VERIFYING CELL LOSS REQUIREMENTS
IN HIGH-SPEED COMMUNICATION NETWORKS

KERRY W. FENDICK

AT&T

Room 1L-425

Holmdel, NJ 07733-3030 USA

kerry@qsun.ho.att.com

WARD WHITT

AT&T Labs

Room A117, 180 Park Avenue

Building 103

Florham Park, NJ 07932-0971 USA

wow@research.att.com

In high-speed communication networks it is common to have requirements of very small cell loss probabilities due to buffer overflow. Losses are measured to verify that the cell loss requirements are being met, but it is not clear how to interpret such measurements. We propose methods for determining whether or not cell loss requirements are being met. A key idea is to look at the stream of losses as successive clusters of losses. Often clusters of losses, rather than individual losses, should be regarded as the important "loss events". Thus we propose modeling the cell loss process by a batch Poisson stochastic process. Successive clusters of losses are assumed to arrive according to a Poisson process. Within each cluster, cell losses do not occur at a single time, but the distance between losses within a cluster should be negligible compared to the distance between clusters. Thus, for the purpose of estimating the cell loss probability, we ignore the spaces between successive cell losses in a cluster of losses. Asymptotic theory suggests that the counting process of losses initiating clusters often should be approximately a Poisson process even though the cell arrival process is not nearly Poisson. The batch Poisson model is relatively easy to test statistically and fit; e.g., the batch-size distribution and the batch arrival rate can readily be estimated from cell loss data. Since batch (cluster) sizes may be highly variable, it may be useful to focus on the number of batches instead of the number of cells in a measurement interval. We also propose a method for approximately determining the parameters of a spe-

cial batch Poisson cell loss with geometric batch-size distribution from a queueing model of the buffer content. For this step, we use a reflected Brownian motion (RBM) approximation of a G/D/1/C queueing model. We also use the RBM model to estimate the input burstiness given the cell loss rate. In addition, we use the RBM model to determine whether the presence of losses should significantly affect the estimation of server utilization when both losses and utilizations are estimated from data. Overall, our analysis can serve as a basis for determining required observation intervals in order to reach conclusions with a solid statistical basis. Thus our analysis can help plan simulations as well as computer system measurements.

Key words: Cell Loss Probabilities, Traffic Measurements, Communication Networks, Batch Poisson Process, Reflected Brownian Motion, Reverse Engineering, Planning Simulations.

AMS subject classifications: 60K25, 60K30.

1. Introduction

We are pleased to be able to participate in this tribute to our colleague Ryszard Syski. His classic book, Syski [22], showed that many practical engineering problems in communication systems can be understood and systematically analyzed by applying the theory of stochastic processes. His continued enthusiasm for this program, and his enthusiasm more generally, have been an inspiration to us.

In this paper we hope to show once again how stochastic processes can provide useful insight into an engineering problem associated with communication systems. In particular, we are concerned with cell loss measurements and how to interpret them. Emerging high-speed communication networks are being designed to have very small cell loss probabilities. The proportion of arriving cells that are allowed to be lost due to buffer overflow is typically in the range $10^{-6} - 10^{-10}$. A challenging problem in this context is determining from simulations or actual system measurements whether or not these stringent requirements are being met. To illustrate, consider the following example.

Example 1: Suppose that the line speed is 10^5 cells per second. Suppose that a buffer with capacity 4000 cells is deemed adequate to provide a cell loss probability of 10^{-7} at an average utilization of 0.5 (an arrival rate of 0.5×10^5 per second). Suppose that we measure the cells lost over an hour. If the cell loss probability is indeed 10^{-7}, then the expected cell loss per hour is 18 cells. However, we cannot expect that exactly 18 cells will be lost each hour. The number of cells lost each hour should be regarded as a random quantity which has expected value 18. Even if the cell loss requirement is being met, some hours will have more losses and some fewer. We want to know if observations are inconsistent with a random process having a mean cell loss of 18 per hour. If 400 cells are lost in a given hour, can we conclude that the requirement is not being met? On the other hand, if only 2 cells are lost in an hour, can we conclude that the requirement is being met?

We develop two approaches for analyzing this problem. The first approach depends on knowing when individual cells have been lost, but not on other characteristics of the traffic or the strategy used for controlling the queue. Because it requires statistics not measured today by switches in operating networks, it probably is most

useful for interpreting the results of simulation experiments or for special studies.

The second approach, though not as precise as the first, does not depend on knowing when individual cells have been lost, but only on the measured number of cells arriving and discarded over an interval. Such measurements are generally available from switches today, so that the second approach is more readily applied in verifying whether an operating network is meeting cell loss objectives. However, the second approach assume that all data entering a queue is treated with the same discard priority and is covered by the same cell loss objective. In practice, cell loss objectives typically apply only for a subset of traffic and are supported by mechanisms that reserve buffer space for this traffic. Extensions to the second method to account for such mechanisms are possible, but not considered here.

Despite its simplifying assumptions, the second method provides qualitative insights that we expect to hold more generally. It suggests that if low cell loss ratios are achieved at queues with large buffers, then the cell loss process tends to be very bursty with long periods in which no loss occurs. At current trunks rates (e.g., 10^5 cells per second), it implies that data collected over tens of hours is needed to verify cell loss objectives.

Previous studies of cell loss processes in the context of video teleconferencing have been made by Cohen and Heyman [7], Hughes, Changari, Pearson, Seferdis and Xiong [14], and Heyman and Reibman [13]. They also conclude that the cell loss process is quite busty.

We begin our analysis by introducing a probability model for the counting process recording the number of lost cells in an interval of length t, for any t, which we denote by $\{N(t): t \geq 0\}$. In particular, we model the cell loss process by a *batch Poisson stochastic process* (described in Section 2). We justify this model as follows: A key idea is that cell losses are typically not isolated events. A period of overload or component failure typically leads to a cluster of cell losses, so that the appropriate "loss event" is the cluster rather than the single lost cell. When there are losses, the spaces between losses in a cluster should not greatly exceed typical cell interarrival times. However, since the cell loss rate is much smaller than the cell arrival rate, there should be much larger times between successive clusters of losses. If we focus on the larger time scale (regarding the times between clusters as of order 1), then the spaces between cell losses within clusters should be negligible. It is from this perspective that we model the cell loss process as a batch arrival process. Since we are interested in the total number of losses in the entire measurement interval, the local behavior within clusters of losses is unimportant to us. If at some time we wish to develop a higher fidelity model, which may be of interest to analyze the loss process as an arrival process elsewhere, then we could introduce spaces between successive losses within each cluster. But we do not focus on that issue here.

A second key idea is that, since losses should be rare events, the counting process of losses initiating clusters should be approximately a Poisson process, regardless of the form of the cell arrival process. There is a substantial theoretical basis for the Poisson property; e.g., see Lewis [18], Leadbetter [16], Chapter 8 of Keilson [15], Leadbetter, Lindgren and Rootzén [17], Aldous [4] and references cited there. It is now well known that the Poisson process is often inappropriate to represent traffic processes in communication networks; e.g., see Paxson and Floyd [20]. Certainly a cell arrival process is very unlikely to be nearly a Poisson process. Nevertheless, the asymptotic theory suggests that the counting processes of losses initiating clusters should be nearly Poisson. Thus the rarity of losses tends to make the batch Poisson model

appropriate for analyzing cell losses in communication networks. We provide some further evidence for the Poisson property here, but we primarily are concerned with showing how to work with the batch Poisson model. In applications, the Poisson property needs to be validated.

