Les Cahiers du GERAD ISSN: 0711–2440

An Empirical Comparison of Diffusion Approximations and Simulation in ATM Networks

Christiane Lemieux
Pierre L'Ecuyer

G–98–23

June 1998

An Empirical Comparison of Diffusion Approximations and Simulation in ATM Networks

Christiane Lemieux
and
Pierre L'Ecuyer

Département d'Informatique et de Recherche Opérationnelle
Université de Montréal
lemieux@iro.umontreal.ca, lecuyer@iro.umontreal.ca

June, 1998

Les Cahiers du GERAD

G–98–23

Abstract

We consider an ATM switch model, in which each of N sources is a Markov modulated rate process. We look at some approximations that have been proposed for p, the probability that in steady-state conditions, the buffer content exceeds x. These approximations are compared with the simulation results for p and the cell-loss ratio.

Résumé

Nous considérons un commutateur ATM nourri par N sources de cellules, chaque source étant un processus d'arrivée dont le taux est Markov modulé. Nous examinons quelques approximations proposées pour p, la probabilité qu'à l'état stationnaire, le contenu du tampon (la file d'attente) dépasse x. Ces approximations sont comparées avec des résultats de simulation pour p et pour la fraction des cellules perdues.

1 Introduction

An Asynchronous Transfer Mode (ATM) switch can be modeled as a network of queues with finite buffer sizes. Cells of information join the network in a stochastic manner and some may be lost due to buffer overflow. The steady-state fraction of cells that are lost at a given node is called the cell-loss ratio (CLR) at that node. Another quantity of interest is the probability p that a queue length exceeds a given level x before returning to empty, given that the queue is started from empty, for a single queue (with a buffer of infinite size) with multiple independent arrival sources. These two closely related quantities are typically very small and estimating them by straightforward simulation is very time-consuming. However, efficiency improvement methods can be applied in this situation: In particular, we can use importance sampling (IS), which is specially designed to deal with *rare-event* simulation.

Although IS makes the simulation more efficient, it is still a question of interest to see whether it is possible to use analytical approximations for the network's measures of performance. In the case of a single node with one buffer and i.i.d. arrival sources, approximations and bounds for p have been proposed [1], [4], [5]. The common idea behind these approximations is to look at the multidimensional process giving the number of sources in each state, a state being characterized by a certain emission rate of cells for the source. By letting the total number of sources, N, go to infinity and after an appropriate normalization, this process converges weakly to a multidimensional Gaussian process. In the special case of exponential sojourn times in each state, we get an Ornstein-Uhlenbeck (O-U) process.

This kind of process is more tractable analytically and so, using different approaches, one can get an approximation for p: Kobayashi and Ren [4] approximate the buffer content process by a diffusion process, which makes it very easy to estimate p, while in [1] and [5], p is directly written as a function of the limiting Gaussian process and a linear logarithmic upper bound for p, valid as $x \to \infty$, is derived. Since p is easier to work with than the CLR, approximations like these would be more difficult to derive for the CLR or would require too many simplifying assumptions or approximations to be accurate and useful in practice.

The goal of this paper is to study empirically the accuracy of these approximations. To do this, we compare them with accurate simulation results for p obtained using IS, for a discrete version of the same model: This implies that the difference observed will be due to the error of the approximations for p (in the continuous models they have been developed for) *and* the difference between the two models (discrete vs continuous). We find that for the approximation of p to be good, the number of sources must be large and a parameter related to the variance of the accumulated superposed traffic process must be small. The approximation generally deteriorates when p gets too small. In [4], the parameters were chosen so that the approximation was quite accurate. Moreover, results were given only for $p > 10^{-5}$. Here, we investigate many sets of parameters and exhibit examples where the approximation is bad, so that practitioners have an idea of when these approximations are dangerous. We also use simulation to estimate the CLR, in order to see how close it is to p and if the analytic approximations for p can be used to approximate the CLR as well. We find that the CLR is always smaller than p, sometimes by up to a factor of 10. In particular, the analytic approximations turn out to be bad for both p and the CLR when these quantities are around 10^{-9}, which is a typical value of interest in the context of ATM networks.

The rest of the paper is organized as follows. The next section explains the different approximations. Section 3 contains numerical results and a conclusion is made in Section 4.

2 The approximations for p

We first present the model under consideration, then we show how the limiting Gaussian process is obtained. After that, we explain different approaches used to get an approximation for the probability p of overflow.

