Loss Correlation for Queues with Single and Multiple Input Streams

H. Schulzrinne, J.F. Kurose, D. Towsley

COINS Technical Report 92-53
July 1992

Loss Correlation for Queues with Single and Multiple Input Streams*

(presented in part at ICC '92, Chicago, June 1992)

Henning Schulzrinne
Department of
Electrical and Computer Engineering

James F. Kurose and Donald F. Towsley
Department of
Computer Science

University of Massachusetts
Amherst, MA 01003
hgschulz,kurose,towsley@cs.umass.edu

August 5, 1992

Abstract

The loss probability of a queueing system provides in many cases insufficient information for performance evaluation, for example, of file transfer retransmission protocols and real-time applications with signal reconstruction.

This paper evaluates and characterizes the correlation between packet losses for three queueing systems in discrete time that are motivated by BISDN applications. In the first, a single-stream batch-arrival queue operating in discrete time, we find that the service discipline and buffer management policy for FIFO and LIFO do not influence the loss correlation. The second, a two-class discrete-time queueing system, approximates the output queue of an ATM switch. The queue serves periodic foreground traffic and random background traffic. The background traffic is modeled as i.i.d. batches of arbitrary distribution. It is shown that the conditional loss probability is independent of the buffer size if the buffer size is at least as large as the period of the foreground traffic. Example calculations indicate that losses occur essentially randomly as long as the foreground traffic uses less than 10% of the channel capacity.

The final analysis derives the conditional loss probability a selected stream in a slotted finite-buffer system with a superposition of interrupted Poisson sources, where the total number of arrivals may be correlated from slot to slot. Traffic correlation is seen to have a strong influence on loss correlation, while buffer size has virtually none.

1 Introduction

With the advent of high-bandwidth wide area networks and their high ratio of propagation to transmission and queueing delay, loss probability rather than average waiting time or throughput has become the dominant performance metric. However, since losses in queueing systems can be strongly correlated, the loss probability of a queueing system provides in many cases insufficient information for performance evaluation, for example, of file transfer retransmission protocols

*This work is supported in part by the Office of Naval Research under contract N00014-90-J-1293, the Defense Advanced Research Projects Agency under contract NAG2-578 and a National Science Foundation equipment grant, CERDCR 8500332.

and real-time applications with signal reconstruction. Bursty traffic, as generated by multimedia applications, have made concerns about loss correlation more urgent.

The investigation of loss correlation has a long history in the context of bit errors in data communication [1], but has only recently attracted stronger interest in the area of packet networks.

We briefly summarize related work, extending the review in [2]. By rate conservation methodology, Ferrandiz and Lazar developed a theoretical framework for studying loss correlation in queueing systems [3–5]. Our focus is more on deriving computable results, relying on well-known Markov chain methods.

A number of authors have quantified the influence of loss correlation for network performance. For example, it has been shown [6–9] how the throughput of go-back-N ARQ increases with positive loss correlation. Similar results for the N-packet transfer time and go-back-∞ and blast protocols are presented in [10]. All papers[1] assume a two-state Markovian error model with geometrically distributed run lengths for losses and successes, which will be seen to be appropriate for some, but not all queueing systems studied below. Thus, our results can be used in conjunction with the work cited here to directly predict ARQ performance.

In some simulation studies, cell loss correlation has been investigated [11, 12] and methods for the compensation of correlated cell loss have been proposed [13, 14].

This paper continues the work begun in [2], extending it to derive the loss correlation seen by packets of a selected stream in queues with *several* arrival streams. The goal of this work is to provide insights as to the structure and magnitude of loss correlation, so that the protocol and system architecture designer can judge when loss correlation will significantly effect system performance. The designer can judge whether either ameliorating measures are called for (in the case of real-time sequence reconstruction) or the benefits of loss correlation (in the case of retransmission algorithms) can be reaped.

Throughout the paper, our interest in broadband ISDN and ATM-like systems motivates a discrete-time perspective, with packet losses caused by buffer overflow[2] Loss correlation is measured here by the conditional loss probability, the conditional probability that a packet from a stream is lost given that its immediate predecessor was also lost. The conditional loss probability will be denoted as $P[L(n)|L(n-1)]$. The loss run length $E[C_L]$, that is, the average number of consecutively lost packets from a stream, is related to the conditional loss probability by

$$E[C_L] = \frac{1}{1 - P[L(n)|L(n-1)]}.$$

The paper is divided into three major sections. First, in section 2, we find, somewhat surprisingly, that for a discrete-time queue with general batch arrivals, LIFO and FIFO combined with different discarding policies exhibit the same loss correlation.

In sections 3 and 4, we turn away from single-source systems and allow for background or interfering traffic. First, in section 3, we focus on a system where a foreground stream with random or deterministic interarrival time competes with a batched background stream for buffer space. We allow general batch arrivals for the background stream and general, but i.i.d., interarrival times for the foreground stream. This model allows us to judge the dependence of the loss correlation as a function of the contribution of the foreground stream, the burstiness of the background stream, the buffer size and the space priority order of the streams. We find first that the buffer size has no appreciable influence for all but the smallest buffers. Also, it appears that as long as the

[1] except for [8], which allows a more general Markovian error model

[2] It appears to be generally accepted that buffer overflow will be the dominant loss mode in wide-area high speed networks.

contribution of the foreground stream to the total traffic is sufficiently small, the losses of that stream can be considered independent.

Then, in section 4, we address the issue of the influence of traffic correlation or burstiness on the loss correlation. Here, the input traffic consists of a superposition of identical sources, so that the total input traffic may be correlated. Since they are commonly used to model voice and video traffic, we choose the superposition of interrupted Poisson processes in the $N \cdot \mathrm{IPP}/D/c/K$ queue. We find that buffer size plays almost no role in the value of the loss correlation, while the burstiness affects both loss probability and conditional loss probability in a similar way.

2 Buffer Overflow in Single-Stream Discrete-Time Queues

2.1 First-Come, First-Served

In this section, we will derive properties of the loss correlation for a class of discrete-time queues with restricted buffer size. As our queueing model, we consider a FIFO single-server discrete-time queue where arrivals occur in i.i.d. batches of general distribution with mean batch size λ. Each arrival requires exactly one unit of service. For short, we will refer to this system as $D^{[G]}/D/1/K$ [15]. Let K denote the system size, that is, the buffer capacity plus one. Arrivals that do not find space are rejected, but once a customer enters the system, it will be served. This buffer policy will be referred to as rear dropping in section 2.2. Arbitrarily, arrivals are fixed to occur at the beginning of a time slot and departures at the end, creating, in Hunter's terminology [16], an early arrival system. This model is used to represent the output queue of a fast packet switch, for example [17].

The waiting time and loss probability for this model have been analyzed by a number of authors [18, 15, 19, 17, 20–22]. For Poisson-distributed batches, Birdsall *et al.* [18, p. 392] computes the conditional probability of a run of exactly n slots in which one or more arrivals are rejected given that an arrival was rejected in the preceding slot. We will call it $P[C_R = n]$. The quantity is seen to be the product of the probability that two or more arrivals occur during the next $n-1$ slots and the probability of zero or one arrivals occurs in the terminating interval.

$$P[C_R = n] = e^{-\lambda}(1+\lambda)\left[1-(1+\lambda)e^{-\lambda}\right]^{n-1}.$$

Birdsall *et al.* [18, Eq. (11)] also compute the probability that exactly d arrivals are rejected in the next slot, provided that one or more was rejected in the previous slot. Their result is related to a relation we will derive later (Eq. (3)).

We define Q_k to be the event that the first customer in an arriving batch sees k customers already in the system and q_k to be the probability of that event. For general batch size probability mass function (pmf) a_k, the q_k's are described by the following recursive equations [17]:

$$\begin{aligned} q_1 &= \frac{q_0}{a_0}(1-a_0-a_1) \\ q_n &= \frac{1}{a_0}\left[q_{n-1} - \sum_{k=1}^{n} a_k q_{n-k}\right], \; 2 \le n < K \\ q_0 &= 1 - \sum_{n=1}^{K-1} q_n = \left[1 + \sum_{n=1}^{K-1} q_n/q_0\right]^{-1} \end{aligned}$$

The probability that a packet joins the queue, $P[J]$, is given by

$$P[J] = \frac{1-q_0 a_0}{\lambda},$$

since $1 - a_0 q_0$ is the normalized throughput. Note that q_n, $n = 1, 2, \ldots$, depends only through the factor q_0 on the buffer size, i.e., q_n/q_0 is independent of the buffer size [16, p. 236].

For later use, let us compute the probability $P[S]$ that one or more losses occurs during a randomly selected time slot. By conditioning on the system state probability q_k, we can write

$$P[S] = \sum_{k=0}^{K-1} q_k P[S|Q_k] = \sum_{k=0}^{K-1} q_k \sum_{j=K-k+1}^{\infty} a_j = \sum_{k=0}^{K-1} q_k \left(1 - \sum_{j=0}^{K-k} a_j\right)$$

Let the random variable C_C be the number of consecutively lost customers. The distribution of loss run lengths follows readily,

$$\begin{aligned} P[C_C = n] &= \sum_{s=1}^{K} P[s \text{ spaces available} \mid \text{loss occurs in slot}] \cdot a_{n+s}, \\ &= \frac{1}{P[S]} \sum_{s=1}^{K} q_{K-s} a_{n+s} \end{aligned} \tag{1}$$

as the number of arrivals in a slot is independent of the system state.

The expected number of consecutive losses can be computed from Eq. (1) or directly by observing that loss runs are limited to a single slot since the first customer in a batch will always be admitted. The expected loss run length is simply the expected number of customers lost per slot, given that a loss did occur in that slot. The expected number of customers lost in a slot is given by $\lambda(1 - P[J])$, so that

$$E[C_C] = E[\text{losses per slot} \mid \text{loss occurs in slot}] = \frac{\lambda(1 - P[J])}{P[S]} = \frac{\lambda - 1 + q_0 a_0}{P[S]}. \tag{2}$$

Numerical computations show that influence of K on the distribution of C_C is very small (see Table 1). The table also shows that $E[C_C]$ is roughly a linear function of λ.

Losses that occur when a batch arrives to a full system, i.e., a system with only one available buffer space, are independent of the system size and can thus be used to approximate the distribution of C_C quite accurately. For $K = 1$, n consecutive losses occur if and only if $n + 1$ packets arrive during the slot duration, conditioned on the fact that two or more packets arrived. Thus,

$$P[C_C = n] \approx P[C_C = n | K = 1] = \frac{a_{n+1}}{1 - a_0 - a_1} \tag{3}$$

$$\begin{aligned} E[C_C] &\approx \frac{1}{1 - a_0 - a_1} \sum_{n=1}^{\infty} n a_{n+1} \\ &= \frac{1}{1 - a_0 - a_1} \left[\sum_{n=2}^{\infty} n a_n - \sum_{n=2}^{\infty} a_n\right] \\ &= \frac{1}{1 - a_0 - a_1} [\lambda - a_1 - (1 - a_0 - a_1)] \\ &= \frac{1}{1 - a_0 - a_1} [\lambda - 1 + a_0]. \end{aligned} \tag{4}$$

As can be seen readily, the above agrees with Eq. (1) and Eq. (2) for $K = 1$.

Also, by the memorylessness property of the geometric distribution, Eq. (3) and Eq. (4) hold exactly for geometrically distributed batches with parameter p and evaluates to

$$P[C_C = n] = \frac{p(1 - p)^{n+1}}{1 - p - p(1 - p)} = p(1 - p)^{n-1}$$

$$E[C_C] = \frac{1}{p} = 1 + \lambda$$

(Regardless of what system occupancy an arriving batch sees, the packets left over after the system is filled are still geometrically distributed.)

a_k	K	$\lambda = 0.5$	$\lambda = 0.8$	$\lambda = 1$	$\lambda = 1.5$
Poisson	1	1.18100	1.30397	1.39221	1.63540
	2	1.15707	1.27511	1.36201	1.60574
	3	1.15707	1.27158	1.35911	1.60403
	4	1.15226	1.27153	1.35910	1.60403
	5	1.15238	1.27159	1.35914	1.60404
	6	1.15242	1.27160	1.35914	1.60404
	∞	1.15242	1.27160	1.35914	1.60404
Geo	any	1.5	1.8	2.0	2.5

Table 1: Expected loss run length ($E[C_C]$) for $D^{[G]}/D/1/K$ system

2.2 Influence of Service and Buffer Policies

It is natural to ask how the burstiness of losses is affected by different scheduling and buffer management policies. As scheduling policies, FIFO (first-in, first-out) and non-preemptive LIFO (last-in, first-out) are investigated. For either policy, we can either discard arriving packets if the buffer is full (*rear discarding*) or push out those packets that have been in the buffer the longest (*front discarding*). Note that this dropping policy is independent of the service policy. From our viewpoint, LIFO serves the packet at the rear of the queue. Obviously, only systems with K greater than one show any difference in behavior.

The analysis of all but FIFO with rear discarding appears to be difficult. Let us briefly discuss the behavior of FIFO and LIFO, each either with front or rear discarding.

FIFO with rear discarding: The first customer in an arriving batch always enters the buffer and will be served eventually. Thus, a loss run never crosses batch boundaries.

FIFO with front discarding: Here, a batch can be completely lost if it partially fills the buffer and gets pushed out by the next arriving batch. However, if a batch was completely lost, the succeeding batch will have at least one of its members transmitted since it must have "pushed through" until the head of the buffer.

LIFO with rear discarding: The first packet in a batch will always occupy the one empty buffer space and be served in the next slot. Again, loss runs are interrupted by packet boundaries.

LIFO with front discarding: A run of losses can consist of at most than one less than arrive in a single batch since the last customer in the batch will be served during the next slot. A loss run never straddles batch boundaries.

For all four systems, indeed over all work-conserving disciplines, the queue state distribution (and, thus, the loss probability) are the same [23–25]. The mean waiting time results favoring front dropping agree with those of [25] for general queueing systems. Clare and Rubin [24] show that the minimum mean waiting time for non-lost packets is obtained using LCFS with front dropping (referred to as preemptive buffering in [24]).

For all systems, a batch arrival causes the same number of lost packets. If there are q packets in the buffer $(0 \geq q < K)$ and a arrive in a batch, $[(q+a)-K]^+$ will be lost.

For rear dropping, at least the first packet in the batch will always enter the system, interrupting any loss run in progress. Thus, we have:

Lemma 1 *The same packets (as identified by their order of generation) will be dropped for all work-conserving service policies and rear dropping.*

Here, we visualize packets within the same batch generated sequentially in the interval $(t, t+0)$.

Lemma 2 *The distributions of loss runs for FIFO with rear and front dropping are the same.*

PROOF We number packets in the order of generation and arbitrarily within a batch so that they are served in order of increasing sequence number. We note first that the buffer for front dropping always contains an uninterrupted sequence of packets. Assume that the buffer contains an uninterrupted sequence. A service completion removes the first element of the sequence, without creating an interruption. A batch arrival that is accepted completely or will also not create a gap. A batch that pushes out some of the customers likewise will continue the sequence. Finally, a batch that pushes out all customers certainly does not create a gap. Note that this property does not hold for rear dropping, as the example of two successive batch arrivals with overflows demonstrates.

