Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Performance Evaluation 52 (2003) 237–267

PERFORMANCE EVALUATION
An International Journal

www.elsevier.com/locate/peva

$M/G/c/K$ blocking probability models and system performance

J. MacGregor Smith*

Department of Mechanical and Industrial Engineering, University of Massachusetts, Marston Hall, Room 111B, Amherst, MA 01003, USA

Received 14 August 2001; received in revised form 6 September 2002

Abstract

An exact solution for the $M/G/c/K$ model is only possible for special cases, such as exponential service, a single server, or no waiting room at all. Instead of basing the approximation on an infinite capacity queue as is often the case, an approximation based on a closed-form expression derivable from the finite capacity exponential queue is presented. Properties of the closed-form expression along with its use in approximating the blocking probability of $M/G/c/K$ systems are discussed. Extensive experiments are provided to test and verify the efficacy of our approximate results.

Keywords: $M/G/c/K$; Blocking probability; Approximations

1. Motivation and purpose

Most physical systems cannot exist without finite buffers. For example, in manufacturing systems, assembly lines, material handling systems, and cellular manufacturing systems all harbor finite buffers. The buffers are important in de-coupling the machines, making them act independently of one another, and avoiding starvation and blocking of the raw materials flowing through them.

In telecommunication systems, there are finite capacity telephone lines, computer networks, and capacitated ATM switches [6,7]. Finally, in service systems such as facilities, there are limited space circulation systems (elevators, stairways and corridors), capacitated activity spaces, and finite storage areas. Fig. 1 illustrates a typical finite network topology which often occurs in practice.

In all these system applications, it is crucial to compute the performance measures of these systems with finite buffers. It is also important to determine the optimal configuration of these systems with these finite buffers in mind. In this paper, closed-form expressions are developed for the blocking probability

* Tel.: +1-413-545-4542; fax: +1-413-545-1027.
E-mail address: jmsmith@ecs.umass.edu (J.M. Smith).

PII: S0166-5316(02)00190-6

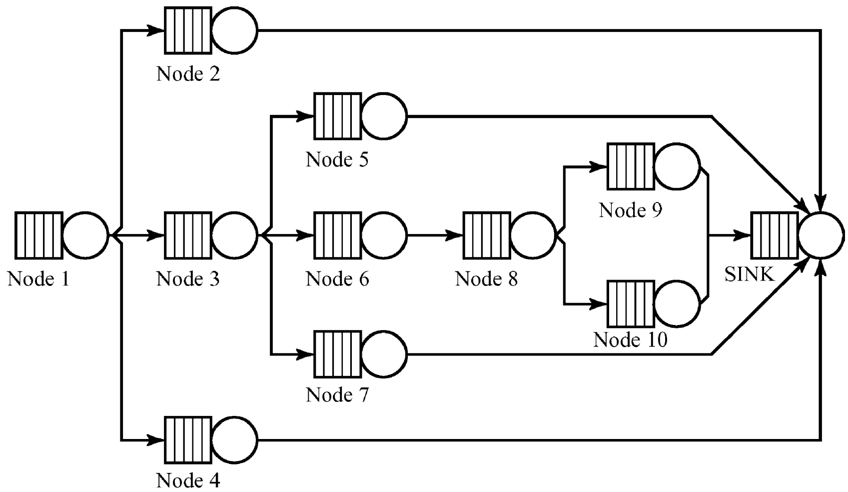


Fig. 1. Finite network geometries.

of these general service systems and use of these expressions in the optimization of these systems is also demonstrated.

1.1. *Outline of paper*

Section 2 of the paper provides necessary background and a brief literature review of the problem. Section 3 is concerned with developing a closed-form expression of the optimal buffer size in $M/M/1/K$ and $M/M/c/K$ systems which forms the basis of the approximations in the remaining sections. Section 4 describes experimental results of the closed-form expression for the optimal buffer size and blocking probability in $M/G/1/K$ systems, and Section 5 describes the approximation for the blocking probability and optimal buffer size in $M/G/c/K$ systems. Extensive computational results are included in Sections 4 and 5 to show the efficacy of the approach while Section 6 closes out the paper.