Given that the average spacing between points within a cluster is several orders of magnitude smaller than the average spacing between clusters, it should not be difficult to reasonably identify the clusters from data. It suffices to regard all interval lengths above an intermediate value as intervals between clusters. In this way, we can easily obtain estimates of the batch-size distribution and the Poisson batch arrival rate from data. In particular, once the batches have been determined, the batch-size distribution can be estimated by its empirical distribution function and the batch arrival rate can be estimated by the observed number of batches divided by the length of the total observation interval. Moreover, to better understand performance, we believe that it should be useful to focus separately upon the rate clusters occur and the size and structure of these clusters.

There are several complications, however. First, network and source controls may make appropriate batches more difficult to identify. For example, the first cell losses in a cluster may cause sources (on their own initiative or due to the network controls) to temporarily reduce their offered load and then increase it again slightly later. This phenomenon could cause clusters to consist of several subclusters. Clearly it is important to properly identify the larger clusters. Since the space between clusters is so large, we anticipate that it should be possible to distinguish new clusters from subclusters. However, the subcluster phenomenon just discussed indicates that we should be careful to allow enough space between what are considered bonafide separate clusters. This discussion also indicates that it may well be difficult to model the behavior within a cluster, but we avoid most of this complexity here by focusing only on the total number of cell losses in the cluster. As a topic for future research, the subcluster phenomenon should also be worth attention to better understand the controls.

A second complication is a time-varying load. Typically the load on the network will vary substantially over the long time scale encompassing several clusters of losses. In that event, it is natural to regard the cluster process as a *nonhomogeneous* Poisson process, i.e., with a time-dependent rate. It is significant that the successive intervals between points are still independent in a nonhomogeneous Poisson process model, but with time-dependent distributions. Alternatively, it is natural to look at data over a collection of separate intervals for which the load is nearly constant. Given the nonhomogeneous Poisson process model, it is significant that the cluster counting process constructed by combining separate intervals that have essentially the same rate is again a Poisson process, but now essentially homogeneous, i.e., with constant rate. This is the approach we advocate. Since we have a homogeneous Poisson process, there is only a single parameter (the rate) of the Poisson process to estimate.

A complication with the procedure just described for real systems (but not for simulations) is that we must measure the load as well as the losses. For example, in the setting of Example 1 we must estimate the traffic intensity (which coincides with the server utilization, i.e., the probability that the server is busy) by the ratio of the observed average input rate to the known output rate. In our treatment of losses, it is natural to regard this estimate of traffic intensity as exact. However, the presence of losses could conceivably bias our estimate. If the presence of losses is coupled with

higher average input rates, then we will think that the traffic intensity in the intervals with losses is higher than it actually is. To illustrate, consider Example 1 where the actual traffic intensity is 0.5. The question is whether, in that model, the estimated traffic intensity over hours with 100 losses, say, is significantly greater than 0.5, e.g., 0.7. We contend that in the setting of Example 1 it is reasonable to regard the estimated traffic intensity over one hour as exact, and we present a method for analyzing this question more generally (Section 8).

For modeling with limited data, and even for modeling with ample data, it is often convenient to have a parsimonious model with few parameters. For this purpose, we also consider the special case of a batch Poisson process with geometric batch-size distribution. The model then has just two parameters, which can be taken to be the mean batch size and the batch arrival rate. The product of the two parameters of this batch Poisson arrival process − the Poisson arrival rate and the mean batch size - is determined by the overall cell loss rate. Without data, we obtain the missing information to determine both parameters by introducing a probability model for the buffer content over time. We initially model the buffer content as the queue length process in a G/D/1/C queue, which has a single server, additional waiting space of size $C - 1$ (and thus total capacity C), deterministic service times and a general stationary arrival process. We then approximate this system by reflected Brownian motion (RBM), where there is reflection at an upper barrier as well as a lower barrier. We choose the drift and diffusion coefficient of RBM to match the rate and asymptotic variances of the net input process (arrival process minus the service process). We then approximate the cell loss process by the upper barrier local time process, and apply results for it derived by Berger and Whitt [5]. This enables us to carry out a normal approximation and specify both parameters of the batch Poisson process with a geometric batch-size distribution.

The two-parameter RBM process provides a greatly simplified view of the queueing process, which usually has a highly complex superposition arrival process. The simplicity is useful because exact expressions are available for the rate and asymptotic variance of the loss associated with RBM. This enables us to do a first-order analysis relating the rate (characterized by the drift) and the burstiness (characterized by the diffusion coefficient) to the buffer size and the loss probability. Indeed, we show how the RBM model can be used to reverse engineer the burstiness; i.e., given the buffer capacity, utilization and loss rate, we can estimate the input burstiness (Section 7).

Here is how the rest of the paper is organized: We specify the batch Poisson model and indicate how to calculate the probability distribution of the number of losses in an interval of any specified length in Section 2. That probability distribution can be used to build confidence intervals for cell loss estimates and to test the hypothesis that the cell loss requirements are being met. We give procedures for estimating parameters of the batch Poisson process model in Section 3. We briefly discuss long-range dependence there; e.g., see Willinger [25]. It is significant that our proposed estimation procedure for the batch Poisson model offers a way to avoid statistical problems caused by long-range dependence if it is present.

We introduce the G/G/1/C queueing model in Section 4. We develop the RBM approximation for the G/G/1/C queueing model in Section 5. We indicate how to fit the two-parameter batch Poisson model of the loss process to the RBM model in Section 6. In Section 7, we show how to do reverse engineering to determine an approximation for the asymptotic variance of the arrival process from given buffer

capacity and loss probability. We complete the analysis of Example 1 there. Finally, in Section 8, we discuss measurement schemes in which the traffic intensity and losses are estimated simultaneously. Consistent with intuition, the traffic intensity estimates are significantly affected by losses in short measurement intervals, but the traffic intensity estimates tend to be unaffected by losses in suitably long measurement intervals.

This work was motivated by the problem of interpreting existing network measurements (about which we intend to write more in a subsequent paper), but our analysis with RBM can also be used to help design simulation experiments involving single-server queues with finite waiting rooms by determining the run lengths required to estimate loss probabilities with desired statistical precision. In this regard, this paper is a sequel to Whitt [23] and Srikant and Whitt [21]. Those papers determine required simulation run lengths for different queueing models, in particular, for a single-server queue with an infinite waiting room and for a multi-server queue without any waiting room, respectively.

2. The Batch Poisson Process Model

As we have indicated above, it is natural to regard the rate of the cell loss process over long intervals (e.g., over a day) as time-dependent. However, by collecting data over separate intervals with essentially the same load, we can obtain an approximately constant cell loss rate, and we only consider the case of a constant arrival rate here. For instance, in Example 1 we have assumed that we are working with a collection of hours with estimated utilizations all about 0.5. (Throughout this paper, we use "server utilization" and "traffic intensity" interchangeably). Hence, we consider a homogeneous Poisson process.