As pointed out before, the loss runs for rear dropping are confined to a single arriving batch and comprise $[q+a-K]^+$ packets. For front dropping, the losses are made up of packets already in the buffer and possibly the first part of the arriving batch. By the sequence property shown in the preceding paragraph, all these form a single loss run (again of length $[q+a-K]^+$), which is terminated by serving the next customer. Thus, while the identity of packets dropped may differ, the lengths of the loss runs are indeed the same for both policies. □

The issue is more complicated for front dropping and general service disciplines. A number of simulation experiments were performed to investigate the behavior of the four combinations of service and buffer policies, with results collected in Table 2. (The rows labeled "theory" correspond to the values for FIFO and rear dropping, computed as in discussed in the previous section.) These experiments suggest the following conjecture:

Conjecture 1 *The distribution of loss runs is the same for all combinations of FIFO, LIFO and rear and front dropping. The sample path of loss run lengths is the same for all systems except for LIFO with front dropping.*

Let us briefly outline a possible approach to a more formal analysis. We focus on one batch and construct a discrete-time chain with the state

$$(i, j) = (\text{left in buffer, consecutive losses in batch}) \qquad i \in [1, K-1];\; j \in [0, \infty)$$

The initial state, that is, the state immediately after the arrival of the batch of interest, is determined by the batch size and the system state and should be computable. The states $(0, j)$ are absorbing.

arrivals	service	dropping	$E[W]$	$1-P[J]$	$E[C_C]$
geometric	theory		1.985	0.3839	2.500
	FIFO	rear	1.984...1.987	0.3833...0.3840	2.496...2.502
		front	1.349...1.351	0.3833...0.3840	2.496...2.502
	LIFO	rear	1.984...1.987	0.3833...0.3840	2.496...2.502
		front	0.867...0.869	0.3833...0.3840	2.495...2.500
Poisson	theory		2.417	0.341	1.604
	FIFO	rear	2.416...2.418	0.340... 0.341	1.602...1.604
		front	1.582...1.584	0.340... 0.341	1.602...1.604
	LIFO	rear	2.416...2.418	0.340... 0.341	1.602...1.604
		front	0.549...0.553	0.340... 0.341	1.603...1.604

Table 2: Performance measures for geometric and Poisson arrivals, $\lambda = 1.5$, $K = 4$, 90% confidence intervals

The transition matrix is given by:

$$\begin{array}{lcll}
(i,j) \to (i,j) & = & a_1 & \forall i,j; \\
(i,j) \to (i-1,j) & = & a_0 & j > 0; \\
(i,j) \to (i-1,j+1) & = & a_2 & i,j > 0; \\
(i,j) \to (i-2,j+2) & = & a_3 & i > 1, j > 0; \\
\ldots & & & \\
(i,j) \to (i-k,j+k) & = & a_{k+1} & i > k, j > 0; \\
(i,j) \to (0,i+j) & = & \sum_{k=i+1}^{\infty} a_k & i,j > 0; \\
(0,j) \to (0,j) & = & 1 &
\end{array}$$

All other entries are zero.

The probability of j losses given initial state (i_0, j_0) is the probability distribution of being absorbed in state $(0, j)$.

Intuitively, random dropping, i.e., selecting a random packet from among those already in the buffer, should reduce the loss run lengths, particularly for large buffers. However, this policy appears to be difficult to implement for high-speed networks. Simulation results for geometric arrivals support this result, as shown in Table 3 as a front or rear dropping would result in an average loss run length of 2.5 for $\lambda = 1.5$ and 1.8 for $\lambda = 0.8$.

It should be noted that average run lengths exceed a value of two only for extremely heavy load. Thus, on average, reconstruction algorithms that can cope with two lost packets in a row should be sufficient.

3 Superposition of Periodic and Random Traffic

Up to this point, we have assumed that the stream analyzed is the only one arriving at a queue. Clearly, this is not realistic in many network situations. As a simple first model closer to actual network traffic, we again consider a discrete-time system similar to the one studied in [2]: single, first-come, first-serve queue with batch arrivals at every time slot and unit service time. The stream of interest is modeled as periodic, superimposed on background traffic with general, but i.i.d., batch arrivals. The background traffic is envisioned as the superposition of a large number of sources. For

	$\lambda = 1.5$			$\lambda = 0.8$		
K	$E[W]$	$1-P[J]$	$E[C_C]$	$E[W]$	$1-P[J]$	$E[C_C]$
3	0.913	0.4148	2.304	0.750	0.1731	1.669
4	1.509	0.3833	2.175	1.120	0.1213	1.573
6	2.842	0.3534	2.008	1.771	0.0659	1.451
8	4.304	0.3415	1.915	2.300	0.0386	1.382
10	5.851	0.3365	1.849	2.721	0.0234	1.328
15	9.884	0.3331	1.755	3.399	0.0071	1.238

Table 3: Effect of random discarding for system with geometrically distributed batch arrivals

brevity, the two streams are abbreviated as "foreground traffic" (FT, stream 1) and "background traffic" (BT, stream 0). If the foreground traffic is periodic, it also referred to in the literature as constant bit-rate (CBR) [26] or continuous bit-stream oriented (CBO) [27,28] traffic. This traffic model is motivated by voice and video applications, where the source of interests emits packets periodically during a talk spurt, a video scan line or even a whole image[3]. To gain insight into the loss behavior, we require transient rather than steady-state results.

A number of authors have investigated the issue of superimposing different traffic streams. Kaplan [29] develops steady-state waiting time results in the transform domain for a continuous-time system with a single deterministic server with infinite buffer used by a primary, deterministic (periodic) stream and a secondary or background stream with Poisson characteristics. A more general continuous-time model is treated in [30], where the primary stream can have any random interarrival and service distribution and the secondary stream is a sequence of deterministic arrivals whose interarrival times are not necessarily identical. The steady-state distribution of the virtual waiting time for the infinite-buffer case is described by an integral equation. For a secondary stream with random arrivals, consult [31]. A number of authors [32–34] analyze the superposition of independent periodic arrival streams at a single server with infinite waiting space. Bhargava *et al.* analyze the same system using the ballot theorem [28]. In general, the queue length survivor function for infinite queues can be used to approximate the cell loss probability for finite queues for sufficiently low (order of 10^{-9}) loss probabilities [34].

Closely related to the issue of superimposed heterogeneous traffic streams is that of overflow problems, covered extensively in the literature. While many overflow models, motivated by call set up in the switched telephone network, assume loss systems, i.e., no buffer, (among others, [35,36]), a number of authors consider delay systems. For example, Matsumoto and Watanabe [37] derive approximations for the individual and combined mean waiting time and overflow probability in a finite system with two overflow streams, modeled as interrupted Poisson processes (IPPs), and one Poisson stream, which is a special case of the more general model of a $MMPP/M/c/c+k$ queue analyzed by Meier-Hellstern [38]. In most cases, the distribution of the number of busy servers in an infinite-server system accepting the overflow input is evaluated. In [39], the Laplace transform of the inter-overflow time matrix probability density function is provided. In [40], the related model of alternating voice and data input to a finite buffer is treated.

Kuczura [41] treats infinite queues (in continuous time) with one Poisson and one renewal input, in particular the $GI+M/M/1$ and $GI+M/M/\infty$ systems. In the context of discrete-time single-server queues with infinite buffer, correlated inputs and unity service time, Gopinath and

[3] e.g., video sources with compression on whole image and/or smoothing buffer

Morrison [42] include an example of a two-class model with different arrival statistics.

Superimposed traffic streams of the same statistical description are found in the analysis of packet voice multiplexers [43], usually in the form of approximations [44, 45].

In the following sections, we set out to analyze in some detail three implementations of the model described at the beginning of this section, increasing in generality as we progress through the section. The first two models assume a constant, deterministic interarrival time for the foreground traffic. The first model (Section 3.1) assumes that the foreground traffic arrives at periodic intervals. It always either attempts to enter the queue before all background traffic arriving during the same slot (*first-admittance system*) or attempts to enter after all background traffic has competed for buffer space (*last-admittance system*). These policies provide bounds when the system practices source discrimination, i.e., has sources compete for buffer space in a specific order as is often the case in packet switches. The more general model of section 3.2 allows these two admittance policies, but also the more interesting case where the foreground packet is positioned randomly among the cells from the background stream. Finally, the third model in section 3.3 straightforwardly extends the analysis to the case where the interarrival time of the foreground stream is a random variable.

3.1 Constant Period, Fixed Position: First and Last Admittance Systems

For simplified analysis and to obtain best or worst-case results, we first assume that the cells that are part of the periodic stream arrive as either the first (early arrival) or last cell (late arrival) during a slot. The case when the FT cell always arrives first is of no particular interest as it will always get admitted. To evaluate performance for the pessimistic late-arrival assumption, we embed a discrete-time Markov chain at the (imagined) instant just after the arrival of the background traffic and just before the arrival of the cell from the periodic stream.

We first consider the simple $D^{[X]}/D/1/K$ system with batch size pmf a_k and transition matrix Q embedded after the arrivals, with states labeled 0 through K.

$$Q = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{K-1} & 1-\sum_{j=0}^{K-1} q_{0j} \\ a_0 & a_1 & a_2 & \cdots & a_{K-1} & 1-\sum_{j=0}^{K-1} q_{1j} \\ 0 & a_0 & a_1 & \cdots & a_{K-2} & 1-\sum_{j=0}^{K-1} q_{2j} \\ 0 & 0 & a_0 & \cdots & a_{K-3} & 1-\sum_{j=0}^{K-1} q_{3j} \\ \cdots \\ 0 & 0 & 0 & \cdots & a_1 & 1-\sum_{j=0}^{K-1} q_{K-1,j} \\ 0 & 0 & 0 & & a_0 & 1-\sum_{j=0}^{K-1} q_{Kj} \end{bmatrix}$$

Given this transition matrix for the background traffic only, the transition matrix for the combined system, embedded just before arrivals of the periodic traffic with period τ, is given by

$$\tilde{Q} = \begin{bmatrix} h_{10} & h_{11} & h_{12} & \cdots & h_{1,K-1} & 1-\sum_{j=0}^{K-1} \tilde{q}_{0j} \\ h_{20} & h_{21} & h_{22} & \cdots & h_{2,K-1} & 1-\sum_{j=0}^{K-1} \tilde{q}_{1j} \\ \cdots \\ h_{K,0} & h_{K,1} & h_{K,2} & \cdots & h_{K,K-1} & 1-\sum_{j=0}^{K-1} \tilde{q}_{K-1,j} \\ h_{K,0} & h_{K,1} & h_{K,2} & \cdots & h_{K,K-1} & 1-\sum_{j=0}^{K-1} \tilde{q}_{K-1,j} \end{bmatrix}$$

where $h_{ij} = [Q^\tau]_{ij}$. Note that the last two rows are identical since the customer from the periodic stream will not be admitted if the system is already in state K prior to its arrival.

After the usual computations, we arrive at the vector of steady-state probabilities as seen by a periodic arrival, $\gamma = (\gamma_0, \ldots \gamma_K)$. Then, the loss probability of periodic arrivals is simply γ_K and the expected waiting time for admitted customers is

$$E[W_1] = \frac{\sum_{k=0}^{K-1} k\gamma_k}{1-\gamma_K}.$$

Finally, $\tilde{q}_{K,K}$ is the conditional loss probability (given that the previous customer was lost). The expected number of consecutively lost periodic customers is a function of the conditional loss probability and given by the properties of the geometric distribution as

$$E[C_C] = \frac{1}{1-\tilde{q}_{K,K}}.$$

As an example, consider a system under overload, with a geometrically distributed batch size of mean 0.7, a period τ of 3 (for a combined average load of 1.033) and a system size of 3. The periodic traffic experiences a loss of 15.74% and an average wait of 0.875. The conditional loss probability is 20.03%, with an average number of 1.25 consecutively lost customers (loss run length). If losses were independent, the average loss run length would be 1.187. Thus, even under heavy load and a periodic stream that contributes a significant fraction of the total traffic, the loss run lengths in that system are very close to one. The calculations were confirmed by simulations.

3.2 Constant Period, Random Position

In this variation, called the random-arrival system, the periodic cell is randomly interspersed with the background traffic. Again, a periodic arrival occurs every τ slots. We will refer to a sequence of τ slots commencing with a slot with a periodic arrival as a *cycle*. The first slot in a cycle bears the index zero.

3.2.1 FT Loss Probability and Waiting Time

The probability that k BT cells enter the queue before the FT cell is given by

$$b_k = \sum_{j=k}^{\infty} \frac{a_j}{j+1} \text{ for } k = 0, 1, 2, \ldots, \tag{5}$$

since the cell has an equal chance to be at any position of the total batch arriving to the queue. The expected number of cells entering the queue prior to the FT cell can be computed by rearranging the order of summation as

$$E[B] = \sum_{k=1}^{\infty} kb_k = \sum_{k=1}^{\infty} k \sum_{j=k}^{\infty} \frac{a_{j+1}}{j} = \sum_{j=1}^{\infty} \frac{a_j}{j+1} \sum_{k=1}^{j} k = \sum_{j=1}^{\infty} a_j \frac{j(j+1)}{2(j+1)} = \frac{1}{2} \sum_{j=1}^{\infty} ja_j = \frac{1}{2} E[A],$$

as intuition suggests. Here, $E[A]$ denotes the average batch size. By virtue of its random position, the FT cell is (in distribution) followed by as many BT cells as precede it. However, these two quantities are not independent within a single slot.

Through the remainder of this section, we will see that the results for the late-arrival system follow from those of the random-arrival system by simply replacing b_k by a_k.

As a preliminary step in the transient analysis, we write down the Markov chain for the background traffic only, embedded just before the first arrival in a batch. Here, the state space encompasses states 0 through $K-1$.

$$P = \begin{bmatrix} a_0 + a_1 & a_2 & a_3 & \dots & a_{K-1} & 1 - \sum_{j=0}^{K-2} p_{0j} \\ a_0 & a_1 & a_2 & \dots & a_{K-2} & 1 - \sum_{j=0}^{K-2} p_{1j} \\ 0 & a_0 & a_1 & \dots & a_{K-3} & 1 - \sum_{j=0}^{K-2} p_{2j} \\ \dots & & & & & \\ 0 & 0 & 0 & \dots & a_0 & 1 - \sum_{j=0}^{K-2} p_{K-1,j} \end{bmatrix} \tag{6}$$

Similar to the case of fixed position arrivals, we now embed a Markov chain with state transition probability matrix $\tilde{P}^{(j)}$ just prior to the first arrival in the jth slot of a cycle. This first arrival could be from either the periodic or the background stream. Since we embed the chain just after a departure, the state space reaches up to $K-1$, not K. The matrix $\tilde{P}^{(j)}$ is given by

$$\tilde{P}^{(j)} = \begin{cases} P^{\tau-j} P' P^{j-1} & \text{for } 0 < j \leq \tau \\ P' P^{(\tau-1)} & \text{for } j = 0. \end{cases} \tag{7}$$

Here, P' is the transition matrix computed with the shifted batch size distribution that takes account of the single foreground arrival: $\mathbf{a}'$, $a'_j = a_{j+1}, j = 1, \dots, K$, $a'_0 = 0$.

