CIPP - A Versatile Analytical Model for VBR Traffic in ATM Networks

R. Manivasakan, U. B. Desai and Abhay Karandikar
Department of Electrical Engineering, Indian Institute of Technology,
Powai, Mumbai - 400 076. INDIA.

Abstract

Correlated Interarrival time Poisson Process (CIPP) has been proposed in [1], for modeling both the composite arrival process of packets in broadband networks and the individual video source modeling. The CIPP - a generalization of the Poisson process - is a stationary counting process and is parameterized by correlation parameter 'ρ', the degree of correlation in adjacent interarrivals and 'λ', the intensity of the process. In this paper, we develop the theory for CIPP/M/1 queue and undertake the performance modeling of statistical multiplexer with VBR video traffic in broadband networks, using the CIPP/M/1 queue. We first derive the expressions for stationary distributions for queue length and waiting time in a CIPP/M/1 queue. Then, we derive the queueing performance measures of interest. For reasons of feasibility of theoretical performance modeling and realistic compulsions, we propose a deterministic smoothing with random (geometrically distributed) packet sizes. We simulate a queue with thus *smoothed* VBR video trace data as input to compare with the theoretical performance measures derived above. Experimental results show that the CIPP/M/1 queue, models well the statistical multiplexer performance with the *real world* MPEG-1 VBR video traffic input.

1 Introduction

Digital video communications (like video phone, video conferencing, television distribution etc.) are the major class of services provided by broadband networks. Accordingly, the video transport over broadband (in particular, ATM) networks is of significant interest. Most of the video encoding are done using the MPEG standards. The output of the encoder in MPEG systems can be either Constant Bit Rate (CBR) or Variable Bit Rate (VBR). A constant quality video signal is delivered at the receiver by VBR, but at the cost of complex transmission. The ATM network takes advantage of the bit rate variations of individual sources by *statistical multiplexing*. In this way, individual (video) 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 Quality of Service (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 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 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 or 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. We intend here to use the fact that the interarrival correlation plays a major role in *imitating the self-similar behavior* in increments of the corresponding counting process and to make a contribution towards *this* key aspect of video traffic modeling.

Two classes of traffic models have been proposed in literature, for modeling the bit rate of a VBR video source, namely, Short Range Dependent (SRD) models [2, 3, 4, 5] and Long Range Dependent (LRD) (or self-similar) models [6, 7]. SRD models addressed in literature, in turn, broadly falls into two categories: single source models [2, 3, 4, 8] and multiplexed source models [2, 5]. For low motion video telephony and medium-motion video conferencing, there exist the autoregressive (AR) model [2, 4], the discrete autoregressive (DAR(1)) model [3] and the discrete state continuous time Markov model [2, 4]. For full motion MPEG video source, a three class AR model with time-varying coefficients has been developed in [8]. Xu *et al* [9] proposed a GAR(1) model, for VBR traffic. In [10], a model based on arrival rate histograms of VBR traffic has been proposed and a quasi-static approximation made for buffer occupancy distributions and cell loss rates. Most of the SRD models proposed for VBR traffic are analytically *intractable* [3, 5, 8, 10]. Moreover, those SRD models which are analytically tractable [2, 4] predict the queueing performance measures over optimistically. Note that LRD (or self-similar) models lack analytical tractability, except that few bounds are obtained for some queueing measures [11, 12].

The above models are intended to capture the correlation in the counts sequence (cells per frame) by a DAR(1) model as in [3] or in the rate process as in ARMA models [2, 3] or by the modulating Markov chain as in SRD models. In models based on self-similarity (or LRD) or AR models, correlations are in rate process (equivalently *cells per unit time*). In [1], we proposed a model called Correlated Interarrival time Poisson Process (CIPP) for composite traffic modeling and individual video source modeling. There, we argued that, it is more appropriate to consider a model which captures the correlation in the interarrival sequence, rather than the correlation in counts sequence as in the above models. In CIPP, the interarrivals form a first order Markov sequence. In [1], it is shown that the CIPP does exhibit self-similarity over a range of time scales of interest. It is also analytically tractable (see for instance [13, 14]). In this paper, after giving a brief account of CIPP, we derive the stationary distributions in a CIPP/M/1 queue and then undertake the performance modeling of statistical multiplexer with VBR video traffic input using the CIPP/M/1 queue. The rest of the paper is organized as given below. In Section 2 we give a short mathematical account of the CIPP process. Section 3 gives analytical results for the CIPP/M/1 queue. In Section 4 we present the performance modeling of VBR traffic using CIPP/M/1 queue. In Section 5 simulation results corresponding to MPEG-1 *real world* VBR traffic traces are given. Section 6 concludes the paper.

2 Correlated Interarrival Time Poisson Process (CIPP)

In this section, we give the axiomatic derivation of CIPP. Some remarks are given to highlight some important features of the CIPP model.

2.1 Mathematical Background and Properties 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 autoregressive process (EAR(1))) developed by Gaver and Lewis in [15], 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,

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

   where, $\{\epsilon_n\}_{n=2}^{\infty}$ is an *iid* sequence, with $\epsilon_n, \ n > 1$, 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 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 $r = 1, 2, \dots$ the joint distribution of $\{N(h+t_{u1}) - N(t_{l1}), N(h+t_{u2}) - N(t_{l2}), \dots, N(h+t_{ur}) - N(t_{lr})\}$ is independent of h, for any $h > 0$ and $0 < t_{l1} < t_{u1} < t_{l2} < t_{u2} < \dots < t_{lr} < t_{ur} < \infty$

Definition: A counting process $\{N(t), t > 0\}$ satisfying Axioms 1 through 3 is called Correlated Interarrival time Poisson Process (CIPP) with parameter 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. For any $t > 0, \quad 0 < Pr\{N(t) > 0\} < 1$. In words, in any interval (no matter how small) there is a non-zero probability that an arrival will occur.

2. For any $t \geq 0$,

$$\lim_{h \to 0} \frac{Pr\{N(t+h) - N(t) \geq 2\}}{Pr\{N(t+h) - N(t) = 1\}} = 0 \tag{2}$$

   In words, in sufficiently small intervals, at most one arrival can occur: i.e., it is not possible for arrivals to happen simultaneously. The process is *orderly*, in the sense of Daley *et al* [16].