2. Background

We will examine $M/G/c/K$ systems where we have Poisson input with rate λ, a general service time S, c identical servers, and $K \geq c$ waiting places in a queue for customers (including those in service).

Arriving customers who find K other customers in the system are turned away or are lost. The long-run probability that arrivals are rejected (blocking probability) is p_K.

An exact solution for the $M/G/c/K$ model is only possible for special cases, such as for exponential service, a single server, or no waiting room at all. Given the complexity of this problem, a solution would be to develop robust and efficient approximations for the optimal buffer size and blocking probabilities of these systems. In this paper, we will explore the development of an expression for p_K which is arrived at by first looking at the optimal buffer size of the system, then inverting the buffer expression to yield the expression for p_K.

2.1. The problem

There are essentially two problems of interest in this paper. The first is how to estimate the blocking probability p_K and the second problem concerns the allocation of buffers so that the loss/delay blocking probability will be below a specific threshold. In one sense, we will treat the second problem initially, then show how after arriving at a closed-form expression for the optimal buffer size, we can arrive at an estimate of the blocking probability.

For the latter problem, it can be argued that the problem in its simplest form is to find the smallest integer $K \geq 0$ for which $p_K \leq \epsilon$ for any acceptable threshold level $\epsilon \in (0, 1)$. For the most part, we will assume that $\rho < 1$ since if $\rho \geq 1$ there may not exist an optimal value of K [14]. If we have more than one finite queue in a network of queues, Fig. 1, the topology of the queueing network can become a difficult problem to model because of the interdependence of the blocking of one workstation upon another.

In principle, for the $M/G/c/K$ system, one could develop the blocking probabilities for fixed K and all the other probabilities, but this would be tedious and not as useful when it comes to modeling the design of queueing networks with varying buffer sizes, see Chapter 5, Section 5.1.8 in [5]. One of the main objectives of this paper, however, is to develop closed formed expression for p_K which is more easily manipulated and more practical and does not depend upon the computation of probabilities of infinite systems.

2.2. $M/G/1/K$ blocking probability expressions

There are numerous approximations for the blocking probability possible for $M/G/1/K$ systems. One survey paper by Springer and Makens [24] analyzes in some detail five different approximation

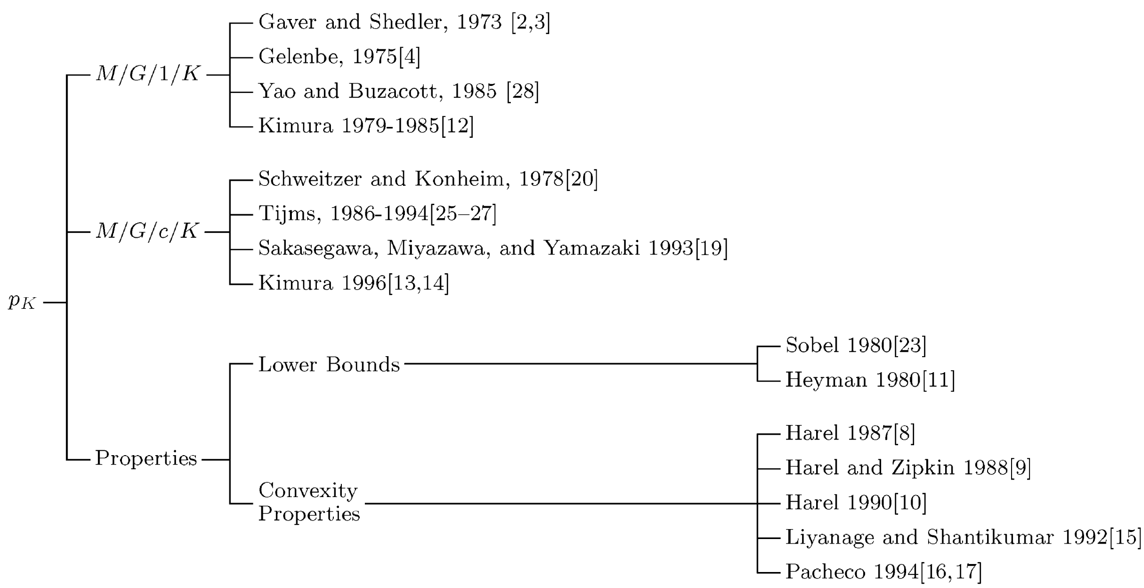


Fig. 2. p_K morphological tree.