The steady state probabilities seen by the first customer in the jth slot of a cycle are denoted by the vector $\tilde{\pi}^{(j)} = (\tilde{\pi}_0^{(j)}, \dots, \tilde{\pi}_{K-1}^{(j)})$. The steady-state probability of loss for periodic arrivals is given by the convolution of position and state distributions,

$$\begin{aligned} P[L_1] &= P[\text{before} + \text{state} \geq K] = \sum_{k=1}^{K} \tilde{\pi}_{K-k}^{(0)} P[\text{before} \geq k] = \sum_{k=1}^{K} \tilde{\pi}_{K-k}^{(0)} \left(1 - \sum_{j=0}^{k-1} b_j\right) \\ &= 1 - \sum_{k=1}^{K} \tilde{\pi}_{K-k}^{(0)} \sum_{j=0}^{k-1} b_j, \end{aligned} \tag{8}$$

since position and state are independent. Alternatively, given the above measures, the distribution of the state seen by a periodic arrival is given by

$$\pi_k = \sum_{j=0}^{k} \tilde{\pi}_j^{(0)} b_{k-j}.$$

Then, the loss probability $P[L_1]$ for periodic traffic is seen to be equal to π_K. $P[L_1]$ and π_k for the last-admittance system can be computed by replacing b_k by a_k. The first-admittance system has $b_0 = 1, b_k = 0$ for $k > 0$, so that $\pi_k = \tilde{\pi}_k^{(0)}$.

An arriving periodic customer has to wait for those already in the queue at the beginning of the slot and those customers within the same batch that enter the queue ahead of the periodic customer. Since these two components are independent, the distribution of the waiting time, $P[W = k]$, for the periodic traffic is given by the convolution

$$P[W = k] = \frac{1}{1 - P[L_1]} \sum_{j=0}^{k} \tilde{\pi}_j^{(0)} b_{k-j} = \frac{\pi_k}{1 - P[L_1]} \text{ for } k = 0, \dots, K-1.$$

Again, the last-admittance and first-admittance quantities are computed by replacing b_k as described above.

3.2.2 BT Loss Probability

The loss probability for BT cells, $P[L_0]$, is computed indirectly through the channel utilization. The channel idle probability, without regard to traffic type, is computed as

$$P[\text{idle}] = \frac{a_0}{\tau} \sum_{j=1}^{\tau-1} \tilde{\pi}_0^{(j)}$$

since the slot steady-state probabilities seen by an arriving batch differ for each slot within a cycle of length τ. The term for j equal to zero is omitted since the FT arrival assures that the channel is never idle. From the idle probability, the loss probability for the combined stream, called $P[L]$, follows from flow-conservation as

$$P[L] = \frac{1 - P[\text{idle}]}{\lambda_0 + 1/\tau}.$$

Background traffic constitutes a fraction of

$$\frac{\lambda_0}{\lambda_0 + 1/\tau} = \frac{1}{1 + 1/(\lambda_0 \tau)},$$

while foreground traffic contributes

$$\frac{1/\tau}{\lambda_0 + 1/\tau} = \frac{1}{1 + \lambda_0 \tau}.$$

Given these fractions, the total loss probability can be expressed as the weighted sum of the stream loss probabilities, $P[L_0]$ and $P[L_1]$, as

$$P[L] = \frac{P[L_0]}{1 + 1/(\lambda_0 \tau)} + \frac{P[L_1]}{1 + \lambda_0 \tau}.$$

Then, the loss probability for the background traffic can be solved for:

$$P[L_0] = (1 + 1/\lambda\tau) \left[P[L] - \frac{P[L_1]}{1 + \lambda\tau} \right] \tag{9}$$

The same results for the BT loss probability $P[L_0]$ can also be obtained by direct methods, similar to those used to compute $P[L_1]$. First, we need to compute the probability mass function of the number of cells that arrive in the batch prior to a randomly selected customer. Since arrivals occur in batches, we are dealing with a random incidence problem. For all but slot zero in a cycle, we have that the distribution of the batch size that a random customer finds itself in is given by

$$\tilde{a}_n = \frac{n a_n}{\lambda_0}$$

Thus, the probability that a random customer finds k other customers ahead of it in its batch, β_k, is found by observing that a random customer is equally likely to occupy any position within the arriving batch. For all but the first slot (index j) within a cycle, we have

$$\beta_k^{(j)} = \sum_{i=k+1}^{\infty} \frac{\tilde{a}_i}{i} = \frac{1}{\lambda_0} \sum_{i=k+1}^{\infty} a_i \text{ for } 0 < j < \tau - 1$$

Similarly, for slot zero in a cycle, the BT cell is equally likely to occupy positions 1 through $n+1$ given that the batch it arrived in had size n. Thus,

$$\beta_k^{(0)} = \frac{1}{\lambda_0} \sum_{j=k}^{\infty} \frac{j a_j}{j+1}$$

The loss probability for BT cells, $P[L_0]$, is then given by a relation analogous to Eq. (8), except that we average over all τ slots in a cycle:

$$P[L_0] = \frac{1}{\tau} \sum_{j=0}^{\tau-1} \left(1 - \sum_{k=1}^{K} \tilde{\pi}_k^{(j)} \sum_{i=0}^{k-1} \beta_i^{(j)}\right) = 1 - \frac{1}{\tau} \sum_{j=0}^{\tau-1} \left\{ \sum_{k=1}^{K} \tilde{\pi}_k^{(j)} \sum_{i=0}^{k-1} \beta_i^{(j)} \right\} \tag{10}$$

Both methods of computing the BT loss, Eq. (9) and Eq. (10) require approximately the same computational effort that increases linearly in τ, assuming matrix powers are computed through eigenvalue methods.

It is tempting to embed a Markov chain just before FT arrivals as it would yield both the unconditional and conditional loss probability for FT packets. This, however, is difficult for random FT arrivals since the number of BT arrivals after the FT arrival depends on the number that have arrived before (in the same slot). The system state seen by the FT arrival in turn depends on the number of BT arrivals that precede the FT arrival.

3.2.3 Conditional Loss Probability

In order to compute the conditional loss probability, we note that the state at the next slot after a FT loss is always $K-1$. Also, all BT cells arriving within the same slot and following a FT cell that is lost are also lost and can thus be ignored. Thus, the dependence of the number of arrivals after the FT cell on the state is no longer a problem. Then, the conditional loss probability is given by

$$P[L_1(n)|L_1(n-1)] = \left[\hat{P}^{\tau-1} B\right]_{K-1,K},$$

where $L_1(n)$ indicates the event that FT arrival n is lost. P denotes the transition matrix defined in Eq. (6), but padded to K rows and columns by zero elements. The K by K transition B describes the state transition caused by the arrivals preceding the FT cell in the same slot. For randomly placed FT cells, the elements b_k are computed according to Eq. (5).

$$B = \begin{bmatrix} b_0 & b_1 & b_2 & \dots & 1-\sum_{j=0}^{K-1} b_j \\ 0 & b_0 & b_1 & \dots & 1-\sum_{j=0}^{K-2} b_j \\ 0 & 0 & b_0 & \dots & 1-\sum_{j=0}^{K-3} b_j \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

For late-arrival FT, the matrix B contains the batch probabilities:

$$B = \begin{bmatrix} a_0 & a_1 & a_2 & \dots & 1-\sum_{j=0}^{K-1} a_j \\ 0 & a_0 & a_1 & \dots & 1-\sum_{j=0}^{K-2} a_j \\ 0 & 0 & a_0 & \dots & 1-\sum_{j=0}^{K-3} a_j \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

After having evaluated the conditional loss probability algebraically, we briefly highlight two structural properties. For this queue, the independence of loss correlation and buffer size observed earlier [2] holds exactly for certain (common) combinations of τ and K.

Theorem 1 *The conditional loss probability for FT packets is independent of K for values of $K \geq \tau$.*

Intuitively, this result holds since for $\tau \leq K$, the buffer that was full on a FT arrival will not empty until the next FT arrival slot, if at all. In other words, there will be a departure at every slot until the next FT arrival after a FT arrival has been lost. The probability that the next FT arrival again sees K in the system depends only on the arrival probabilities, not on the system size. We prove the result formally by sample path arguments.

PROOF We track the number of available spaces rather than the system occupancy. After a FT loss, there will be zero available slots. Thus, the conditional loss probability can be viewed as the probability of reaching state zero at the next FT arrival given that the (lost) FT arrival left state zero behind. At the end of each slot, the number of available spaces increases by one. Here, we make use of the fact that the condition $\tau \leq K$ guarantees a departure. At the beginning of each slot, it decreases by the number of BT arrivals, but not below zero. This behavior is completely independent of the system size. □

Theorem 2 *(i) The conditional loss probability is greater than or equal to the unconditional loss probability. (ii) The conditional loss probability decreases monotonically in K.*

PROOF The following argument shows part (i). For any customer, the probability that it is lost depends only on the state after the arrivals of the previous slot ("state") and the number of arrivals that precede the arrival during the same slot ("arrivals"). The latter quantity is by system definition independent of the state of the queue. Clearly, the loss probability for the next customer increases with increasing state[4]. For the loss probability, we consider a random customer. The state lies somewhere between 0 and K. In contrast, for the conditional loss probability, we only consider those customers that follow a lost customer of the same stream. Their state is known to be K. Since, in particular, some non-zero number of random customers will see states of less than K, their loss probability will be lower than those seeing K as state.

Part (ii) follows from a sample-path argument. Consider two systems that are distinguished only by their buffer size: one has a buffer of K slots, the other $K + 1$. Since we are interested in the conditional loss probability, we assume that a loss has occurred in both systems. The arrival process in the following slots is independent of the loss event. Consider the state just after a batch arrival. As long as the $K + 1$ system occupancy stays above zero, the state of the $K + 1$ system is simply that of the K plus one. In that case, the probability that the next FT arrival finds the system full is the same for both systems. (This is the argument used in the proof of theorem 1.)

On the other hand, if the $K + 1$ system reaches zero, we know that the K system will also be at zero since it must have been at zero at the previous slot already. (The occupancy can only decrease by at most one in this single-server system). Thus, from that point on the two systems share the same occupancy. The next FT arrival will only be lost if the occupancy reaches K or $K + 1$, respectively. Since the arrivals are the same, the probability of that event for the smaller system is larger. □

3.2.4 Numerical Examples

To gain a better understanding of the dependence of conditional and unconditional loss probabilities on system structure and parameters, we graph and discuss a number of numerical examples below. All graphs display unconditional and conditional loss probabilities for FT packets as well as the

[4] Formal proof by sample path arguments

unconditional loss probability for the BT stream. (The conditional loss probability for BT is less interesting since we assumed that the BT stream is the aggregate of several streams.)

In Fig. 1, the loss behavior for a BT stream consisting of either geometrically and Poisson distributed batches (mean batch size of 0.8) is shown as a function of the system size K. The FT period is held constant at $\tau = 10$. We see that the unconditional loss probabilities for the BT and FT stream differ, at least for larger values of K, by a roughly constant factor. That factor is larger for geometric than for Poisson arrivals, as the burstiness of the latter is smaller (and thus closer to that of the periodic traffic). In comparing unconditional loss probabilities across BT batch distributions, we note that the Poisson arrivals are associated with loss probabilities that decrease much more rapidly with increasing system size, again reflecting the lower burstiness. As expected, the conditional loss probability monotonically decreases as K increases, but is independent of K for $K \geq \tau$, as predicted by Theorem 1 and shows very little dependence for values of $K \geq \tau/2$. (This behavior seems to be valid across the range of K and τ, judging from sampling experiments not shown here.) For interesting loss probabilities, the conditional loss probability is several orders of magnitude larger than the unconditional loss probability, but the average number of consecutively lost packets is still very close to 1 (1.049 for geometric arrivals and the given set of parameters).

Plotting loss against FT period τ, as in Fig. 2, yields similar conclusions. In addition, we note that the conditional loss probability approaches the unconditional loss probability from above as τ increases, reflecting the fact that as the FT arrivals are spaced further and further apart, the congestion experienced by one arrival is dissipated by the time of the next FT arrival. The asymptotic behavior as τ approaches infinity will be discussed shortly, in Section 3.2.5. In the figure, the asymptotic loss probabilities are marked with a filled circle and rhombus. It is worth noting that a single 64 KBit/second voice connection on a 150 MBit/second data link corresponds to a value of τ of approximately 2300. At that value, losses are virtually uncorrelated (according to our definition of loss correlation.)

Rather than comparing batch size distributions, Fig. 3, Fig. 4, and Fig. 5 elucidate the connection between loss and the buffer admission policy for FT cells. Placing the FT arrivals randomly among the BT arrivals yields a conditional and unconditional probability that is better by factor of about five than that for the late-admittance system. The factor stays rather constant for all but the smallest values of K. Except for values of τ of less than, say, ten, the BT loss probability is largely unaffected by the treatment of FT arrivals. Beyond a certain value of τ, the overall system load and therefore the FT loss probability change relatively little.

In Fig. 6, the FT period τ is varied while the overall system load remains constant at $\lambda = 0.8$. As the contribution of periodic traffic diminishes and random traffic increases, the increased overall burstiness leads to increased loss probability. At the same time, the increased spacing of FT arrivals has the conditional loss probability approach the unconditional one as losses become independent. Again, the spacing by a constant factor of roughly five in both conditional and unconditional loss is observed between the two FT arrival patterns.

Almost all queueing systems show a roughly exponential rise in loss probability as the system load approaches one. This behavior is also found in Fig. 7 for the loss probabilities of both traffic types as the background traffic is varied from $\lambda = 0.2$ to 0.9. The buffer size and FT period are held constant at $K = 10$ and $\tau = 10$, respectively.[5] The flattening of the loss curve seen here is also common to all finite-buffer systems. The more interesting observation in our context, however, is that the expected loss run length varies relatively little, staying loss to the value of one associated with random loss. Note the vastly different scales of the left and right ordinate in Fig. 7. While the loss probability ranges over eleven orders of magnitude, the mean loss run length changes by

[5]Thus, the expected run length would remain the same for all values of K greater or equal to 10.

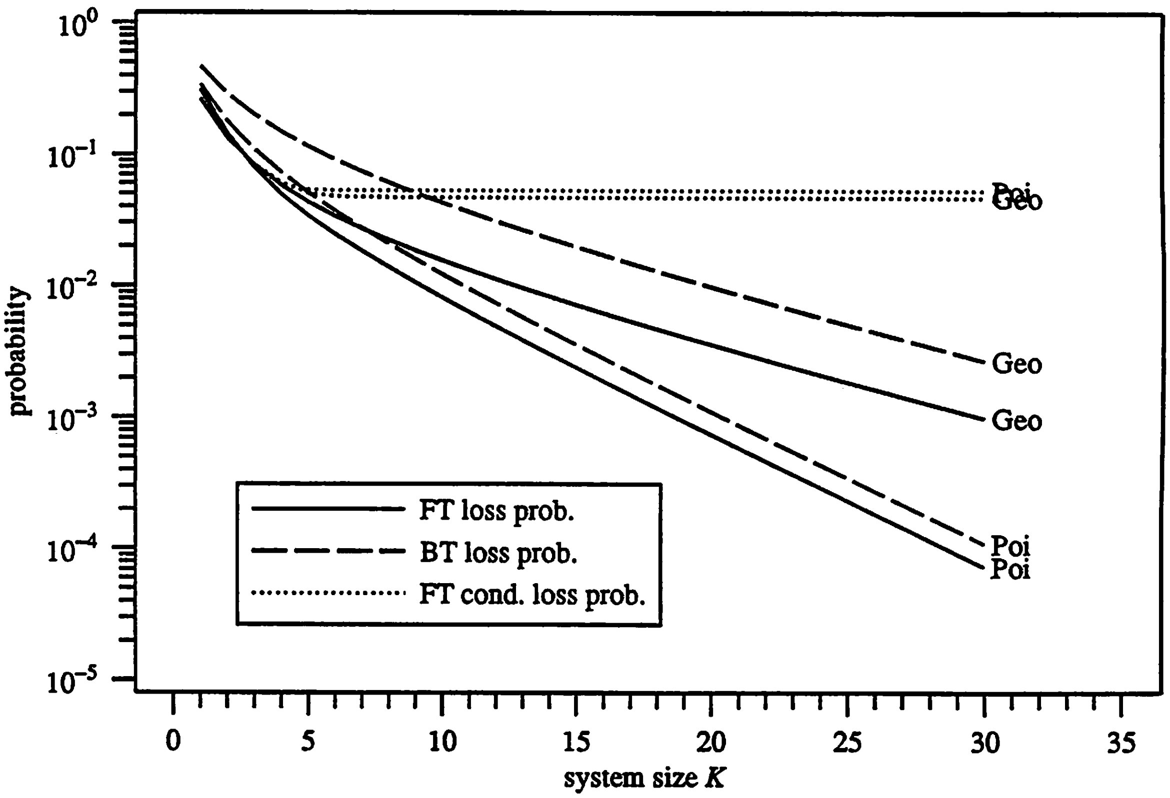

Figure 1: Loss probability and conditional loss probability for Poisson or geometrically distributed background traffic BT with $\lambda_0 = 0.8$ and periodic traffic FT with period $\tau = 10$, as a function of system size K

less than ten percent.