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,*

$$P_n(t) = \begin{cases} (1+\rho-(1-\rho)\alpha_n) \\ \times \sum_{j=0}^{n} \frac{\alpha_j \prod_{k=0}^{n}[\alpha_j - \rho\alpha_k]}{(\rho(\alpha_j+1)-\alpha_n+1)(\alpha_j-\alpha_n+1)(\alpha_j-\rho\alpha_n)\prod_{k=0,k\neq j}^{n}[\alpha_j-\alpha_k]} e^{-\frac{\lambda t}{\alpha_j}} & 0 \leq \rho < 1 \end{cases} \tag{3}$$

for $\quad n = 0, 1, 2, \dots\ , \ \lambda > 0, \ \ t \geq 0, \quad$ *and* $\quad \alpha_j = \sum_{i=0}^{j} \rho^i$

Proof: For proof of the Lemma and other details about CIPP, please see [1].

It is worth noting that the mean of the process, $\mathbf{E}\{N(t)\}$, can be shown to be λt [17].

Remark 3: The prime reason for assuming (1) for the interarrival sequence, is to mimic the self-similar behavior in increments for the corresponding counting process. Note that to exhibit self-similar behavior (over a range of time scales) with $0.5 < H < 1$, it is necessary that the increments should have positive correlations [18]. The reason for this particular correlation is that, (1) introduces strong positive correlation than other models based on moving average structure like EMA(1) proposed by Lawrance [19] and the one proposed by Finch [20]. It is shown in [1] that, as ρ increases, the estimated local Hurst parameter $(0.5 < H < 1)$ increases. Secondly, the correlation structure (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. Thirdly, the analytical simplicity of the correlation structure (1), allows one to develop the corresponding queueing theory [14]. 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: In (3), as the limit $\rho \to 0$, we obtain the Poisson distribution. Thus, the Poisson distribution is approached continuously from the CIPP distribution.

Remark 5: 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 [3, 9] 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. Note that the discrete AR (DAR(1)) process is used in [3, 9] to model the *frame size* of VBR traffic, where the counts (rather than interarrivals) form a Markov chain. The authors would like to emphasize that, the introduction of correlation in the interarrival sequence in the context of VBR traffic modeling is a new concept. Thus, the main contribution of this paper is the performance modeling of VBR traffic by a queue with the above process as input.

Remark 6: The number of parameters needed to mimic the self-similar behavior (to a range of time scale) is minimal in CIPP as compared to the conventional (SRD) short range dependent models like MMPP, DBMAP etc. For example in MMPP, the number of parameters are 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. The authors believe that the reason for this phenomena is that in CIPP, the better imitation of self-similar behavior emanates from the interarrival correlation rather than from the correlation due to the modulating Markov chain.

Remark 7: For many reasons, using an excessive number of parameters is undesirable, especially because 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 both of them have a nice physical interpretation. λ denotes the arrival rate, and ρ indicates the degree of interarrival correlation. While λ can be estimated from the mean of the process, ρ can be estimated by matching the higher order statistics of the CIPP model and the data.

Remark 8: It has been shown in [21] empirically, that there exists a monotonic relationship between the correlation parameter ρ of CIPP and the estimated local Hurst parameter (a measure of degree of self-similarity) of the corresponding CIPP trace as shown in [21]. On the contrary, in SRD models addressed in literature, 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 9: We derived CIPP, by first generating correlated exponential sequence and then constructing a stationary counting process over it. The whole exercise can be repeated in discrete time domain, by considering correlated geometric sequence generated by discrete AR(1) model [22]. But the result does not seem to yield a closed form expression and will not be quoted any further in this paper.

2.2 CIPP and the *Neuts* Process

In this section we compare CIPP with Neuts process - a process which encompasses many processes (including MMPP) which are of so central to modern teletraffic modeling. Here, we establish that CIPP is *not* a special case of Neuts process [23].

In the Neuts process [23], 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[1]. 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.

3 CIPP/M/1 Queue

We consider the single server FCFS queue with CIPP arrivals (with intensity λ) and exponential service times with parameter μ. We obtain analytical expressions for the waiting time distribution and buffer occupancy distribution in Appendix A.

Before going into the details of CIPP/M/1 queue, we would like to mention some points about the CIPP/D/1 queue. In [24], the approach of Benes is used to obtain the approximate expression for the buffer occupancy distribution in a CIPP/D/1 queue.

[1]Explicit correlations in interarrival may occur in MMPP also. But such events have very low probability due to the *assumption* that the rate of state changes of the modulating Markov chain is much less than the arrival rate of the Poisson process within a sojourn interval of the Markov chain. This is explained as below: Due to the above assumption, the probability of event that a single arrival occurring in each of the three consecutive states is low. It is obvious that whenever such event occurs, the interarrivals are correlated since the interarrival time depends on the sojourn interval of the states.

Since, this (approximate) analytical expression for the buffer occupancy distribution poses a numerical problem for large values of ρ, one has to estimate the distribution of the CIPP, by suitable simulations. This leads to the estimation of density, which is computationally intensive. This factor does not allow the applicability of CIPP/D/1 for performance modeling purposes.

3.1 Buffer Occupancy Distribution

Let Q denote the stationary random variable denoting the number of customers in the system (queue and server) as seen by an arriving customer. We denote the stationary probability of $Q = j$ number of customers in the queue by P_j. The expression for P_j is given by (65) and is repeated here for convenience.

$$P_j = \begin{cases} 1 - \frac{TM(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda+a}\right) & j = 0 \\ \frac{(1-T)T^j M(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda+a}\right) & j > 0 \end{cases}$$

where $T = \frac{1-\rho+\rho(\frac{\mu}{\lambda})}{\frac{\mu}{\lambda}}$, $a = \frac{\mu\rho}{1-\rho}$. λ and μ are arrival and service rate respectively. $M(a,1)$ and $M(0,1)$ are given by,

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

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

3.2 Waiting Time Distribution

Let $Pr\{W \leq x\}$ be the stationary waiting time distribution. Then, it is given by (72), which is repeated here.

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

with $\lambda, \mu, a, T, M(0,1)$ and $M(a,1)$ have the same meaning as above.

3.3 Derivation of Queueing Performance Measures

We derive the expressions for the queueing performance measures corresponding to CIPP/M/1 queue from the stationary distributions given above. In this section and the rest of the paper, we would freely use the term packet to denote the arriving entities.

We define

$$\beta \triangleq \frac{M(a,1)}{M(0,1)}\frac{\lambda}{\lambda+a} \tag{4}$$

where $\lambda, \mu, a, T, M(0,1)$ and $M(a,1)$ have the same meaning as given in Section 3.1.

