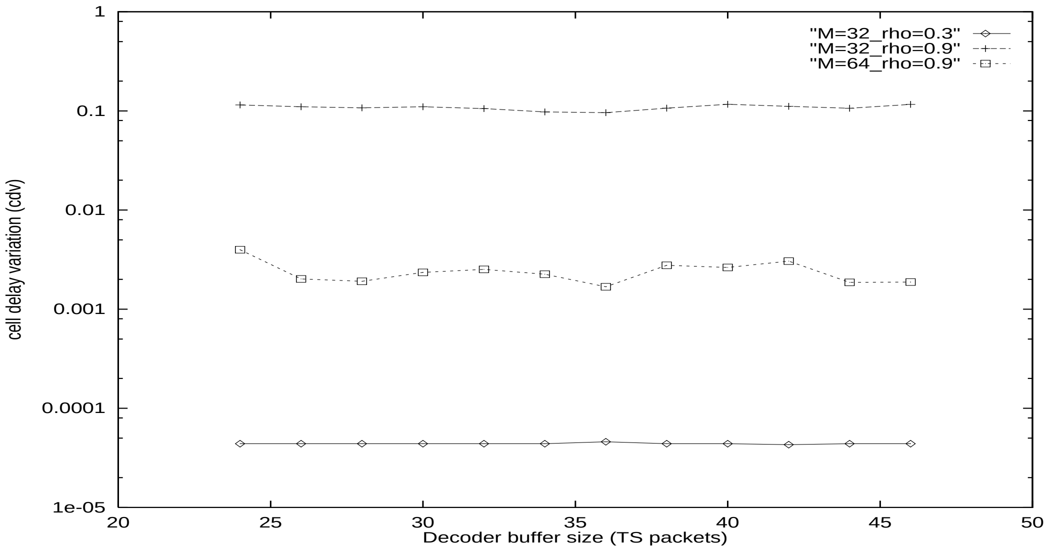


(b)

Figure 8.6: Decoder performance measures versus D for *terminator* video sequence (a) Delay losses (b) Overflow losses.

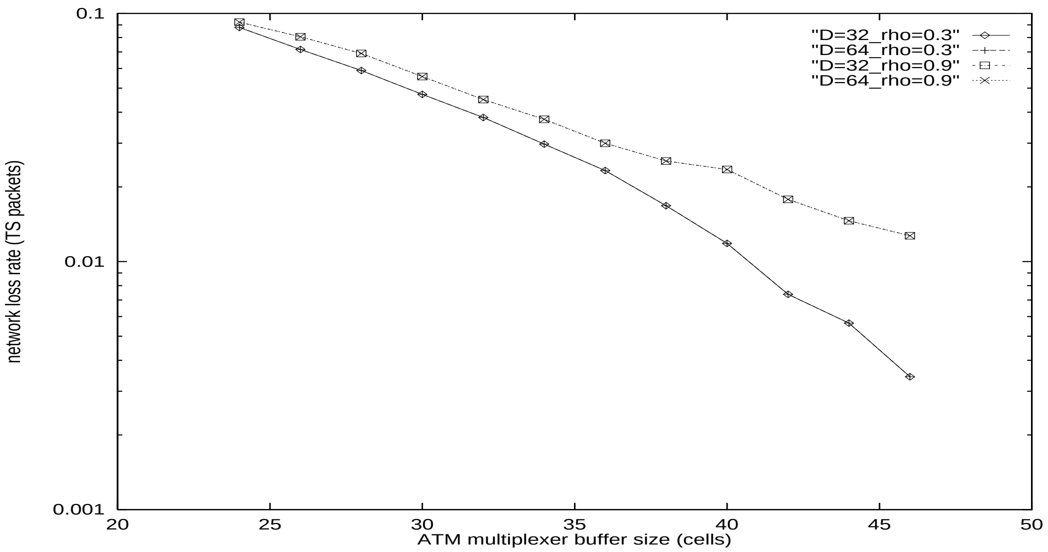


(a)

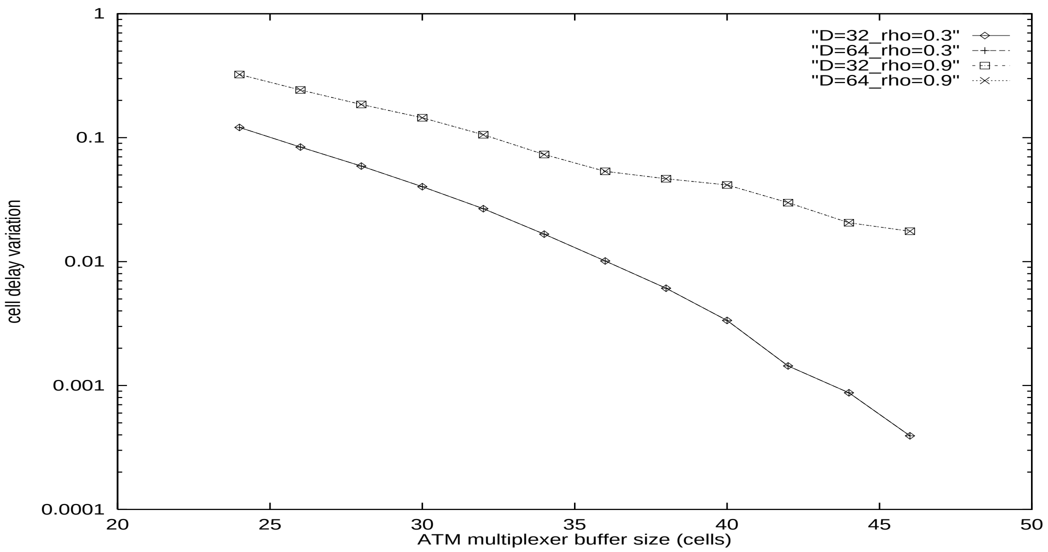


(b)

Figure 8.7: Multiplexer performance measures versus D for *terminator* video sequence (a) Network losses (b) cdv.



(a)



(b)

Figure 8.8: Multiplexer performance measures versus M for *terminator* video sequence (a) Network losses (b) cdv.

traffic process increases, the delay loss rate, overflow loss rate, network losses and *cdv* (corresponding to the VBR video traffic under study) increase. This means that as the background network traffic becomes more variable (corresponding to high ρ), it has destructive effect on the VBR video stream under study. This is the another manifestation of the fact that the variability in the arrival process degrades the queueing performance. Since CIPP has the monotonic non-linear empirical relationship between ρ and the estimated Hurst parameter H as given in 4.5.3, we conclude that the above observation as the effect of high degree of self-similarity in the background arrival process on the VBR video traffic under study.

We noted that the delay loss rate is independent of the decoder buffer size D. The overflow losses decrease as D increases. We also observed that the network losses and the cell delay variation (*cdv*) are independent of D. Network losses decrease as statistical multiplexer size M increases. The cell delay variation decreases as M increases.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

The integrated broadband networks are expected to carry a broad mix of packet traffic ranging from data to voice and video. Understanding the nature of such a traffic, is essential for performance evaluation, design, deployment and real-time control of these networks. Consequently, an important and challenging topic of research in computer networks, is to characterize this heterogeneous packet arrival process by a (parsimonious) stochastic model with minimum number of parameters, which is analytically tractable.

In literature, Poisson, Bernoulli and discrete-batch Markovian arrival process (D-BMAP) models are proposed for cell level traffic and on-off processes, non-renewal Markov chain (MC) and Markov modulated Poisson process (MMPP) for burst level traffic are used. For modeling VBR video traffic, researchers have used auto-regressive (AR) model, the first order discrete auto-regressive (DAR(1)) model and the three class AR model with time-varying coefficients. The above models are essentially, short-range dependent (SRD).

A significant contribution in the area of teletraffic modeling is the discovery of long-range dependence (LRD) in Ethernet LAN traffic in Bellcore [24], metropolitan area network (MAN) traffic [25], wide area network (WAN) traffic and variable bit rate (VBR) video traffic [1, 26, 27]. Moreover, teletraffic researchers have observed that there is a *consistent positive correlation* in the packet interarrivals of Ethernet traces (p. 229 of [32]) and in the cell interarrivals from video codecs [30]. All these led Poisson process obsolete for modeling the broadband teletraffic (despite its analytical simplicity) and over and above, forced researchers to discard the conventional short-range dependent (SRD) models and look for models based on LRD and self-similarity. But, it is to be noted that models based on self-similarity or LRD are analytically *intractable*. Hence, researchers were looking for models, which exhibit self-similarity to a range of time scales on one-end and on the other hand are analytically tractable. But, this pseudo-self-similarity could be brought in by introducing strong correlation in increments. In light of the above, the models which have correlations in increments are of particular interest. The correlation in increments could be brought in a number of ways: (i) by considering correlations in the counts sequence, (ii) by considering correlations in rate process, (iii) by introducing correlations by modulating the underlying arrival process through a Markov chain, or (iv) by considering correlations in the interarrival sequence itself. In literature, researchers have considered models which capture correlation in the counts sequence (packets per unit time) as in discrete autoregressive (DAR(1) model [18]), or in the (traffic) rate process as in autoregressive (AR) model [16] or self-similar model [24, 28], or the correlation introduced by the modulating Markov chain as in Markov modulated models [31].

In light of the above, it is of interest to come up with a counting process whose interarrivals are correlated and develop the corresponding queueing theory. In this thesis, we have defined such a process, derived its distribution and developed the corresponding queueing theory. The above process was applied successfully, to the broadband teletraffic modeling and its corresponding queue to the performance modeling of statistical multiplexer with broadband teletraffic input. We now summarize the contributions made in this thesis.

To start with, we have defined a process, with correlated interarrivals - herein, called Correlated Interarrival time Poisson Process (CIPP)) - and derived its distribution. The CIPP - a stationary counting process - is parameterized by a correlation parameter 'ρ' which represents the degree of correlation in adjacent interarrivals in addition to 'λ' the intensity of the process.

We have proposed the CIPP model for modeling broadband teletraffic.We then undertook an experimental check to justify CIPP as a "good" model for the individual source traffic, namely, the VBR video traffic and the composite Ethernet traffic. Specifically, we carried out the "goodness of fit" test of CIPP to the trace data based on the following second order measures [81]: (1) autocorrelation of counts sequence (2) index of dispersion for intervals (IDI) and (3) index of dispersion for counts (IDC). We have used the MPEG-1 VBR movie sequences and Bellcore Ethernet traffic datasets for our experimental purpose. We observed that for low Hurst parameter MPEG-1 video sequences, both the CIPP and MC models show almost the same approximation quality corresponding to the measure of correlation coefficient. For high Hurst parameter video sequences, the CIPP performs substantially better than the MC model, due to the non-exponential decay of correlation coefficient of the counts sequence. This is in sharp contrast to very fast decay of the correlation coefficient corresponding to the MC model. Although CIPP saturates beyond a certain lag with respect to the measures IDI and IDC, it performs substantially better than that of the Poisson process.