Finally, Fig. 8 provides a graphic illustration of the effect of the conditional loss probability r on the distribution of loss run lengths. Since the loss run lengths are geometrically distributed, the tail of the run length distribution decays rapidly even for values of r significantly larger than "typical" loss probabilities.

3.2.5 Asymptotic Analysis in τ

It is instructive to take a closer look at the case when τ approaches infinity. We will show in this section that all three performance measures (loss of FT and BT, conditional loss of FT) converge to the same value for Poisson BT, while FT and BT loss probabilities for geometric BT converge to different values. The convergence of the unconditional and conditional loss probability for all batch distributions was pointed out earlier and is to be expected since the state seen by one FT arrival is less and less correlated to that seen by the next FT arrival as the spacing between FT arrivals increases.

The unconditional loss probabilities for both types of traffic as τ approaches infinity is computed intuitively by ignoring the FT arrivals. The loss probability can easily be derived formally. We

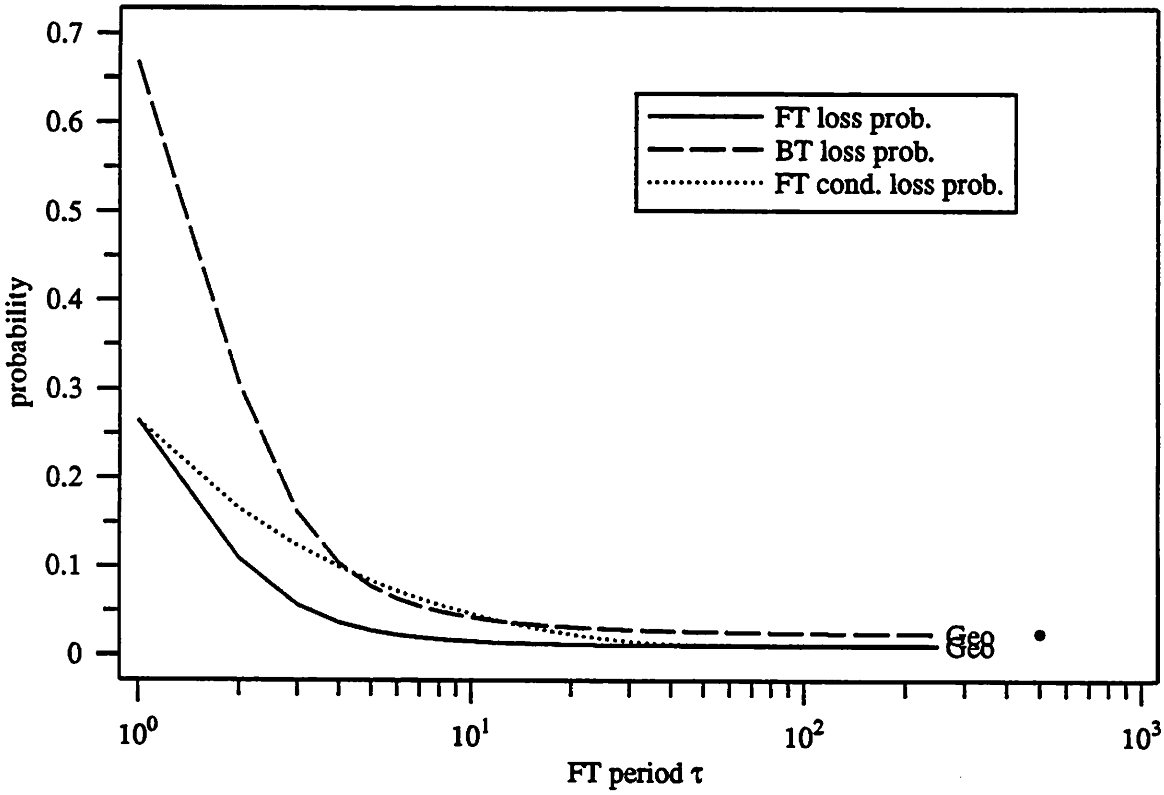

Figure 2: Loss probability and conditional loss probability for Poisson or geometrically distributed background traffic BT with $\lambda_0 = 0.8$ and periodic traffic FT (random arrival), as a function of τ; system size $K = 10$

define

$$b_j = \sum_{i=j}^{\infty} \frac{a_i}{i+1}$$
$$\beta_j = \sum_{i=j+1}^{\infty} \frac{a_i}{\lambda}.$$

Also, $\bar{\pi}_k$ is defined as the steady-state distribution computed from $\bar{P}$. Using standard arguments [46, p. 45], we take the limit of Eq. (10):

$$\lim_{\tau\to\infty} P[L_0] = 1 - \sum_{k=1}^{K}\left\{\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{j=0}^{\tau-1}\bar{\pi}_k^{(j)}\sum_{i=0}^{k-1}\beta_i^{(j)}\right\} = 1 - \sum_{k=1}^{K}\left\{\lim_{\tau\to\infty}\bar{\pi}_k^{(\tau)}\sum_{i=0}^{k-1}\beta_i^{(\tau)}\right\} = 1 - \sum_{k=1}^{K}\left\{\bar{\pi}_k\sum_{i=0}^{k-1}\beta_i\right\}$$

Here, $\bar{\pi}_k^{(j)}$ is the steady state distribution for the jth slot, which can be computed as one row of $[P^{(j)}]^{\infty}$. Taking the limit for the state transition matrix used to compute $\bar{\pi}$, we obtain, $\lim_{j\to\infty}\bar{P}^{(j)} = P^0 P' P^{\infty} = P' P^{\infty} = P^{\infty}$. The last step follows since P' is stochastic and all rows are equal to the steady-state state distribution in matrix P^{∞}.

Hence, every slot has the same transition matrix and state distribution. The derivation for the limit of $P[L_1]$ proceeds similarly. Therefore, the loss probabilities for background and foreground

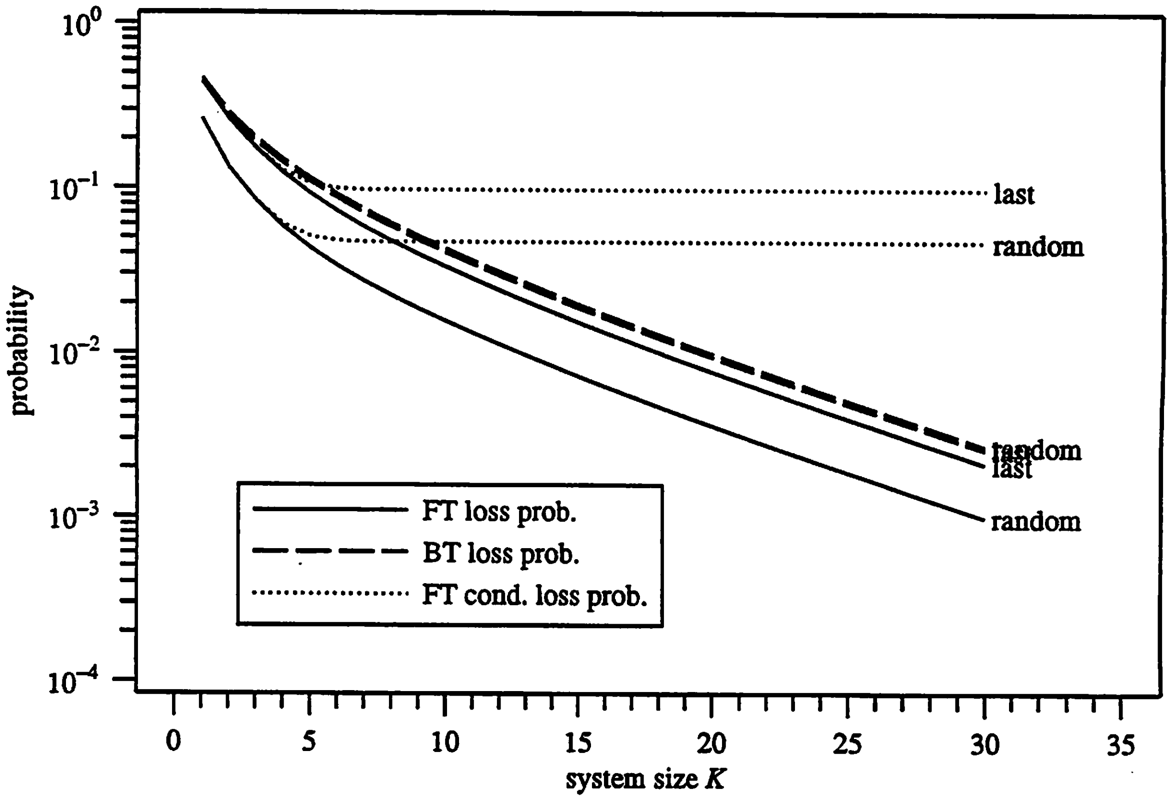

Figure 3: Loss probability and conditional loss probability for geometrically distributed background traffic BT with $\lambda_0 = 0.8$ and periodic traffic FT with period $\tau = 10$, as a function of system size K

traffic are given by, respectively,

$$P[L_0] = 1 - \sum_{k=1}^{K} \tilde{\pi}_{K-k} \sum_{i=0}^{k-1} \beta_i \quad (11)$$

$$P[L_1] = 1 - \sum_{k=1}^{K} \tilde{\pi}_{K-k} \sum_{i=0}^{k-1} b_i. \quad (12)$$

$$(13)$$

The two loss probabilities are equal if and only if b_k equals β_k. For the geometric distribution with parameters p and q, due to its memorylessness property, β_j equals a_j, so that

$$\sum_{j=0}^{k-1} \beta_j = q^{k-1}.$$

No closed-form expression appears to exist for the corresponding sum of b_j. Clearly, β_j is not equal to b_j for the geometric distribution. For the Poisson distribution, however, we see that β_j and b_j, and therefore the loss probabilities, are indeed the same:

$$b_j = \sum_{i=j}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i+1)i!} = e^{-\lambda} \sum_{i=j}^{\infty} \frac{\lambda^i}{(i+1)!}$$

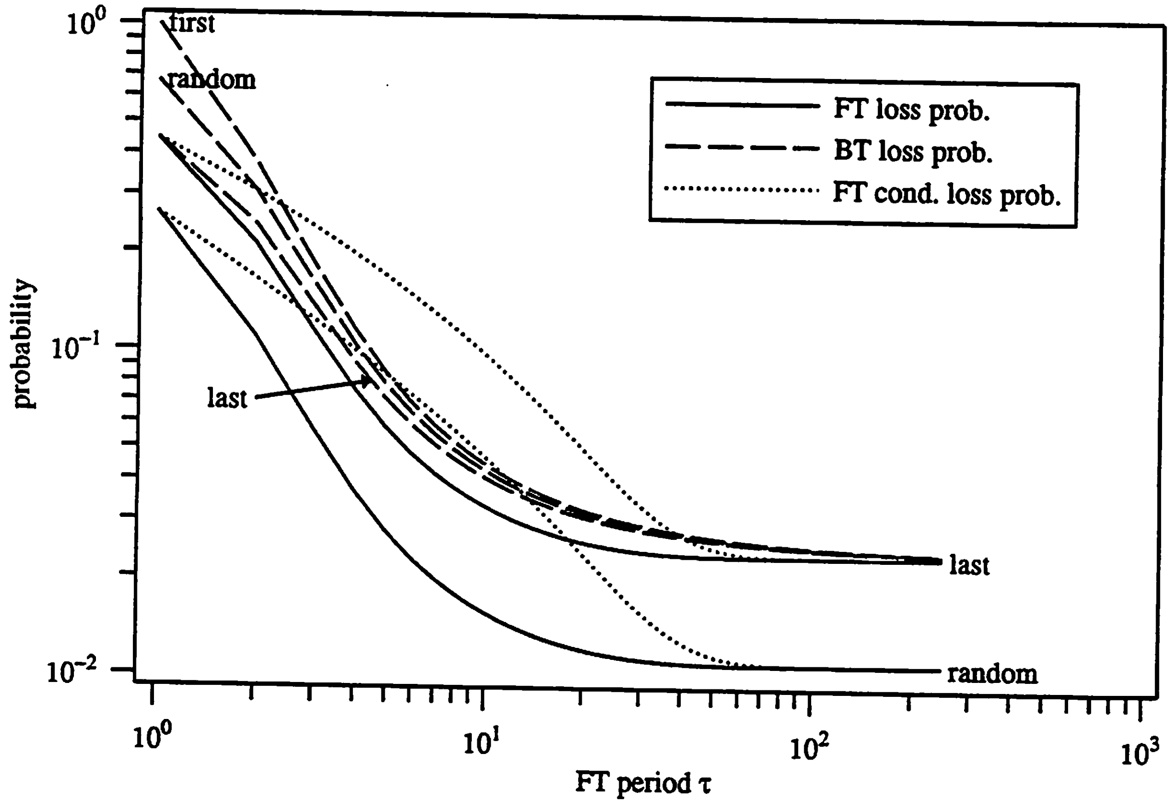

Figure 4: Loss probability and conditional loss probability for geometrically distributed background traffic BT with $\lambda_0 = 0.8$ and periodic traffic FT, as a function of τ; system size $K = 10$

$$\beta_j = \frac{1}{\lambda} \sum_{i=j+1}^{\infty} \frac{e^{-\lambda} \lambda^i}{i!} = e^{-\lambda} \sum_{i=j}^{\infty} \frac{\lambda^i}{(i+1)!}$$

The results are illustrated by Fig. 2, where the asymptotic values for $\tau \to \infty$ are indicated by a bullet or a rhombus.

3.3 Random Period, Random Position

Previously, the foreground source was assumed to transmit at periodic intervals of constant duration τ. We now relax this restriction and allow the interarrival time of the foreground stream, τ, to be a random variable. By simple conditioning, the results for the conditional loss probability in the previous section can be extended to cover this case. If τ can assume only values strictly greater than zero, that is, the foreground cells do not arrive in batches, we can write

$$P[L_1(n)|L_1(n-1)] = \sum_{k=1}^{\infty} P[L_1(n)|L_1(n-1), \tau] P[\tau = k].$$

The evaluation of this sum is made easier by the fact that the conditional loss probability approaches the unconditional loss probability as τ gets large.