formula and concludes that the formula by Gelenbe is the most accurate and robust. We shall focus on Gelenbe's formula and compare it with our approximation in the experimental sections of the paper. The other approximations of some note which are based on a diffusion approximation approach that were compared by Springer and Makens include Gaver and Shedler [2,3], Kimura [12], and Yao and Buzacott [28].

2.3. $M/G/c/K$ blocking probability expressions

Most of the expressions for computing p_K in multi-server systems are similar to the following formula developed by Tijms [26]

$$\tilde{p}_K = \frac{(1-\rho)\sum_{j=K+c}^{\infty}\pi_j^{(\infty)}}{1-\rho\sum_{j=K+c}^{\infty}\pi_j^{(\infty)}},$$

where $\pi_j^{(\infty)}$ denotes the long-run probability of finding upon arrival j customers present. Other researchers have developed similar formulas based upon the conditional probabilities and the tail probabilities of the infinite capacity model including those by Schweitzer and Konheim [20] and also the work of Sakasegawa et al. [19].

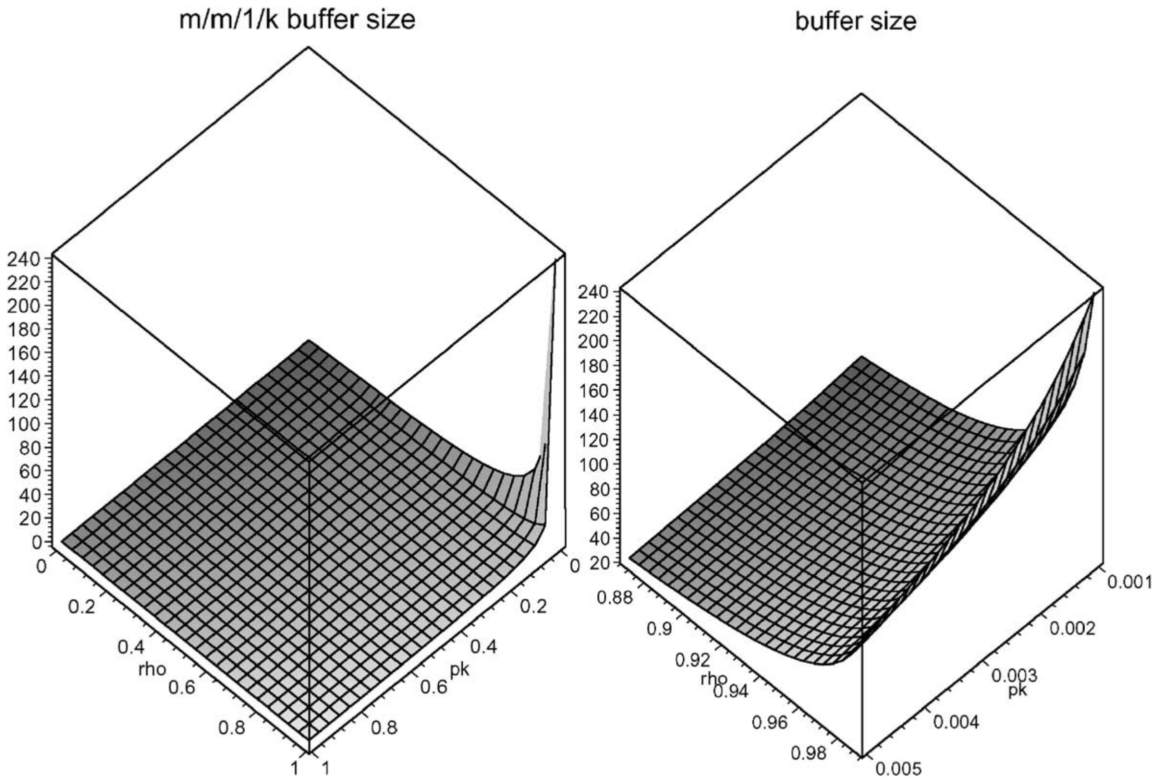


Fig. 3. $M/M/1/K$ function properties.