If we indeed collect intervals with a common utilization, then our primary goal is to determine the probability distribution of the number of lost cells associated with a given utilization. We can identify the overall cell loss probability distribution as the weighted average of cell loss probability distributions associated with the different utilizations, where the weights are the frequency of each utilization level. Alternatively, we could choose to have a more stringent requirement, in which we focus only on the intervals with high utilizations (busy-hour engineering). We can use the cell loss probability distribution to construct confidence intervals for cell loss estimates and to test the hypothesis that the cell loss requirements are being met.

Given a (homogeneous) batch Poisson process $\{N(t): t \geq 0\}$ with batch size probability mass function $p_k \equiv P(X = k)$, $k \geq 0$, and batch arrival rate β, the probability of having k losses in an interval of length t has the compound Poisson probability mass function

$$q_k \equiv P(N(t) = k) = \sum_{j=0}^{\infty} \frac{e^{-\beta t}(\beta t)^j}{j!} p_k^{j*}, \qquad (2.1)$$

where p_k^{j*} is the j-fold convolution of the batch-size probability mass function $\{p_k: k \geq 0\}$. These probabilities can be quite daunting to compute directly, but they are easily computed (exactly) by numerically inverting the generating function

$$\widehat{q}(z) \equiv \sum_{k=0}^{\infty} q_k z^k = e^{-\beta t(1 - \widehat{p}(z))}, \qquad (2.2)$$

where

$$\widehat{p}(z) \equiv \sum_{k=0}^{\infty} p_k z^k; \tag{2.3}$$

e.g., see Abate and Whitt [3]. Since we are interested in tail probabilities, we directly invert the generating function of $P(N(t) > k)$, which is

$$\widehat{Q}^c \equiv \sum_{k=0}^{\infty} P(N(t) > k) z^k = \frac{1 - \widehat{q}(z)}{1 - z} \tag{2.4}$$

for $\widehat{q}(z)$ in (2.2). In the rest of this paper we do not focus on the exact distribution in (2.1)-(2.4), but it is readily computed in applications to fully describe the implications of the batch Poisson model.

Alternatively, for greater simplicity, a normal approximation of the cell loss probability distribution can be used, which only requires β and first two moments of X. The mean and variance of $N(t)$ are:

$$E[N(t)] = \beta t E[X] \quad \text{and} \quad Var[N(t)] = \beta t E[X^2], \tag{2.5}$$

where Var is the variance. With the normal approximation, we treat $N(t)$ as normally distributed with mean and variance in (2.5). A normal approximation is justified for suitably large t, provided that the batch-size distribution has finite variance, because the distribution of $N(t)$ is asymptotically normal as $t \to \infty$. Since $\{N(t) : t \geq 0\}$ has stationary and independent increments, the ordinary central limit theorem can be applied to substantiate this claim. The central limit theorem is very important because it shows that only two parameters are relevant when t is suitably large. Indeed, with large t the batch-Poisson model plays a minor role, only providing one of many sufficient conditions justifying the normal approximation.

However, it should be noted that if the expected number of batches is small, then the normal approximation is likely not to be accurate. Then it should be much better to calculate the exact batch Poisson probabilities than to employ the normal approximation. Moreover, for small values of βt, the actual batch-size distribution can play a very important role.

The approaches above have been chosen largely because of their simplicity, to be used with hourly summary statistics. Given full loss data, it is natural to consider more general models, e.g., as in Section VI.C of Fendick, Saksena and Whitt [10]. One simple generalization is to make the successive idle and cluster periods an alternating renewal process. From Cohen and Heyman [7], we would be led to consider Weibull idle periods instead of exponential ones. Ignoring the cluster lengths, we could consider a batch renewal process instead of a batch Poisson process.

3. Estimating the Parameters of the Batch Poisson Model

In this section we discuss three possible methods for estimating the parameters of the batch Poisson process model. The first method — batch means — is a natural simple way to identify the two parameters in the normal approximation. However, we contend that a second method — identifying clusters — should be preferable.

Batch means

Suppose that we intend to use the normal approximation described in Sections 2, so that we only need the first two moments of $N(t)$, as in (2.5). Given a sample path of

the cell loss process over an interval $[0, t]$, we can divide the interval into n equal subintervals of length t/n and count the number of losses in each subinterval. Let x_i denote the observed number of losses in subinterval i. We can estimate $E[N(t/n)^k]$ for any k by the sample means

$$\bar{x}_n^k \equiv n^{-1} \sum_{i=1}^n x_i^k. \tag{3.1}$$

We then estimate $\beta E[X]$ and $\beta E[X^2]$ in (2.5) by $(n/t)\bar{x}_n^1$ and $(n/t)(\bar{x}_n^2 - (\bar{x}_n^1)^2)$, respectively.

If $\{N(t): t \geq 0\}$ actually were a batch Poisson process, then the observations x_i, $1 \leq i \leq n$, would be independent and identically distributed (i.i.d.). If, in addition, the batch-size distribution has finite variance and t is suitably large, then x_i is approximately normally distributed as well. Under these assumptions, we can estimate the statistical precision of our estimates, i.e., we can form confidence intervals for the mean and variance using standard statistical techniques; e.g., see Sections 11.2 and 11.3 of Mood and Graybill [19].

The major difficulty with this approach is that the underlying assumptions may not nearly hold. The batch Poisson model may be reasonable, but it may well be that t is not big enough for the observations to be approximately normally distributed. The main difficulty is that the Poisson mean βt may be too small. This is likely to occur because β may be very small. It may be significantly less than the cell loss probability.

Moreover, high statistical precision will not be possible if the variance of the batch-size distribution, $Var\, X$, is large. In particular, this estimation procedure would break down completely if $Var\, X = \infty$, which is a possibility that should be considered.

Let the number of cell arrivals to the queue in the interval $(0, t]$ be denoted by $A(t)$. We partially characterize the general arrival process $\{A(t): t \geq 0\}$ by its *rate*

$$\lambda \equiv \lim_{t \to \infty} t^{-1} A(t) \tag{3.2}$$

and its *asymptotic variance*

$$\sigma_a^2 \equiv \lambda c_a^2 \equiv \lim_{t \to \infty} t^{-1} Var\, A(t), \tag{3.3}$$

which we assume are well defined, positive and finite. It is useful to focus on the variability independent of the rate via the variability parameter c_a^2 in (3.3); it is the limiting value of the index of dispersion; e.g., see Fendick and Whitt [11].

Recent studies have shown that cell arrival processes may exhibit long-range dependence, e.g., see Willinger [25]. This means that the asymptotic variance of the cell arrival process (c_a^2 in (3.3)) may actually be infinite. Large bursts in the arrival process may then contributed to large bursts of losses, i.e., so that $Var\, X$ may become large or even infinite. This implies that the normal approximation and associated standard statistical techniques focusing on sample averages, such as just described, may not apply, and that we should consider other methods. In particular, better estimates should be possible if we can exploit the full batch-Poisson model, including the batch-size distribution.

To some extent, we avoid difficulties caused by long-range dependence by basing our analysis on a collection of intervals with similar utilizations. The observed long-range dependence may result from the change in utilization over a longer time scale; e.g., see Bhattacharya, Gupta and Waymire [6]. The change in utilization that usually occurs in a longer time scale is avoided by focusing on the collection of intervals

with common utilization. Nevertheless, we believe that the long-range dependence tends to make the batch means approach more suspect.