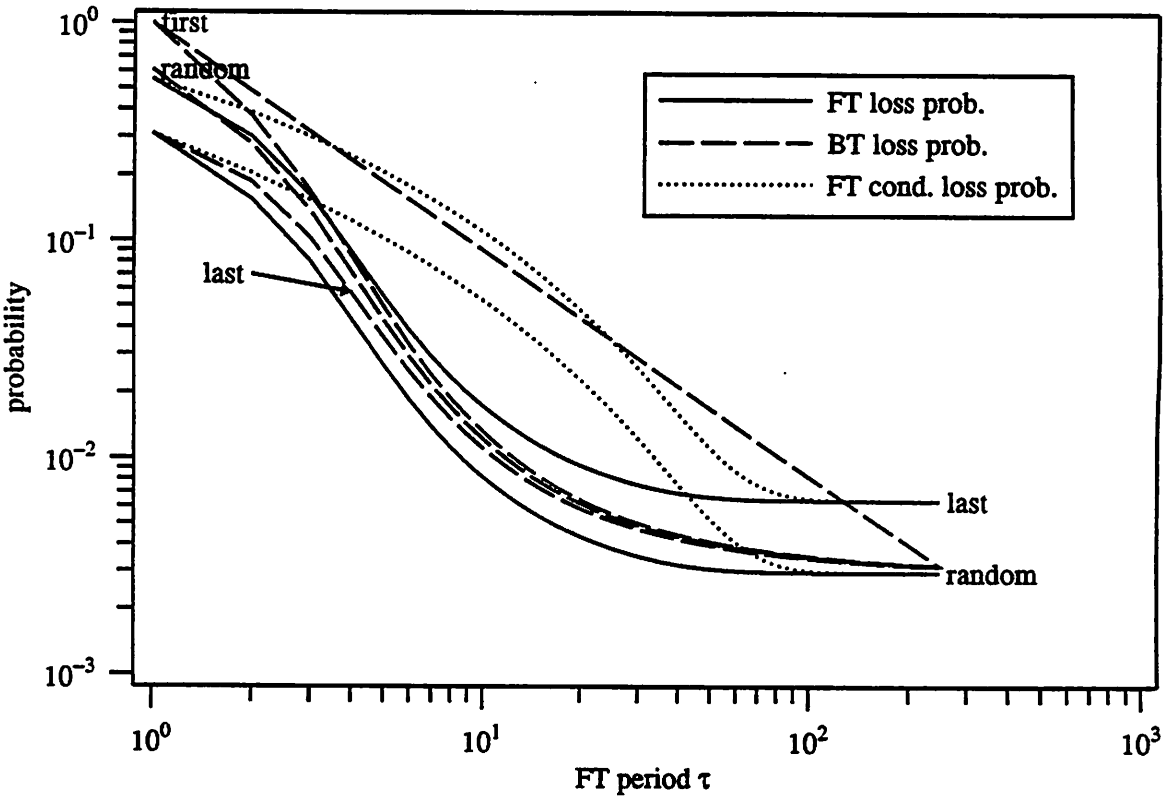

Figure 5: Loss probability and conditional loss probability for Poisson distributed background traffic BT with $\lambda_0 = 0.8$ and periodic traffic FT, as a function of τ; system size $K = 10$

If batch arrivals are permitted for the foreground stream[6], renewal-type effects need to be taken into account and the conditional loss probability is written as

$$P[L_1(n)|L_1(n-1)] = P[\tau^* = 0] + \sum_{k=1}^{\infty} P[L_1(n)|L_1(n-1), \tau]P[\tau^* = k]. \tag{14}$$

Here, τ^* denotes the distribution as seen by a random foreground arrival, which is related to the interarrival distribution seen by a random observer by [48]

$$P[\tau^* = k] = \frac{kP[\tau = k]}{E[\tau]}$$

The first term in the sum of Eq. (14) is due to the fact that once an arrival is lost within a slot, all subsequent arrivals within the same slot will also be lost.

Computation of the unconditional loss probability is more difficult, requiring a two-dimensional Markov chain, except for the case where the interarrival distribution translates into a distribution of the number of FT packets per slot that is identical and independent from slot to slot. This is satisfied for the geometric interarrival distribution with support $[1, \infty]$, where each slot sees zero or one FT arrivals according to a Bernoulli process.

[6] In most applications, batch arrivals from the foreground stream seem unlikely unless its peak rate exceeds the channel rate or a fast link feeds a switch with a slow outgoing link. The Knockout switch [47], for example, allows at most one packet from each input to reach the concentrator. Depending on the switch hardware to be modeled, the FT cells can either be treated as arriving consecutively or randomly distributed within the total arrivals in a slot.

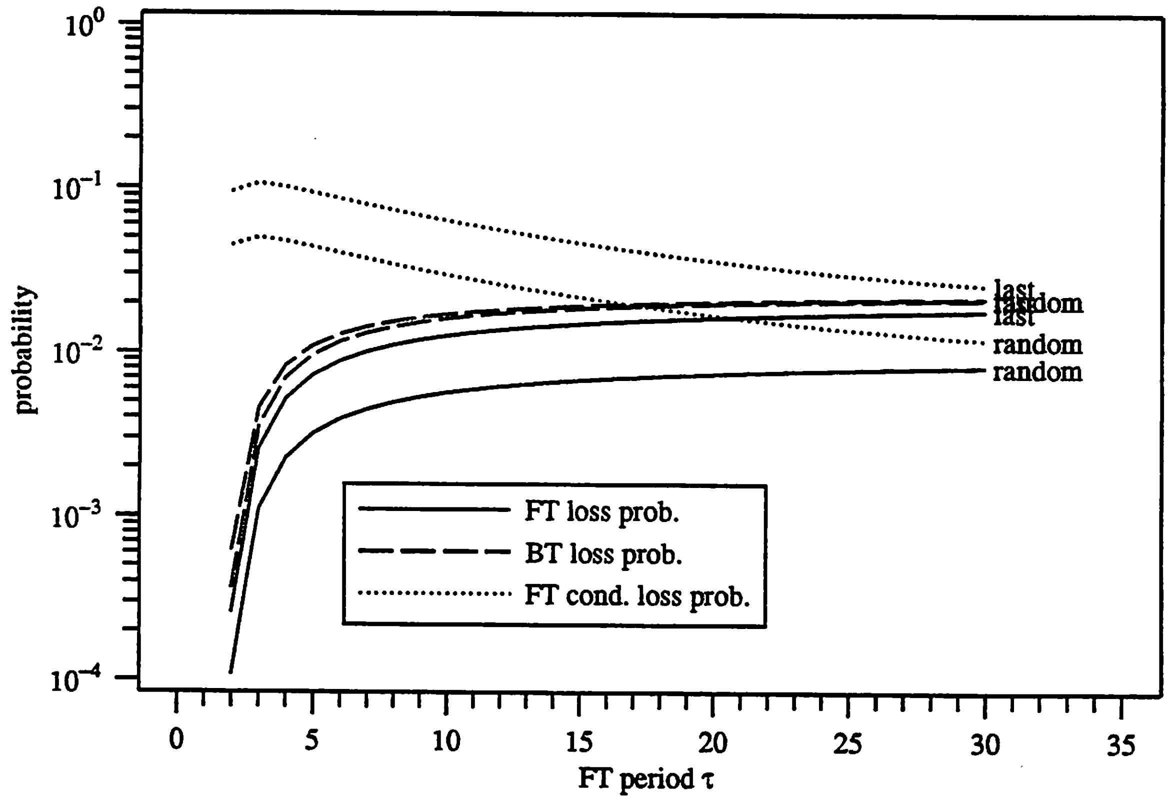

Figure 6: Loss probability and conditional loss probability for geometrically distributed background traffic BT with total load of $\lambda = 0.8$ and periodic traffic FT, as a function of τ; system size $K = 10$

3.4 Future work

A number of extensions to the model treated in this section suggest themselves. First, a stream could receive preferential treatment if it has suffered a loss at its last arrival, reducing the average loss run length to very close to one. However, implementation costs may not make this practical except in the case when the number of streams is very small (which is where the issue of long loss runs is of interest in any event).

In the analysis, it was assumed that the foreground stream contributes at most one packet per slot, which is a reasonable assumption as long as the outgoing link speed is at least as large as the access link speeds. The effect of batch foreground traffic on loss correlation could be investigated within the same analytical framework.

4 Bursty Traffic Sources

There appears to be general agreement that traditional, Poissonian traffic models do not capture the bursty arrival process that characterizes many proposed sources of BISDN traffic, for example packet voice, still and compressed full motion video. Since the exact behavior of these sources is ill understood and may depend strongly on the source material, several generic bursty-source models have been proposed and analyzed. These models differ not only in their generality, but also seem to be appropriate for different parts of the network. Note that altogether different models are to

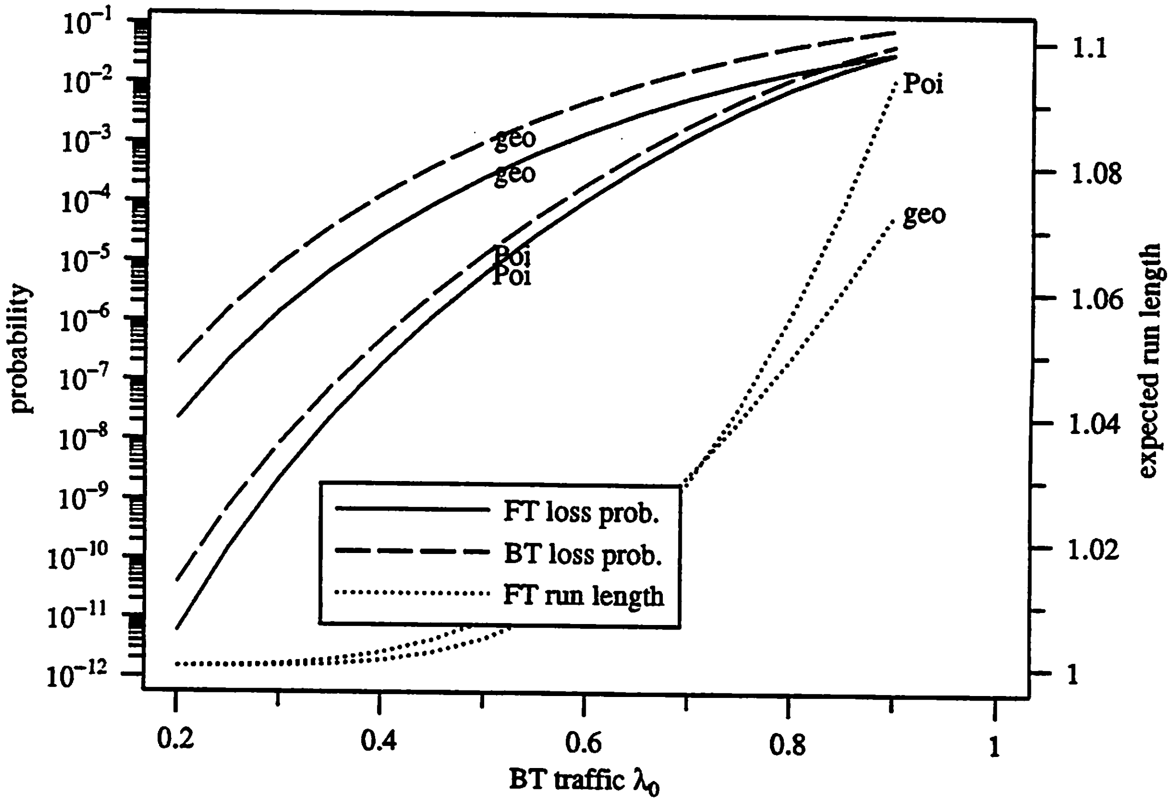

Figure 7: Loss probabilities and expected run length for Poisson and geometrically distributed background traffic BT and periodic traffic FT (random arrival), as a function of BT intensity λ_0; system size $K = 10$, period $\tau = 10$

be used for aggregate traffic. We will emphasize discrete-time models.

The first model, commonly referred to as an *on/off source*, features periodic arrivals during a burst and no arrivals during the silence periods separating bursts. For analytical tractability, it is often assumed that the burst and silence duration are exponentially (continuous time) or geometrically (discrete-time) distributed. The model was probably first used for packet voice sources with speech activity detection. Examples can be found in [49,50,44,51–53]. For packet video, Maglaris *et al.* [54] model each source as the superposition of a number of on/off minisources. For their picturephone source, about ten identical minisources provided sufficient accuracy.

Li also investigated the loss behavior for a voice source model [52] related to the $N\cdot$IPP$/D/c/K$ queue studied here. However, he considered the temporal properties of the aggregate loss rate for all sources, not the behavior of individual sources. In agreement with our findings, Li also arrives at the conclusion that the buffer size does not materially effect the loss behavior once an overload period sets in.

Cidon *et al.* independently investigate the number of lost packets due to buffer overflow in a frame of deterministic size or random size for a single Poisson or IPP source [55] and as well as multiple sources [56], in both discrete and continuous time.

Multiplexers and other packet switch queues introduce delay jitter, thus, packet arrivals within a burst even of a on/off source will no longer be strictly periodic within the network. The interrupted Poisson process (IPP) source model attempts to capture this. In continuous time [57,58], the on/off

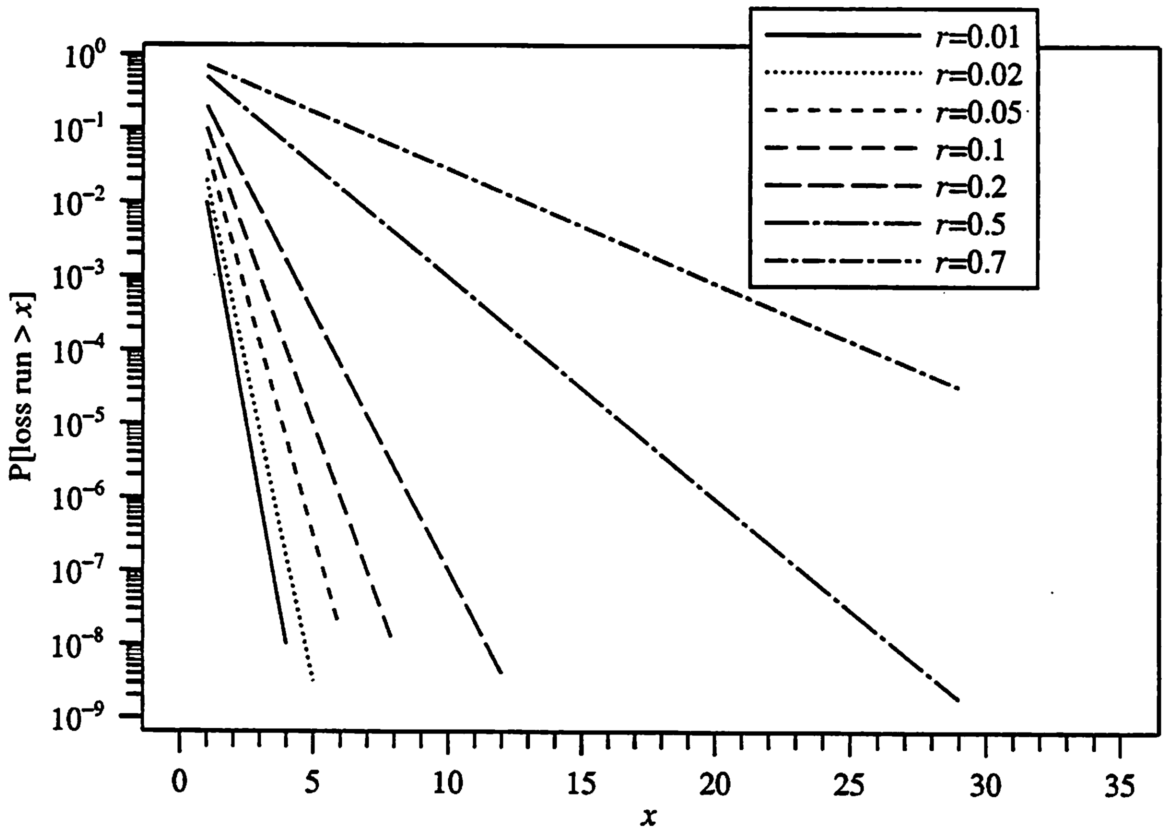

Figure 8: Probability that a loss run exceeds length x for given conditional loss probability r

time periods are exponentially distributed and independent of each other. A fixed-rate Poisson source emits packets during the on periods. In discrete-time [59, 60], the on/off periods are taken to be geometrically distributed. A packet arrives at each slot during the on period with probability λ. The designation as a discrete-time IPP appears misleading since the source does not generate a Poisson-distributed number of cells in a time interval during on periods. Interrupted Bernoulli process (IBP) seems more descriptive.