3.3.1 Mean Waiting Time

From (65), we have

$$\mathbf{E}\{W\} = \frac{TM(a,1)}{\mu(1-T)M(0,1)}\left(\frac{\lambda}{\lambda+a}\right). \tag{5}$$

3.3.2 Mean Number of Packets in the System

From (65), we have

$$\mathbf{E}\{Q\} = \frac{TM(a,1)}{(1-T)M(0,1)}\left(\frac{\lambda}{\lambda+a}\right). \tag{6}$$

3.3.3 Packet Delay Variation

In a broadband network, for a given connection, cells 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 [25]. 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. Since the W and S are independent, using the sum rule for variance, it is straight forward to obtain the expression for the *pdv* as,

$$\tau = \frac{2}{\mu^2}\left[1-\beta+\frac{\beta}{(1-T)^2}\right] + \left[\frac{1}{\mu}\left(1+\frac{\beta T}{(1-T)}\right)\right]^2 \tag{7}$$

where, β is as given above.

3.3.4 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 [25], 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[2] for $K > 0$, is given by,

$$Pr\{Q \geq K+1\} = \frac{M(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda + a}\right) T^{K+1} \tag{8}$$

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

4 Performance Modeling of Statistical Multiplexer with VBR Traffic Input by CIPP/M/1 Queue

Before discussing the performance modeling of statistical multiplexer with real-world VBR video traffic by CIPP queues, we present here the simulation methodology for our performance modeling study.

4.1 Simulation Methodology

In this subsection, we discuss the methods to handle the *real-world* MPEG-1 VBR video traffic data for our simulation purposes.

There is a problem (discussed in detail in Section 4.1.1) 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 the previous section 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 Section 4.1.2.

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 the previous section 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 the previous section), corresponding to this "best" fit CIPP.

[2]The unit "packet" refers to the mean packet length (in bits).

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.

4.1.1 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 [10]. 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.

4.1.2 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 (9)). 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[3]. The purpose of the whole procedure is two-fold, one being feasibility of *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 4.1 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 believe 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,

$$N_{T_f}(i) = \overset{argmax}{k} \; \{k | \sum_{j=1}^{k} V_i(j) < u_{T_f}(i)\} \tag{9}$$

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,

$$T_k \;=\; (i-1)T_f + \frac{N_{T_f}(i)}{T_f} j \quad 1 \leq j \leq N_{T_f}(i) \tag{10}$$

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

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,

$$S(j + \sum_{l=1}^{i-1} N_{T_f}(l)) = V_i(j) \quad 1 \leq j \leq N_{T_f}(i) \tag{11}$$

for $2 \leq i \leq N$ and

$$S(j) = V_i(j) \quad 1 \leq j \leq N_{T_f}(i) \tag{12}$$

for $i = 1$.

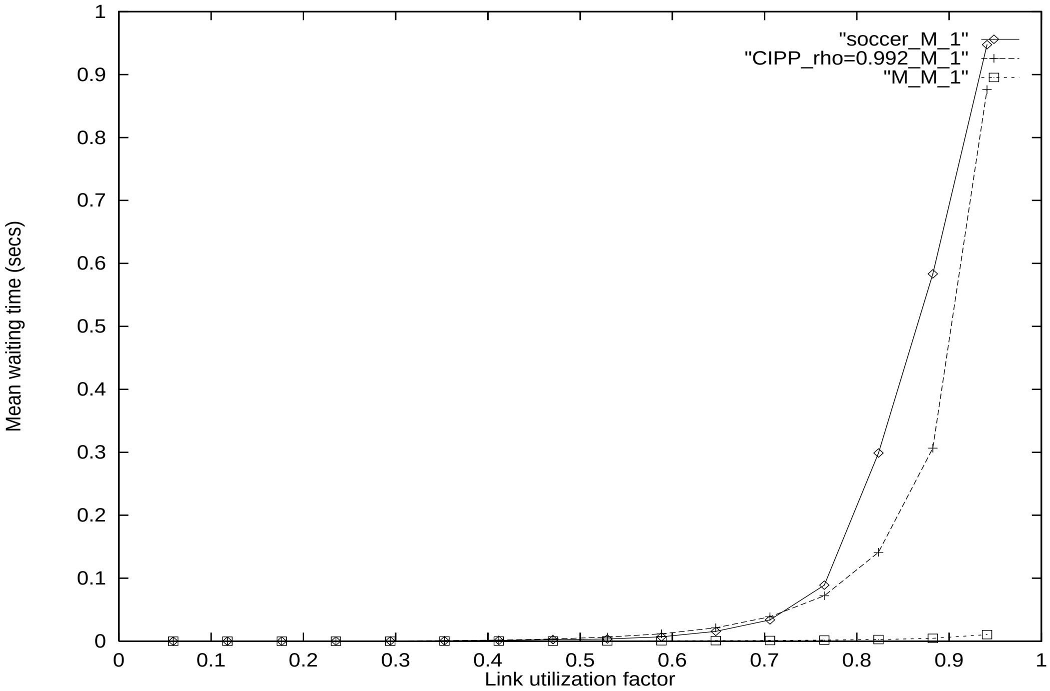
5 Simulation Results

In this subsection, we consider the performance modeling of statistical multiplexer with *real-world* MPEG-1 VBR video traffic using CIPP/M/1 queue. Queueing performance measures[4] considered here are mean waiting time, packet delay variation (*pdv*) and overflow probability which are given by (5), (7) and (8) respectively. We consider the *soccer* and *starwars* MPEG-1 encoded VBR video sequences [26] for our experiments.