Heuristics based upon this approach have one computational advantage since the probabilities of the infinite model only have to be computed once. However, for certain situations, computing these probabilities can be problematic when the buffer size gets large [27].

2.3.1. *Bounds on the blocking probability*

Lower bounds are very desirable but are very scarce for this situation. The simplest lower bound which is the inequality

$$p_K \geq 1 - \frac{1}{\rho}$$

is due to Sobel [23] and Heyman [11]. This bound, however, is not accurate unless $\rho \gg 1$. While we will not develop a lower bound, we will develop a close approximation of p_K as we shall see.

Upper and lower bounds based on $M/M/c/K$ systems are generally pretty good if c and ρ is also large. We shall experimentally compare our approximation with $M/M/c/K$ systems to illustrate the quality of these bounds.

2.3.2. *Properties of the blocking probability*

In certain buffer design contexts, one may use an expression for the optimal size as an objective function. In this sense, the convex nature of the objective function is important to understand. If one examines the literature on the convexity of queueing systems, one will come across the work of Harel and Zipkin [8],

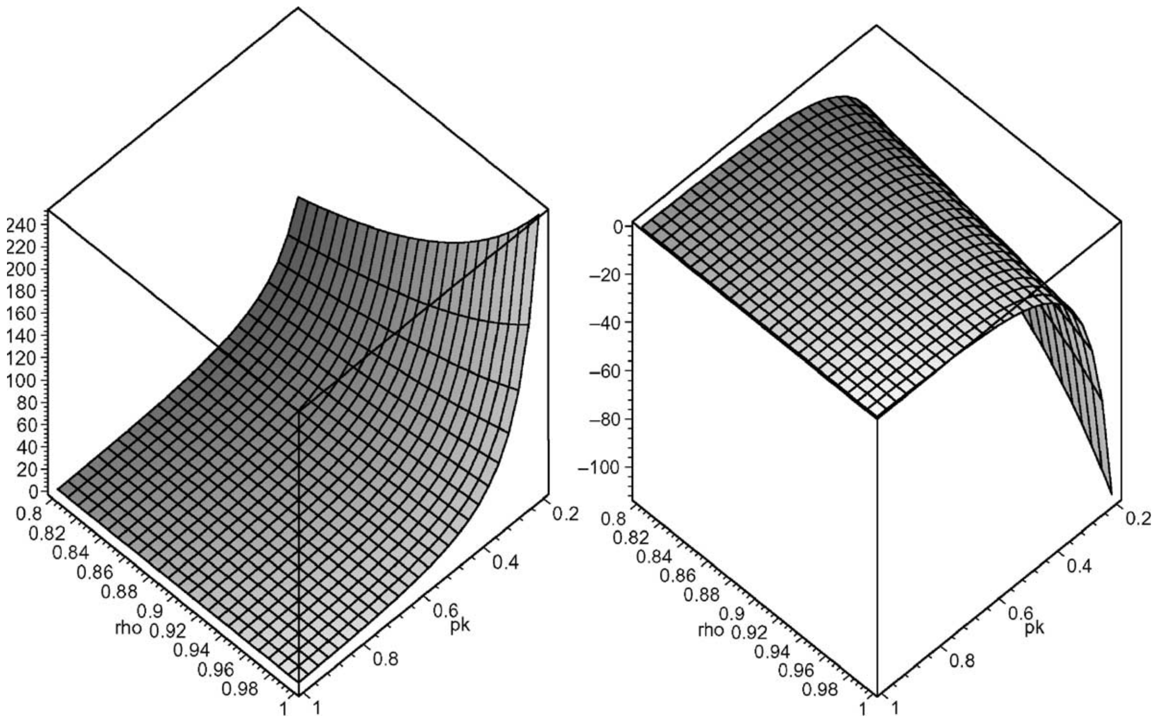


Fig. 4. First diagonal H_{11} terms and off-diagonal Hessian terms H_{12}, H_{21}.