When the peak rate of sources or input channels exceeds the output rate, a batch arrival model is more appropriate. We can generalize the IPP model by allowing batch arrivals at each time slot during the on period. No performance evaluations using this model are known to the authors.

For the aggregate stream, a number of simplifying models can be used. The most popular in continuous time is probably the Markov-modulated Poisson process (MMPP), typically with two states. A simpler model aggregates all arrivals into batches, where batch arrivals occur as Bernoulli events with a given probability in a slot [59]. Thus, batches are separated by a geometrically distributed number of slots.

In this section, we limit our discussion to the IPP model. Possible extensions are discussed at the end of the section. In the subsequent discussion we will neglect source discrimination effects and assume random admittance, as defined in Section 3.

We begin in section 4.1 by reviewing some fundamental results for a single interrupted Poisson process (IPP) and then use these in section 4.2 first to describe the queue arrival process as a superposition of IPPs, proceed to find the loss probability and finally, the conditional loss probability. Numerical examples in section 4.4 illustrate the influence of buffer size and burstiness on loss correlation.

4.1 The Interarrival Time Distribution in an IPP

An interrupted Poisson process (IPP) is defined as follows:

Definition 1 *A discrete-time interrupted Poisson process (IPP) alternates between active (on) and silence (off) periods. At each slot, the state moves from active to silent with probability* $1-\gamma$ *and from silence to active with probability* $1-\omega$*. Thus, silence and activity duration are geometrically distributed. In the active state, a packet is generated independently from slot to slot with probability* λ.

We assume that a state transition takes place just prior to the end of a time slot and that a packet is generated at the beginning of a slot with probability λ if the new state is the active state. Below, the silence and active states will be indicated by subscripts zero and one, where necessary.

For ease of reference, we reproduce the results for the interarrival (distance) random variable D presented by Murata *et al.* [59]. The average dwell times, T_i, for the off and on state, respectively, are given by $T_0 = 1/(1-\omega)$ and $T_1 = 1/(1-\gamma)$, or, conversely, $\omega = 1 - 1/T_0$ and $\gamma = 1 - 1/T_1$. We also define the activity factor α,

$$\alpha = \frac{T_1}{T_0 + T_1} = \frac{1-\omega}{2-\omega-\gamma}.$$

The z-transform of the interarrival times is computed [59, Appendix B] by recursively writing down the probability that no arrival occurred in k slots since the last arrival, given that the system is in burst or silence state after k steps. These probabilities are denoted by $p_1(k)$ and $p_0(k)$, respectively. The recursive equations are

$$\begin{aligned}
p_1(0) &= 1 \\
p_0(1) &= (1-\gamma)p_1(0) \\
p_1(1) &= \gamma p_1(0) \\
p_0(k) &= \omega p_0(k-1) + (1-\lambda)(1-\gamma)p_1(k-1), \; k = 2,3,\ldots \\
p_1(k) &= (1-\omega)p_0(k-1) + (1-\lambda)\gamma p_1(k-1), \; k = 2,3,\ldots
\end{aligned}$$

Arrivals occur with probability λ if the system is in burst state, so that the z-transform can be derived as

$$\Delta_{\mathrm{IPP}}(z) = \lambda P_1(z) = \frac{\lambda[(1-\gamma-\omega)z^2 + \gamma z]}{1 - (\gamma(1-\lambda)+\omega)z - (1-\lambda)(1-\gamma-\omega)z^2}$$

Inversion of A_{IPP} yields the probability density function of the interarrival time, $\Delta_{\mathrm{IPP}}(k)$:

$$\Delta_{\mathrm{IPP}}(k) = d(1-r_1)r_1^{k-1} + (1-d)(1-r_2)r_2^{k-1} \tag{15}$$

where

$$\begin{aligned}
r_{1,2} &= \frac{1}{2}\left\{\gamma(1-\lambda) + \omega \pm \sqrt{(\gamma(1-\lambda)+\omega)^2 + 4(1-\lambda)(1-\gamma-\omega)}\right\} \\
d &= \frac{1-\lambda\gamma - r_2}{r_1 - r_2}
\end{aligned}$$

By differentiating the z-transform, or by taking the ratio of the burst over the total cycle time, we obtain the average interarrival time

$$E[\Delta] = \left.\frac{\partial \Delta_{\mathrm{IPP}}(z)}{\partial z}\right|_{z=1} = \frac{1/(1-\gamma)+1/(1-\omega)}{\lambda/(1-\gamma)} = \frac{1}{\lambda} + \frac{1-\gamma}{\lambda(1-\omega)}$$

and its inverse, the arrival rate

$$\rho = \frac{\lambda \cdot 1/(1-\gamma)}{1/(1-\gamma)+1/(1-\omega)} = \lambda\frac{1-\omega}{2-\gamma-\omega}.$$

The first moment can also be expressed in the quantities $r_{1,2}$ and d:

$$E[\Delta] = \frac{d}{1-r_1} + \frac{1-d}{1-r_2}$$

With the aid of the relation [61, p. 72]

$$\sum_{k=1}^{\infty} k^2 x^{k-1} = \frac{x+1}{(1-x)^3},$$

we obtain the second moment from the probability density function,

$$E[\Delta^2] = d\frac{1+r_1}{(1-r_1)^2} + (1-d)\frac{1+r_2}{(1-r_2)^2}.$$

The autocorrelation function of this process was derived by Gihr [62]. Other metrics, and methods of matching this process to an observed random process, are provided by Gilbert [1].

4.2 The $N \cdot \mathrm{IPP}/D/c/K$ queue

We analyze the $N \cdot \mathrm{IPP}/D/c/K$ queue operating in discrete time. N IPP sources, as described above, feed into the queue, which has K buffer positions, not including the customer being served. There are c identical servers. Each customer requires a service time of one time slot. All customers arrive at the beginning of a slot (just after the slot boundary), while the departure, if any, occurs at the end of the slot, just before the slot boundary. This is commonly refered to as an *early arrival system* [16]).

We work towards the goal of computing the conditional loss probability[7] in several steps. First, immediately below, we characterize the aggregate arrival process. Then, we write down the transition matrix, which also yields the (unconditional) loss probability. Through a number of intermediate quantities, we finally arrive at the desired conditional loss probability.

To set the stage, we review the behavior of the aggregate arrival process. The superposition of N i.i.d. IPPs is described in [60]. The transition probability that j sources are active in the current slot, given that i were active in the previous slot can be computed by [60, Eq. (3)]

$$S_{i,j} = \sum_{m=0}^{i} \underbrace{\binom{i}{m}\gamma^m(1-\gamma)^{i-m}}_{i-m \text{ go off}, m \text{ stay on}} \underbrace{\binom{N-i}{j-m}(1-\omega)^{j-m}\omega^{N-i-(j-m)}}_{j-m \text{ go on}, N-i-(j-m) \text{ stay off}} \quad \text{for } i,j \in [0,N], 0 \le j-m \le N-i$$

[7] As usual, we mean the conditional probability that customer n is lost given that customer $n-1$ from the same source was also lost.

The steady-state distribution s of the number of active sources is, as usual, computed as $s = s\mathbf{S}$, but can be written down immediately:

$$s_i = \binom{N}{i}\alpha^i(1-\alpha)^{N-i}.$$

The mean and variance of the number of active sources are $N\alpha$ and $N\alpha(1-\alpha)$. The correlation coefficient r is given by [50] $r = \omega + \gamma - 1$.

Because of batch-effects, the activity distribution seen by a random arrival, $\mathbf{s}'$, is a scaled version of $\mathbf{s}$ [63, p. 69]:

$$s'_i = \frac{is_i}{E[i]}$$

Given that i sources are active, the number of packets generated during a slot is binomially distributed with pmf

$$a_{i,k} = \binom{i}{k}\lambda^k(1-\lambda)^{i-k} \text{ for } k = 0, 1, \ldots, i$$

and zero otherwise. The distribution of batch sizes seen by a random observer is given by

$$a_k = \sum_{i=1}^{N} a_{k,i}s_i,$$

while the distribution of batch sizes seen by a random *arrival* is given by $a'_k = ka_k/E[A]$, where $E[A]$ is the expected batch size as seen by a random observer. Naturally, it equals ρ.

4.2.1 The Loss Probability

As a first performance metric, we derive the customer loss probability of this queueing system through the state transition matrix. The state of the system is described by the tuple (k, i), i.e., k cells in the system (waiting and in service), with i sources active and $N - i$ silent. For brevity, k and i will be referred to as the occupancy and activity, respectively. The state is tracked on slot boundaries, with state transitions occurring on the boundary and batch arrivals immediately afterwards. As in all finite-capacity, early-arrival systems with deterministic service time and batch arrivals, the first cell in a batch always enters the system. At most a single cell departs immediately prior to the slot boundary.

The $(N, K) \times (N, K)$ transition matrix is given by

$$\begin{array}{lcll}
P_{(i,q);(j,r)} & = & S_{i,j}a_{i,r-q+1} & \text{for } 0 \le q \le K, 0 < r < K \\
P_{(i,q);(j,K)} & = & S_{i,j}\sum_{m=K-q+c}^{\infty} a_{i,m} & \text{for } 0 < q \le K \\
P_{(i,q);(j,0)} & = & S_{i,j}\sum_{m=0}^{c-q} a_{i,m} & \\
P_{(i,q);(j,r)} & = & 0 & \text{otherwise}
\end{array}$$

where we define $a_{i,k} = 0$ for $k < 0$ and $k > i$. We denote the corresponding steady-state distribution as $\mathbf{p}$ with elements $p_{i,k}$.

The probability that a customer is lost, $P[L]$, is most readily computed by conservation-of-flow arguments as the ratio of channel occupancy, called ν here, to arrival rate:

$$1 - P[L] = \frac{\nu}{\rho} = \frac{1 - \sum_{i=0}^{N} p_{i,0}a_{i,0}}{\rho} \text{ for } c = 1.$$

For general values of c, the mean channel occupancy is given by

$$\nu = \sum_{i=1}^{N}\left[\sum_{k=1}^{c-1}\sum_{j=0}^{i} p_{i,k-j}a_{i,j} + c\sum_{j=0}^{i} p_{i,k-j}\sum_{m=j}^{i} a_{i,m}\right]$$

4.3 The Conditional Loss Probability

We designate an arbitrary source as the test source. The conditional loss probability is computed by tracing the state (activity and occupancy) of the system from the time some random customer, called the test customer C_1, is lost to the time when the next customer from that same test source, C_2, arrives. The computation is only possible because arrivals from a *single* source are i.i.d., even though the aggregate arrival process is correlated.

The computation proceeds in four major steps. First, we compute the number of active sources seen by C_1 as it failed to find buffer space. At this instant the system occupancy is $K+1$. Secondly, we compute the activity without the test source on the next slot boundary immediately following the loss. At this point just before the next set of arrivals, the occupancy is K. Then, we condition on the interarrival time A of the test source. For each possible value of the interarrival time, we compute the state, that is, occupancy and number of active sources, at the slot boundary of the slot where C_2 arrives. Finally, we arrive at the probability that C_2 is also lost by computing the number of customers from other sources that arrive in that slot.

Even though not directly needed for the final result, the conditional probability that one or more customers are lost during a slot given that i sources are active is given by

$$P[\Omega|i] = \sum_{k=0}^{K} P[p=k|i] \sum_{j=K-k+2}^{i} a_{i,j} = \sum_{k=0}^{K} \frac{p_{i,k}}{s_i} \sum_{j=K-k+2}^{i} a_{i,j}.$$

Here, we designate the overflow event, i.e., one or more customers are lost, by Ω.

Through Bayes' rule, the conditional probability that i sources are active given that one or more losses occurred during a slot can be computed given that the converse probability is available:

$$P[i|\Omega] = \frac{s_i P[\Omega|i]}{\sum_{i=1}^{N} P[\Omega|i] s_i}$$

For the first step, we need the probability $P[i|L]$ that a lost customer sees i active sources (including itself), which is most readily derived by Bayes' rule,

$$P[i-1|L] = \frac{s'_i P[L|i]}{\sum_{i=1}^{N} P[L|i] s'_i},$$

where

$$P[L|i] = \sum_{k=0}^{K} \frac{p_{i,k}}{s_i} \sum_{j=K-k+1}^{i-1} b_{i,j},$$

and s'_i is the activity seen by a random arrival, with $s'_i = i s_i / E[s]$.

$P[L|i]$ depends in turn on the number of arrivals that precede a randomly chosen customer in the same slot, conditioned on the number of active sources i in a slot. Since a customer is equally likely to occupy every position within the batch, we have that

$$b_{i,k} = \begin{cases} \sum_{j=k}^{i} \frac{a_{i-1,j}}{j+1}, & \text{for } k < i, \\ 0, & \text{otherwise.} \end{cases}$$

For convenience, we define $b_{0,0} = b_{1,0} = 1$.

Naturally, $P[L|i]$ is related to the loss probability:

$$P[L] = \sum_{i=1}^{N} P[L|i] s'_i,$$

The initial state vector $\mathbf{q_0}$ seen by the test customer is written as

$$q_0(i,k) = \begin{cases} P[i-1|L], & \text{for } k = K \\ 0, & \text{otherwise.} \end{cases}$$

Note that here and in the following steps, we condition on the interarrival time so that the test source is known to be inactive up to the last slot of that interarrival time.

The transition matrix $\tilde{P}_0$ from the loss of the test customer to the next slot has to reflect the fact that the test source is inactive. Any arrivals that occur after the test customer are lost and do not affect the system state. Thus,

$$\tilde{P}_{0(i,k);(j,l)} = \begin{cases} \tilde{S}_{i,j}, & \text{for } k = l \\ 0, & \text{otherwise.} \end{cases}$$

Here, $\tilde{S}$ and $\tilde{P}$ are the source activity and queue state transition matrices for $N-1$ sources, respectively.