Identifying clusters

The second approach is to directly estimate the batch-size distribution and the batch arrival rate. The key is to identify the clusters in the loss process. An infinite variance for the batch-size distribution presents no difficulty if the batch-size distribution is estimated from data by its empirical distribution. In that way we can avoid anomalies in sample means caused by fat tails.

Let $\{L_i : i \geq 0\}$ be the sequence of spaces between successive losses, with L_i being the space between losses $i - 1$ and i for $i > 1$. A second estimation procedure is based on first estimating the clusters. Let the first loss belong to the first cluster. Let the ith loss belong to the last cluster if $L_i \leq \tau$ and let the ith loss initiate a new cluster if $L_i > \tau$, where τ is an appropriately chosen threshold. Let each cluster be located at the loss initiating the cluster.

In this case, we are assuming that the length of a cluster, i.e., the time from the first loss until the last loss within a cluster, is negligible compared to the intervals between clusters. If this were not so, then we could delete the cluster subintervals. Equivalently, we can think of the new cluster process being turned off in the middle of a cluster, so that a new cluster can begin only after an old one has ended. However, our overall analysis is largely based on the length of clusters being small compared to the intervals between clusters. Several methods for estimating the parameters of a cluster Poisson process in more general situations are proposed by Lewis [18].

We also need to account for the fact that the average space between clusters should be substantially greater than the average space between losses, which is r^{-1}. If the average cluster size is m, then the average space between clusters is m/r. Of course, we do not known m in advance. In general, we could proceed iteratively starting with a tentative estimate of m, and refining it after analysis.

Given the congestion control effect discussed in the introduction, we should be careful not to make the threshold too short. Since congestion controls tend to operate in the scale of an end-to-end roundtrip time, the threshold should be substantially greater than a few roundtrip times for distant sources in the network. For example, the propagation delay across the United States is of the order of 50 ms.

We want to choose the threshold in between the average cell interarrival time λ^{-1} and the average space between clusters m/r. A reasonable approach appears to be to average the exponent but tending to the long side; e.g., we could let

$$\tau = 10^{2\log_{10}(m\lambda/r)/3}\lambda^{-1} = (m^2/r^2\lambda)^{1/3}. \tag{3.4}$$

For example, in the setting of Example 1, $\lambda^{-1} = 0.00002$ secs. Suppose that we estimate m as $m \approx 10^3$. (Our rough analysis in Section 7 gives $m \approx 250$.) Since $m\lambda/r \approx 10^{10}$, formula (3.4) dictates having $\tau \approx 10^{6.7}\lambda^{-1} \approx 100$ secs. In general, the appropriate choice should depend on the situation, after examining the data.

Once the clusters have been identified, we can estimate the batch-size distribution by the empirical batch-size distribution and the batch arrival rate by the observed number of batches divided by the total length of the observation interval. We can estimate statistical precision of the batch-size distribution by assuming successive observations are i.i.d. Moreover, standard statistical tests for independence can be used.

Alternatively, we can estimate the batch arrival rate β by estimating its reciprocal β^{-1}. The successive spaces should be mutually independent and exponentially distributed with mean β^{-1}. With this framework, we can easily estimate the statistical precision and directly test the batch Poisson assumption.

It is also natural to use parametric models for the batch-size distribution. An obvious candidate is the geometric distribution. If X is a geometrically distributed random batch size, then its probability mass function is

$$P(X = k) = (1 - p)p^{k-1}, \ k = 1, 2, \ldots, \tag{3.5}$$

which has mean

$$m \equiv EX = 1/p. \tag{3.6}$$

A natural way to estimate the parameter p in (3.5) is to let $\widehat{p} = 1/\widehat{m}$, where $\widehat{m}$ is the sample mean, which of course is the natural estimator for m in (3.6).

We conclude this subsection by pointing out that there can be some censoring problems associated with clusters occurring at the ends of measurement intervals. A conservative approach is to include in such edge batches all losses in adjacent intervals that would be considered part of the batch if the measurement interval were extended. Clusters would then be counted twice at the end of one interval and at the beginning of another interval. We do not dwell on this effect because we believe it should be minor. In practice that should be checked.

Identifying clusters from first passage times

The method above identifies the clusters by only looking at the observed losses. We now propose an alternative method that use the observed buffer content process as well. We now modify the previous procedure by stipulating that a loss initiates a new cluster if and only if the buffer has completely emptied in the interval since the last loss.

This procedure is appealing because it is not necessary to identify a threshold. Moreover, in many cases this procedure should actually be very close to the previous one. Assuming that buffer overflow is indeed a rare event, the buffer content process should move relatively quickly from the upper barrier to the lower barrier, while moving relatively slowly from the lower barrier to the upper barrier. We demonstrate this phenomenon for RBM below.

At first glance, this first-passage-time criterion would seem to be more conservative (lead to longer clusters), but it actually might not be in practice (in communication networks as opposed to queueing models). The buffer might empty completely due to sources temporarily reducing input in response to losses, but then the buffer might soon fill again and overflow when these same sources ramp up again slightly later. So, for communication networks, the first method would actually seem preferable. However, a combination could be used: a new cluster might be stipulated to begin the first time the buffer is full, after it has been empty after a threshold τ.

4. The G/G/1/C Queueing Model

We now introduce a more detailed queueing model in order to relate the cell arrival process to the cell loss process. Our queueing model is a standard single-server queue with a finite waiting room (buffer) with total capacity C. Waiting customers begin

service in order of arrival as soon as the server becomes free. An arrival finding all servers busy and the waiting room full is blocked and lost, without affecting future arrivals. As in Section 3, let the arrival process by $A(t)$, partially characterized by its arrival rate λ and asymptotic variance $\sigma_a^2 = \lambda c_a^2$.

We are interested in deterministic service times, but the results apply more generally. Let the server have a potential service process $\{S(t): t \geq 0\}$. The variable $S(t)$ is the number of service completions during the first t units of time in which the server is busy (excluding idle periods). We partially characterize the potential service process $\{S(t): t \geq 0\}$ by its rate

$$\mu \equiv \lim_{t \to \infty} \frac{S(t)}{t} \tag{4.1}$$

and asymptotic variance

$$\sigma_s^2 \equiv \mu c_s^2 \equiv \lim_{t \to \infty} t^{-1} Var S(t), \tag{4.2}$$

which again we assume are well defined and finite. The deterministic (D) special case of primary interest here is covered by letting $\sigma_s^2 = c_s^2 = 0$ in (4.2).

We can use the queueing model to provide some theoretical support for a geometric batch-size distribution. For a GI/M/1/C model (with renewal arrival process and exponential service times), successive losses are regeneration points for the model. Thus the number of successive losses until there is an interval between losses greater than any threshold τ indeed has a geometric distribution, where p is the probability that the interval from one loss to the next exceeds τ. Moreover, for the GI/M/1/C model the sizes of the clusters and the spaces exceeding τ are all mutually independent. Hence, even though we do not have exactly a GI/M/1/C model, because service is D and arrivals typically are not nearly GI, these provide supporting motivation for our approach. In the next section, we use RBM to provide additional support for this estimation procedure.