Harel [9], and Harel [10], and more recently Liyanage and Shantikumar [15]. Liyanage and Shantikumar compile comprehensive results for many systems including some new results on finite systems. Harel and Zipkin examined the convexity properties of queueing systems and generally focused on infinite queueing systems except for their examination of Erlang loss systems. Pacheco [16,17] has generalized the analysis of Erlang loss systems and some of the same results we arrive at are echoed in his paper. However, to the author's knowledge, no one has really examined the convexity properties of the blocking probabilities relaxing the integrality of K starting from the $M/M/1/K$ system and then on to the $M/M/c/K$ systems.

If one relaxes this integrality as we shall see, then a closed-form expressions for K can be realized. Also, when examining this closed-form expression, the convexity properties will be examined and ultimately, with this closed-form expression, we will develop our approximation for p_K.

A morphological tree which summarizes many of the approaches for constructing closed-form expressions and the properties of these blocking probabilities is described in Fig. 2.

3. Mathematical models

In this section of the paper, we will develop our approximation models for the threshold buffer size and blocking probabilities of finite single and multi-server queueing systems.

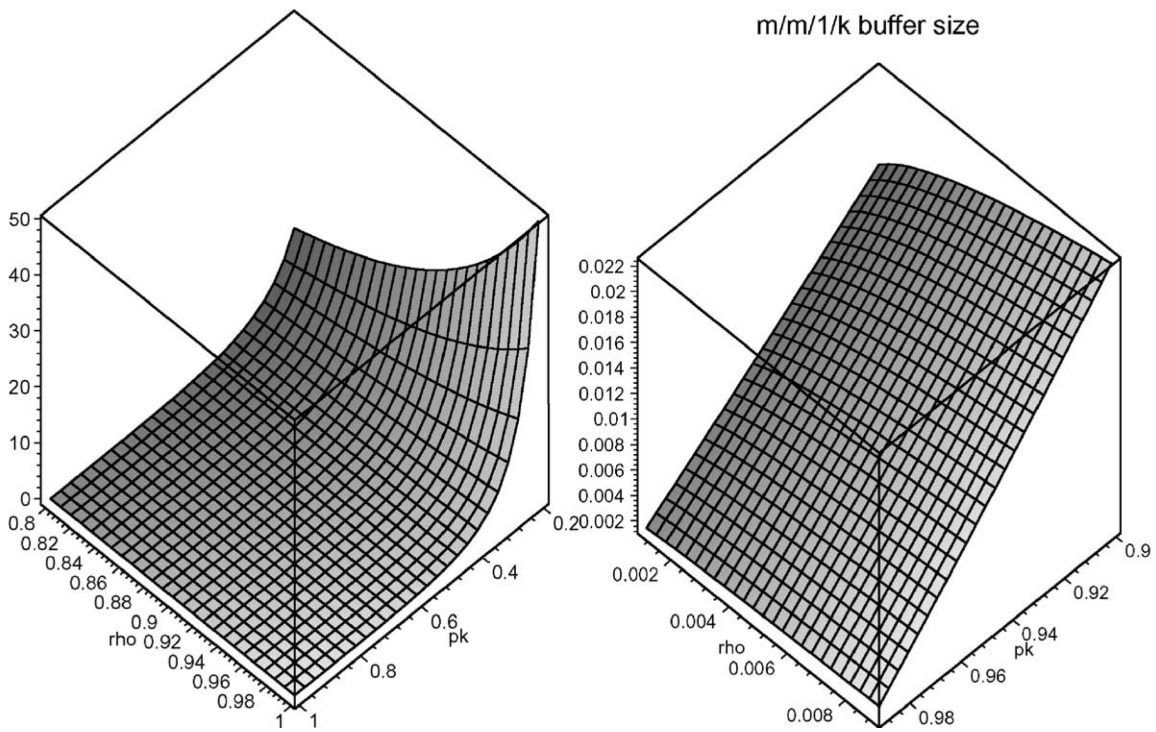


Fig. 5. Hessian term H_{22} on top and non-convex region on the bottom.