2.1 How to get the limiting Gaussian process

We consider a very simple ATM switch with a single node and one buffer of infinite size. The node is equipped with a transmission link that outputs cells at a rate of C per second. There are N i.i.d. sources feeding this node, each one modeled by a *Markov modulated rate process*: A source can be in M different states and when it is in state $m \in \{1, \ldots, M\}$, it produces cells of information at a rate of R_m cells per second. The sojourn time in state $m \in \{1, \ldots, M\}$ is a random variable with mean α_m^{-1} and variance σ_m^2. These random variables are independent within each source and between sources. The state transitions follow a Markov chain with irreducible transition probability matrix $\mathcal{P}$, whose elements p_{ij} satisfy $p_{ii} = 0$ for all i, and with (unique) stationary distribution $\boldsymbol{\pi}$. Under the additional assumption that each sojourn time in a state is exponentially distributed, we will call this model an *exponential model*.

Let $\{\mathbf{n}(t),\, t \geq 0\} = \{[n_1(t), \ldots, n_M(t)]^T,\, t \geq 0\}$ be the M-dimensional process giving the number of sources in each state, where the superscript T means transpose, and $\bar{\mathbf{n}}(t) = N^{-1/2}(\mathbf{n}(t) - \mathrm{E}(\mathbf{n}(t)))$. The process $\{\bar{\mathbf{X}}(t),\, t \geq 0\} = \{[\bar{X}_1(t), \ldots, \bar{X}_M(t)]^T,\, t \geq 0\}$ represents the limit of $\{\bar{\mathbf{n}}(t),\, t \geq 0\}$, i.e. $\bar{\mathbf{n}}(\cdot) \stackrel{w}{\to} \bar{\mathbf{X}}(\cdot)$ as $N \to \infty$ where $\stackrel{w}{\to}$ means weak convergence. Finally, $\mathbf{X}(t) = N^{1/2}\bar{\mathbf{X}}(t) + \mathrm{E}(\mathbf{n}(t))$.

In the case of an exponential model, we can use limit theorems for generalizations of the multi-urn Ehrenfest model, [3], [7], to show that $\{\bar{\mathbf{X}}(t), t \geq 0\}$ is an O-U process. In the general case of a Markov modulated rate process, a result stated in [2] tells us that $\{\bar{\mathbf{X}}(t), t \geq 0\}$ is a Gaussian process, not necessarily Markov. In fact, the only stationary and continuous process that is Gaussian and Markov is the O-U process and in the exponential model, this property follows from the assumption of exponential sojourn times. In the general case, this property is no longer true. However, dealing with an O-U process is easier than dealing with a more general Gaussian process and this is probably what motivated Kobayashi and Ren to use such a limiting process for $\{\bar{\mathbf{n}}(t),\, t \geq 0\}$, even when the length of time spent in each state follows a general distribution. Their result can only be proven rigorously in the case of an exponential model.

2.2 The approximation of Kobayashi and Ren

Once we have an O-U process for $\bar{\mathbf{X}}$, two approaches can be used to approximate p. The first one is the idea used in [4]. From $\{\mathbf{X}(t), t \geq 0\}$, they define the superposed Gaussian traffic process $R(t) = \sum_{i=1}^{M-1} R_m X_m(t)$ (the non-restrictive assumption that $R_M = 0$ is also made). Then, they approximate the accumulated traffic process $\{\int_0^t R(s)\,\mathrm{d}s, t \geq 0\}$ by a shifted Brownian motion $\{A(t), t \geq 0\}$ such that $\mathrm{d}A(t) = b\,\mathrm{d}t + \sqrt{a}\,\mathrm{d}W(t)$, where W is a standard Brownian motion, $b = N[\pi_1, \ldots, \pi_{M-1}]\mathbf{R}$ is the mean stationary traffic rate, $a = \mathbf{R}^T \mathcal{B}^{-1} \mathcal{A} (\mathcal{B}^{-1})^T \mathbf{R}$ is such that the variance of the accumulated traffic process at time t, $A(t)$, is at, $\mathbf{R} = [R_1, \ldots, R_{M-1}]^T$, and $\mathcal{A}$ and $\mathcal{B}$ are two $(M-1) \times (M-1)$ matrices used in the definition of $\{\bar{\mathbf{X}}(t), t \geq 0\}$. Finally, they approximate the buffer content $\{Q(t), t \geq 0\}$ by a diffusion process $\{q(t), t \geq 0\}$ such that

$$\mathrm{d}q(t) = \begin{cases} \mathrm{d}A(t) - C\,\mathrm{d}t & \text{if } q(t) > 0 \\ \mathrm{d}A(t) - \eta\,\mathrm{d}t & \text{if } q(t) \leq 0, \end{cases}$$

where $\eta = \mathrm{E}(R_\infty \mid R_\infty < C)$ and $R_\infty = \lim_{t\to\infty} R(t)$. Remark that: (1) $\eta\,\mathrm{d}t$ is an approximation for $R(t)\,\mathrm{d}t$, (2) $q(t)$ is allowed to take negative values. The approximation for p is now easy to derive:

Proposition 1 *For a Markov modulated rate process,*

$$p = P(Q^* > x) \approx P(q^* > x) = \frac{b - \eta}{C - \eta} e^{-2(C-b)x/a},$$

for $x > 0, C > b$, where $Q^* = \lim_{t\to\infty} Q(t)$, $q^* = \lim_{t\to\infty} q(t)$ and η, b and a are defined as above.

2.3 The bound of Debiçki and Rolski

The second idea is the one presented in [1]. After having obtained an approximation $\{\bar{\mathbf{X}}(t), t \geq 0\}$ for $\{\bar{\mathbf{n}}(t), t \geq 0\}$, they directly write for $q(t)$,

$$\mathrm{d}q(t) = \begin{cases} (Z(t) - C_1)\,\mathrm{d}t & \text{if } q(t) > 0 \\ (Z(t) - C_1)^+\,\mathrm{d}t & \text{if } q(t) = 0, \end{cases}$$

where Z is a stationary Gaussian process with mean function 0 and covariance $K(t) = \mathrm{E}(Z(0)Z(t))$, and C_1 is an appropriate constant related to the output capacity C (in fact, $C_1 = C - b$, with b defined as in Section 2.2). Then $p \approx \lim_{t\to\infty} P(q(t) > x) = P(q^* > x)$, where $q^* \stackrel{d}{=} \sup_{u\geq 0} \int_0^u (Z(s) - C_1)\,\mathrm{d}s$.

Proposition 2 *If the following assumptions for the covariance function K hold: (i) K is continuous, (ii) $0 < K(0) < \infty$, (iii) $0 < \int_0^\infty K(s)\,\mathrm{d}s < \infty$ and (iv) $\int_0^\infty s^2 |K(s)|\,\mathrm{d}s < \infty$, then*

$$P(q^* > x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\psi} e^{-\gamma^2 \beta} e^{-\gamma x} + o(e^{-\gamma x}) \;\; \text{as } x \to \infty,$$

where $\gamma = C_1 / \left(\int_0^\infty K(s)\,\mathrm{d}s\right)$, $\beta = \int_0^\infty sK(s)\,\mathrm{d}s$ and $\psi = C_1/(K(0))^{1/2}$.

So, for some unknown constant $D > 0$, $P(q^* > x) \leq De^{-\gamma x}$, $x > 0$. The quantities ψ, γ and β are computed easily when $\{\bar{\mathbf{X}}(t), t \geq 0\}$ is an O-U process.

To conclude this section, we point out the similarities between the two results when we use the same O-U process $\bar{\mathbf{X}}(\cdot)$: In Proposition 1, we have $p \approx G_1 e^{-2C_1 x/a}$ whereas in Proposition 2, we have an *asymptotic upper bound* $G_2 e^{-2C_1 x/a} + o(e^{-2C_1 x/a})$, where G_1 and G_2 are two positive constants independent of x.

3 Numerical results

The results that follow are for a model with N ON-OFF homogeneous sources: The sojourn times in the ON and OFF states are geometric random variables with means κ_1 and κ_0, respectively. These random variables are mutually independent. The quantity $\rho = \kappa_1/(\kappa_1 + \kappa_0)$ is the average production rate at each source and C gives the capacity of the transmission link.

Because the simulation is easier to handle for a discrete-time model, we use the geometric distribution, which is the discrete counterpart of the exponential. For this reason, we have to use the O-U limit process $\bar{\mathbf{X}}$ suggested by Kobayashi and Ren, since the assumptions of the exponential model do not hold. However, the durations being geometric suggests that the limit process is Markov, so representing $\bar{\mathbf{X}}$ by an O-U process is heuristically justified.

The results of the simulation (SIM) give the stationary probability $p = P(Q^* > x)$ and the CLR for a buffer of capacity x. The asymptotic upper bound of Debiçki and Rolski (DR) and the approximation of Kobayashi and Ren (KOB) are for p. In (DR), we only give the first term of the bound. Therefore, (DR) is not really an upper bound for p, because the $o(\cdot)$ term is missing.