Apart from statistical properties, we also explored other properties of CIPP, which are relevant to modeling broadband traffic. In particular, we investigated burstiness factor [35] and self-similar behavior. A notable property of CIPP which needs a special mention is that, the pre-history of the process extends up to the arrival instant previous to the latest arrival. We believe that this is one of the factors responsible for the self-similar behavior (to a range of time scales). We then examined the widely addressed property in broadband teletraffic literature, namely, the self-similar behavior, corresponding to CIPP. We first studied the "pictorial justification" of self-similar behavior. In order to explore the relation between the positive correlation in interarrivals and the degree of self-similarity, we used the wavelet based estimator for H proposed by Abry *et al* [46] to estimate the Hurst parameter of CIPP of given parameters. Interestingly, this experiment showed that, the degree of self-similarity is independent of λ and depends only on ρ. Accordingly, we arrived at the empirical relationship between ρ, the correlation parameter of CIPP and the estimated Hurst parameter $\hat{H}$. We have presented a variance based method of investigating the degree of self-similarity namely the log(variance) versus log(aggregation level) plot. Finally, we described the wavelet analysis of CIPP process, which explores the log(variance of wavelet coefficient) versus log2(scale) plot. From all of the above experiments, we found that the CIPP exhibits self-similarity over a range of time scales.

The simulation study of autocorrelation function of CIPP at various scales revealed that, as (block size) increases, the autocorrelation tends to approach that of the Poisson case (that is, zero correlation for non-zero lag). We studied the self-similar nature of superposed process of two independent CIPPs, by estimating the Hurst parameter of the superposed process and the component processes. The estimated Hurst parameter of the superposed process is found to lie in between the Hurst parameter of the component processes.

Below, we summarize some findings of the CIPP process:

1. With strong positive correlation in interarrivals (by Equation (4.1)) one gets strong positive correlation in the corresponding counts sequence $u_T(i)$. The magnitude of the correlation depends on ρ.

2. The above strong positive correlation in the interarrival sequence (as given by Equation (4.1)), results in some complex correlation structure for the correspond-

ing counts process $u_T(i)$, rather than exponential decay, despite the fact that the interarrivals form a first order Markov sequence. We believe that this feature is responsible for large scale correlations and subsequently, the self-similar behaviour of CIPP over a range of time scales.

3. The range of time scale of self-similar behavior depends on the parameter ρ of the CIPP process.

4. Since, the interarrivals in CIPP can be expressed as additive random linear combination of random variables, it is easy to generate on the computer.

5. The relationship between the 'ρ' of CIPP and the estimated local Hurst parameter (given in Section 4.5.3), will be very useful in the study of performance evaluation and in real-time traffic management of broadband networks.

6. CIPP is *not* a special case of Neuts process. There is no renewal instant in the CIPP process as in the Neuts process. This is a unique feature of CIPP, which is responsible for its self-similar property to a range of time scales.

7. CIPP is a *parsimonious* model. It is parameterized by λ, which denotes the intensity of the process and ρ, which denotes the degree of interarrival correlation.

After deriving the distribution for CIPP and validating the CIPP model for broadband teletraffic, the next task, we undertook was to consider, queues with CIPP arrivals where service times are exponential (CIPP/M/1 queue). Buffer occupancy distribution and waiting time distribution have been derived. In sequel, we have derived analytical expressions for various queueing performance measures of interest in the context of teletraffic modeling, namely, mean waiting time, mean number of packets in the system, packet delay variation and overflow probability for the CIPP/M/1 queue. The principle contribution here, is to obtain the *analytical* expression relating correlation in interarrivals and the queueing performance measures. This adds to the current efforts of gaining a better understanding of the impact of correlation in interarrivals on the queueing performance.

Below, we summarize some salient features of the CIPP/M/1 queue:

1. The queueing theory results which we have derived are consistent with the Loynes results [88]. Loynes has proved that, for the stability of the infinite capacity queue the ratio of arrival rate to service rate should be less than 1. He assumed more general correlation structure for interarrival and service interval sequences within

themselves (non-zero autocorrelation for non-zero lags) and across them (cross correlation). The correlation structure within the interarrival sequence (or for that matter, correlation within the service intervals and the cross correlation between the interarrival sequence and the service sequence) has nothing to do with the stability of the queue. In CIPP/M/1 queue case, one can see from (5.55) and (5.62) that, ρ has no role in deciding the stability of the queue. It is only the ratio of λ to μ that decides the stability of the queue.

2. The queueing performance measures (corresponding to CIPP/M/1 queue) degrade as the correlation parameter ρ increases. This is consistent with the empirical observation by Livny *et al* [80] and Patuwo *at al* [103]. In our study, the queueing performance measures (for the CIPP/M/1 queue) considered are the mean waiting time, the mean number of packets, the packet delay variation and the overflow probability. We have verified the above theoretical results by simulations.

3. The queueing performance measures for CIPP/M/1 queue have explicit expressions, unlike in the SRD models (for example, MMPP) wherein, the performance measures are often obtained from a recursive equation involving matrices.

4. There is a strong nexus between positive correlation, degree of self-similarity and the queueing performance. CIPP brings out this relationship.

5. By virtue of the analytical tractability, parsimony, and ease with which the queueing performance measures can be computed, CIPP can be used as a full-fledged parsimonious model in a "measurement based system".

In order to take into account the constant switching time of a *real-world* broadband switch, we have considered the CIPP queues with deterministic service intervals (CIPP/D/1 queue). We used the approach of Benes, to obtain the approximate expression for buffer occupancy distribution. The (approximate) analytical expression for the buffer occupancy distribution for CIPP poses a numerical problem for large values of ρ. Hence, we estimated the expressions for the mean delay, by suitable simulations. The simulation results were found to agree well with the theoretical results even though the expression is an approximate one.

Since all realistic systems have finite storage, we considered finite capacity CIPP queues to get more realistic results. We carried out a simulation based study of CIPP/M/1/K queue, since the corresponding analysis, appears to be intractable. The performance measures investigated are packet loss probability and conditional packet loss probability. We observed from simulations that the packet loss probability increases as ρ increases. We also noted that the conditional packet loss probability decays *slowly* with respect to k (denoting the number of *consecutive* packet losses), when the arrivals are CIPP with high ρ. This is in sharp contrast with the M/M/1/K queue case where this probability decays very fast. We demonstrated that the buffer occupancy distribution for CIPP case, peaks at either extremes (empty state and fully occupied state) for high ρ and at high load. This behavior is consistent with that observed in literature for highly correlated arrivals for high loads. From above experiments, we conclude that the correlation in interarrivals degrades the queueing performance in the *finite capacity* queue case *also*.

Having developed the queueing theory for CIPP/M/1 queue, we apply these results for studying the queueing situation as encountered by composite Ethernet traffic in a statistical multiplexer. We undertake the performance modeling of infinite capacity statistical multiplexer with composite traffic input and exponential service using CIPP/M/1 queue. For our experiments, we have used Bellcore Ethernet traffic datasets namely, `pAug.TL` and `pOct.TL`. The exponential service times are assumed in order to use the theoretical results obtained for queueing performance measures corresponding to CIPP/M/1 queue. We estimated the queueing performance measures from simulations of queue with Bellcore Ethernet traffic dataset. To get the theoretical result, we 'fit' the CIPP to the data with autocorrelation function as a measure. Then, the theoretical performance measures are computed corresponding to this 'best fit' CIPP. Our experimental results show that the CIPP/M/1 queue models well the queueing situation as encountered by the the Ethernet traffic in a statistical multiplexer with respect to the queueing performance measures mentioned above.

We used the queueing performance measures, namely, packet loss probability and conditional packet loss probability for performance modeling of finite capacity statistical multiplexer with composite traffic input and exponential service using CIPP/M/1/K queue. We then compared the buffer occupancy distribution of the finite capacity queue with Bellcore Ethernet traffic and the "best fit" CIPP/M/1/K queue with K arbitrary. Experimental results show that the CIPP/M/1/K queue makes a good approximation to the performance of finite buffer statistical multiplexer with real-world Ethernet traffic.

A new process, called Markov modulated CIPP (MMCIPP) has been defined to overcome a limitation of CIPP. This process possesses correlation in interarrivals in addition to correlations introduced due to the modulating Markov chain. It has been shown by various methods that this process is self-similar over a wider scale.

Next, we undertook the performance modeling of statistical multiplexer with VBR video traffic input using CIPP/M/1 queue. Here again, we assumed exponential services, to make use of the analytical results. Queueing performance measures considered here are mean waiting time, mean number of packets in the system, packet delay variation (*pdv*) and overflow probability. We have used a variation of deterministic smoothing of video traces used in [39] to get packets of variable sizes in the smoothed version.

We have considered *soccer, starwars, atp, lambs* and *MrBean*, MPEG-1 encoded VBR video sequences [2] for our simulation purposes. To get the results corresponding to data, the smoothed VBR video trace is fed to a single server queue with exponential service. The above mentioned queueing performance measures are estimated corresponding to data. To get the *theoretical* performance results, we fit the CIPP to video trace data by autocorrelation matching and compute the queueing performance measures, corresponding to this "best" fit CIPP. Simulation results show that the CIPP/M/1 queue models very well, the queueing situation encountered by VBR video traffic in broadband networks.

We undertook the performance modeling of more realistic finite buffered statistical multiplexer with *smoothed* VBR video traffic input using CIPP/M/1/K queue. Queueing performance measures considered here are, packet loss probability and conditional packet loss probability. The buffer occupancy distribution in a queue with Bellcore Ethernet traffic is compared with that of its "best fit" CIPP/M/1/K. The conclusions from the above experiments are two fold: (i) the CIPP/M/1/K queue models quite well the queueing situation as encountered by VBR traffic in broadband networks with respect to all the measures mentioned above, (ii) the correlation in interarrivals degrades the queueing performance in the finite capacity queue case *also*.

Finally, we considered the fluid-flow approximation of queues with CIPP arrivals to study the behavior of statistical multiplexer with VBR video traffic input. We have taken a frame interval of the *real world* MPEG-1 VBR video sequence to be our unit time slot, over which we performed our simulations. Comparisons with other models like, Markov chain, histogram model are also undertaken. The simulation results showed that for low buffer sizes, CIPP outperforms the MC and the histogram models. For larger buffer sizes, CIPP marginally overestimates the cell loss probability. This is desirable since, for network dimensioning purpose, it is more convenient to use a model which marginally overestimates the traffic.