5. The RBM Approximation

Since the queueing model in Section 4 is complicated to describe and analyze in detail, we approximate the number of cells in the system and the overflow process by regulated or *reflected Brownian motion* (RBM) and its associated upper-barrier regulator process. The RBM process $Z \equiv \{Z(t): t \geq 0\}$ is ordinary Brownian motion modified by reflection at the barriers 0 and C. These processes are specified by a Brownian motion $B \equiv \{B(t): t \geq 0\}$ and a two-sided regulator map, mapping B into the triple $(Z, L, U) = \{(Z(t), L(t), U(t)): t \geq 0\}$; see p. 22 of Harrison [12] and Berger and Whitt [5]. Intuitively, the regulator (local time) processes L and U provide the necessary corrections at the lower and upper barriers to keep the buffer content process Z within the interval $[0, C]$. Formally, the two-sided regulator is the (unique) map taking a continuous function x with $0 \leq x(0) \leq C$ into the triple of continuous functions (z, ℓ, u), where $0 \leq z(t) \leq C$, $z(t) = x(t) + \ell(t) - u(t)$, $t \geq 0, \ell$ and u are nondecreasing with $\ell(0) = u(0) = 0$, $\ell(t)$ increases only when $z(t) = 0$ and $u(t)$ increases only when $z(t) = C$, i.e.,

$$\int_0^T z(t) d\ell(t) = \int_0^T [z(t) - C] du(t) = 0. \tag{5.1}$$

The goal of the approximation is to choose the Brownian motion $\{B(t): t \geq 0\}$ to

approximate the net input process $\{A(t) - S(t): t \geq 0\}$, such that the RBM $\{Z(t): t \geq 0\}$ accurately approximates the buffer content process. Then we use the upper-barrier regulator process $\{U(t): t \geq 0\}$ to approximate the cell loss process $\{N(t): t \geq 0\}$.

In the queueing model the net input process is $A(t) - S(t)$. We follow standard practice by choosing the drift δ and diffusion coefficient σ^2 of our driving Brownian motion B to match the long run mean and variance of the net input process; e.g.,

$$\delta \equiv \lim_{t \to \infty} \frac{B(t)}{t} \approx \lambda - \mu = \lim_{t \to \infty} \frac{A(t) - S(t)}{t} \tag{5.2}$$

and

$$\sigma^2 \equiv \lim_{t \to \infty} \frac{Var\, B(t)}{t} \approx \lambda c_a^2 + \mu c_s^2 = \lim_{t \to \infty} \frac{Var\,(A(t) - S(t))}{t}. \tag{5.3}$$

The parameters λ, c_a^2, μ and c_s^2 are as specified in Section 3.

As indicated above, we approximate the cell loss process by the upper barrier local time process $\{U(t): t \geq 0\}$. The *loss rate* is

$$r \equiv \lim_{t \to \infty} \frac{U(t)}{t} = \begin{cases} \dfrac{\sigma^2}{2C} & \text{if } \delta = 0 \\[2ex] \dfrac{-\delta}{e^{-\theta C} - 1} & \text{if } \delta \neq 0 \end{cases} \tag{5.4}$$

and the *asymptotic variance* is

$$\sigma_L^2 \equiv r c_L^2 \equiv \lim_{t \to \infty} t^{-1} Var\, U(t) \tag{5.5}$$

where the associated asymptotic variability parameter is

$$c_L^2 = \begin{cases} \dfrac{2C}{3} & \text{if } \delta = 0 \\[2ex] \dfrac{2(1 - e^{2\theta C}) + 4\theta C e^{\theta C}}{-\theta(1 - e^{\theta C})^2} & \text{if } \delta \neq 0, \end{cases} \tag{5.6}$$

and

$$\theta \equiv \frac{2\delta}{\sigma^2}; \tag{5.7}$$

see Section 4 of Berger and Whitt [5]. The parameters r and c_L^2 provide all that is needed to analyze the normal approximation in Section 2 and the batch means estimation procedure in Section 4.

The asymptotic variability parameter c_L^2 in (5.6) is a deceasing function of θ. For $|\theta C| < 1$, $c_L^2 \approx 2C/3$; while for $|\theta C| > 4$, $c_L^2 \approx -2/\theta$. A convenient rough approximation overall is $c_L^2 \approx \min\{2C/3, -2/\theta\}$.

It is interesting to compare the asymptotic variance of the cell loss process to the asymptotic variance of the cell arrival process. Since the rates are different, it is useful to focus on the associated variability parameters c_a^2 and c_L^2. The RBM approximation provides a way to do this. To focus on the application of interest, consider deterministic service times (so that $c_s^2 = 0$) and assume that $|\theta C| > 4$, so that $c_L^2 \approx -2/\theta$. Combining this with (5.2), (5.3) and (5.7), we obtain

$$c_L^2 \approx \frac{\rho}{1 - \rho} c_a^2. \tag{5.8}$$

For $\rho \geq 1/2$, formula (5.8) implies that $c_L^2 > c_a^2$, which is consistent with intuition. However, for $\rho \leq 1/2$, formula (5.8) implies that $c_L^2 \leq c_a^2$.

An interesting feature of (5.8) is that it does not depend on the capacity C. In

the parameter region of interest, the cell loss rate r depends strongly upon C, as indicated by (5.4), but the approximation for the variability parameter c_L^2 in (5.8) depends only upon c_a^2 and the server utilization ρ.

We can use the RBM formulas to roughly relate the lengths of the measurement intervals required to estimate the two parameters r and λ. The lengths should be proportional to rc_L^2 and λc_a^2, respectively; e.g., see Section 2 of Whitt [23]. Now suppose that $r = \lambda\epsilon$ for $\epsilon < 1$. We anticipate that the measurement interval required to estimate r accurately should be about $1/\epsilon$ times as long as the interval required to estimate λ. However, an extra correction is required to account for differences in the asymptotic variance. Formula (5.8) gives a rough indication of this difference.

Berger and Whitt [5] showed, through examples, that the Brownian approximation of queueing models is not especially accurate for the case of a low traffic intensity and a small loss probability, but we believe that it nevertheless provides the basis for important insights for the estimation problem. The Brownian approximation provides a simple way to capture the effect of the (often high) variability in the arrival process (as measured by c_a^2).

As reviewed in Section 2 of Abate and Whitt [1], any RBM with negative drift can be rescaled to be represented as canonical RBM, which has drift -1 and diffusion coefficient 1. To express the result, let $R(t; \mu, \sigma^2, C, X)$ denote an RBM with drift μ, diffusion coefficient σ^2, upper boundary at C and initial random position X. Then

$$\{aR(bt; \mu, \sigma^2, C, X): t \ge 0\} \stackrel{d}{=} \{R(t; -1, 1, aC, aX): t \ge 0\} \tag{5.9}$$

where $a = |\mu|/\sigma^2$, $b = \sigma^2/\mu^2$ and $\stackrel{d}{=}$ means equality in distribution. Hence, it suffices to give formulas only for canonical RBM.