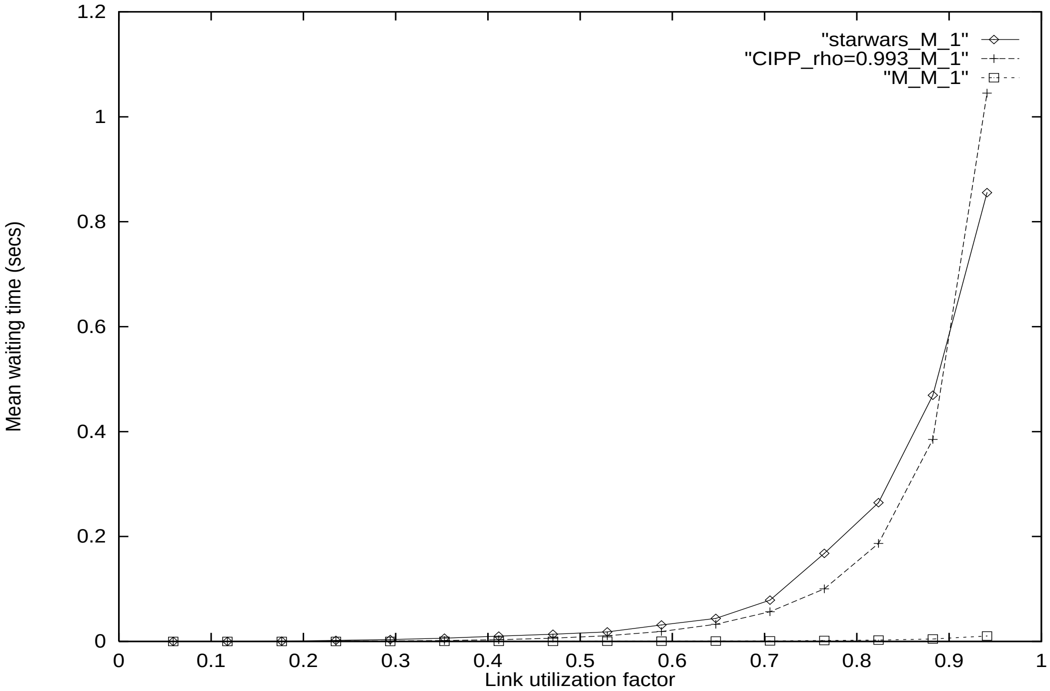
5.1 Mean Waiting Time

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 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 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* [27] that the Poisson process poorly models the traffic. Moreover, the M/M/1 queue seriously underestimates the queueing performance measures [28], 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.

[4]Since the expressions for distributions for the mean number of packets in the system and the mean waiting time are very similar (except for the factor $1/\mu$) we would not give the simulation results for the performance measure of mean number of packets.



(a)



(b)

Figure 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	1 (a)	96.94	0.36	99.31	0.45	7.51	0.48	52.72	0.42	0.992
starwars	1 (b)	98.15	0.33	99.03	0.45	17.97	0.45	40.24	0.39	0.993

Table 1: Percentage error between the mean waiting time of packets in queues with *soccer* (Figure 1 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 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	2 (a)	99.9	0.48	99.95	0.45	27.87	0.42	37.95	0.45	0.992
starwars	2 (b)	99.95	0.48	99.99	0.36	0.99	0.48	88.08	0.33	0.993

Table 2: Percentage error between the packet delay variation in queues with *soccer* (Figure 2 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 2 (b)). In both the plots corresponding error for the approximating M/M/1 queue is shown for comparison

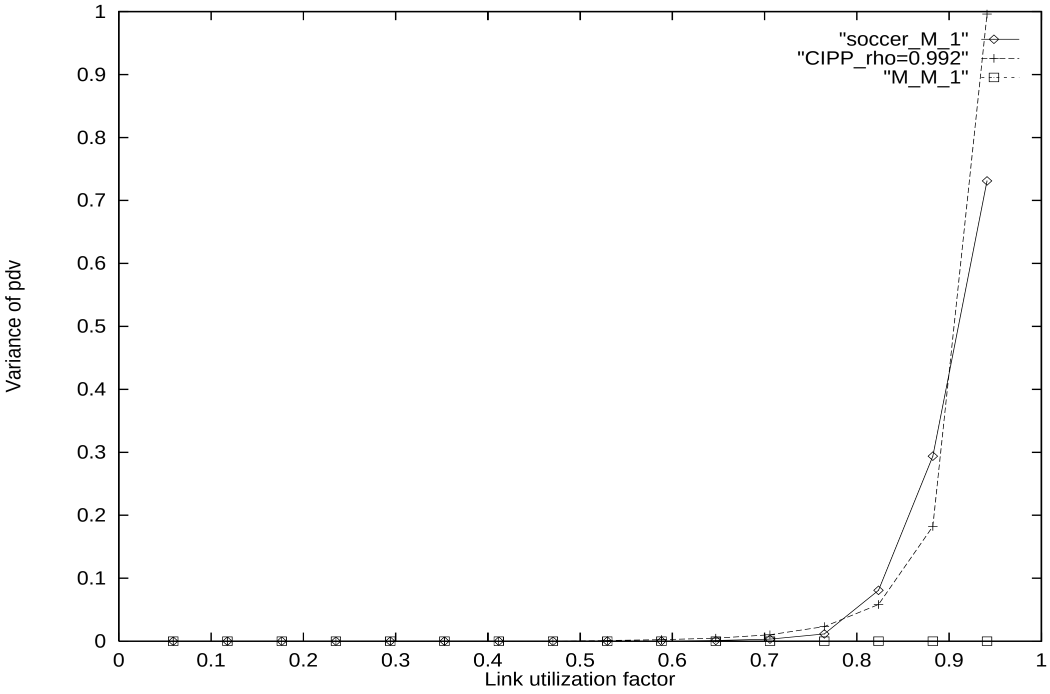
.

5.2 Packet Delay Variation (pdv)

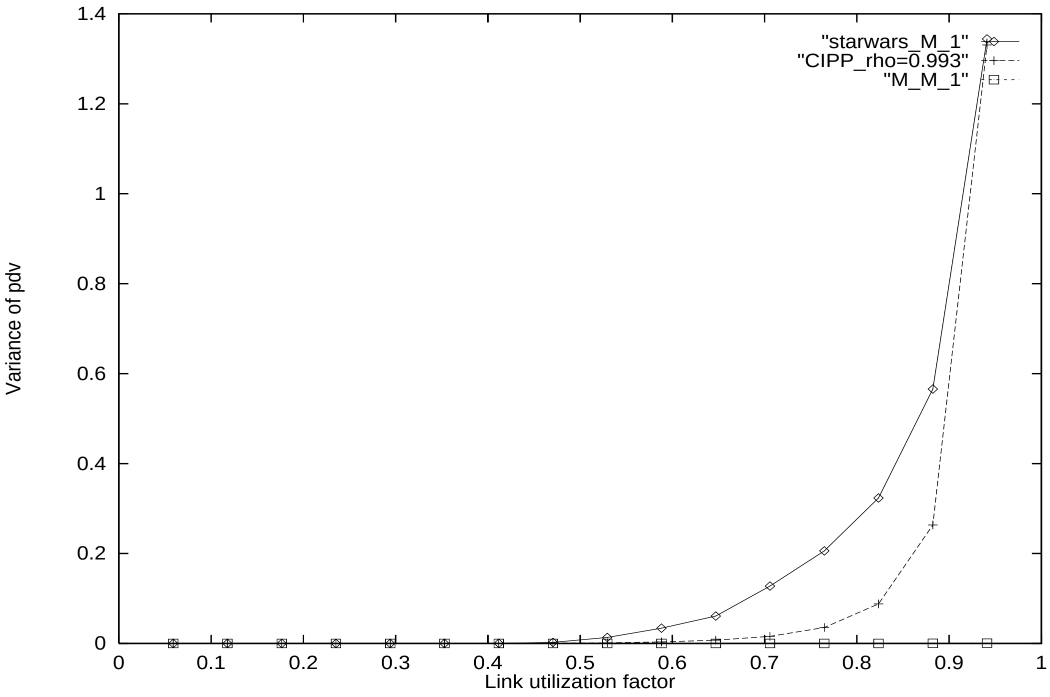
Here, our queueing performance measure will be packet delay variation (*pdv*). Our results corresponding to the video sequences *soccer* and *starwars* are depicted in Figure 2. Table 2 gives the maximum and minimum percentage error between the theoretical performance measure, the *pdv* corresponding 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.