In order to study the interference effects on MPEG video traffic in ATM networks by the background network traffic, we simulated the end-to-end transport of *real world*

MPEG-2 video traffic over ATM networks using TS packets. The background network traffic was modeled by CIPP model.

We found through experiments that as the variability of background traffic process increases, the delay loss rate (probability of loss of TS packets due to delay at the decoder buffer), overflow loss rate (probability of loss of TS packets due to decoder overflow), network losses (probability of loss of (equivalent) TS packets due to cell losses in the ATM multiplexer) and *cdv* (corresponding to the VBR video traffic under study) increase. This means that as the background network traffic becomes more variable (corresponding to high ρ), it has destructive effect on the VBR video stream under study. This is the another manifestation of the fact that the variability in the arrival process degrades the queueing performance. Since CIPP has the monotonic non-linear empirical relationship between ρ and the estimated Hurst parameter ($\hat{H}$) (as given in Section 4.5.3), we conclude that the above observation as the effect of high degree of self-similarity in the background arrival process.

We also repeated the above experiment to explore the effect of ATM multiplexer size M and the decoder buffer size D on the above mentioned performance measures. We noted that the delay loss rate is independent of D. The overflow losses decrease as D increases. We also observed that the network losses and cell delay variation (*cdv*) are independent of D. Network losses decrease as M increases. The cell delay variation decreases as M increases.

Limitations of CIPP Model:

1. Both the forms of CIPP (4.3) and (4.4) pose a numerical problem for large values of n and ρ due to the appearance of large alternating, very nearly canceling terms in the sum. This poses a very serious limitation on the use of (4.3) and (4.4). For example, one has to use a precision of up to 40 decimal places in order to computate $P_{10}(t)$ in (4.3). But, the form (4.3) is comparatively better than (4.4), as the form (4.3) requires less number of decimal places of precision needed to compute $P_{10}(t)$.

2. Although the phase plot of CIPP exhibits diagonal lines (as in the phase plot of Bellcore Ethernet traffic), the points below the line of slope ρ are absent. This means that this feature of points appearing everywhere in the phase plot of real world traffic, is poorly captured by CIPP.

To conclude: while we agree with the fact that the CIPP is not a purely a self-similar process (or LRD process) and broadband teletraffic can be better modeled by fGn and the buffer occupancy process in an M/G/∞ queue, the main thrust of this work is the demonstration of self-similarity of CIPP over a range of time scale of practical interest and the *theoretical* performance modeling of broadband teletraffic. In other words, what we have shown is that, while CIPP is a reasonable approximation for LRD processes, like other models MHOP [84] and ON-OFF processes (but with added advantages like simplicity and analytical tractability), the multiplexer behaviour can be studied analytically as well as experimentally and the two results closely approximate the real-life behavior. On the contrary, if one considers fGn or buffer occupancy process of $M/G/\infty$ as an input to the multiplexer, then the study of the system becomes complex. To this extent, our model is superior and a good way to get around with fGn.

9.2 Future Work

As mentioned in the thesis, self-similar nature of Ethernet traffic has been discovered recently. Hence, the modeling and analysis of traffic process in broadband networks by self-similar processes (which are analytically intractable) is relatively new. In this context, we have come up with a process (namely, the CIPP) which is self-similar over a range of time scales of interest and still is analytically tractable. On the theoretical side, we have made a significant contribution of generalizing the Poisson process and developing the corresponding queueing theory. This has a significant implications in queueing theory, where there is an explicit expression relationship between the degree of correlation in interarrivals and the corresponding queueing performance. Thus, this model can be used in any practical situation (where there is an evidence of correlations in interarrival times). On the engineering side, validating the CIPP as a 'good' model for broadband teletraffic and our demonstration that it is self-similar over a range of time scales, opens up a new window in analyzing the real world traffic by an analytically tractable model. Moreover, the existence of corresponding queueing theory (for CIPP/M/1 queue) allows one to *theoretically* predict the performance of such broadband network systems. We would like to emphasize that capturing the correlation in the interarrival sequence (of the *real world data* in the context of broadband teletraffic modeling) is a new concept and is the main contribution of this thesis. Moreover, CIPP succinctly captures the relationship between the strong positive (interarrival) correlation, the degree of self-similarity and the queueing performance measures. Hence, some of the important problems in traffic modeling and analysis of ATM networks can be given a fresh look, using this new model to see the effect of interarrival correlation. We indicate a few of them here.

- **Priority Queues With CIPP arrivals:** ATM cells have an explicit cell loss priority bit in the header so that two different ATM priority classes can be distinguished. In many situations, a single ATM connection can comprise both priority classes. For example, when the information to be transmitted is classified by the user into more and less important cells as in layered video coding or in the policing function (where the cells which are violating the contract on which the admission of the corresponding connection is based, will be marked as low priority cells). In Kroner *et al* [117], the authors consider space priority mechanisms with Poisson arrivals (more specifically, the M/G/1/N system). They obtain the loss probabilities for both the high and low priority traffic classes. It would be interesting to study the above priority queueing situations for CIPP/M/1/N or CIPP/G/1/N system. The above study would bring out the effect of variability in the arrival process (captured by ρ of the corresponding CIPP) of high (or low) priority cells on the loss rates of low and high priority cells.

- **Effective Bandwidth:** In [118], the authors show that, when traffic sources share a buffer with deterministic service rate, a constraint on the tail of the buffer occupancy distribution is a linear constraint on the number of sources. That is, for a small loss probability one can assume that each source transmits at a fixed rate called its effective bandwidth. They also show the existence of effective bandwidth for more general source models under some assumptions. When traffic parameters of the sources are known, effective bandwidth can be calculated and may be used to obtain a circuit-switched style call acceptance and routing algorithm for ATM networks. The important feature of the effective bandwidth of a source is that it is a characteristic of that source and the acceptable loss probability only. Thus the effective bandwidth of a source does not depend on the number of sources sharing the buffer or the model parameters of other types of sources sharing the buffer. It will be interesting to see how the effective bandwidth for Poisson sources generalizes to CIPP.

- **Connection Admission Control (CAC):** Connection admission control is defined as the set of actions taken by the network at the call set up phase (or during call renegotiation phase) in order to decide whether a connection can be accepted. A connection can only be accepted if sufficient network resources are available to establish the connection end to end at its required QoS without affecting the agreed QoS of already existing connections in the network. Two classes of parameters are necessary for CAC, namely, a set of parameters specifying the traffic source characteristics and a set of parameters identifying the required QoS. In literature, various traffic source models are used to evaluate the performance of CAC. It would be interesting to see how a given CAC mechanism performs when the variability of the source traffic increases. Here, one can use the CIPP model for the source traffic to explore the effect of the variability in the individual source traffic on the performance of any proposed CAC.

- **Exploring the multiplexing gain versus variability in the traffic using CIPP model:** In an ATM network, a necessary number of cells is allocated on demand, which allows the network to take advantage of the bit rate variations of individual sources by *statistical multiplexing*. In this way, individual sources share a link of capacity less than the sum of individual peak rates, achieving significant *multiplexing gain*. This multiplexing gain is restricted by the QoS requirements of the sources. Now, the number of sources that can be accommodated, depends on their BW, QoS requirements and how precarious or how variable the individual traffic is. Hence, it would be interesting to how the multiplexing gain varies when the traffic is CIPP.

- **Concept of Correlated Markov Chain (CMC):** We know that for a discrete state stochastic process to be a Markov chain, the holding times should be independent and exponentially distributed. We define a Correlated Markov chain (CMC), as a discrete state stochastic process, in which, the sequence of holding times are exponentially distributed and form a AR(1) process as given by (4.1). Note that, in a CMC, the holding times form a *stationary* sequence. Observe that, this discrete time stochastic process (CMC) is *not* a *pure* Markov chain, since it looses the memoryless property, despite its holding times being exponential. We also define a modification of CIPP, herein called CMCCIPP as a counting process in which the underlying CIPP process is modulated by CMC. Now, consider the following problem of modeling the bursty arrivals in broadband teletraffic: burst arrivals are usually modeled in literature by ON-OFF process, with *independent* and exponential holding times. CMCCIPP can be used to model the correlated arrival of bursts. Here, not only the arrivals in a burst are correlated, but also the *sojourn intervals of burst* are also correlated here.

- **An Analogous PASTA property for CIPP:** It would be interesting to see how PASTA (Poisson Arrivals See Time Average) property extends to CIPP.

Based on the advantages described in this thesis (including queueing theory), we believe that CIPP is a better model for broadband teletraffic and it will attract the attention of teletraffic researchers to this extent, in the years to come.

Appendix A

Construction of CIPP process

For stationarity in the counting variable, the initial point of the interval of the counting must be chosen in a particular probabilistic way[1]. These are the stationary initial conditions which we will give; they can be used to construct a count stationary point process. By doing so, we are exploiting Axiom 3.

We next use Axiom 2. Since our aim is to construct a process whose interarrivals are correlated, we first generate correlated exponential sequence $\{X_0, X_2, X_3, \ldots, \}$ as given[2] in [69] where X_0 is the length biased distribution with respect to exponential distribution with parameter λ and $X_1 = UX_0$ where again U is uniformly distributed in $[0,1]$. Now form the partial sum,

$$T_n = X_1 + X_2 + \ldots + X_n, \quad n = 1, 2, 3 \ldots \tag{A.1}$$

Note that, the sequence $\{T_n\}$ gives the arrival instants. Using

$$X_{n+1} = \rho X_n + \epsilon_{n+1} \qquad 0 \leq \rho < 1 \quad n = 1, 2, 3 \ldots \tag{A.2}$$

and the relation $X_1 = UX_0$ and X_1 we get,

$$T_n = (U + \alpha_{n-1} - 1)X_0 + \alpha_{n-2}\epsilon_2 + \ldots + \alpha_0 \epsilon_n \tag{A.3}$$

where, X_0 is the length biased distribution with respect to the exponential distribution with parameter λ and $\alpha_j = \sum_{i=0}^{j} \rho^i$ for $0 \leq \rho < 1$, and $j = 1, 2, 3, \ldots$. We use $\Phi_W(s)$ to denote the moment generating function of the random variable W. We note from [69],

$$\Phi_X(s) = \rho + \frac{(1-\rho)\lambda}{\lambda + s} \tag{A.4}$$

$$\Phi_\epsilon(s) = \rho + \frac{(1-\rho)\lambda}{\lambda + s} \tag{A.5}$$

Define, $Y = (U + \alpha_{n-1} - 1)X_0$, then,

$$\Phi_Y(s) = \frac{\lambda^2}{(\lambda + s(\alpha_{n-1} - 1))(\lambda + s\alpha_{n-1})} \tag{A.6}$$

[1] see for instance [78, 119, 120]

[2] Note the absence of X_1 in the sequence.