In each of the following tables, ρ and κ_1 are constants, two or three different values of N and C are chosen and for each one of these combinations, x varies. We chose values of x so that p is approximately between 10^{-6} and quantity $\hat{\Delta}$ is the relative half-width of a 99% confidence interval for p by simulation, i.e., $\hat{p} \pm \hat{\Delta}\hat{p}$ is a 99% confidence interval for p, where $\hat{p}$ is the estimator for p. These confidence intervals were computed as in [6]. $\hat{\mu}$ is the estimator for the CLR. The numbers in parentheses give the number of batches and of cycles per batch used in the simulation (see [6]). On the graphs, the small vertical bars crossing the continuous line represent the 99% confidence intervals for p obtained by simulation. Remark that these bars are not crossed by the continuous line at their respective middle-point since we use a logarithmic scale for the vertical axis. The continuous line has been drawn using linear interpolation. Also, when (DR) is very far from the simulation results, it does not appear on the graphs (see Table 1 and 2 for example).

Table 1: Comparison for $\boldsymbol{\rho}$=0.1 and $\boldsymbol{\kappa_1}$=100

N=15, C=4, (200,200)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
450	1.63E-6	1.51E-9	2.58E-5	6.14E-5	0.21
550	2.05E-7	1.91E-10	1.09E-5	2.42E-5	0.39
650	2.59E-8	2.41E-11	3.29E-6	7.16E-6	0.22
750	3.28E-9	3.05E-12	1.18E-6	2.56E-6	0.24
850	4.14E-10	3.85E-13	3.63E-7	8.24E-7	0.14
$a = 241.6$		$b = 1.5$		$b - \eta = 4.58E - 2$	
$G_1 = 9.73E - 2$		$G_2 = 9.04E - 5$			
N=30, C=8, (200,150)(and (250,300) for $x \geq 550$)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
200	1.98E-5	1.71E-11	3.97E-6	9.59E-5	0.61
300	2.50E-6	2.15E-12	1.27E-6	4.60E-6	0.77
400	3.16E-7	2.72E-13	2.35E-7	8.29E-7	0.28
550	1.42E-8	1.22E-14	4.80E-8	1.88E-7	0.53
600	5.05E-9	4.34E-15	2.60E-8	9.03E-8	0.30
$a = 483.3$		$b = 3.0$		$b - \eta = 6.22E - 3$	
$G_1 = 9.51E - 3$		$G_2 = 8.18E - 9$			
N=45, C=12, (100,100)(and (200,250) for $x \geq 450$)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
200	1.58E-6	1.26E-15	2.88E-8	1.26E-7	0.41
300	1.99E-7	1.59E-16	4.90E-9	2.02E-8	0.32
400	2.52E-8	2.01E-17	1.89E-9	8.18E-9	0.63
450	8.95E-9	7.14E-18	8.40E-10	4.36E-9	0.56
550	1.13E-9	9.01E-19	3.60E-10	1.51E-9	0.59
650	1.43E-10	1.14E-19	7.57E-11	3.40E-10	0.64
750	1.80E-11	1.44E-20	2.08E-11	7.57E-11	0.33
850	2.28E-12	1.82E-21	7.34E-12	2.66E-11	0.33
$a = 724.9$		$b = 4.5$		$b - \eta = 7.43E - 4$	
$G_1 = 9.27E - 4$		$G_2 = 7.39E - 13$			
N=60, C=16, (100,100)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
200	1.33E-7	9.86E-20	2.30E-10	1.19E-9	0.37
256	4.19E-8	3.09E-20	7.26E-11	3.58E-10	0.36
300	1.68E-8	1.25E-20	4.10E-11	2.23E-10	0.60
350	5.99E-9	4.43E-21	2.03E-11	1.01E-10	0.48
400	2.13E-9	1.57E-21	9.22E-12	4.22E-11	0.41
$a = 966.6$		$b = 6.0$		$b - \eta = 8.36E - 5$	
$G_1 = 9.04E - 5$		$G_2 = 6.69E - 17$			