5.3 Overflow Probability

We now consider the queueing measure, the overflow probability. We take the threshold of 32 packets ($K = 32$) for our experiments. Figure 3 illustrates our results for over-



(a)



(b)

Figure 2: 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*.

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	3 (a)	95.58	0.76	100.0	0.24	3.6	0.3	35.99	0.33	0.992
starwars	3 (b)	99.99	0.94	100.0	0.88	15.53	0.94	34.56	0.88	0.993

Table 3: Percentage error between the overflow probability in queues with *soccer* (Figure 3 (a)) input and the approximating CIPP/M/1 queue and the corresponding error for *starwars* (Figure 3 (b)). In both the plots corresponding error for approximating M/M/1 is shown for comparison. A threshold of 32 packets is considered here

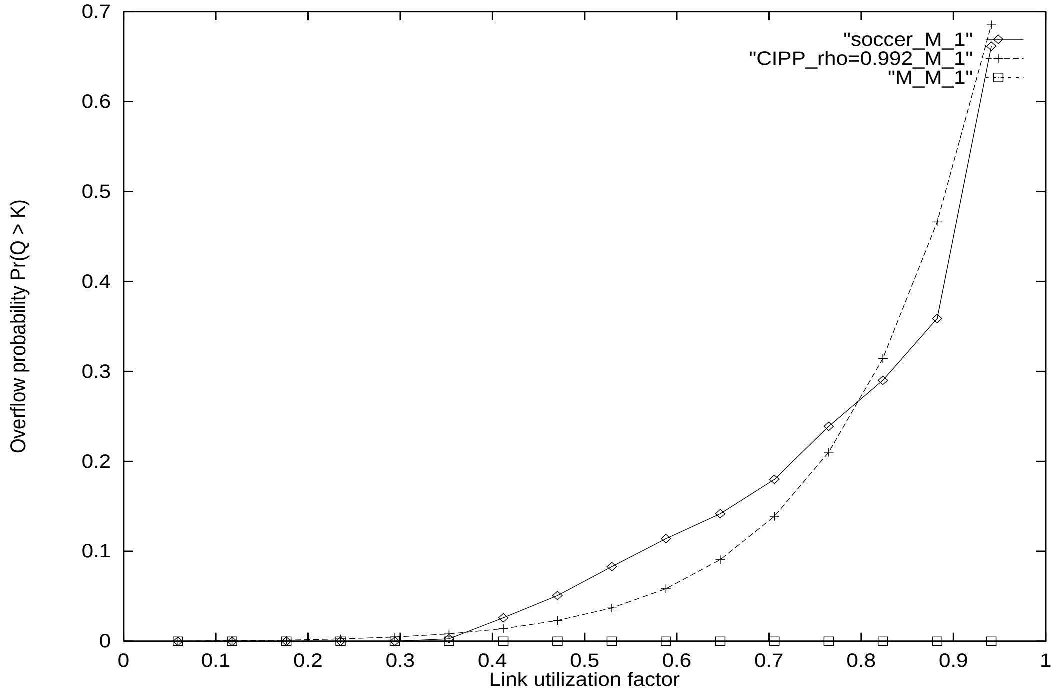
.

flow 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 3 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* [28].

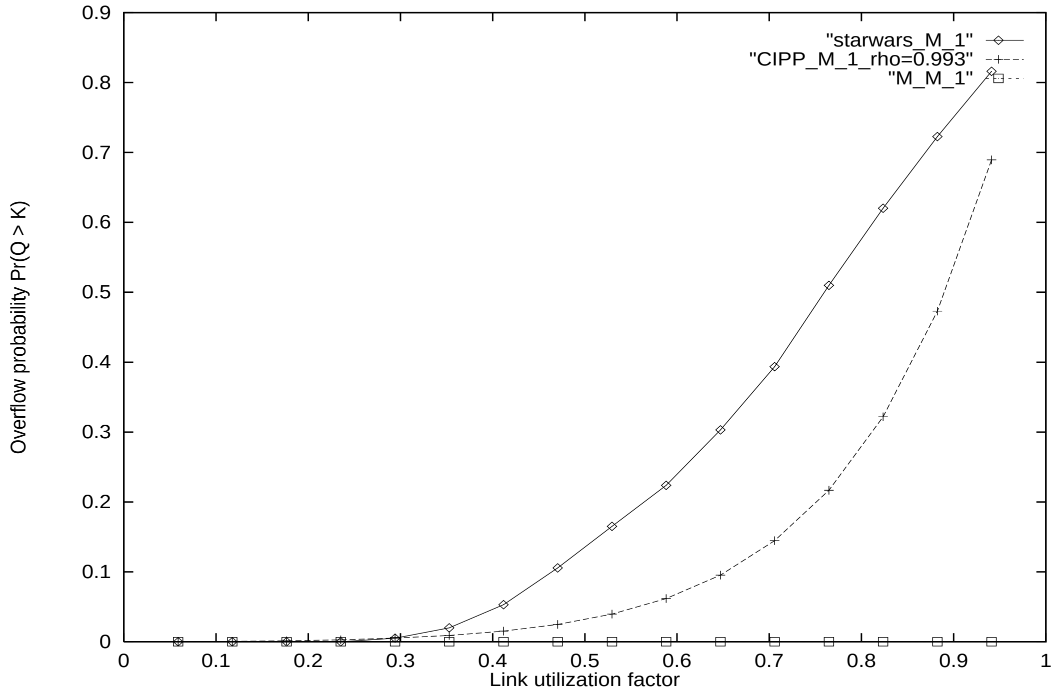
6 Conclusions

In this paper, we have developed the theory for CIPP/M/1 queue and used it to characterize the queueing situation encountered by the VBR video traffic in a statistical multiplexer. We noted that, not only the first order statistics of queueing performance corresponding to data namely, the mean waiting time for packets is 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 noted 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). 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*



(a)



(b)

Figure 3: 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.

sequence is 88 and occurs for the measure of *pdv*. 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.