Then exploiting independence among random variables,

$$\Phi_{T_n}(s) = \frac{\lambda}{(\lambda + s(\alpha_{n-1} - 1))(\lambda + s\alpha_{n-1})} \prod_{i=2}^{n} \Phi_\epsilon(\alpha_{n-i}s) \tag{A.7}$$

Substituting this, and upon some manipulation, we get,

$$\Phi_{T_n}(s) = \frac{\lambda}{(\lambda + s\rho\alpha_{n-2})(\lambda + s\alpha_{n-1})} \prod_{i=0}^{n-2} \Phi_\epsilon(\alpha_i s) \tag{A.8}$$

and upon further reducing , we have,

$$\Phi_{T_n}(s) = \frac{(\lambda)^2 \prod_{i=0}^{n-3}(\lambda + s\rho\alpha_i)}{\prod_{i=0}^{n-1}(\lambda + s\alpha_i)} \tag{A.9}$$

Rewriting the above,

$$\Phi_{T_n}(s) = \frac{\lambda^2 \rho^{n-2} \prod_{k=0}^{n-3}(s + \frac{\lambda}{\rho\alpha_k})}{\alpha_{n-2}\alpha_{n-1} \prod_{k=0}^{n-1}(s + \frac{\lambda}{\alpha_k})} \tag{A.10}$$

Inverting the above expression to get the density function $f_{T_n}(t) \triangleq \mathcal{L}^{-1}(\Phi_{T_n}(s))$ and $F_{T_n}(t) \triangleq \int_0^t f_{T_n}(\tau)d\tau$ we have for distribution $F_{T_n}(t)$,

$$F_{T_n}(t) = \left\{ \begin{array}{ll} 1 - \sum_{j=0}^{n-1} \frac{\alpha_j \prod_{k=0}^{n-1}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \rho\alpha_{n-2})(\alpha_j - \rho\alpha_{n-1}) \prod_{k=0,k\neq j}^{n-1}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} & 0 \leq \rho < 1 \end{array} \right. \tag{A.11}$$

Next we derive the distribution (see page 133 [121])

$$\begin{aligned} P_n(t) &= F_{T_n}(t) - F_{T_{n+1}}(t) \\ &= \left\{ \begin{array}{ll} \sum_{j=0}^{n} \frac{\alpha_j \prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \rho\alpha_{n-1})(\alpha_j - \rho\alpha_n) \prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} & \\ - \sum_{j=0}^{n-1} \frac{\alpha_j \prod_{k=0}^{n-1}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \rho\alpha_{n-2})(\alpha_j - \rho\alpha_{n-1}) \prod_{k=0,k\neq j}^{n-1}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} & 0 \leq \rho < 1 \end{array} \right. \end{aligned}$$

After some manipulations, we get[3] the required result,

$$P_n(t) = \left\{ \begin{array}{ll} (1 + \rho - (1 - \rho)\alpha_n) & \\ \sum_{j=0}^{n} \frac{\alpha_j \prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\rho(\alpha_j + 1) - \alpha_n + 1)(\alpha_j - \alpha_n + 1)(\alpha_j - \rho\alpha_n) \prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} & 0 \leq \rho < 1 \end{array} \right. \tag{A.12}$$

for $\quad n = 0, 1, 2, 3 \ldots$, $\lambda > 0$, $\ t \geq 0$, $\quad$ and $\quad \alpha_j = \sum_{i=0}^{j} \rho^i$

[3]It is worth noting that the value of $P_n(t)$ at $\rho = 0$ is obtained by the limiting value $\lim_{\rho \to 0} P_n(t)$(which is the Poisson process)

Appendix B

Construction of Interval Stationary CIPP process

For stationarity in the intervals, the initial point of the interval of the counting starts at origin. A counting process constructed in this way is called interval stationary in the sense of Lawrance. The counting variable may not have stationary increments in this case. Hence it may not satisfy Axiom 3 in Section 4.1.2 and the form given by A.12 is of more practical interest. Nevertheless, we present the interval stationary CIPP[1] here for academic reasons.

Again, we follow more or less the same steps as we did for construction of CIPP. We first generate correlated exponential sequence $\{X_1, X_2, X_3, \ldots,\}$ as given in [69]. Now form the partial sum,

$$T_n = X_1 + X_2 + \ldots + X_n \quad n = 1, 2, 3 \ldots \tag{B.1}$$

Note that, the sequence $\{T_n\}$ gives the arrival instants. Substituting,

$$X_{n+1} = \rho X_n + \epsilon_{n+1} \qquad 0 \leq \rho < 1 \quad n = 1, 2, 3 \ldots \tag{B.2}$$

and eliminating all X_i's, except X_1 we get from (B.1),

$$T_n = \alpha_{n-1} X_1 + \alpha_{n-2}\epsilon_2 + \ldots + \alpha_0 \epsilon_n \tag{B.3}$$

where, X_1 is exponentially distributed with parameter λ, $\alpha_j = \sum_{i=0}^{j} \rho^i$ for $0 \leq \rho < 1$, and $j = 1, 2, 3, \ldots$. We use $\Phi_W(s)$ to denote the moment generating function of the random variable W.

[1]Strictly speaking this process has nothing to do with CIPP. It is just that instead of choosing the origin at an *arbitrary instant* inorder to construct CIPP, if we choose it at an *arbitrary event*, we land up with this process.

Then exploiting independence among random variables, we have for moment generating funtion of the random variable T_n,

$$\Phi_{T_n}(s) = \frac{\lambda}{(\lambda + s\rho\alpha_{n-1})} \prod_{i=0}^{n-1} \Phi_\epsilon(\alpha_i s) \tag{B.4}$$

where,

$$\Phi_\epsilon(s) = \rho + \frac{(1-\rho)\lambda}{\lambda + s}$$

See for details [69].

Inverting the above expression to get the density function $f_{T_n}(t) \triangleq \mathcal{L}^{-1}(\Phi_{T_n}(s))$ and $F_{T_n}(t) \triangleq \int_0^t f_{T_n}(\tau)d\tau$ we have for distribution $F_{T_n}(t)$,

$$F_{T_n}(t) = 1 - \sum_{j=0}^{n-1} \frac{\prod_{k=0}^{n-1}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \rho\alpha_{n-1}) \prod_{k=0,k\neq j}^{n-1}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \quad 0 \leq \rho < 1 \tag{B.5}$$

Next we derive the distribution, (please see page 133 [121])

$$\begin{aligned} P_n(t) &= F_{T_n}(t) - F_{T_{n+1}}(t) \\ &= \begin{cases} \sum_{j=0}^{n} \frac{\prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \rho\alpha_n)\prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \\ -\sum_{j=0}^{n-1} \frac{\prod_{k=0}^{n-1}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \rho\alpha_{n-1})\prod_{k=0,k\neq j}^{n-1}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \quad 0 \leq \rho < 1 \end{cases} \end{aligned}$$

After some manipulations, we get[2] the required result.

$$P_n(t) = \sum_{j=0}^{n} \frac{\prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\alpha_j - \alpha_n + 1)(\alpha_j - \rho\alpha_n) \prod_{k=0,k\neq j}^{n}[\alpha_j - \alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} \quad 0 \leq \rho < 1 \tag{B.6}$$

for $\quad n = 0, 1, 2, 3 \ldots\ , \ \lambda > 0, \ \ t \geq 0, \quad$ and $\quad \alpha_j = \sum_{i=0}^{j} \rho^i$

[2]It is worth noting that the value of $P_n(t)$ at $\rho = 0$ is obtained by the limiting value $\lim_{\rho\to 0} P_n(t)$(which is the Poisson process)

Appendix C

Derivation of IDI for CIPP

Consider the partial sum of n interarrival times starting from $i+1$,

$$M_n = X_{i+1} + X_{i+2} + \ldots + X_{i+n} \quad n = 1, 2, 3 \ldots \tag{C.1}$$

Note that, the sequence $\{X_i\}$ forms stationary sequence for CIPP. Using the correlation structure of interarrival sequence,

$$X_{n+1} = \rho X_n + \epsilon_{n+1} \qquad 0 \leq \rho < 1 \quad n = 1, 2, 3 \ldots \tag{C.2}$$

in (C.1) and eliminating all X_{i+j} for $2 \leq j \leq n$, except X_{i+1} we get,

$$M_n = X_{i+1} \left(\frac{1-\rho^n}{1-\rho} \right) + \sum_{j=1}^{n} \epsilon_{n_j} \left(\frac{1-\rho^n}{1-\rho} \right) \tag{C.3}$$

We use $\Phi_W(s)$ to denote the moment generating function of the random variable W. After exploiting independence among random variables ϵ_n's, we have for the moment generating funtion of the random variable M_n,

$$\Phi_{M_n}(s) = \Phi_{X_n} \left[s \left(\frac{1-\rho^n}{1-\rho} \right) \right] \prod_{j=1}^{n} \Phi_\epsilon \left[s \left(\frac{1-\rho^{n-j}}{1-\rho} \right) \right] \tag{C.4}$$

Now the variance of M_n can be obtained from its moment generating function as given by the relation,

$$Var\{M_n\} = \Phi''_{M_n}(0) - \left(\frac{n}{\lambda} \right)^2 \tag{C.5}$$

It is straight forward to obtain the following expression for variance of W_n by using (C.5), (C.4), (A.4) and (A.5).

$$\begin{aligned} Var\{M_n\} \quad = \quad & \tfrac{1}{\lambda^2} \Big[\left(\tfrac{1-\rho^n}{1-\rho} \right)^2 + \left(\tfrac{1+\rho}{1-\rho} \right) \times \\ & \Big\{ n - 1 - \left(\tfrac{2\rho(1-\rho^{n-1})}{1-\rho} \right) + \left(\tfrac{\rho^2(1-\rho^{2(n-1)})}{1-\rho^2} \right) \Big\} \Big] \end{aligned} \tag{C.6}$$

Upon further reducing, we have finally,

$$Var\{M_n\} = \frac{n(1-\rho^2) + 2\rho^{n+1} - 2\rho}{\lambda^2 \, (1-\rho)^2} \tag{C.7}$$

The expression for IDI corresponding to CIPP follows directly.