Figure1: Results from Table 1 with ***N***=15

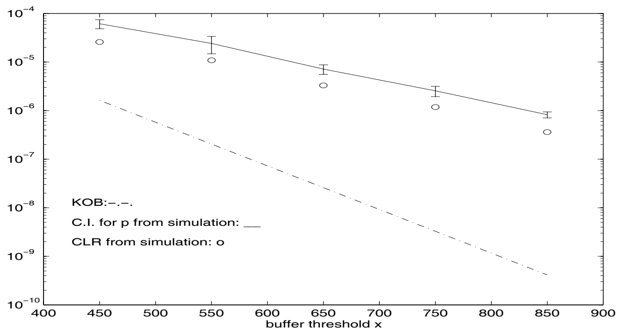


Figure 2: Results from Table 1 with ***N***=30

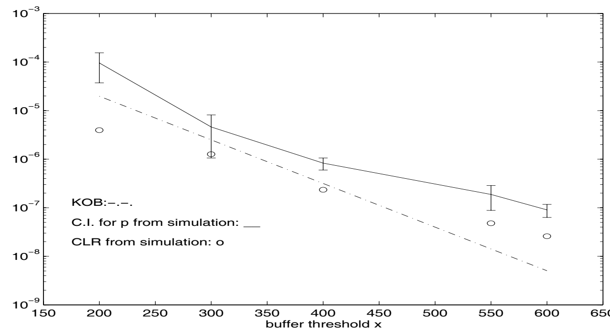


Figure 3: Results from Table 1 with ***N***=45

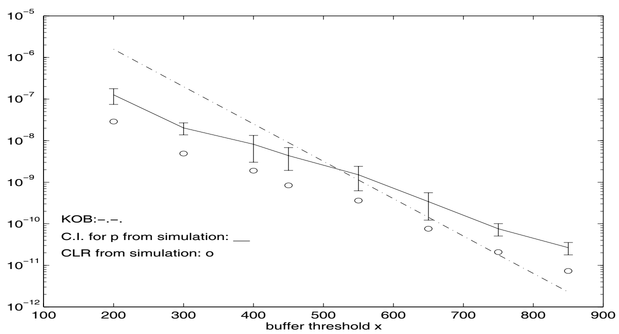


Figure 4: Results from Table 1 with ***N***=60

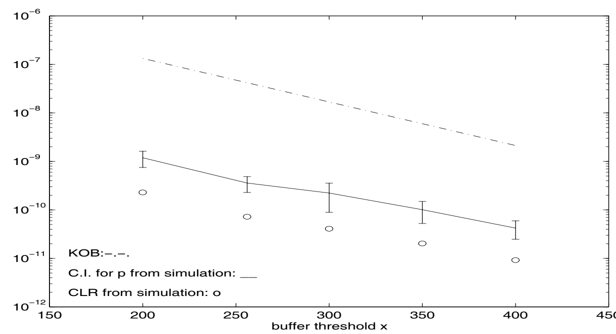


Table 1 gives some results for an average burst length of $\kappa_1 = 100$. We first observe that the (DR) asymptotic bound underestimates p by several orders of magnitude. It is practically useless in this case. We also see that (KOB) *underestimates* p by large factors for $N = 15$, *overestimates* it by roughly a factor of 10 for $N = 60$, and gives the best results when $N = 30$ and p is between 10^{-7} and 10^{-5} or $N = 45$ and p is between 10^{-11} and 10^{-9}. The underestimation for small N comes from the fact that the approximation is valid only as $N \to \infty$. When $N = 60$, what happens is that the gain in accuracy provided by a bigger N is diminished by an increase in the value of a, which makes the convergence of $\bar{\mathbf{n}}(\cdot)$ to $\bar{\mathbf{X}}(\cdot)$ slower. Also, one may infer that for a fixed value of a, there is some N_0 such that $\bar{\mathbf{X}}(\cdot)$ is a good approximation of $\bar{\mathbf{n}}(\cdot)$ for $N \geq N_0$, at least sufficiently good for the purpose of the approximation of p by (KOB). As soon as N is above this N_0, increasing it further does not improve enough the accuracy of $\bar{\mathbf{X}}(\cdot)$ to have a significant impact on the quality of (KOB). Remember that replacing $\bar{\mathbf{n}}(\cdot)$ by $\bar{\mathbf{X}}(\cdot)$ is not the only source of error for (KOB), other approximations are made as well, like when $R(t)\,\mathrm{d}t$ is replaced by $\eta\,\mathrm{d}t$.

When we change the value of κ_1, the results are approximately the same: For example, the only difference we observed by changing κ_1 from 100 to 25 was that the approximation (KOB) for $N = 60$ was slightly better, but still inaccurate.