The results corresponding to M/M/1 queue in all the experiments corroborate with the earlier work [27]. To the author's knowledge, CIPP is the first model which models well the *real world* VBR video traffic in broadband networks, as well as *analytically tractable*. We believe that, the theoretical approach proposed in this paper, will be a useful tool in supporting real-time ATM network traffic management. Currently we are working on the theoretical aspects of priority queues with CIPP input.

A Derivations of Stationary Distributions in a 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 derived.

A.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 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,

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{13}$$

and

$$G(s, x) = \frac{\lambda}{\lambda + s} \tag{14}$$

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

$$f(t/x) = \rho\delta(t - \rho x) + (1 - \rho)\lambda e^{-\lambda(t-\rho x)}U(t - \rho x) \tag{15}$$

and the corresponding conditional LST is given by

$$G(s, x) = \left(\rho + \frac{(1-\rho)\lambda}{\lambda + s}\right) e^{-\rho x s} \tag{16}$$

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 A.2 and A.3.1, we investigate transient and limiting behavior of the process $\{Q_n, T_n\}$ respectively. The method is similar to that used by Tin [29] (see also Pearce [30] and Finch and Pearce [31] for a moving-average interarrival process). In Section A.3.2, the waiting time distribution is derived.

A.2 Transient Behavior of the Queue Size

Let us define,

$$P_{ij}^{(n)}(t) \triangleq Pr\{Q_n = j, X_n \leq t | Q_0 = i\}. \tag{17}$$

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

$$\begin{aligned} k_j(t, x) &= \int_0^t \frac{e^{-\mu y}(\mu y)^j}{j!} f(y|x) dy & (18) \\ k_j^*(s, x) &= \int_0^\infty e^{-st} d_t k_j(t, x) & (19) \end{aligned}$$

Throughout this section, we use the following notation,

$$\begin{aligned} P_{ij}^*(s; n) &= \int_0^\infty e^{-st} d_t(P_{ij}^{(n)}(t)) & (20) \\ P_i^*(s, z; n) &= \sum_{j=0}^{\infty} P_{ij}^*(s; n) z^j & (21) \\ P_i^*(s, z, w) &= \sum_{n=0}^{\infty} P_i^*(s, z; n) w^n & (22) \\ &= \sum_{j=0}^{\infty} P_{ij}^*(s, w) z^j, \quad (|z| \leq 1, |w| < 1, s \geq 0) & (23) \\ C_m^*(s; n) &= \int_0^\infty \sum_{l=0}^{\infty} k_{l+m+1}^*(s, x) d_x(P_{il}^{(n)}(x)) & (24) \\ C(s, z, w) &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} (1 - z^{-m}) C_m^*(s; n) w^n, \quad (|z| \geq 1, |w| < 1, s \geq 0) & (25) \end{aligned}$$

Fundamental Equations: If a customer arriving at T_{n+1} finds $j \geq 1$ individuals in the system, the customers arriving at T_n must have found at least $(j-1)$ customers already present. If there were $j-1+l \quad (l \geq 0)$ 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,

$$P_{ij}^{n+1}(t) = \int_0^\infty \sum_{l=0}^\infty k_l(t,x) d_x(P_{i,j+l-1}^{(n)}(x)), \quad (j \geq 1) \tag{26}$$

$$P_{i0}^{n+1}(t) = \int_0^\infty \sum_{l=0}^\infty \sum_{m=l+1}^\infty k_m(t,x) d_x(P_{il}^{(n)}(x)) \tag{27}$$

Hence we have,

$$P_{ij}^*(s;n+1) = \int_0^\infty \sum_{l=0}^\infty k_l^*(s,x) d_x(P_{i,j+l-1}^{(n)}(x)), \quad (j \geq 1) \tag{28}$$

$$P_{i0}^*(s;n+1) = \int_0^\infty \sum_{l=0}^\infty \sum_{m=l+1}^\infty k_m^*(s,x) d_x(P_{il}^{(n)}(x)) \tag{29}$$

By substituting (19) into (28) and (29) and then taking generating functions with respect to j, we finally obtain after using (18), (20) and (21),

$$P_i^*(s,z;n+1) = zB(s,z)P_i^*(A(s,z),z;n) + \sum_{m=0}^\infty (1-z^{-m})C_m^*(s;n) \tag{30}$$

where,

$$A(s,z) = \rho(s+\mu-\mu z^{-1}) \tag{31}$$

$$B(s,z) = \rho + \frac{(1-\rho)\lambda}{\lambda+\mu+s-\mu z^{-1}} \tag{32}$$

Once again taking the generating functions with respect to n we have, from (30),

$$P_i^*(s,z;w) = wzB(s,z)P_i^*(A(s,z),z;w) + z^i + wC(s,z,w) \tag{33}$$

$A(s,z) = s$ has exactly one positive root for s. Denoting the root by ξ and putting $s=\xi$ in (33), we have,

$$P_i^*(\xi,z,w) = \frac{z^i + C(\xi,z,w)}{1-wzB(\xi,w)} \tag{34}$$

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

$$\Phi(s) = P_i^*(s,z,w) - P_i^*(\xi,z,w) \tag{35}$$

Equation (33) then reduces to,

$$\Phi(s) = h(s, \Phi(A(s,z))) \tag{36}$$

where

$$\begin{aligned} h(s,y) &= wzB(s,z)[y + H(s)] && (37)\\ H(s) &= [wzB(s,z)]^{-1}\left[z^i + wC(s,z,w) - \frac{1 - wzB(s,z)}{1 - wzB(\xi,z)}[z^i + wC(\xi,z,w)]\right] && (38)\\ H(\xi) &= 0 && (39) \end{aligned}$$