Appendix D

Derivation of Correlation Coefficient of Interarrivals for CIPP

Owing to the relation (A.2), X_{i+k} can be written as,

$$X_{i+k} = \rho^k X_i + \sum_{j=1}^{k} \rho^{j-1} \epsilon_{i+k-j+1} \quad i \geq 0 \tag{D.1}$$

Now, let us consider the correlation $\mathbf{E}\{X_i X_{i+k}\}$. Using D.1, we have,

$$\begin{aligned}
\mathbf{E}\{X_i X_{i+k}\} &= \mathbf{E}\left\{X_i \left(\rho^k X_i + \sum_{j=1}^{k} \rho^{j-1} \epsilon_{i+k-j+1}\right)\right\} && \text{(D.2)} \\
&= \rho^k \mathbf{E}\{X_i^2\} + \sum_{j=1}^{k} \rho^{j-1} \mathbf{E}\{X_i\} \mathbf{E}\{\epsilon_{i+k-j+1}\} \\
&= \frac{2\rho^k}{\lambda^2} + \sum_{j=1}^{k} \rho^{j-1} \left(\frac{1-\rho}{\lambda^2}\right) && \text{(D.3)} \\
&= \frac{1+\rho^k}{\lambda^2} && \text{(D.4)}
\end{aligned}$$

Plugging into the expression for correlation coefficient for interarrival intervals, we get the required result.

Appendix E

Derivation of Equations in Section 5.2

E.1 Derivation of Equation (5.20)

First, let us prove that $P^*_{i0}(s;n+1) = \sum_{m=0}^{\infty} C^*_m(s,n)$. We have the LHS,

$$P^*_{i0}(s;n+1) = \int_0^{\infty} \sum_{l=0}^{\infty} \sum_{m=l+1}^{\infty} k^*_m(s,x) d_x(P^{(n)}_{il}(x))$$

Changing the index of inner summation,

$$\begin{aligned} P^*_{i0}(s;n+1) &= \int_0^{\infty} \sum_{l=0}^{\infty} \sum_{h=0}^{\infty} k^*_{l+h+1}(s,x) d_x(P^{(n)}_{il}(x)) \\ &= \sum_{h=0}^{\infty} \int_0^{\infty} \sum_{l=0}^{\infty} k^*_{l+h+1}(s,x) d_x(P^{(n)}_{il}(x)) \end{aligned}$$

after changing the order of summations. Using the definition of $C^*_h(s,n)$, we have,

$$P^*_{i0}(s;n+1) = \sum_{h=0}^{\infty} C^*_h(s,n) \tag{E.1}$$

which is what we wanted to show. In order to show, (5.20), we start with

$$P^*_i(s,z;n+1) = P^*_{i0}(s;n+1) + \sum_{j=1}^{\infty} P^*_{ij}(s;n+1) z^j$$

Using the definition of $P^*_{ij}(s;n+1)$, and then changing the order of integration and summation, we have,

$$\begin{aligned} P^*_i(s,z;n+1) &= P^*_{i0}(s;n+1) + \sum_{j=1}^{\infty} \int_0^{\infty} \sum_{l=0}^{\infty} k^*_l(s,x) d_x(P^{(n)}_{i,j+l-1}(x)) z^j \\ &= P^*_{i0}(s;n+1) + \int_0^{\infty} \sum_{l=0}^{\infty} k^*_l(s,x) \sum_{j=1}^{\infty} d_x(P^{(n)}_{i,j+l-1}(x)) z^j \\ &= P^*_{i0}(s;n+1) + \int_0^{\infty} \sum_{l=0}^{\infty} k^*_l(s,x) \sum_{m=l}^{\infty} d_x(P^{(n)}_{i,m}(x)) z^{m-l+1} \end{aligned}$$

after changing the running variable in the inner-most summation. Splitting z, we have,

$$\begin{aligned} P_i^*(s,z;n+1) &= P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x)\sum_{m=l}^{\infty} d_x(P_{i,m}^{(n)}(x))z^m \\ &= P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\sum_{m=0}^{\infty} z^m \sum_{l=0}^{m} z^{-l}k_l^*(s,x)d_x(P_{i,m}^{(n)}(x))z^m \end{aligned}$$

after changing the order of summations. Writing the inner-most partial sum, in terms of infinite summation, we have,

$$\begin{aligned} P_i^*(s,z;n+1) &= \begin{cases} P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x)) \\ \left[\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x) - \sum_{l=m+1}^{\infty} z^{-l}k_l^*(s,x)\right] \end{cases} \\ &= \begin{cases} P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x) \\ -z\int_{x=0}^{\infty}\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\sum_{l=m+1}^{\infty} z^{-l}k_l^*(s,x) \end{cases} \\ &= \begin{cases} P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x) \\ -\int_{x=0}^{\infty}\sum_{m=0}^{\infty} d_x(P_{i,m}^{(n)}(x))\sum_{l=m+1}^{\infty} z^{m-l+1}k_l^*(s,x) \end{cases} \end{aligned}$$

Now, changing the running variable in the innermost summation, and then interchanging the order of integration and summation in the third term, we have,

$$\begin{aligned} P_i^*(s,z;n+1) &= \begin{cases} P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right)\left(\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x)\right) \\ -\int_{x=0}^{\infty}\sum_{m=0}^{\infty} d_x(P_{i,m}^{(n)}(x))\sum_{h=0}^{\infty} z^{-h}k_{m+h+1}^*(s,x) \end{cases} \\ &= \begin{cases} P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right)\left(\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x)\right) \\ -\sum_{h=0}^{\infty}\left[\int_{x=0}^{\infty}\sum_{m=0}^{\infty} d_x(P_{i,m}^{(n)}(x))k_{m+h+1}^*(s,x)\right]z^{-h} \end{cases} \end{aligned}$$

Using the definition of $C_h^*(s;n)$,

$$P_i^*(s,z;n+1) = \begin{cases} P_{i0}^*(s;n+1) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right)\left(\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x)\right) \\ -\sum_{h=0}^{\infty} C_h^*(s;n)z^{-h} \end{cases}$$

$$P_i^*(s,z;n+1) = \begin{cases} \sum_{h=0}^{\infty}(1-z^{-h})C_h^*(s;n) \\ +z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right)\left(\sum_{l=0}^{\infty} z^{-l}k_l^*(s,x)\right) \end{cases}$$

where, we have made use of (E.1). Again using the definition of $k_l^*(s,x)$, we have,

$$P_i^*(s,z;n+1) = \begin{cases} \sum_{h=0}^{\infty}(1-z^{-h})C_h^*(s;n) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right) \\ \left(\sum_{l=0}^{\infty} z^{-l}\int_{t=0}^{\infty}\frac{exp(-(s+\mu)t)(\mu t)^l}{l!}f(t/x)dt\right) \end{cases}$$

$$= \begin{cases} \sum_{h=0}^{\infty}(1-z^{-h})C_h^*(s;n) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right) \\ \int_{t=0}^{\infty} exp(-(s+\mu)t)f(t/x)\left(\sum_{l=0}^{\infty}\frac{(\mu z^{-1}t)^l}{l!}dt\right) \end{cases}$$

$$= \begin{cases} \sum_{h=0}^{\infty}(1-z^{-h})C_h^*(s;n) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right) \\ (\int_{t=0}^{\infty} exp(-(s+\mu-\mu z^{-1})t)f(t/x)dt) \end{cases}$$

$$= \begin{cases} \sum_{h=0}^{\infty}(1-z^{-h})C_h^*(s;n) + z\int_{x=0}^{\infty}\left(\sum_{m=0}^{\infty} z^m d_x(P_{i,m}^{(n)}(x))\right) \\ \left(\rho + \frac{(1-\rho)\lambda}{\lambda+\mu+s-\mu z^{-1}}\right) exp\left(-\rho x(s+\mu-\mu z^{-1})\right) \end{cases}$$

where use has been made of (5.6). Using (5.10) and (5.11) we get the result (5.20).

E.2 Derivation of Equation (5.23)

Taking the generating functions with respect to n we have, from (5.20),

$$\begin{aligned} \sum_{n=0}^{\infty} P_i^*(s,z;n+1)w^n &= \sum_{n=0}^{\infty} zB(s,z)P_i(A(s,z),z;n)w^n + \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}(1-z^{-m})C_m^*(s;n)w^n \\ \frac{1}{w}\sum_{n=1}^{\infty} P_i^*(s,z;n)w^n &= \sum_{n=0}^{\infty} zB(s,z)P_i^*(A(s,z),z;n)w^n + C(s,z,w) \end{aligned}$$

Multiplying throughout by w, and adding $P_i^*(s,z,0)$ on either side, we have,

$$\sum_{n=0}^{\infty} P_i^*(s,z;n)w^n = wzB(s,z)\sum_{n=0}^{\infty} P_i^*(A(s,z),z;n)w^n + wC(s,z,w) + P_i^*(s,z;0);$$

$$P_i^*(s,z,w) = wzB(s,z)\sum_{n=0}^{\infty} P_i^*(A(s,z),z;n)w^n + wC(s,z,w) + \sum_{j=0}^{\infty} P_{ij}^*(s;0)z^j$$

using the definition of $P_i^*(s,z;0)$. Again, using the definition of $P_{ij}^*(s;0)$, we have,

$$P_i^*(s,z,w) = wzB(s,z)\sum_{n=0}^{\infty} P_i^*(A(s,z),z;n)w^n + wC(s,z,w) + \sum_{j=0}^{\infty} z^j\int_{t=0}^{\infty} e^{-st}d_t(P_{ij}^{(0)}(t))$$

$$P_i^*(s,z,w) = \begin{cases} wzB(s,z)\sum_{n=0}^{\infty} P_i^*(A(s,z),z;n)w^n + wC(s,z,w) \\ +\sum_{j=0}^{\infty} z^j\int_{t=0}^{\infty} e^{-st}d_t(Pr\{Q_0=j, T_0\le t/Q_0=i\}) \end{cases}$$

$$P_i^*(s,z,w) = \begin{cases} wzB(s,z)\sum_{n=0}^{\infty} P_i^*(A(s,z),z;n)w^n + wC(s,z,w) \\ +\sum_{j=0}^{\infty} z^j\int_{t=0}^{\infty} e^{-st}d_t(\delta(i-j)U(t)) \end{cases}$$

$$P_i^*(s,z,w) = wzB(s,z)\sum_{n=0}^{\infty} P_i^*(A(s,z),z;n)w^n + wC(s,z,w) + z^i$$

E.3 Derivation of Equation (5.44)

Taking z tranform of (5.43), we have,

$$\sum_{n=0}^{\infty} P^*_{0j}(s; n+1)w^n = \left\{ \begin{array}{l} \sum_{l=0}^{\infty} \frac{(-\lambda)^l}{l!} \frac{\partial^l}{\partial y^l} \left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)} \right]^r \right. \\ \left. P^*_{0,j+l-1} \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) \right]_{y=\lambda} \quad (j \geq 1) \end{array} \right.$$

$$\sum_{k=1}^{\infty} P^*_{0j}(s; k)w^k = \left\{ \begin{array}{l} w \sum_{l=0}^{\infty} \frac{(-\lambda)^l}{l!} \frac{\partial^l}{\partial y^l} \left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)} \right]^r \right. \\ \left. P^*_{0,j+l-1} \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) \right]_{y=\lambda} \quad (j \geq 1) \end{array} \right.$$

$$P^*_{0j}(s, w) = \left\{ \begin{array}{l} w \sum_{l=0}^{\infty} \frac{(-\lambda)^l}{l!} \frac{\partial^l}{\partial y^l} \left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)} \right]^r \right. \\ \left. P^*_{0,j+l-1} \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) \right]_{y=\lambda} + P^*_{0j}(s; 0) \quad (j \geq 1) \end{array} \right.$$