Table 2: Comparison for $\boldsymbol{\rho}$=0.1 and $\boldsymbol{\kappa_1}$=100

N=25, C=9, (200,300)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
75	6.47E-7	3.24E-19	4.73E-7	1.97E-6	0.70
150	5.75E-8	2.88E-20	5.93E-8	2.18E-7	0.56
225	5.11E-9	2.56E-21	1.21E-8	3.92E-8	0.73
300	4.54E-10	2.27E-22	3.02E-9	8.12E-9	0.62
400	1.80E-11	9.00E-24	5.42E-10	1.61E-9	0.81
500	7.14E-13	3.57E-25	8.66E-11	2.29E-10	0.41
$a = 402.7$		$b = 2.5$		$b - \eta = 4.74E - 5$	
$G_1 = 7.94E - 5$		$G_2 = 3.97E - 17$			
N=50, C=18, (200,150)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
16	2.44E-10	6.10E-35	1.40E-9	5.39E-9	0.52
32	1.46E-10	3.64E-35	1.04E-11	8.85E-11	0.43
64	5.18E-11	1.30E-35	1.84E-12	1.05E-11	0.29
80	3.09E-11	7.73E-36	9.88E-13	5.70E-12	0.31
128	6.57E-12	1.64E-36	1.86E-13	1.03E-12	0.47
200	6.43E-13	1.61E-37	2.92E-14	1.35E-13	0.49
$a = 805.5$		$b = 5.0$		$b - \eta = 1.00E - 7$	
$G_1 = 6.30E - 9$		$G_2 = 1.58E - 33$			

Figure 5: Results from Table 2 with $\boldsymbol{N}$=25

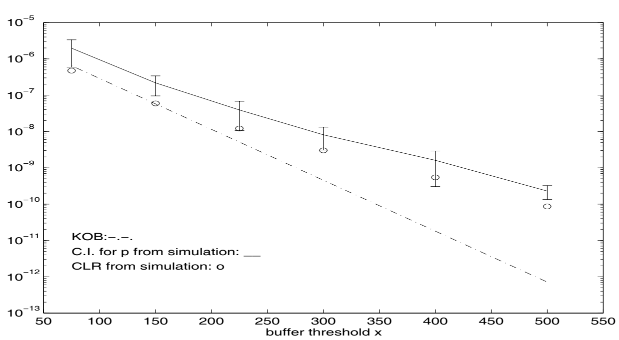
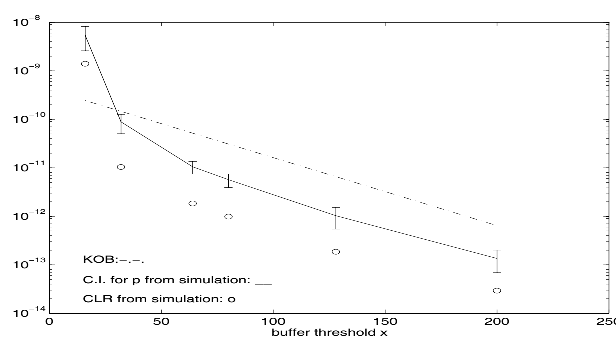


Figure 6: Results from Table 2 with $\boldsymbol{N}$=50



In Table 2, the value of a is large and (KOB) underestimates p for $N = 25$ and overestimates p for $N = 50$. For (DR), the problem is that G_2 is very small, so the term $o(e^{-2C_1 x/a})$ is probably bigger than the first term in the bound, which means that the latter is very far from the asymptotic upper bound.

Table 3: Comparison for $\boldsymbol{\rho = 0.2}$ and $\boldsymbol{\kappa_1 = 4}$

N=15, C=4, (200,300)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
75	3.47E-6	1.99E-6	6.44E-6	3.35E-5	0.27
100	7.32E-8	4.20E-8	2.57E-7	1.31E-6	0.21
125	1.55E-9	8.86E-10	1.57E-8	6.93E-8	0.55
150	3.26E-11	1.87E-11	4.85E-10	2.47E-9	0.25
175	6.89E-13	3.95E-13	1.96E-11	1.08E-10	0.32
$a = 12.96$		$b = 3.0$		$b - \eta = 0.584$	
$G_1 = 0.65$		$G_2 = 0.37$			
N=30, C=8, (200,200)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
50	1.01E-4	2.48E-5	7.30E-6	4.79E-5	0.77
75	2.12E-6	5.24E-7	2.29E-7	1.53E-6	0.35
100	4.48E-8	1.11E-8	9.83E-9	6.07E-8	0.31
150	2.00E-11	4.93E-12	1.77E-11	1.20E-10	0.43
$a = 25.92$		$b = 6.0$		$b - \eta = 0.583$	
$G_1 = 0.56$		$G_2 = 0.14$			
N=45, C=12, (200,200)					
x	KOB	DR	SIM $\hat{\mu}$	SIM $\hat{p}$	$\hat{\Delta}$
50	6.67E-5	7.55E-6	3.18E-7	4.09E-6	0.25
75	1.41E-6	1.59E-7	4.12E-9	3.56E-8	0.30
100	2.97E-8	3.36E-9	1.57E-10	1.23E-9	0.20
125	6.27E-10	7.10E-11	7.29E-12	5.39E-11	0.24
150	1.32E-11	1.50E-12	3.87E-13	2.84E-12	0.61
$a = 38.88$		$b = 9.0$		$b - \eta = 0.528$	
$G_1 = 0.46$		$G_2 = 5.17E - 2$			