We now exhibit first passage time distributions in order to obtain insight into the estimation procedures in Section 3. A great appeal of the RBM model is that for it we can actually perform these calculations. Let $T(a, b)$ denote the first passage time to b from a for canonical RBM. Then the Laplace transforms of $T(0, C)$ and $T(C, 0)$ are

$$Ee^{-sT(0,C)} = \frac{e^{-C}}{\cosh(\gamma C) - \dfrac{\sinh(\gamma C)}{\gamma}} \tag{5.10}$$

$$Ee^{-sT(C,0)} = \frac{e^{C}}{\cosh(\gamma C) + \dfrac{\sinh(\gamma C)}{\gamma}}, \tag{5.11}$$

where $\gamma = \sqrt{1 + 2s}$, while the first two moments are

$$E[T(0, C)] = \frac{e^{2C} - 1 - 2C}{2} \tag{5.12}$$

$$E[T(C, 0)] = \frac{2C - 1 + e^{-2C}}{2} \tag{5.13}$$

$$E[T(0, C)^2] = \frac{e^{4C} + e^{2C}(1 - 6C) + 2(C^2 - 1)}{2} \tag{5.14}$$

$$E[T(C, 0)^2] = \frac{2(C - 1) + (1 + 6C)e^{-2C} + e^{-4C}}{2}; \tag{5.15}$$

see p. 151 of Abate and Whitt [2] and Williams [24]. From (5.10) and (5.11), it follows that $T(0, C)/ET(0, C)$ converges in distribution to an exponential distribution with mean 1, thus showing that the batch Poisson process model is asymptotical-

ly correct as $C \to \infty$ for the upper barrier local time process of RBM. Also, $T(C,0)/C$ converges in distribution to a unit point mass at 1 as $C \to \infty$, so that $T(0,C) \gg T(C,0)$ for C large, as indicated in Section 3. (However, the RBM model does not account for congestion controls.)

Moreover, the amount of increase in U during a cycle (first passage from C to 0 and back to C) is exponentially distributed, see Remark 2 of Williams [24], which is consistent with the geometric batch-size distribution. (The exponential distribution is the continuous analog of the geometric distribution.) The exponential distribution could also be deduced from the known geometric distribution in the GI/M/1/C models by taking limits of queueing models approaching RBM, as in Theorem 4.5 of Berger and Whitt [5].

6. Using RBM to Fit the Batch Poisson Process

We can use the RBM approximation to generate a batch Poisson process for the cell losses. Then we model the loss process as a batch Poisson process with a geometric batch-size distribution. Let β be the rate of the Poisson process of batch arrivals and let m be the mean batch size. We show how to express the pair (β, m) in terms of the rate r and variability parameter c_L^2 of the RBM upper barrier local-time process, whose formulas are given in terms of the RBM parameters λ, μ and c_a^2 in (5.4) and (5.5) respectively.

First, the overall rate is obviously

$$r = \beta m. \tag{6.1}$$

Next, this special batch Poisson process is a renewal process, so that the asymptotic variance coincides with the squared coefficient of variation (variance divided by the square of the mean) of an interarrival time (here a mixture of an exponential and a point mass at 0). Denoting the loss process by $\{N(t): t \geq 0\}$ as before, this asymptotic variance is

$$r c_L^2 \equiv \lim_{t \to \infty} t^{-1} Var\, N(t)$$

where

$$c_L^2 = 2m - 1. \tag{6.2}$$

We obtain the parameters r and c_L^2 from the parameters of the RBM upper barrier local time process $\{U(t): t \geq 0\}$ in Section 5. Then we solve for m and β, obtaining

$$m = \frac{c_L^2}{2} + 1 \text{ and } \beta = \frac{r}{m}. \tag{6.3}$$

If more general models than the batch Poisson process are deemed necessary to faithfully represent the loss process, then formulas for the asymptotic variability parameters in Section VI of Fendick, Saksena and Whitt [10] may be useful.

7. Reverse Engineering to Determine the Variability

The batch Poisson process parameters β and m in Section 6 are obtained from the cell arrival rate λ, the cell service rate μ, and the arrival process asymptotic variance $\sigma_a^2 \equiv \lambda c_a^2$. Unfortunately, however, the variability parameter c_a^2 may not be readily available, although it can be estimated from data; e.g., see Cox and Lewis [8] and

Fendick and Whitt [11].

Here we suggest an indirect way to estimate the arrival-process variability parameter c_a^2 under the assumption that the cell loss probability is ϵ. We assume that the relevant case is $\delta \neq 0$, where δ is the drift in (5.2). Then, by (5.4),

$$\theta = \frac{-\log\left(1 - \frac{\lambda - \mu}{r}\right)}{C},$$ (7.1)

so that, by (5.7),

$$c_a^2 = \frac{2(1 - \rho)C}{\rho \log\left(1 + \frac{(1 - \rho)}{\rho\epsilon}\right)},$$ (7.2)

where $\rho = \lambda/\mu$ may be estimated from measurements. When ϵ is small, the 1 can be ignored inside the logarithm in (7.2).

To illustrate how we can apply the reverse engineering with RBM, we consider Example 1 again.

Example 1 Revisited: Supposing that the system is indeed properly designed to achieve a 10^{-7} loss probability at utilization 0.5, we do reverse engineering using the RBM model to determine an estimate of the asymptotic variance c_a^2 of the cell arrival process. The given parameters are $\lambda = 0.5 \times 10^5$, $\mu = 10^{-5}$, $C = 4000$, $r = \lambda 10^{-7} = .005$. Hence, by (5.2), $\delta = 0.5 \times 10^5$ and, by (7.2), $c_a^2 = 496.3 \approx 500$. Then, by (5.8) and (6.3), $m \approx 250$ and $\beta = 2 \times 10^{-5}$. Hence batches of mean size $m = 250$ should arrive at rate 2×10^{-5} batches per second or $(2 \times 10^{-5}) \times (3600) = 0.072$ batches per hour. The average interval between clusters of losses should thus be about 14 hours.

These parameters clearly indicate that an hour is not long enough to be confident that the requirement is being met under the assumption that the loss probability is indeed 10^{-7}. We anticipate that 93% of all hours will have no losses at all. Moreover, if an hour does have losses, then 400 is not exceptionally large. On the other hand, any losses at all should be regarded as a bad sign. Hence, even though 2 losses is well below the mean of 18, it is more a bad sign than a good one. These calculations also indicate what an appropriate threshold for identifying clusters might be. Since the average time between clusters should be about 840 minutes, a good threshold might be about 10 minutes. For the second method in Section 3, we estimated 100 seconds.

These parameters also indicate what are appropriate measurement periods. For the example, an appropriate measurement period might be 200 hours. Then the mean number of batches is 14.4. Since the distribution is approximately Poisson, the variance is also 14.4. The number of cells lost in this longer period is approximately normally distributed with mean 300 and standard deviation 1341. We would thus regard more than 6282 losses (two standard deviations above the mean) as occurring under the model assumptions only with probability 0.023.

We emphasize that our analysis assumes stationary steady-state conditions, in which the arrival rate is constant over time. That situation is hardly likely to prevail in real networks over 200 consecutive hours. However, as we have indicated, this difficulty can be addressed by selecting a subset of hours with similar operating conditions. Our assumed conditions are routinely satisfied directly in computer simulations and they can often be achieved in testbed networks.

It is also instructive to see how the RBM model looks when we transform it to canonical RBM (with drift 1 and diffusion coefficient 1). For canonical RBM, the buffer capacity becomes $C = 8$ and the hour becomes 36. Recalling that the steady-state distribution of canonical RBM is exponential with mean $1/2$, we can see that

the level 8 is unlikely to be reached within the interval $[0, 36]$.