By substituting $P^*_{0j}(s, w)$ from (5.42) in (5.43) for $i = 0$, we have for $j > 0$,

$$T^{j-1}(w)(T(w)I_0(s, w) + I_1(s, w)) = \left\{ \begin{array}{l} w \sum_{l=0}^{\infty} \frac{(-\lambda)^l}{l!} \frac{\partial^l}{\partial y^l} \left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)} \right]^r \right. \\ \left[T^{j+l-1}(w) I_0 \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) + \right. \\ \left. \left. T^{j+l-2}(w) I_1 \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) \right] \right]_{y=\lambda} \\ +P^*_{0j}(s; 0) \quad (j \geq 1) \end{array} \right.$$

$$T(w)(T(w)I_0(s, w) + I_1(s, w)) = \left\{ \begin{array}{l} w \sum_{l=0}^{\infty} \frac{(-\lambda T(w))^l}{l!} \frac{\partial^l}{\partial y^l} \left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)} \right]^r \right. \\ \left[T(w) I_0 \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) + \right. \\ \left. \left. I_1 \left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}, w \right) \right] \right]_{y=\lambda} \\ +P^*_{0j}(s; 0) \quad (j \geq 1) \end{array} \right.$$

Using Taylor's series theorem, and noting that $P^*_{0j}(s; 0) = 0$ for $j > 0$ (for this result, please refer the derivation of (5.20)), we obtain the result.

E.4 Derivation of Equation (5.48)

We wish to evaluate

$$V_n(s) = \prod_{i=0}^{n-1} (T(w)/w)[B(g^i, T(w)^{-1})]^{-1} \tag{E.2}$$

at $s = 0$ and $w = 1$, where g^i satisfies $g^0 = s$ & $g^{i+1} = A(g^i, T^{-1}(w))$ and $T(w)$ is as in (5.35). Working out the details, one gets for g^i,

$$g^i = \left\{ \begin{array}{ll} s & i = 0 \\ \rho^i s + \mu\rho(1 - T(w)) \frac{(1-\rho^i)}{1-\rho} & i > 0 \end{array} \right. \tag{E.3}$$

But we want,

$$V_n(s)|_{s=0,w=1} \quad = \quad \frac{T(1)}{\left(\rho + \frac{\lambda(1-\rho)}{\lambda+\mu(1-T(1))}\right)} \prod_{i=1}^{n-1} \frac{T(1)}{\left(\rho + \frac{\lambda(1-\rho)}{\lambda+(1-T(1))\mu\frac{(1-\rho^{i+1})}{(1-\rho)}}\right)} \tag{E.4}$$

where $T(1) = \frac{1-\rho+\frac{\mu}{\lambda}\rho}{\frac{\mu}{\lambda}}$ and $\xi = \frac{\rho\mu(1-z^{-1})}{1-\rho}$. After some manipulation we get the result. It is straight forward to see that the term inside the product term, goes to 1 as $i \to \infty$ and that ensures convergence of $V_n(s)$

E.5 Derivation of Equation (5.49)

We use the method of proof by induction. We assume that,

$$L(s_n/t_0) \quad = \quad \frac{\lambda + \rho^n s_n}{\lambda + s_n} e^{-\rho^n t_0 s_n}$$

In the following we prove that,

$$L(s_{n+1}/t_0) \quad = \quad \frac{\lambda + \rho^{n+1} s_{n+1}}{\lambda + s_{n+1}} e^{-\rho^{n+1} t_0 s_{n+1}}$$

We start with,

$$\begin{aligned} L(s_{n+1}/t_0) \quad &= \quad \int_{t_n \geq 0} L(s_{n+1}/t_n) f(t_n/t_0) dt_n \\ &= \quad \left\{ \begin{array}{l} \int_{t_n \geq 0} \frac{\lambda+\rho s_{n+1}}{\lambda+s_{n+1}} e^{-\rho t_0 s_{n+1}} \\ \left[\rho^n \delta(t_n - \rho^n t_0) + (1-\rho^n)\lambda e^{-\lambda(t_n-\rho^n t_0)} U(t_n - \rho^n t_0)\right] dt_n \end{array} \right. \end{aligned} \tag{E.5}$$

where we get the expression for $f(t_n/t_0)$ by taking the inverse Laplace transform of assumed expression for $L(s_n/t_0)$.

After evaluating the integral we get the result. Hence, our assumption is true for any n. Putting $t_0 = 0$ we get (5.49).

E.6 Derivation of Equation (5.52)

From (5.45) and (5.47), we get,

$$T(w)I_0(s,w) + I_1(s,w) = \frac{D}{M(s,w)} \tag{E.6}$$

Eliminating D from (5.40) and (E.6),

$$D = T(w)I_0(a,w)M(a,w)$$

. Eliminating $I_0(a,w)$ from (5.40) and (5.50) after putting $s = a$ in (5.50), we get,

$$I_0(a,w) = (1 - T(w))\sum_{n=0}^{\infty}\frac{\lambda + \rho^n a}{\lambda + a}w^n$$

Putting back the value of D in (E.6) and using the above value of $I_0(a,w)$, we have,

$$T(w)I_0(s,w) + I_1(s,w) = \frac{M(a,w)T(w)(1 - T(w))}{M(s,w)}\sum_{n=0}^{\infty}\frac{\lambda + \rho^n a}{\lambda + a}w^n \qquad \text{(E.7)}$$

Eliminating $I_0(s,w)$ and $I_1(s,w)$, from (E.7) and (5.50), and substituting them in (5.42), we get the result.

Appendix F

Some Important Theorems

F.1 Read's Theorem

Theorem 1 *Let ζ be a double point of the function $g(z)$ such that $g(z)$ is regular at ζ and $0 < \left|g^{'}(\zeta)\right| < I$; and let $w = \beta$ be a solution of the Equation,*

$$f(\zeta, w) = w \tag{F.1}$$

such that (i) $f(z, w)$ is regular in a domain of the place (ζ, β), and (ii)$\left|\frac{\partial f}{\partial w}\right| < \frac{I}{\left|g^{'}(\zeta)\right|}$ when $z = \zeta,\ w = \beta$. Then the functional Equation,

$$f[z, \phi(g(z))] = \phi(z) \tag{F.2}$$

has just one solution $\phi(z)$ which is regular at z and takes the value β there. This solution can be given as $\lim_{n\to\infty} \phi_n(z)$, where the functions $\phi_n(z)$ are defined by the recurrence relation,

$$\phi_{n+1}(z) = f[z, \phi_n(g(z))] \tag{F.3}$$

and $\phi_1(z)$ is an arbitrary function regular at ζ and taking the value β there.

F.2 Abel's Theorem

If $\lim_{n\to\infty} a_n = a$, then,

$$\lim_{x\to 1}(1-x)\sum_{n=0}^{\infty} a_n\, x^n \;=\; a \tag{F.4}$$

F.3 Rouche's Theorem

If $f(z)$ and $g(z)$ are regular inside and on a closed contour C, and $|g(z)| < |f(z)|$ on C, then $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside C.

F.4 Liouville's Theorem

Let a function $f(z)$ be analytic throughout the complex plane, and let its modulus be uniformly bounded. Then $f(z)$ is identically equal to a constant.

F.5 A theorem of Kuczma

In this Section, we give the results of Kuczma. For more details reader is referred to [101].

We consider the problem of solutions to the following homogeneous functional Equation,

$$\phi[f(x)] = g(x)\psi(x) \tag{F.5}$$

We assume that the following conditions hold:
Hypothesis 1.1 $f \in \mathbf{R}^0_\psi[I], \quad \psi \in I$
Hypothesis 1.2 $g \in \mathbf{\Phi}[I], g(x)$ is continuous in I and $g(x) \not= 0$ for $x \in I$, $x \not= \psi$.

F.5.1 The Main Theorem

Theorem 1 *Let the hypothesis D.1 and D.2 be fulfilled.*

In case (i) Equation F.5 has one parameter family of continuous solutions in I: for every $\eta \in \Xi$ there exists exactly one continuous function $\varphi \in \Phi[I]$ satisfying Equation F.5 and fulfillign the condition

$$\varphi(\xi) = \eta \tag{F.6}$$

This solution is given by

$$\varphi(x) = \eta/G(x) \tag{F.7}$$

In case (ii) Equation (F.5) has in the class $\Phi[I]$ a continuous function depending on an arbitrary function, In this case every continuous solution $\varphi(x)(x)$ of Equation (F.5) in I fulfills the condition

$$lim_{x\to\xi}\varphi(x) = 0 \tag{F.8}$$

In case (iii) the funtion $\varphi(x) \equiv 0$ is only continuos solution of Equation (F.5) in the function class $\Phi[I]$.

Appendix G

The Derivation of *pdv* for CIPP

Noting that, the density function for stationary waiting time (W) (corresponding to 5.62),

$$f_W(x) = (1 - T\ \beta)\delta(x) + \mu\beta T(1-T)\left(e^{-\mu(1-T)x}\right) \tag{G.1}$$

where,

$$\beta \triangleq \frac{M(a,1)}{M(0,1)}\frac{\lambda}{\lambda + a}. \tag{G.2}$$

The density of $Y \triangleq W + S$, is given by the convolution integral,

$$f_Y(y) = \int_{u=0}^{\infty} f_W(y-u)\ f_S(u)\ du. \tag{G.3}$$

We have for $f_Y(y)$,

$$\begin{aligned} f_Y(y) &= \int_{v=0}^{\infty} \left[(1-\beta T)\delta(y-v) + \mu\beta T(1-T)e^{-\mu(1-T)(y-v)}\right] U(y-v)\mu e^{-\mu v}U(v)\ dv \\ &= \mu(1-\beta T)\int_{v=0}^{y} \left[e^{-\mu v}\delta(y-v)\right]\ dv + \mu^2\beta T(1-T)\int_{v=0}^{y} e^{-\mu(1-T)(y-v)}e^{-\mu v}\ dv \end{aligned}$$

Evaluation of the integral and further reduction yields,

$$f_Y(y) = \mu\left[(1-\beta)e^{-\mu y} + \beta(1-T)e^{-\mu(1-T)y}\right] \tag{G.4}$$

for $0 \leq y$. The *pdv* (τ) is given by the variance of the delay (Y), i.e.,

$$\begin{aligned} \tau &= I\mathbf{E}\{(Y-\bar{Y})^2\} \\ &= \int_{-\bar{Y}}^{\infty} y^2\ f_Y(y+\bar{Y})\ \ dy \end{aligned} \tag{G.5}$$

Upon changing the variable and using (G.4), after some manipulation we have,

$$\tau = \frac{2}{\mu^2}\left[1 - \beta + \frac{\beta}{(1-T)^2}\right] - \left[\frac{1}{\mu}\left(1 + \frac{\beta T}{(1-T)}\right)\right]^2 \tag{G.6}$$

Bibliography

[1] J. Roberts, U. Mocci and J. Virtamo, *"Broadband Network Teletraffic: Performance Evaluation and Design of Broadband Multiservice Networks"*, Springer-Verlag, 1996.