Figure 7: Results from Table 3 with $\boldsymbol{N}$=15

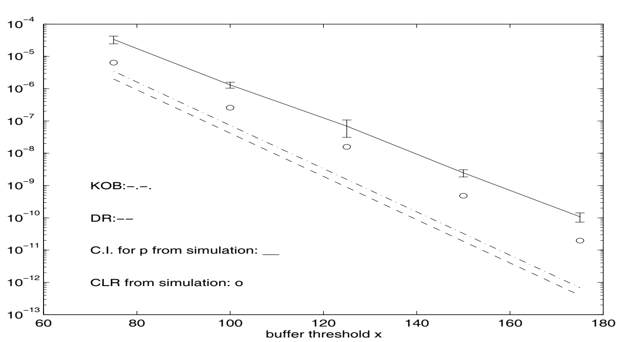
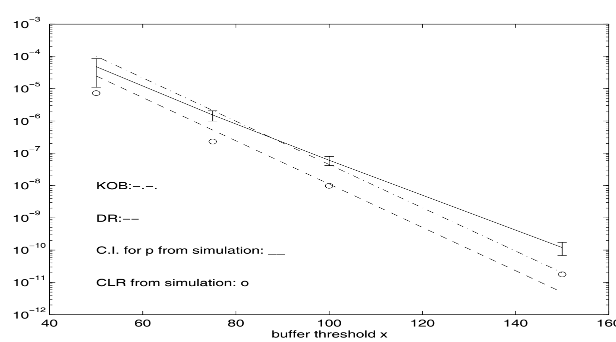
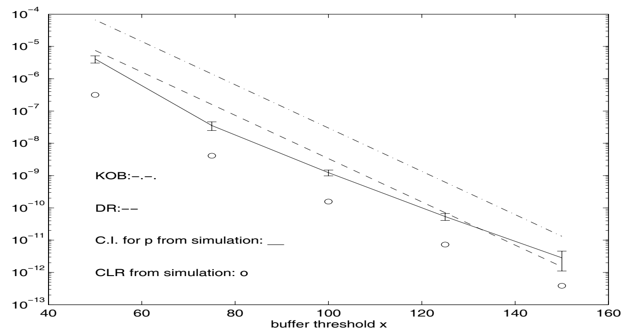


Figure 8: Results from Table 3 with $\boldsymbol{N}$=30



In Table 3, ρ is twice as large as in the first two tables. This means a higher traffic intensity ($b/C = 0.75$ instead of 0.375 in Table 1 and 5/18 in Table 2), a larger value of $b - \eta$, and a smaller variance parameter a. These facts imply that the approximation of $R(t)\,\mathrm{d}t$ by $\eta\,\mathrm{d}t$ is more accurate here than in Tables 1 and 2: For example, in Table 2, for $N = 50$, $b - \eta = 1.00\mathrm{E} - 7$, which means that the approximation of $R(t)\,\mathrm{d}t$ by $\eta\,\mathrm{d}t$ is almost equivalent to replace $A(t)$ by ηt in $q(t)$, since the buffer is empty very often. But $a = 805.5$ in this case: This very large value implies that replacing the process $\{A(t), t \geq 0\}$ by the constant process $\{\eta t, t \geq 0\}$ is rather inaccurate.