3.1. Notation

The following section presents some of the notation we need for the paper:

a^2	squared coefficient of variation of the arrival process
B_j	buffer capacity at node j *excluding* those in service
c	number of servers
K_j	buffer capacity at node j *including* those in service
p_K	blocking probability of finite queue of size K
p_0^j	unconditional probability that there is no customer in the service channel at node j (either being served or being held after service)
s^2	squared coefficient of variation of the service process
$\epsilon \in (0, 1)$	threshold for the blocking probability
Θ	mean throughput rate
λ_j	Poisson arrival rate to node j
Λ	external Poisson arrival rate to the network
μ_j	exponential mean service rate at node j
$\rho = \lambda/\mu c$	the traffic intensity

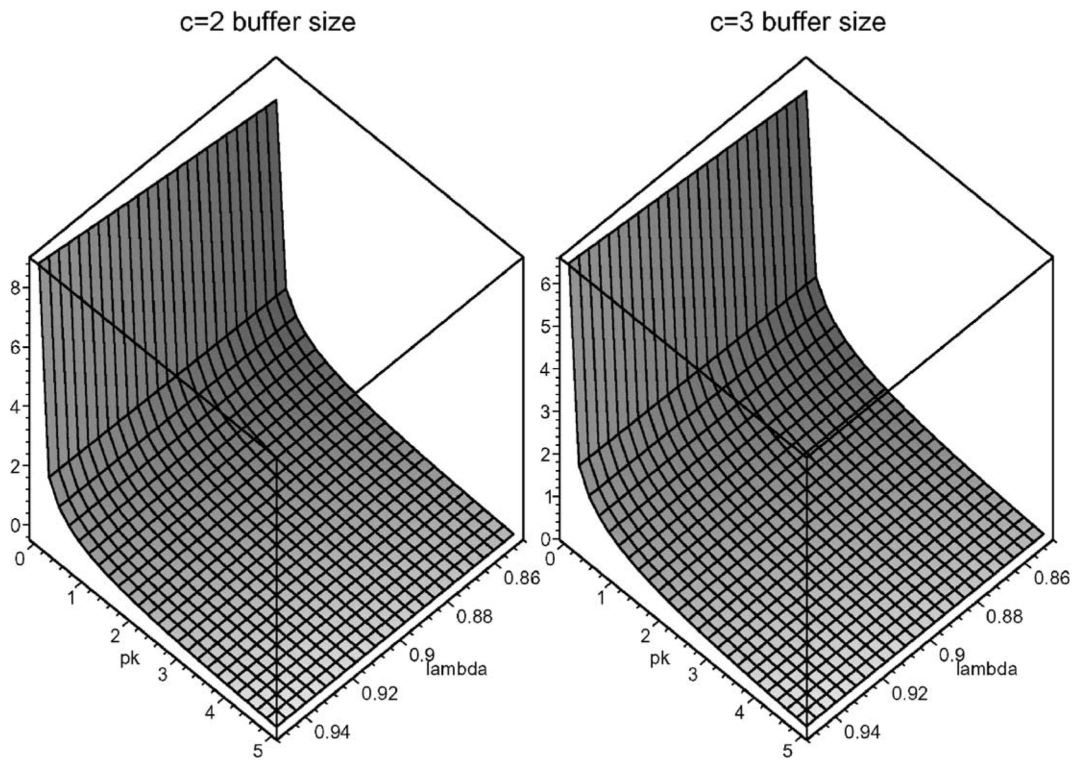


Fig. 6. $M/M/2/K$ closed-form expression on the left and $M/M/3/K$ on the right.

3.2. Assumptions and definitions

We will confine ourselves to Markovian arrival processes since exact results can be derived for these systems.