[2] *"ftp-info3.informatik.uni-wuerzburg.de /pub/mpeg/"*, 1995.

[3] M. Kuczsura, "The Interrupted Poisson Process as an overflow process", *Bell Sys. Tech. J.*, vol. 52, no. 3, pp. 437–448, Mar. 1973.

[4] S. Q. Li and J. W. Mark, "Traffic characterisation of integrated services network", *IEEE Trans. commun.*, vol. 3, pp. 1231–1243, Aug. 1990.

[5] H. Heffes, "A class of data traffic processes: Covariance function characterisation and related queueing results", *Bell Syst. Tech. J.*, vol. 59, pp. 897–929, July/Aug 1980.

[6] N. Abramson, "ALOHA", *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 119–123, Mar. 1985.

[7] H. Michiel and K. Laevens, "Teletraffic Engineering in a Broadband Era", *Proc. of the IEEE*, vol. 85, no. 12, pp. 2007–2033, 1997.

[8] M. F. Neuts, "A versatile Markovian point process", *J. Appl. Prob.*, vol. 16, pp. 764–779, 1979.

[9] V. Ramaswami, "The N/G/1 queue and its detailed analysis", *Adv. Appl. Prob.*, vol. 12, pp. 222–261, 1980.

[10] M. F. Neuts, *"Matrix-Geometric Solutions in Stochastic Models "*, Baltimore, John-Hopkins Univ. Press., 1981.

[11] M. F. Neuts, *"Structured Stochastic Matrices of M/G/1 Type and Their Applications "*, Marcel and Dekkar, NewYork, NY, 1989.

[12] H. Heffes and D. M. Lucantoni, "A Markov modulated characterisation of packetised voice and data traffic and related statistical multiplexer performance", *IEEE Jr. Select. Areas in Commun.*, vol. SAC-4, pp. 856–868, Sept. 1986.

[13] D. Anick, D. Mitra and M. M. Sondhi, "Stochastic theory of a data-handling system", *Bell Sys. Tech. J.*, vol. 61, no. 8, pp. 1871–1894, Oct 1982.

[14] J. N. Daigle and J. D. Langford, "Models for analysis of packet voice communication systems", *IEEE J. Select. Areas. Commun.*, vol. SAC-4, pp. 847–855, Sept. 1986.

[15] T. Stern, "A queueing analysis for packetised voice", *Proc. GLOBECOM'83*, pp. 2.5.1–2.5.6, Dec 1983.

[16] B. Maglaris, B. Anastassiou, P. Sen, G. Karlsson and J. D. Robbins, "Performance models for statistical switching in packet video communications", *IEEE Trans. commun.*, vol. 16, pp. 834–843, July 1988.

[17] C. Blondia and O. Casals, "Statistical multiplexing of VBR sources: A matrix analytical approach", *Performance evaluation*, vol. 16, pp. 5–20, Nov. 1992.

[18] D. P. Heyman, A. Tabatabai and T.V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks", *IEEE Trans. Circuits. Sys. Video Tech.*, vol. 2, no. 1, pp. 49–59, Mar. 1992.

[19] P. A. Jacob and P. A. W. Lewis, "Time series generated by mixtures", *J. Time Series Analysis*, vol. 4, no. 1, pp. 19–36, 1983.

[20] R. G. Addie and M. Zukerman, "An approximation for performance evaluation of stationary single server queues", *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3150–3160, Dec. 1994.

[21] S. Q. Li, "A general solution technique for discrete queueing analysis of multimedia traffic on ATM", *IEEE Trans. commun.*, vol. 39, no. 7, pp. 1115–1132, July 1991.

[22] S. Q. Li and C-L. Hwang, "Queue Response to Input Correlation Functions: Discrete Spectral Analysis", *IEEE/ACM Trans. Networking*, vol. 1, no. 5, pp. 522–533, Dec. 1993.

[23] S. Q. Li and C-L. Hwang, "Queue Response to Input Correlation Functions: Continuous Spectral Analysis", *IEEE/ACM Trans. Networking*, vol. 1, no. 6, pp. 678–692, Dec. 1993.

[24] W. E. Leland, M. S. Taqque, W. Willinger and D.V. Wilson, "On the self-similar nature of ethernet traffic (extended version)", *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.

[25] R. G. Addie, M. Zukerman and T. Neame, "Fractal traffic: Measurements, modeling and performance evaluation", *Proc. IEEE INFOCOM'95.*, Apr. 1995.

[26] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Variable bit rate video traffic and long range dependence", *IEEE Trans. Commun.*, 1993.

[27] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of slf-similar VBR video traffic", *preprint*, 1994.

[28] B. Tsybakov and N. D. Georganas, "On Self-similar Traffic in ATM Queues: Definitions, Overflow Probability Bound and Cell Delay Distribution ", *IEEE/ACM Trns. Networking*, vol. 5, no. 3, pp. 397–409, Oct. 1997.

[29] A. Erramili, O. Narayan and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic", *IEEE/ACM Trans. Networking*, vol. 4, no. 2, pp. 129–164, April 1996.

[30] S. K. Chan and A. Leon-Garcia, "Analysis of cell inter-arrival from VBR video codecs", *Proc. INFOCOM94*, pp. 350–357, 1994.

[31] S. V. Frost and E. Melamed, "Traffic modeling in telecommunication networks", *IEEE Comm. Mag.*, pp. 70–81, Mar. 1994.

[32] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling", *IEEE/ACM Trans. Networking*, vol. 3, no. 3, pp. 226–243, June 1995.

[33] S. Xu, Z. Huang and Y. Yao, "A theoretic analysis model for VBR video traffic in ATM networks", *International Conference on Communications, 1997 (ICC'97)*, vol. 2, pp. 844–848, 1997.

[34] R. Manivasakan, U. B. Desai and A. Karandikar, "Broadband Teletraffic Characterization using Correlated Interarrival time Poisson Process (CIPP)", *J. Indian Inst. Sci.,*, vol. 79, March-April 1999.

[35] H. Saito, M. Kawarasaki and H. Yamada, "An analysis of statistical multiplexing in an ATM transport network", *IEEE Jr. Select. Areas. Commun.*, vol. 9, no. 3, pp. 359–367, April 1991.

[36] R. Manivasakan, U. B. Desai and A. Karandikar, "CIPP/M/1 queue", Tech. Rep., SPANN lab., IIT, Bombay, May. 1998, Technical Report no: spann.mani.1998.2.

[37] R. Manivasakan, U. B. Desai and A. Karandikar, "CIPP/M/1 queue", *Proceedings of the 5th National Conference on Communications (NCC'99)*, pp. 458–461, January, 29-31, IIT, Kharagpur, India 1999.

[38] A. A. Lazar, G. Pacific and J. S. White, "Real-time traffic measurements on MAGNET II", *IEEE J. Select. Areas. Commun.*, vol. 8, pp. 467–483, 1990.

[39] P. Skelly, M. Schwartz and S. Dixit, "A histogram based model for video traffic behaviour in an ATM multiplexer", *IEEE/ACM Trans. Networking.*, vol. 1, no. 4, pp. 446–459, Aug. 1989.

[40] R. Manivasakan, U. B. Desai and A. Karandikar, "CIPP based characterisation of VBR traffic and related statistical multiplexer performance", *Proc. GLOBECOM'99*, December, 5-9 1999.

[41] U. B. Desai R. Manivasakan and Abhay Karandikar, "The CIPP and the Performance Modeling of ATM Multiplexer with VBR Video Traffic Input by CIPP/M/1/K Queue", *ADCOM'99*, December 1999.

[42] R. Manivasakan, U. B. Desai and A. Karandikar, "End-to-end simulation of VBR traffic over ATM networks using CIPP network traffic model", *Proceedings of 3rd Int. Conf. on Computational Intelligence and Mulitimedia Applications,(ICCIMA'99)*, pp. 338–342, Sept. 23-26, New Delhi 1999.

[43] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications", *SIAM Review*, vol. 10, no. 4, pp. 422–437, Oct. 1968.

[44] H. E. Hurst, R. P. Black and Y. M. Sinaika, *"Long Term Storage in Reservoirs. An An Experimental Study "*, Constable, London, 1965.

[45] G. W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes", *Proceedings of the IEEE*, vol. 105, pp. 15–30, 1995.

[46] P. Abry and D. Veitch, "Wavelet analysis of long range dependent traffic", *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 1–15, Jan. 1998.

[47] L. Kleinrock, *"Queueing Systems Vol.1: Theory 1"*, Marcel and Dekkar, 1975.

[48] A. A. Markov, "Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain", *The Notes of the Imperial Academy of Sciences of St. Petersburg VIII Series*, , no. 9, Dec. 1907.

[49] J. Norros and J. T. Virtamo, "Who loses cells in case of burst scale conjestion", *In 13th Int. Teletraffic Congress proceedings*, 1991.

[50] E. Eberlein and M. S. Taqqu, "Dependence in Probability and Statistics ", *Progress in Prob. and Stat.*, vol. 11, Birkhauser, Bosten, 1986.

[51] G. Samorodnitsky and M. Taqqu, *"Stable No-Gaussian Random Processes"*, Chapman & Hall, 1994.

[52] H-O. Peitgen and D. Saupe, *"Science of fractal images"*, Springer-Verlag, New York, 1988.

[53] C. K. Chui, *"An Introduction to Wavelets"*, Academic Press Inc., USA, 1992.

[54] A. Antoniadis and G. Oppenheim, *"Wavelets and Statistics "*, Springer Verlag, NY, 1995.

[55] J. Beran, *"Statistics for long-memory processes"*, Chapman and Hall, New York., 1994.

[56] M. G. Riley and I. E. G. Richardson, *"Digital Video Communications"*, Springer-Verlag, 1996.

[57] S. S. Huang, "Source modeling for packet video", *Proc. IEEE. Infocom.*, pp. 1262–1267, 1988.

[58] S. S. Huang, "Modeling and analysis for packet video", *Proc. IEEE. GLOBE-COM'89*, pp. 1262–1267, Dallas, TX 1989.