Figure 9: Results from Table 3 with $\boldsymbol{N}$=45



Also, in Table 3, the constant G_2 is larger than in Tables 1 and 2, due to the fact that the traffic is more intense here. This fact implies that the term in $o(e^{-2C_1x/a})$ contributes less to the upper bound, meaning that even if we look only at the first term of the bound, it can give a good idea of the value of p, as we can see in Table 3, $N = 45$ for example. Another observation is that when a is smaller, the fact that N is not very large is less relevant for the quality of the approximation: This is illustrated in Table 3, where (KOB) seems to be almost as good when $N = 15$ than when $N = 45$. The variance parameter a in this table is smaller than in Table 1 and as we mentioned earlier, in this table, (KOB) performs poorly when $N = 15$ compared with $N = 45$. Also, notice that for $N = 45$ in Table 3, (KOB) still overestimates p for values around 10^{-12}. This suggests that the values of x where (KOB) will get close to p are very large or equivalently, will be such that p is small.

Here again, changing the value of κ_1 leads to the same conclusions about the accuracy of (KOB): For example, if $\kappa_1 = 2$ instead of 4, (KOB) gets slightly better for $N = 45$, but all the observations made for the results of Table 3 still hold. We now make some general observations concerning all the sets of parameters.

In every table, the best results are obtained when $N = 30$ or 45. More precisely, we can say that for $N \approx 45$ and a between 10 and 400, (KOB) can estimate p with a relative error of 0.5, at least for a small range of values of x. For each set of parameters, the (KOB) approximation remains close to the true value of the CLR for small values of x, and starts to diverge at a certain point when x increases. For large x, the best we can do is to give $2C_1/a$, the asymptotic rate at which (KOB) goes to 0 as $x \to \infty$.

Concerning the relation between p and the CLR, our results show that the CLR is always smaller than p, but these two quantities never differ by a factor greater than 10. When we compare the simulation results for the CLR with (KOB), we see that the latter can approximate the former with a relative error of 0.5 for $N \approx 30$ and a is between 10 and 400, but is not accurate for $N = 15$ or 45.

4 Conclusion

We have explained how the approximation for p was constructed in [4] and how an upper bound was derived in [1]. In both cases, the key is to approximate $\{\bar{\mathbf{n}}(t), t \geq 0\}$ by a diffusion process $\{\bar{\mathbf{X}}(t), t \geq 0\}$, which must be an O-U process in the former case. We also showed the similarities between these two quantities. The bound given in [1] contains an $o(\cdot)$ term with an unknown constant, which makes it not very useful in practice.

The two factors that influence most the quality of the approximation proposed by Kobayashi and Ren [4] are the number of sources N and the variance parameter a of the accumulated arrival process. These two factors interact together: A small N is less harmful when a is small and when both of them increase, the gain in accuracy obtained by a larger N is diminished or even lost due to the augmentation of the parameter a. The value of $b - \eta$ also seems to have an impact on the quality of the approximation and the order of magnitude of p must also be taken into account.

The sets of parameters for which (KOB) is good for directly estimating small p's appear to be rather limited. An idea for further research would be to explore how this approximation can be used to increase the efficiency of the simulation estimator, for example by using it as an external control variate in the simulation. Another idea would be to work directly on (KOB) i.e. to look at the different approximations that have been made to get it and see how they could be refined.

5 Acknowledgments

This work was supported by the Natural Science and Engineering Research Council of Canada via a scholarship to the first author and grants # ODGP0110050 and SMF0169893 to the second author. We thank Hossain Pezeshki-Esfahani of Newbridge Networks for suggesting this study and Søren Asmussen for pointing out references [1] and [5].

References

[1] K. Debiçki and T. Rolski. A Gaussian fluid model. *Queueing Systems*, 20:443–452, 1995.

[2] P.W. Glynn. Diffusion approximations. In D.P. Heyman and M.J. Sobel, editors, *Handbook in Operations Research and Management Science, II: Stochastic Models*, chapter 4. North-Holland, Amsterdam, 1990.

[3] D.L. Iglehart. Limit theorems for the multi-urn Ehrenfest model. *The Annals of Mathematical Statistics*, 39:864–876, 1968.

[4] H. Kobayashi and Q. Ren. Diffusion process approximations of a statistical multiplexer with Markov modulated bursty traffic sources. In *Proceedings of the Conference on Global Telecommunications (Globecom)*, pages 1100–1106, November 1994.

[5] V. Kulkarni and T. Rolski. Fluid model driven by an Ornstein-Uhlenbeck process. *Probability in the Engineering and Informational Sciences*, 8:403–417, 1994.

[6] P. L'Ecuyer and Y. Champoux. Importance sampling for large ATM-type queueing networks. In *Proceedings of the 1996 Winter Simulation Conference*, pages 309–316. IEEE Press, December 1996.

[7] S. Schach. Weak convergence results for a class of multivariate Markov processes. *The Annals of Mathematical Statistics*, 42:451–465, 1971.