In the slot where the test source emits the next customer, the activity increases deterministically by one. Also, the arrivals before the customer from the test source increase the queue occupancy. Note that the occupancy reaches now from 0 to $K+1$. This $(N, K+1) \times (N, K+1)$ transition matrix is therefore written down as

$$\tilde{P}_\infty(i,k);(j,l) = \begin{cases} \sum_{j=l-k}^{N} b_{i+1,j} & \text{for } k \le l \le K+1, 0 \le i < N, j = i+1 \\ 1 & \text{for } 0 \le k \le K, i = j = N \\ 1 & \text{for } 0 \le i < N, j = i+1, k = l = K+1 \\ 0 & \text{otherwise.} \end{cases}$$

Given all these intermediate quantities, we then uncondition on the interarrival distribution $P[D = n]$ and obtain the state distribution vector seen by the next arrival from the test source as

$$\mathbf{q} = \sum_{n=1}^{\infty} \mathbf{q_0} \tilde{P}_0 \tilde{P}^{n-1} \tilde{P}_\infty P[A = n]. \tag{16}$$

The conditional loss probability is then $\sum_{i=0}^{N} q_{i,k}$. Note that the interarrival time distribution, Eq. (15), is monotonically decreasing, simplifying the truncation of the sum when numerically evaluating Eq. (16).

The computation cost is dominated by the matrix multiplication to compute $\tilde{P}^n$ and the matrix product. Since the computation time is proportional to $(N \cdot K)^3$, and, to make matters worse, larger N tend to increase the number of terms that need to be evaluated in Eq. (16), a direct computation is only feasible for relatively small systems. On the other hand, as pointed out in more detail below, the loss correlation depends very little on K, so that $K = 1$ or even $K = 0$ can be used. Also, non-random behavior is most pronounced when a small number of sources is superimposed. It remains to be determined whether bounding the conditional loss probability is possible, for example by assuming that the number of active sources remains constant in the interarrival interval. Choosing the distribution as seen by a lost customer should provide an upper bound, while choosing the distribution seen by an average arrival should bound the conditional loss probability from below. As a closer approximation, we could treat the number of active sources as a birth-death process, as is commonly done for voice multiplexers [49,50], and then use numerical methods for sparse and block matrices. This approximation would probably improve in accuracy with larger N and longer state holding times.

While the discussion has focused on the $N \cdot \mathrm{IPP}/D/c/K$ system, a number of extensions are immediately apparent. We discuss them briefly in their rough order of complexity. As for the example in Section 3, different ordering of sources in competing for buffer space (priority discarding) can be modeled by appropriately modifying $b_{i,j}$. Sources with more than two states are also readily handled at the expense of a larger state space. Note that as long as the loss probability for the individual source can be determined, only the aggregate batch arrival process of the remaining background sources enters into the calculation.

The case of an on/off source with *periodic* arrivals during the active periods can be approximately treated within this framework as well by making some simplifying assumptions. First, the interarrival time of the test source can be modeled as a constant since most of the consecutive losses will take place within the active period of the source. Then, the superposition of the remaining sources should have roughly the same impact on the test source as the superposition of the sources with Bernoulli arrivals, at least for large N. These assumptions remain to be validated by analysis. Also, the case of several heterogeneous sources and of sources with two different Bernoulli arrival processes (MMBP) rather than on/off sources deserve closer scrutiny (see [64,65] for a loss probability analysis).

In the Gilbert channel model [1], the system alternates between two states in a manner identical to the IPP source. In one state, every packet is lost, in the other, no losses occur. While the loss runs are indeed geometrically distributed, simulations indicate that the success runs (the sequences of customers from a single source without loss) are not. It remains to be investigated whether the model can be applied nevertheless. Naturally, the mean of the success runs can be computed from the loss probability and the loss run length (see [2]).

4.4 Numerical Examples

Compared to earlier systems, the large number of queueing system parameters (λ, γ, ω, N, K, c) covers a wide range of systems, but also makes it difficult to generalize from numerical examples.

The systems studied earlier had the property that the conditional loss probability depended only weakly, if at all, on the system size K. This property carries over into this queueing system, as indicated by Fig. 9, and other examples not shown here. The effect of K will be small, by the argument of Theorem 1, as long as most of the probability mass of the interarrival time is concentrated in the first K slots.

The influence of the channel bandwidth is indicated by Fig. 10, where the ratio of the source on/off durations and the system load are kept constant, but the number of sources, N, is raised, with λ, γ and ω scaled correspondingly. Thus, the source correlation coefficient r increases with N. This leads to a somewhat unexpected result: The conditional loss probability decreases with increasing N, indicating that the effect of burstiness is more than compensated for by the "thinning" during active periods.

As a final parameter study, we investigate the effect of different source behavior, again maintaining the same average load by adjusting λ and α. We also keep the sum of active and silent time constant. Thus, for high values of λ, a source emits a short burst of densely packed packets, followed by a long silence. A lower value of λ, conversely leads to a smoother traffic process. Fig. 11 shows that the loss probability is far more affected by the change in traffic burstiness than is the conditional loss probability.

To provide a real-life example, the loss correlation in a DS1 voice multiplexer was determined. The DS-1/T1 channel offers a user capacity of 1.536 Mb/s. Following Heffes and Lucantoni [44], each source encodes speech in 64 kb/s and generates a 128-byte packet every 16 ms. Talk spurts and silence periods are geometrically distributed with means 352 ms and 650 ms, respectively, for

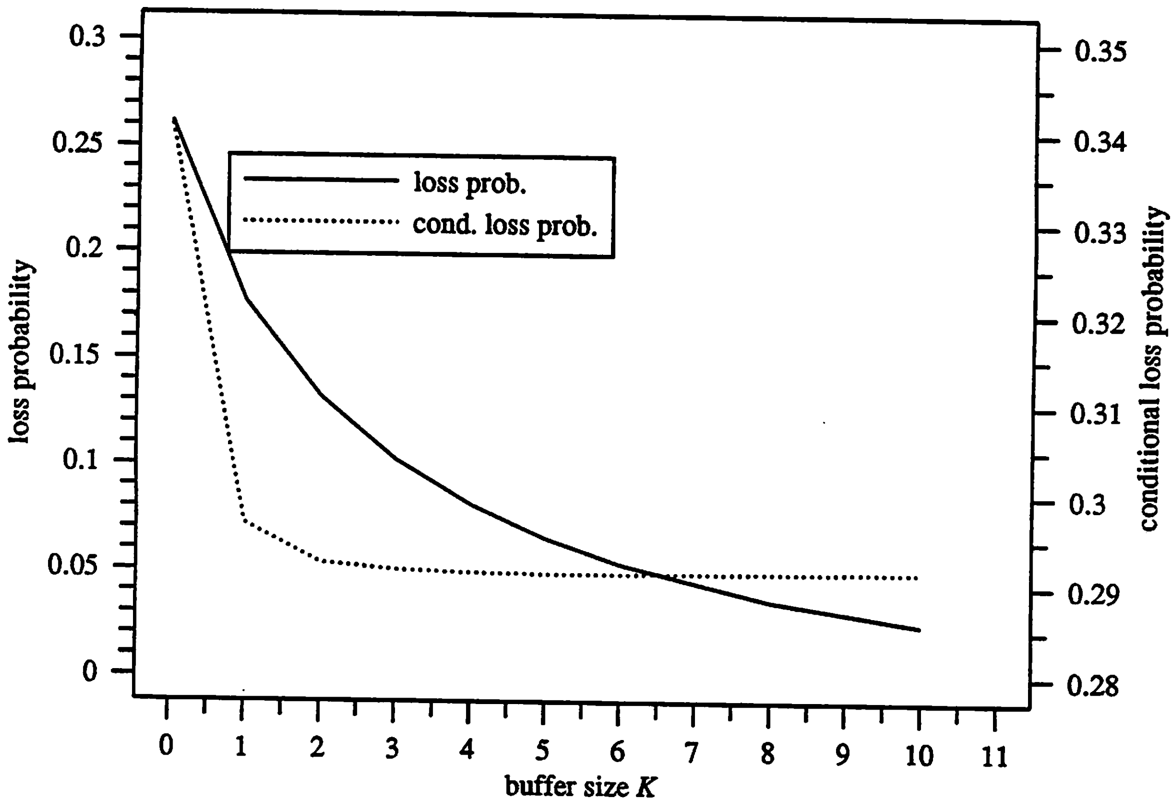

Figure 9: Conditional and unconditional loss probability for $N \cdot \text{IPP}/D/c/K$ system, as a function of the buffer size, K; $N = 4$, $\lambda = 0.8$, $\gamma = 0.7$, $\omega = 0.9$

an activity factor of 35%. When active, this source produces a packet every 24 slots. The channel processes 1324.2 packets per second. These system characteristics translate, at a load of 0.8, to the model parameters $1 - \lambda = 0.04167$, $1 - \gamma = 2.145 \cdot 10^{-3}$, $\omega = 1.162 \cdot 10^{-3}$ and $N = 55$. At a buffer size of $K = 3$, the conditional loss probability evaluates to 10.83%, or a mean loss run length of 1.12. This computation required about 20 hours of DECstation 5000 CPU time, evaluating 1483 terms of the total probability summation. A simulation confirmed these results (90% confidence on intervals 5.93 to 6.06% for loss and 10.78 to 10.99% for conditional loss probability) and showed that the numerical accuracy is satisfactory.

For a smaller multiplexing factor, consider video sources on an STS-3 channel. The STS3 channel has a usable capacity of 150.336 Mb/s or 354566 ATM cells/s. Unlike voice sources, video sources have not been well characterized in terms of standard source models. We approximately model the CATV video coder of Verbiest and Pinnoo [66]. It has an average rate of 26.5 Mb/s and a standard deviation of 3.4 Mb/s. We assume that the coder transmits at channel capacity and that the active and silent period add up to the frame duration of 1/25 s. We choose model parameters $\lambda = 1$, $1 - \gamma = 3.623 \cdot 10^{-4}$, $1 - \omega = 8.755 \cdot 10^{-5}$, and, again for a load of approximately 0.8, $N = 5$. With these parameters and $K = 1$, the conditional loss probability and mean loss run evaluate to the significantly higher values (compared to the voice example) of 0.5658 and 2.303, respectively. Thus, appropriate packet interleaving or priority mechanisms must be used to reduce the incidence of noticeable display artifacts.

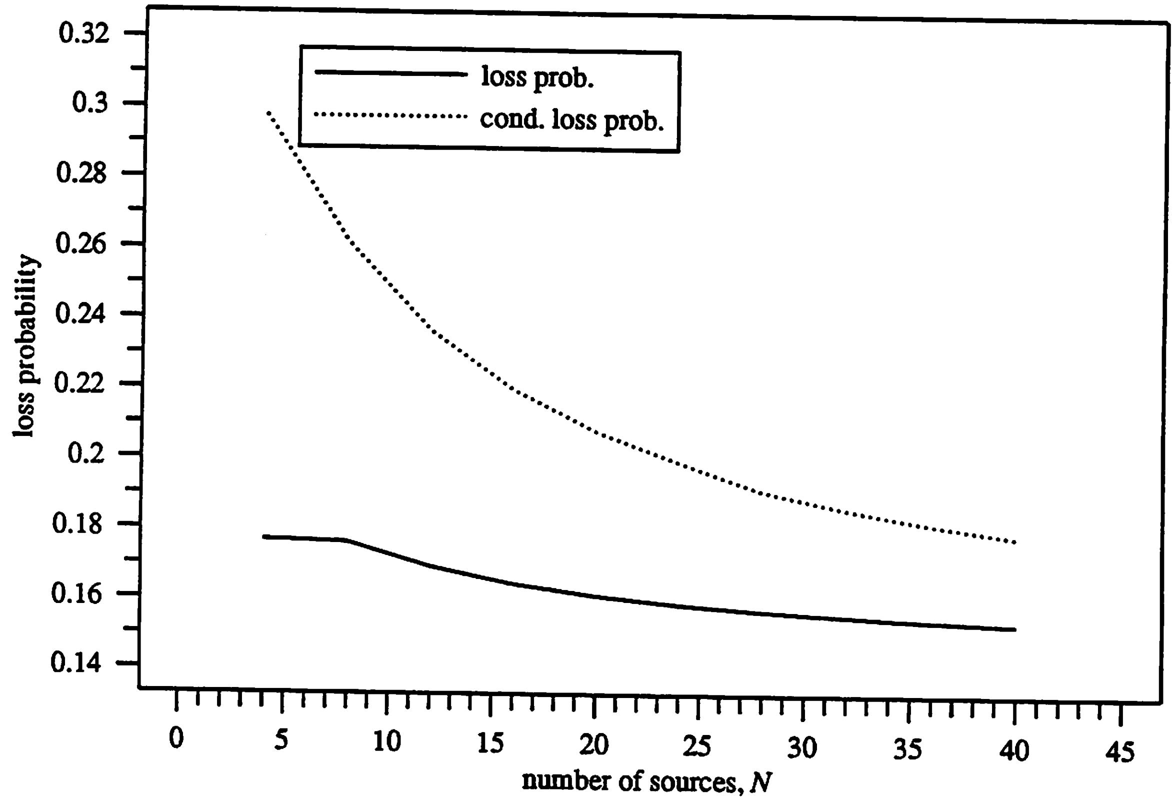

Figure 10: Conditional and unconditional loss probability for $N \cdot \mathrm{IPP}/D/c/K$ system, for constant activity, $\alpha = 0.25$, and load, $\rho = 0.8$, as a function of the number of sources, N; $K = 1$, $\lambda = 3.2/N$

5 Summary and Conclusions

In this paper, we have attempted to provide both methodologies and traffic engineering insight. The tools provided should allow the computation of the loss correlation for a wide variety of sources of practical interest in high-speed networking. Also, for a range of systems, it was shown qualitatively that the buffer size or waiting time limit has virtually no influence on the loss correlation. For the periodic arrival case, we saw that as long as individual sources do not contribute more than, say, 10% of the total bandwidth, the average loss run lengths are not significantly larger than one, even though the conditional loss probability can be orders of magnitude higher than the time-averaged loss probability. As expected, the traffic burstiness and the number of multiplexed sources play significant roles in the loss correlation for the IPP system.

Throughout this paper, we have used the conditional loss probability, or, equivalently, the average loss run length as an indicator of loss correlation. Another useful indicator could be the number of lost packets among a random sequence of packets with a given length. Also, the distribution of success runs deserves closer investigation. Particularly for queues with correlated inputs, Approximation methods need to be investigated that allow systems with large state spaces to be treated.