8. Measured Losses and Utilizations

In actual system operation, it may be necessary to estimate the traffic intensity as well as losses. For example, a knowledge of the queue's traffic intensity is needed for the reverse engineering of Section 7. For fairly long measurement periods, such as hours, it is natural to assume that the actual traffic intensity coincides with the measured utilization. Then one could collect a set of hours with nearly identical traffic intensities and see if the losses with this traffic intensity are consistent with predictions. This reasoning assumes both that the underlying arrival process is stationary within the selected intervals and that the presence of losses does not alter the distribution of its realized traffic intensity significantly. The later assumption would clearly be invalid for short measurement periods, since cell loss is necessarily preceded by a period during which the number of cells arriving exceeds the number of cells served by the size of the buffer. Nevertheless, if the arrival process is stationary, we would expect the dependence between measurements of traffic intensity and cell loss to be small when the numbers of cells arriving and served over the interval are large relative to the size of the buffer.

To understand this phenomenon quantitatively, we again consider the RBM Z from Section 5, this time without an upper boundary ($C = \infty$), and we apply results from Fendick [9], which shows how the realized drift of the underlying Brownian net-input process B builds up prior to the RBM exceeding a given level c. Even though the RBM can fluctuate above c, we equate c with the buffer limit of the queue that we are modeling, since we are primarily interested in how the buffer content reaches c from below. We assume that RBM exceeds c at the end of the interval, rather than at some other point in the interval, because this maximizes the impact on the drift. If, for example, we were to condition on the RBM exceeding c at the beginning of the interval, then the impact on the drift would not be felt over the interval, but over previous intervals. Fixing the time T at the end of an interval of length t and conditioning on the event that $Z(T) > c$, we define the realized drift $\delta_{t,c}$ over the interval (a random variable) as

$$\delta_{t,c} = (B(T) - B(T-t) \mid Z(T) > c).$$

In addition, we let $d_{t,c}(x)$ denote the probability density of the random-variable $\delta_{t,c}$. By Theorem 1 of Fendick [9],

$$d_{t,c}(x) \approx \frac{1}{\sigma}\exp(\delta(x-2c)/\sigma^2 - \delta^2 t/2\sigma^2)t^{-1/2} \tag{8.1}$$

$$\times\left(\left(1 - \exp\left(\frac{2\delta(c-x)^+}{\sigma^2}\right)\right)\phi\left(\frac{(x-2c)}{\sigma t^{1/2}}\right) + \exp\left(\frac{2\delta(c-x)^+}{\sigma^2}\right)\phi\left(\frac{x}{\sigma t^{1/2}}\right)\right)$$

for $T \gg 1$ and all x, where $\phi(w) = (2\pi)^{-1/2}\exp(-w^2/2)$ is the density for the standard normal distribution and $w^+ \equiv \max\{w, 0\}$. Since t is fixed, the condition that $T \gg 1$ implies that $T - t \gg 1$, which assures that only the assumed boundary condition $Z(T) > c$ at the end of the interval, and not the initial buffer content $Z(0)$, affects behavior over the interval $[T - t, T]$.

To apply this result, we again consider the framework of Example 1, in which the service times are fixed (deterministic) with value μ^{-1}. Analogously to (5.2), we

think of the effective drift per unit time $\delta_{t,c}/t$ over the interval as the difference between an effective input per unit time $\lambda_{t,c}$ and output per unit time μ. We then define the effective traffic intensity $\rho_{t,c}$ over the interval (again a random variable) as

$$\rho_{t,c} \equiv \frac{\lambda_{t,c}}{\mu} = \frac{\delta_{t,c}/t + \mu}{\mu}. \tag{8.2}$$

We now can determine the density of $\rho_{t,c}$, say $r_{t,c}$, by setting

$$P(\rho_{t,c} \le x) = P\left(\frac{\delta_{t,c} + \mu t}{\mu t} \le x\right)$$

$$= P(\delta_{t,c} \le \mu t(x-1))$$

to obtain

$$r_{t,c}(x) = \mu t \, d_{t,c}(\mu t(x-1)). \tag{8.3}$$

We now examine the behavior of the random variable $\rho_{t,c}$ in the setting of Example 1. Figure 1 shows the mean $E\rho_{t,c}$ for Example 1 as a function of the length t of the interval preceding the conditioned event that $Z(T) > c$. The expected traffic intensity over the interval is very close to its steady-state value of 0.5 for intervals larger than a few seconds, thus confirming that the presence of cell loss should not significantly bias the estimate of traffic intensity obtained over reasonable measurement intervals such as hours or days.

Figure 2 shows the full density $r_{t,c}$ of the traffic intensity for Example 1 for four different interval lengths. As expected, the density rapidly concentrates around the steady-state mean for intervals larger than a few seconds.
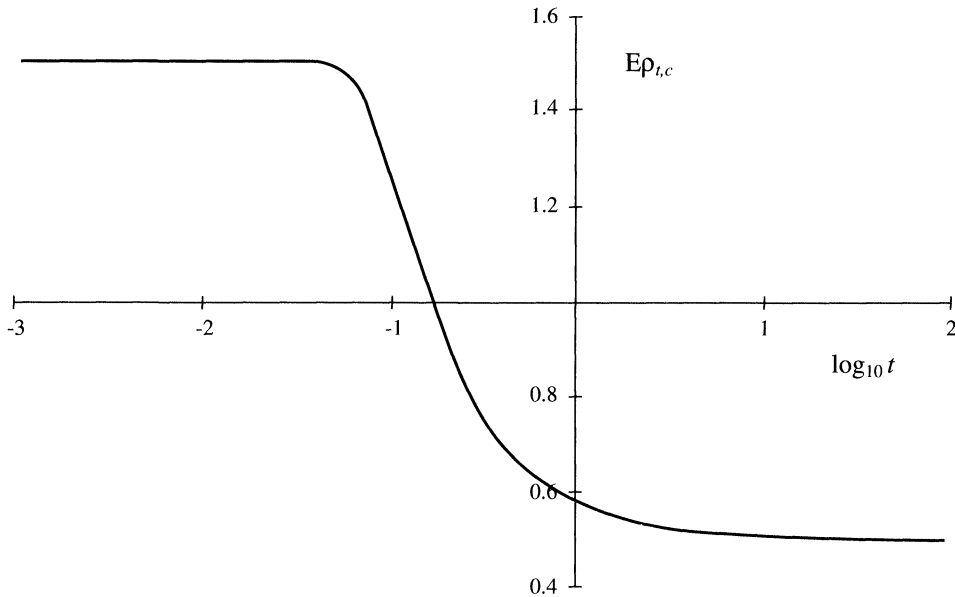