[59] K. Joseph, D. Raychaudhuri and J. Zdepski, "Shared access packet transmission system for compressed digital video.", *IEEE J. Select. Areas. Commun.*, vol. 7, no. 815-825, June 1989.

[60] "ccitt study group xviii ivs baseline document ", 1992.

[61] C. L. Hwang and S. Q. Li, "On the convergence of traffic measurement and queueing analysis: A statistical-match queueing (SMAQ) tool", *Proc. IEEE INFOCOM'95*, pp. 602–612, Boston, MA 1995.

[62] P. Sen, B. Maglaris, N. E. Rikli and D. Anatassiou, "Models for packet switching of variable bit rate video sources", *IEEE Jr. Select. Areas. Commun*, vol. 7, no. 3, pp. 865–869, 1989.

[63] G. E. P. Box and G. M. Jenkins, *"Time Series Analysis: Forecasting and Control"*, Holden Day, San Francisco, CA, 1976.

[64] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing ", *J. Time Series Anal.*, vol. 1, pp. 15–29, 1980.

[65] J. R. M. Hosking, "Fractional Differencing ", *Biometrika*, vol. 68, pp. 165–176, 1981.

[66] Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single server queues with applications", *preprint*, 1993.

[67] R. G. Addie, M. Zukerman and T. Neame, "Broadband Traffic Modeling: Simple Solutions to Hard Problems", *IEEE Communications Magazine*, vol. 36, no. 8, pp. 88–95, Aug 1998.

[68] S. Keshav, *"An Engineering Approach to Computer Networking: ATM Networks, the Internet and the Telephone Network"*, Addison-Wesley, 1997.

[69] D. P. Gaver and P. A. W. Lewis, "First-order autoregressive gamma sequences and point processes", *Adv. Appl. Prob.*, vol. 12, pp. 727–745, 1980.

[70] A. J. Lawrance and P. A. W. Lewis, "An exponential moving average sequence and point process ema1 process", *J. Appl. Prob.*, vol. 14, pp. 98–113, 1977.

[71] P. D. Finch, "The single server queueing system with non-recurrent input-process and Erlang service time", *J. Australian Math. Soc.*, vol. 3, pp. 220–236, 1963.

[72] R. Manivasakan, U. B. Desai and A. Karandikar, "Correlated Interarrival time Poisson Process (CIPP) for modeling broadband teletraffic", *Proceedings of the Fifth Biennial Conference on Signal Processing and Communications (SPCOM'99), IISc, Bangalore*, pp. 43–50, July 1999.

[73] P. C. Consul, *"Generalised Poisson Distribution"*, Marcel and Dekkar, Inc., NY, 1989.

[74] P. C. Consul, "A generalisation of Poisson distribution", *Technometrics*, vol. 15, no. 4, Nov. 1973.

[75] Drezner, "Correlated Poisson Process ", *Commun. Stat.*, vol. B17, pp. 129–164, 1994.

[76] H. Wold, "Sur les processus stationnaires ponctuels", *Colloques Internat.*, vol. CRNS 13, pp. 75–86, 1965.

[77] D. R. Cox, "Some statistical methods connected with series of events", *J.R.Statist. Soc.*, vol. B 17, pp. 129–164, 1955.

[78] D. J. Daley and D. Vere-Jones, *"An Introduction to the Theory of Point Processes"*, Springer-Verlag, 1988.

[79] J. A. McFadden, "On the lengths in a stationary point process", *J.R. Statist. Soc.*, vol. B-24, pp. 364–382, 1993.

[80] M. Livny, B. Melamed and A. K. Tsiolis, "The impact of autocorrelation function on queueing systems", *Management Science*, vol. 39, no. 3, pp. 322–339, Mar. 1993.

[81] R. Gusella, "Characterising the variability of arrival processes with indexes of dispersion", *IEEE Jr. Select. Areas. Commun.,*, vol. 9, no. 2, pp. 203–211, Feb. 1991.

[82] "*http://bellcore.com*", 1992.

[83] S. K. Srinivasan, *"Stochastic point processes and their applications"*, Charles Griffin, London, 1974.

[84] V. Misra and W. B. Gong, "Markovian hierarchical on-off processes for teletraffic modeling", Tech. Rep., Dept. of ECE, University of Massachusetts, Amherst, 1998, Technical Report no:TRCCS9801.

[85] I. Daubecheis, "Orthonormal bases of compactly supported wvaelets", *Communications of Pure and Applied Mathematics*, vol. 41, pp. 909–996, 1988.

[86] P. D. Finch and C. Pearce, "A second look at a queueing system with moving average input process", *J. Australian Math. Soc.*, vol. 5, pp. 100–106, 1965.

[87] P. Tin, "A queueing system with Markov-dependent arrivals", *J. Appl. Prob.*, vol. 22, pp. 668–677, 1985.

[88] R. M. Loynes, "The stability of a queue with non-independent inter-arrival and service times", *Proc. Camb. Philos. Soc.*, vol. 58, pp. 497–520, 1962.

[89] D. V. Lindley, "The theory of queues with a single server", *Proc. Camb. Philos. Soc.*, vol. 48, pp. 227–289, 1952.

[90] W. L. Smith, "On the distribution of queueing times", *Proc. Camb. Philos. Soc.*, vol. 49, pp. 449–461, 1953.

[91] F. Spitzer, "The Wiener-Hopf equation whose kernel is a probability density", *Duke. Math. J.*, vol. 24, pp. 327–347, 1957.

[92] R. M. Loynes, "Stationary waiting time distribution times for single server queues", *Ann, Math. Statist.*, vol. 23, pp. 1323–1339, 1962.

[93] J. Th. Runnenburg, *"On the use of Markov Processes in Server Waiting time problems in Renewal Theory*, PhD thesis, Poortpers, Amsterdam, 1960.

[94] J. Th. Runnenburg, "An example illustrating the possibilities of renewal theory and waiting time theory for markov dependent arrival intervals", *Proc. Ser. A. Kon. Nederl. Akad. Weten.*, 1961.

[95] P. Franken, D. Konig, U. Arndt and V. Schmidt, *"Queues and Point Processes"*, Akademie-Verlag, Berlin, 1982.

[96] A. A. Borovkov, *"Random Processes in Queueing Theory"*, Springer-Verlag, Berlin, 1976.

[97] R. Szekli, R. L. Disney and S. Hur, "A MR/GI/1 queues with positively correlated arrival stream ", *J. Appl. Prob.*, vol. 31, pp. 497–514, 1994.

[98] C. Pearce, "A queueing system with general moving average input and negative exponential service time", *J. Austral. Math. Soc.*, vol. 6, pp. 223–236, June 1966.

[99] A. H. Read, "The solution of a functional equation", *Proc. R. Soc. Edinburgh*, vol. A 63, pp. 336–365, 1952.

[100] R. M. Phatarford, "A queueing system with general moving average input and negative exponential service time", *J. Austral. Math. Soc.*, vol. 6, pp. 223–236, June 1971.

[101] M. Kuczma, "Functional equations in a single variable", *Polska Akademia Nauk, Monografia Matematyczma*, 1968.

[102] V. E. Benes, *"General Stochastic Processes in the Theory of Queues"*, Addison-Wesley, 1963.

[103] B. E. Patuwo and R. L. Disney, "Effect of correlated arrivals on queues", *IEE Trans.*, vol. 25, no. 3, 1993.

[104] V. Ramaswamy and G. Latouche, "Traffic performance modeling for packet communication whence, where and whither, (keynote address), 3rd Australian Teletraffic Seminar", Nov. 1998.

[105] C. Blondia and O. Casals, "Cell loss probabilities in a statistical multiplexer in an ATM network", *Proc. 6th GI/ITG-Fachtagung, Munchen*, vol. 286, pp. 121–136, 1991.

[106] M. D. Prycker, *"Asynchronous Transfer Mode: Solution for Broadband ISDN"*, Prentice-Hall, 1995.

[107] U. Sumita M. F. Neuts and Y. Takahashi, "Renewal characterization of Markov modulated Poisson processes", *J. Appl. Math. Simulation 2 (1989) 53-70*, vol. 2, pp. 53–70, 1989.

[108] W. Fischer and K. Meier-Hellstern, "The Markov modulated Poisson process (MMPP) cookbook ", *Performance Evaluation*, vol. 18, pp. 149–171, 1992.

[109] D. A. Lucantoni, M. F. Neuts and A. R. Reibman, "Methods for performance evaluation of vbr video traffic models", *IEEE/ACM Trans. Networking*, vol. 2, pp. 176–180, 1994.

[110] F. Yegenoglu, B. Jabbari and Y. Q. Zhang, "Motion classified autoregressive modeling and of variable bit rate video", *IEEE Trans. Circuits Syst. Video Tech.*, vol. 3, pp. 42–53, Feb. 1993.

[111] ISO/IEC JTC 1/SC 29/WG 11, *"Generic coding of moving pictures and associate audio information, Part 1: Systems, Part 2: Video, Part 3: Audio"*, CD 138 18, 1995.

[112] A. R. Reibman and B. G. Haskell, "Constraints on variable bit rate video encoders", *IEEE Trans. Circuits Syst. Video Tech.,*, vol. 2, 1992.

[113] K. Joseph and D. Reininger, "Source traffic smoothing for vbr video encoders.", *Proc. 6th Int. Workshop on Packet Video*, 1994.

[114] C. Hsu and A. Ortega, *"Joint encoder and VBR channel optimisation with buffer and leaky bucket constraints in multimedia communications and video coding."*, Plenum, 1996.

[115] B. G. Haskell and A. R. Reibman, "Encoder/decoder buffer control for variable bit rate channel", *U. S. Patent 5 565 924*, Oct 1996.

[116] W. Zhu,Y. T. You,Y. Wang and Ya-Qin Zhang, "End-to-end modeling and simulation of MPEG-2 transport streams over ATM networks with jitter", *IEEE Trans. Circuits Syst. Video Tech.*, vol. 8, no. 1, pp. 9–12, Feb. 1998.

[117] H. Kroner, G. Hebuterne, P. Boyer and A. Gravey, "Priority Management in ATM Switching Nodes", *IEEE J. Select. Areas. Commun.*, vol. 9, no. 3, pp. 10–18, April 1996.

[118] J. Walrand G. Kesidis and C-S. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources", *IEEE/ACM Trans. Networking*, vol. 1, no. 4, pp. 424–427, Aug. 1993.

[119] D. J. Daley, "Stationary point processes with Markov dependent intervals and infinite intensity", *J. Appl. Probab.*, vol. 19A, pp. 313–320, 1982.

[120] D. R. Cox and V. Isham, *"Point Processes"*, Chapman and Hall, London, 1980.

[121] E. Parzen, *"Stochastic Processes"*, Holyden-Day, San Francisco, 1962.