We now solve the functional equation (36) in s (for fixed w and z) by making use of a theorem due to Read [32] 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 (36) is given by,

$$\Phi(s) = \lim_{n\to\infty} \Phi_n(s) \tag{40}$$

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,

$$\Phi(s) = \sum_{j=0}^{\infty} H(f^j) \prod_{i=0}^{j} [wzB(f^i, z)] \tag{41}$$

$$f^{i+1} = A(f^i, z) \tag{42}$$

with $f^0 = s$.
By substituting $H(s)$ from (38) and using (35), we finally obtain,

$$P_i^*(s,z,w) = H_1(s,z,w) + H_2(s,z,w), \tag{43}$$

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,

$$F(s,z,w) = \begin{cases} [1-zT(w)]P_i^*(s,z,w), & (|z|\le 1, |w|<1, s\ge 0) \\ (1-zT(w))Q_i(s,z,w), & (|z|\ge 1, |w|<1, s\ge 0) \end{cases} \tag{44}$$

where, $Q_i(s,z,w)$ denotes the right-hand side of (43) and $T(w)$ is the root of

$$z = wB(\xi, z^{-1}) \tag{45}$$

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

$$T = T(1) \tag{46}$$

$$= \frac{(1-\rho)+\rho(\mu/\lambda)}{\mu/\lambda} \tag{47}$$

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 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 [33], we obtain, for (31),(32), (42),

$$a = \lim_{j\to\infty}a_j(s)$$

$$= \frac{\mu\rho}{1-\rho} \tag{48}$$

and $a_j(a)=a$ for all $j\ge 0$. 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

$$F(s,z,w) = \sum_{m=0}^{i+1} z^m I_m(s,w) \tag{49}$$

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

$$I_{i+1}(a,w) = 0 \tag{50}$$

$$P_i^*(s,z,w) = [1 - zT(w)]^{-1}\sum_{m=0}^{i+1} z^m I_m(s,w) \quad (|z| \le 1, |w| < 1, s \ge 0), \tag{51}$$

giving

$$P_{ij}^*(s,w) = \begin{cases} \sum_{m=0}^{j}[T(w)]^{j-m} I_m(s,w), & (j \le i) \\ \\ \sum_{m=0}^{i+1}[T(w)]^{j-m} I_m(s,w), & (j \ge i+1) \end{cases} \tag{52}$$

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 (19),(28) and (29) we have,

$$P_{0j}^*(s;n+1) = \begin{cases} \sum_{l=0}^{\infty} \frac{(-\lambda)^l}{l!}\frac{\partial^l}{\partial y^l}\left[\left[\frac{1}{1+\alpha(1-\rho)(y+s)}\right]^r \right. & \times \\ \\ \left. P_{0,j+l-1}^*\left(\frac{\rho(y+s)}{1+\alpha(1-\rho)(y+s)}; n\right)\right]_{y=\lambda} & (j \ge 1) \end{cases} \tag{53}$$

By substituting $P_{0j}^*(s,w)$ from (52) in (53) and using Taylor's series theorem, we finally obtain,

$$T(w)U(s,w) = wB(s,T(w)^{-1})U(A(s,T(w)^{-1}),w) \tag{54}$$

where,

$$U(s,w) = T(w)I_0(s,w) + I_1(s,w) \tag{55}$$

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,

$$\lim_{n\to\infty} V_n(s) = \lim_{n\to\infty}\prod_{i=0}^{n-1}(T(w)/w)[B(g^i,T(w)^{-1})]^{-1} \tag{56}$$

tends to a limit, $V(s)$, say. Then, the general solution of the functional equation in this case is given by (Kuczma [34], p.48, Theorem 2.2),

$$U(s,w) = \frac{D}{V(s)} \tag{57}$$

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. It is straight forward to derive the expression for $V_n(s)|_{s=0,w=1}$.

$$V_n(s)|_{s=0,w=1} = \frac{(\rho+\frac{\lambda}{\mu}(1-\rho))(1-\rho+\rho\frac{\lambda}{\mu})}{(1-\rho)(\rho+\frac{\lambda}{\mu})+\rho^2\frac{\lambda}{\mu}} \prod_{i=1}^{n-1} \frac{(\rho+(1-\rho)\frac{\lambda}{\mu})(1-(1-\frac{\lambda}{\mu})\rho^{i+1})}{(1-\rho)\frac{\lambda}{\mu}+\rho(1-(1-\frac{\lambda}{\mu})\rho^{i+1})} \quad (58)$$

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.

Noting that,

$$P_i^*(s,1,w) \;=\; \sum_{n=0}^{\infty} \frac{\lambda+\rho^n s_n}{\lambda+s_n} w^n \quad (59)$$

which is the Laplace-Steiltjes transform of the distribution of T_n given $T_0 = 0$, we have from (51), for $i = 0$,

$$I_0(s,w) + I_1(s,w) = [1-T(w)] \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n s_n}{\lambda+s_n}\right) w^n \quad (60)$$

By solving (54) and (60) for I's and using (50), we obtain D so that I's are fully determined. By substituting these I's in (52), we finally obtain,

$$\begin{aligned} P_{0j}(w) &= P_{0j}^*(0,w) & (61)\\ &= \begin{cases} (1-w)^{-1} - T(w)\frac{M(a,w)}{M(0,w)} \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n a}{\lambda+a}\right) w^n, & j = 0 \\ \\ (1-T(w))[T(w)]^j \frac{M(a,w)}{M(0,w)} \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n a}{\lambda+a}\right) w^n, & j > 0 \end{cases} & (62) \end{aligned}$$

where,

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

A.3 Limiting Behavior

In this section, we derive the stationary distributions for queue length and the waiting time.

A.3.1 Buffer Occupancy Distribution

We have from (16) and (59), that for $|\rho| < 1$,

$$\lim_{w \to 1}(1-w)\sum_{n=0}^{\infty}\left(\frac{\lambda+\rho^n s}{\lambda+s}\right)w^n = \left(\frac{\lambda}{\lambda+a}\right)^r \tag{63}$$