Figure 1. The expected traffic intensity $E\rho_{t,c}$ over the interval of length t seconds that precedes a buffer content in excess of $c = 4000$ cells for Example 1.
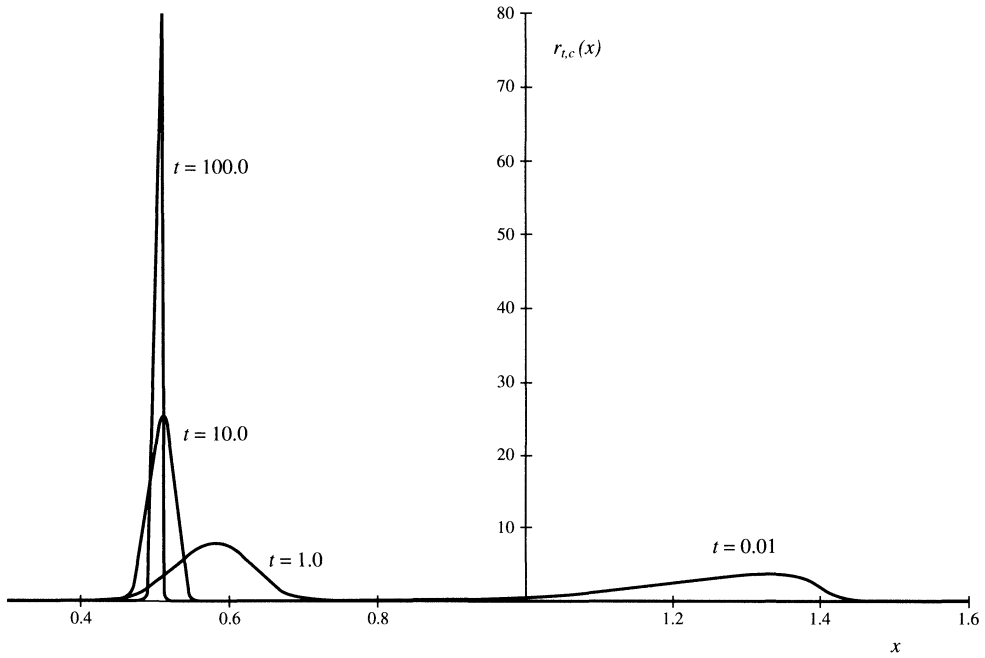
Figure 2. The probability density function $r_{t,c}$ of the traffic intensity over the interval of length t seconds that precedes a buffer content in excess of $c = 4000$ cells for Example 1.

9. Conclusions

We have presented two models for interpreting and statistically analyzing cell losses in communication networks. The first model is the batch Poisson model for cell losses; the second model is the reflected Brownian motion (RBM) model for buffer content and cell loss. The main idea in the batch Poisson model is to focus on successive clusters of losses, and separately consider the size of these clusters and the spaces between them. Since the sizes may prove to be highly variable, the spaces may prove to be the more reliable predictor of performance. We have presented arguments supporting the batch Poisson model and shown how it can be fitted and used. The natural next step is to try our approach on system measurements and simulations.

The batch Poisson model directs attention to identifying loss clusters and their rate and size. For the purpose of verifying whether or not cell loss requirements are being met, we suggest ignoring the structure of losses within clusters. However, the within-cluster structure may prove useful in understanding controls. That is a promising direction for future research.

The RBM model is a relatively simple model characterized by two parameters representing the rate and burstiness. The great appeal of RBM is that the desired calculations can be performed for it. With the RBM model, it is possible to obtain at least rough answers analytically. In Sections 6 and 7 we showed how the RBM model

could be used to fit a batch Poisson model by reverse engineering, given the buffer size, cell arrival rate, cell loss rate and utilization. From this analysis of RBM, we are able to obtain reasonable answers to the questions raised in Example 1 at the outset. These answers may be sufficient for some analyses, e.g., for determining appropriate measurement intervals. Even if more elaborate models are deemed appropriate, RBM is useful as a frame of reference or base case.

Finally, our analysis indicates that fitting measurements to stationary Poisson and RBM models and, more generally, making reliable estimates of cell loss rates, requires long periods of stationary traffic. We conclude that such long intervals of stationary traffic can be obtained by piecing together separate shorter intervals with similar utilization.

Acknowledgement

References

[1]   Abate, J. and Whitt, W., Transient behavior of regulated Brownian motion, I: starting at the origin, *Adv. Appl. Prob.* **19** (1987), 560-598.

[2]   Abate, J. and Whitt, W., Transient behavior of the M/M/1 queue via Laplace transforms, *Adv. Appl. Prob.* **20** (1988), 145-178.

[3]   Abate, J. and Whitt, W., Numerical inversion of probability generating functions, *Opns. Res. Letters* **12** (1992), 245-251.

[4]   Aldous, D., *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York 1989.

[5]   Berger, A.W. and Whitt, W., The Brownian approximation for rate-control throttles and the G/G/1/C queue, *J. Disc. Event Dyn. Syst.* **2** (1992), 7-60.

[6]   Bhattacharya, R.N., Gupta, V.K. and Waymire, E., The Hurst effect under trends, *J. Appl. Prob.* **20** (1983), 649-662.

[7]   Cohen, D.M. and Heyman, D.P., Performance modeling of video teleconferencing in ATM networks, *IEEE Trans. Circuits and Syst. for Video Technology* **3** (1993), 408-420.

[8]   Cox, D.R. and Lewis, P.A.W., *The Statistical Analysis of Series of Events*, Methuen, London 1966.

[9]   Fendick, K.W., An asymptotically exact decomposition of coupled Brownian systems, *J. Appl. Prob.* **30** (1993), 819-834.

[10] Fendick, K.W., Saksena, V.R. and Whitt, W., Dependence in packet queues, *IEEE Trans. Commun.* **37** (1989), 1173-1183.

[11] Fendick, K.W. and Whitt, W., Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue, *Proc. IEEE* **77** (1989), 171-194.

[12] Harrison, J.M., *Brownian Motion and Stochastic Flow Systems*, Wiley, New York 1985.

[13] Heyman, D.P. and Reibman, A.R., *Comparisons Among Models of Cell Losses*

for Video Conferences, AT&T Labs, Holmdel, NJ 1997.

[14] Hughes, C.J., Ghangari, M., Pearson, D.E., Seferdis, V. and Xiong, J., Modeling and subjective assessment of cell discard in ATM video, *IEEE Trans. Image Proc.* **3** (1993), 212-222.

[15] Keilson, J., *Markov Chain Models — Rarity and Exponentially*, Springer-Verlag, New York 1979.

[16] Leadbetter, M.R., Point processes generated by level crossings, In: *Stochastic Processes: Statistical Analysis, Theory and Applications*, (ed. by P.A.W. Lewis), Wiley, New York (1972), 436-467.

[17] Leadbetter, M.R., Lindgren, G. and Rootzén, H., *Extreme and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York 1983.

[18] Lewis, P.A.W., A branching Poisson process model for the analysis of computer failure patterns, *J. Roy. Statist. Soc. Ser.* **B 26** (1964), 398-456.

[19] Mood, A.M. and Graybill, F.A., *Introduction to the Theory of Statistics*, second ed., McGraw-Hill, New York 1963.

[20] Paxson, V. and Floyd, S., Wide-area traffic: The failure of Poisson modeling, *IEEE/ACM Trans. Networking* **3** (1995), 226-244.

[21] Srikant, R. and Whitt, W., Simulation run lengths to estimate blocking probabilities, *ACM Trans. Modeling Computer Sim.* **6** (1996), 7-52.

[22] Syski, R., *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd, Endinburgh 1960 (second ed., North-Holland, Amsterdam 1986).

[23] Whitt, W., Planning queueing simulations, *Management Sci.* **35** (1989), 1341-1366.

[24] Williams, R.J., Asymptotic variance parameters for the boundary local times of reflected Brownian motion on a compact interval, *J. Appl. Prob.* **29** (1992), 996-1002.

[25] Willinger, W., Traffic modeling for high-speed networks: theory versus practice, In: *Stochastic Networks* (ed. by F.P. Kelly and R.J. Williams), IMA Vol. in Math and Appl., Springer-Verlag, New York 1995.