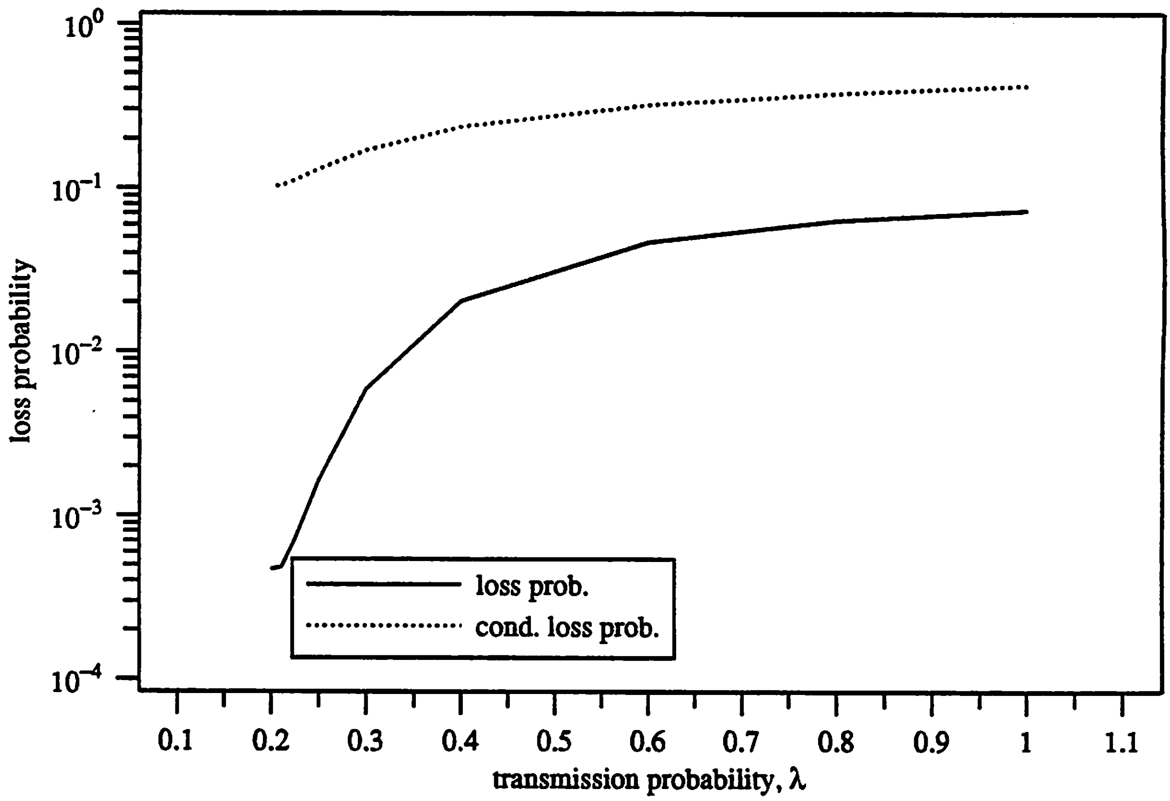

Figure 11: Conditional and unconditional loss probability for $N \cdot \mathrm{IPP}/D/c/K$ system, for constant load ($\rho = 0.8$), cycle time ($T_0 + T_1 = 25$) and buffer size ($K = 1$), as a function of λ

References

[1] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253–1265, Sept. 1960.

[2] H. Schulzrinne and J. F. Kurose, "Distribution of the loss period for some queues in continuous and discrete time," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (Bal Harbour, FL), pp. 1446–1455 (12C.1), Apr. 1991.

[3] J. M. Ferrandiz and A. A. Lazar, "Consecutive packet loss in real-time packet traffic," in *Proceedings of the Fourth International Conference on Data Communication Systems*, (Barcelona), pp. 306–324, IFIP TC6, June 1990.

[4] J. M. Ferrandiz and A. A. Lazar, "A study of loss in $N/GI/1$ queueing systems." Department of Electrical Engineering and Center for Telecommunications Research, Columbia University, New York, NY, Jan. 1990.

[5] J. M. Ferrandiz and A. A. Lazar, "Rate conservation for stationary processes," *Journal of Applied Probability*, Mar. 1991.

[6] A. K. Kaul, "Performance of high-level data link control in satellite communications," *COMSAT Technical Review*, vol. 8, pp. 41–88, Spring 1978.

[7] F. C. Fujiwara, M. Kasahara, K. Yamashita, and T. Namekawa, "Evaluations of error-control techniques in both independent and dependent-error channels," *IEEE Transactions on Communications*, vol. COM-26, pp. 785–793, June 1978.

[8] D. Towsley, "A statistical analysis of ARQ protocols operating in a nonindependent error environment," *IEEE Transactions on Communications*, vol. COM-29, pp. 971–981, July 1981.

[9] C. H. C. Leung, Y. Kikumoto, and S. A. Sorensen, "The throughput efficiency of the go-back-n ARQ scheme under Markov and related error structures," *IEEE Transactions on Communications*, vol. 36, pp. 231–234, Feb. 1988.

[10] G. R. Pieris and G. H. Sasaki, "The performance of simple error control protocols under correlated packet losses," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (Bal Harbour, FL), pp. 764–772 (7C.1), IEEE, Apr. 1991.

[11] H. Ohta and T. Kitami, "Simulation study of the cell discard process and the effect of cell loss compensation in ATM networks," *The Transactions of the IEICE*, vol. E73, pp. 1704–1711, Oct. 1990.

[12] A. A. Lazar, G. Pacifici, and J. S. White, "Real-time traffic measurements on MAGNET II," *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 467–483, Apr. 1990.

[13] N. Shacham and P. McKenney, "Packet recovery in high-speed networks using coding and buffer management," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (San Francisco, CA), pp. 124–131, IEEE, June 1990.

[14] T. Kitami and I. Tokizawa, "Cell loss compensation schemes in an asynchronous broadband ISDN," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (San Francisco, CA), pp. 116–123, IEEE, June 1990.

[15] N. M. Dor, "Guide to the length of buffer storage required for random (Poisson) input and constant output rates," *IEEE Transactions on Electronic Computers*, vol. EC-16, pp. 683–684, Oct. 1967.

[16] J. J. Hunter, *Mathematical Techniques of Applied Probability - Discrete Time Models: Techniques and Applications*, vol. 2. New York, NY: Academic Press, 1983.

[17] M. G. Hluchyj and M. J. Karol, "Queueing in high-performance packet switching," *IEEE Journal on Selected Areas in Communications*, vol. SAC-6, pp. 1587–1597, Dec. 1988.

[18] T. G. Birdsall, M. P. Ristenbatt, and S. B. Weinstein, "Analysis of asynchronous time multiplexing of speech sources," *IRE Transactions on Communication Systems*, vol. CS-10, pp. 390–397, Dec. 1962.

[19] M. J. Karol, M. G. Hluchyj, and S. P. Morgan, "Input versus output queueing on a space-division packet switch," *IEEE Transactions on Communications*, vol. COM-35, pp. 1347–1356, Dec. 1987.

[20] P. Tran-Gia and H. Ahmadi, "Analysis of a discrete-time $G^{[X]}/D/1-S$ queueing system with applications in packet-switching systems," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (New Orleans), pp. 861–870 (9A.1), IEEE, Mar. 1988.

[21] A. Y. M. Lin and J. A. Silvester, "Queueing analysis of an ATM switch with multichannel transmission groups," Tech. Rep. CRI 89-25, Computer Engineering Division, Electrical Engineering-Systems Department, University of Southern California, Los Angeles, CA, 1989.

[22] H. Schulzrinne, J. F. Kurose, and D. Towsley, "Congestion control for real-time traffic in high-speed networks," Technical Report TR 89-92, Department of Computer and Information Science, University of Massachusetts, Amherst, MA, 1989.

[23] L. P. Clare and I. Rubin, "On the design of prioritized multiplexing systems," in *Conference Record of the International Conference on Communications (ICC)*, (Boston, MA), pp. 1344–1348 (E5.3), IEEE, June 1983.

[24] L. P. Clare and I. Rubin, "Preemptive buffering disciplines for time-critical sensor communications," in *Conference Record of the International Conference on Communications (ICC)*, (Toronto, Canada), pp. 904–909, IEEE, June 1986.

[25] N. Yin and M. G. Hluchyj, "Implication of dropping packets from the front of a queue," in *International Teletraffic Congress, Seventh Specialist Seminar*, (Morristown, NJ), p. 10.4, ITC, Oct. 1990.

[26] C. I. T. Study Group XVIII and T. C. Committee), "Study group XVIII - report r 34," June 1990.

[27] P. Joos and W. Verbiest, "A statistical bandwidth allocation and usage monitoring algorithm for ATM networks," in *Conference Record of the International Conference on Communications (ICC)*, (Boston), pp. 415–422 (13.5), IEEE, June 1989.

[28] A. Bhargava, P. Humblet, and M. G. Hluchyj, "Queueing analysis of continuous bit-stream transport in packet networks," in *Proceedings of the Conference on Global Communications (GLOBECOM)*, (Dallas, TX), pp. 903–907, IEEE, Nov. 1989.

[29] M. Kaplan, "The queue $D/D/1$ with a Poisson background," in *Conference Record of the International Conference on Communications (ICC)*, (Toronto, Canada), pp. 36.5.1–36.5.4, IEEE, June 1978.

[30] I. Sahin and U. N. Bhat, "A stochastic system with scheduled secondary inputs," *Operations Research*, vol. 19, pp. 436–446, March-April 1971.

[31] I. Sahin, "Equilibrium behavior of a stochastic system with secondary input," *Journal of Applied Probability*, vol. 8, pp. 252–260, June 1971.

[32] A. E. Eckberg, Jr., "The single server queue with periodic arrival process and deterministic service times," *IEEE Transactions on Communications*, vol. COM-27, pp. 556–562, Mar. 1979.

[33] M. J. Karol and M. G. Hluchyj, "Using a packet switch for circuit-switched traffic: A queueing system with periodic input," in *Conference Record of the International Conference on Communications (ICC)*, (Seattle, Washington), pp. 1677–1682 (48.3), IEEE, June 1987.

[34] J. T. Virtamo and J. W. Roberts, "Evaluating buffer requirements in an ATM multiplexer," in *Proceedings of the Conference on Global Communications (GLOBECOM)*, (Dallas, Texas), pp. 1473–1477, IEEE, Nov. 1989.

[35] A. A. Fredericks, "Congestion in blocking systems — a simple approximation technique," *Bell System Technical Journal*, vol. 59, pp. 805–827, July–August 1980.

[36] A. Kuczura, "Loss systems with mixed renewal and Poisson inputs," *Operations Research*, vol. 21, pp. 787–795, May-June 1973.

[37] J. Matsumoto and Y. Watanabe, "Individual traffic characteristics of queueing systems with multiple Poisson and overflow inputs," *IEEE Transactions on Communications*, vol. COM-33, pp. 1–9, Jan. 1985.

[38] K. S. Meier-Hellstern, "The analysis of a queue arising in overflow models," *IEEE Transactions on Communications*, vol. 37, pp. 367–372, Apr. 1989.

[39] F. Machihara, "On the overflow processes from the $PH_1 + PH_2/M/S/K$ queue with two independent ph-renewal inputs," *Performance Evaluation*, vol. 8, pp. 243–253, Aug. 1988.

[40] H. B. Kekre, C. L. Saxena, and M. Khalid, "Buffer behavior for mixed arrivals and single server with random interruptions," *IEEE Transactions on Communications*, vol. COM-28, pp. 59–64, Jan. 1980.

[41] A. Kuczura, "Queues with mixed renewal and Poisson inputs," *Bell System Technical Journal*, vol. 51, pp. 1305–1326, July–August 1972.

[42] B. Gopinath and J. A. Morrison, "Discrete-time single server queues with correlated inputs," *Bell System Technical Journal*, vol. 56, pp. 1743–1768, Nov. 1977.

[43] C.-S. Chang, X. Chao, and M. Pinedo, "Integration of discrete-time correlated Markov processes in a TDM system," *Probability in the Engineering and Informational Sciences*, vol. 4, pp. 29–56, Jan. 1990.

[44] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on Selected Areas in Communications*, vol. SAC-4, pp. 856–867, Sept. 1986.

[45] R. Nagarajan, J. F. Kurose, and D. Towsley, "Approximation techniques for computing packet loss in finite-buffered voice multiplexers," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (San Francisco, CA), pp. 947–955, IEEE, June 1990.

[46] W. R. Parzynski and P. W. Zipse, *Introduction to Mathematical Analysis*. New York, NY: McGraw-Hill, 1982.

[47] Y.-S. Yeh, M. G. Hluchyj, and A. S. Acampora, "The knockout switch: A simple, modular architecture for high-performance packet switching," *IEEE Journal on Selected Areas in Communications*, vol. SAC-5, pp. 1274–1282, Oct. 1987.

[48] P. J. Burke, "Delays in single-server queues with batch input," *Operations Research*, vol. 23, pp. 830–833, July–August 1975.

[49] K. Sriram, P. K. Varshney, and J. G. Shanthikumar, "Discrete-time analysis of integrated voice/data multiplexers with and without speech activity detection," *IEEE Journal on Selected Areas in Communications*, vol. SAC-1, pp. 1124–1132, Dec. 1983.

[50] S. Li and J. W. Mark, "Performance of voice/data integration on a TDM system," *IEEE Transactions on Communications*, vol. COM-33, pp. 1265–1273, Dec. 1985.

[51] S. Li and M. El Zarki, "Dynamic bandwidth allocation on a slotted ring with integrated services," *IEEE Transactions on Communications*, vol. 36, pp. 826–833, July 1988.

[52] S. Li, "Study of information loss in packet voice systems," *IEEE Transactions on Communications*, vol. 37, pp. 1192–1202, Nov. 1989.

[53] M. Butto´, E. Cavallero, and A. Tonietti, "Effectiveness of the "leaky bucket" policing mechanism in ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 335–342, Apr. 1991.

[54] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Transactions on Communications*, vol. 36, pp. 834–844, July 1988.

[55] I. Cidon, A. Khamisy, and M. Sidi, "On packet loss processes in high-speed networks," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, vol. 1, (Florence, Italy), pp. 242–251 (2C.3), IEEE, May 19922.

[56] I. Cidon, A. Khamisy, and M. Sidi, "Analysis of packet loss processes in high-speed networks," EE Publication 786, Dept. of Electrical Engineering, Technion Israel Institute of Technology, Haifa, Israel, May 1991.

[57] A. Kuczura, "The interrupted Poisson process as an overflow process," *Bell System Technical Journal*, vol. 52, pp. 437–448, Mar. 1973.

[58] J. H. Rath and D. Sheng, "Approximations for overflow from queues with a finite waiting room," *Operations Research*, vol. 27, pp. 1208–1216, November–December 1979.

[59] M. Murata, Y. Oie, T. Suda, and H. Miyahara, "Analysis of a discrete-time single-server queue with bursty inputs for traffic control in ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 8, pp. 447–458, Apr. 1990.

[60] Y. Ohba, M. Murata, and H. Miyahara, "Analysis of interdeparture processes for bursty traffic in ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 468–476, Apr. 1991.

[61] V. Mangulis, *Handbook of Series for Scientists and Engineers.* New York: Academic Press, 1965.

[62] O. Gihr and P. Tran-Gia, "A layered description of ATM cell traffic streams and correlation analysis," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (Bal Harbour, FL), pp. 137–144 (2D.4), IEEE, Apr. 1991.

[63] R. W. Wolff, *Stochastic Modeling and the Theory of Queues.* Englewood Cliffs, NJ: Prentice Hall, 1989.

[64] J. J. Bae, T. Suda, and R. Simha, "Analysis of a finite buffer queue with heterogeneous markov modulated arrival processes: A study of the effect of traffic burstiness on individual packet loss." submitted to *Infocom 1992*, May 1992.

[65] J. J. Bae, T. Suda, and R. Simha, "Heterogeneous arrival streams, burstiness and packet discarding: A study of individual packet loss," Technical Report 91-58, Dept. of Computer and Information Science at the University of California, Irvine, Irvine, CA, July 1991.

[66] W. Verbiest and L. Pinnoo, "A variable bit rate codec for asynchronous transfer mode networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 761–770, June 1989.