It can be seen that from (45), $T(1)$ lies in the unit circle iff $\lambda/\mu < 1$. Thus, we have from (62) that $\lim_{w\to 1}(1-w)P_{0j}(w)$ exists if and only if $\lambda/\mu < 1$ and $|\rho| < 1$. In this situation, we obtain, from (62),

$$P_j = \lim_{w \to 1}(1-w)P_{0j}(w) \tag{64}$$

$$= \begin{cases} 1 - \frac{TM(a,1)}{M(0,1)}\left(\frac{\lambda}{\lambda+a}\right), & j = 0 \\ \\ \frac{(1-T)T^j M(a,1)}{M(0,1)}\left(\frac{\lambda}{\lambda+a}\right), & j > 0 \end{cases} \tag{65}$$

where, $T = T(1)$ is given by (47), which is < 1 for $\mu/\lambda > 1$ and $a = \frac{\mu\rho}{1-\rho}$. $M(a,1)$ and $M(0,1)$ are given by,

$$M(0,1) = \begin{cases} \frac{(\rho+\frac{\lambda}{\mu}(1-\rho))(1-\rho+\rho\frac{\lambda}{\mu})}{(1-\rho)(\rho+\frac{\lambda}{\mu})+\rho^2\frac{\lambda}{\mu}} \times \\ \\ \lim_{n\to\infty}\prod_{i=1}^{n-1}\frac{(\rho+(1-\rho)\frac{\lambda}{\mu})(1-(1-\frac{\lambda}{\mu})\rho^{i+1})}{(1-\rho)\frac{\lambda}{\mu}+\rho(1-(1-\frac{\lambda}{\mu})\rho^{i+1})} \end{cases} \tag{66}$$

$$M(a,1) = \begin{cases} \frac{((1-\rho)\frac{\lambda}{\mu}+\rho)(1-\rho(1-\rho)(1-\frac{\lambda}{\mu}))}{\left[(1-\rho)^2\frac{\lambda}{\mu}+\rho(1-\rho(1-\rho)(1-\frac{\lambda}{\mu}))\right]} \times \\ \\ \lim_{n\to\infty}\prod_{i=1}^{n-1}\frac{\left(\rho+(1-\rho)\frac{\lambda}{\mu}\right)\left[(1-\rho)(1-(1-\frac{\lambda}{\mu})\rho^{i+1})+\rho^{i+1}\right]}{(1-\rho)^2\frac{\lambda}{\mu}+\rho\left[(1-\rho)(1-(1-\frac{\lambda}{\mu})\rho^{i+1})+\rho^{i+1}\right]} \end{cases} \tag{67}$$

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

A.3.2 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 $Pr\{W_n \le x/Q_0 = 0\}$, then we have

$$\Omega_n(s) = \sum_{j=0}^{\infty}\left(\frac{\mu}{\mu+s}\right)^j P_{0j}^{(n)} \tag{68}$$

where $P_{0j}^{(n)} = Pr\{Q_n = j | Q_0 = 0\}$. Hence, by taking generating functions with respect to n and using (65), we obtain,

$$\begin{aligned}\sum_{n=0}^{\infty} \Omega_n(s) w^n &= \sum_{j=0}^{\infty} \left(\frac{\mu}{\mu+s}\right)^j P_{0j}(w) \\ &= \left\{ \begin{array}{l} (1-w)^{-1} - \frac{T(w)M(a,w)}{M(0,w)} \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n a}{\lambda+a}\right) w^n + \\ \\ \frac{(1-T(w))M(a,w)\mu T(w)}{M(0,w)(s+\mu(1-T(w))} \sum_{n=0}^{\infty} \left(\frac{\lambda+\rho^n a}{\lambda+a}\right) w^n \end{array} \right. \end{aligned} \qquad (69)$$

If $\lambda/\mu < 1$ we obtain, from (63) and (69),

$$\lim_{n\to\infty} \Omega_n(s) = \lim_{w\to 1}(1-w) \sum_{n=0}^{\infty} \Omega_n(s) w^n \qquad (70)$$

$$= 1 - \frac{TM(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda+a}\right) + \frac{(1-T)M(a,1)\mu T}{M(0,1)(s+\mu(1-T))} \left(\frac{\lambda}{\lambda+a}\right) \qquad (71)$$

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

$$Pr\{W \le x\} = 1 - \frac{TM(a,1)}{M(0,1)} \left(\frac{\lambda}{\lambda+a}\right) exp(-\mu x(1-T)) \qquad (72)$$

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 (31) and (48). This gives $\xi = 0$ and $a = 0$. Hence, from (65) and (72), we have

$$P_j = (1-T)T^j \quad j \ge 0 \qquad (73)$$

$$Pr\{W \le x\} = 1 - Texp\{-\mu(1-T)x\} \qquad (74)$$

where $T = \lambda/\mu$.

References

[1] R. Manivasakan, U. B. Desai and A. Karandikar, "Broadband Teletraffic Characterization using Correlated Interarrival time Poisson Process (CIPP)", *J. Indian Inst. Sci.*, vol. 79, no. 3, pp. 233 – 249, to appear 1999.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Variable bit rate video traffic and long range dependence", *IEEE Trans. Commun.*, 1993.

[7] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of slf-similar VBR video traffic", *preprint*, 1994.

[8] 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.

[9] 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.

[10] 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.

[11] J. Norros and J. T. Virtamo, "Who loses cells in case of burst scale conjestion", *In 13th Int. Teletraffic Congress proceedings*, 1991.

[12] 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.

[13] R. Manivasakan, U. B. Desai and A. Karandikar, "Correlated Interarrival time Poisson Process (CIPP) and the CIPP/M/1 queue", Tech. Rep., SPANN lab., IIT, Bombay, May. 1998, Technical Report no: spann.mani.1998.2.

[14] 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.

[15] 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.

[16] D. J. Daley and D. Vere-Jones, *"An Introduction to the Theory of Point Processes"*, Springer-Verlag, 1988.

[17] J. A. McFadden, "On the lengths in a stationary point process", *J.R. Statist. Soc.*, vol. B-24, pp. 364–382, 1993.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] M. F. Neuts, "A versatile Markovian point process", *J. Appl. Prob.*, vol. 16, pp. 764–779, 1979.

[24] R. Manivasakan, *"The Correlated Interarrival time Poisson Process (CIPP) for modeling broadband teletraffic"*, PhD thesis, Indian Institute of Technology, Mumbai, 1999.

[25] 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.

[26] "*ftp-info3.informatik.uni-wuerzburg.de /pub/mpeg/*", 1995.

[27] 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.

[28] 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.

[29] P. Tin, "A queueing system with Markov-dependent arrivals", *J. Appl. Prob.*, vol. 22, pp. 668–677, 1985.

[30] 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.

[31] 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.

[32] A. H. Read, "The solution of a functional equation", *Proc. R. Soc. Edinburgh*, vol. A 63, pp. 336–365, 1952.

[33] 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.

[34] M. Kuczma, "Functional equations in a single variable", *Polska Akademia Nauk, Monografia Matematyczma*, 1968.