3.3. $M/M/1/K$ expression

The blocking probability for an $M/M/1/K$ system with $\rho < 1$ is well-known

$$p_K = \frac{(1-\rho)\rho^K}{1-\rho^{K+1}}.$$

If we relax the integrality of K, we can express K in terms of ρ and p_K and arrive at a closed-form expression for the buffer size which is the largest integer as follows:

$$K = \left\lceil \frac{\ln(p_K/(1-\rho+p_K\rho))}{\ln(\rho)} \right\rceil.$$

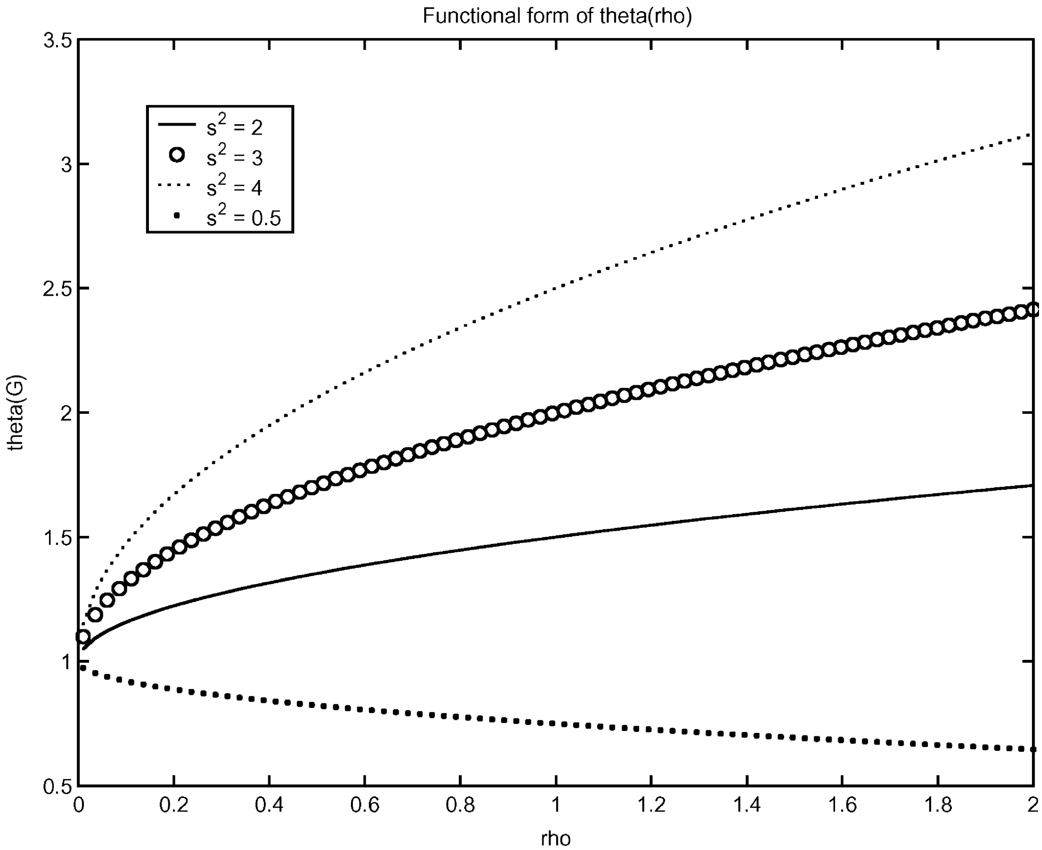


Fig. 7. Kimura's approximation.

What also is interesting about the formula is illustrated in Fig. 3 which shows the smooth monotonic nature of the function over the range of ρ and p_K. The general range of the function for all values of ρ, p_K is in Fig. 3 and the specific function values for K in the range of most practical interest is in Fig. 3 on the right.

3.4. Monotonicity/convexity of buffer formula

Here, we shall relax the integrality property of K, and as is standard practice, we shall take a classical approach in order to examine the convex nature of the closed-form expression.

The Hessian matrix, will be derived, first differentiating with respect to ρ and then p_K. For the sake of the argument and compactness of the presentation, the terms of the Hessian are equal to

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

Let us examine each of the Hessian components, term-by-term. The first diagonal term H_{11} is

$$H_{11} = \frac{(-1+\rho)(2p_K\rho+1-\rho)}{(1-\rho+p_K\rho)^2 p_K^2 \ln(\rho)}.$$

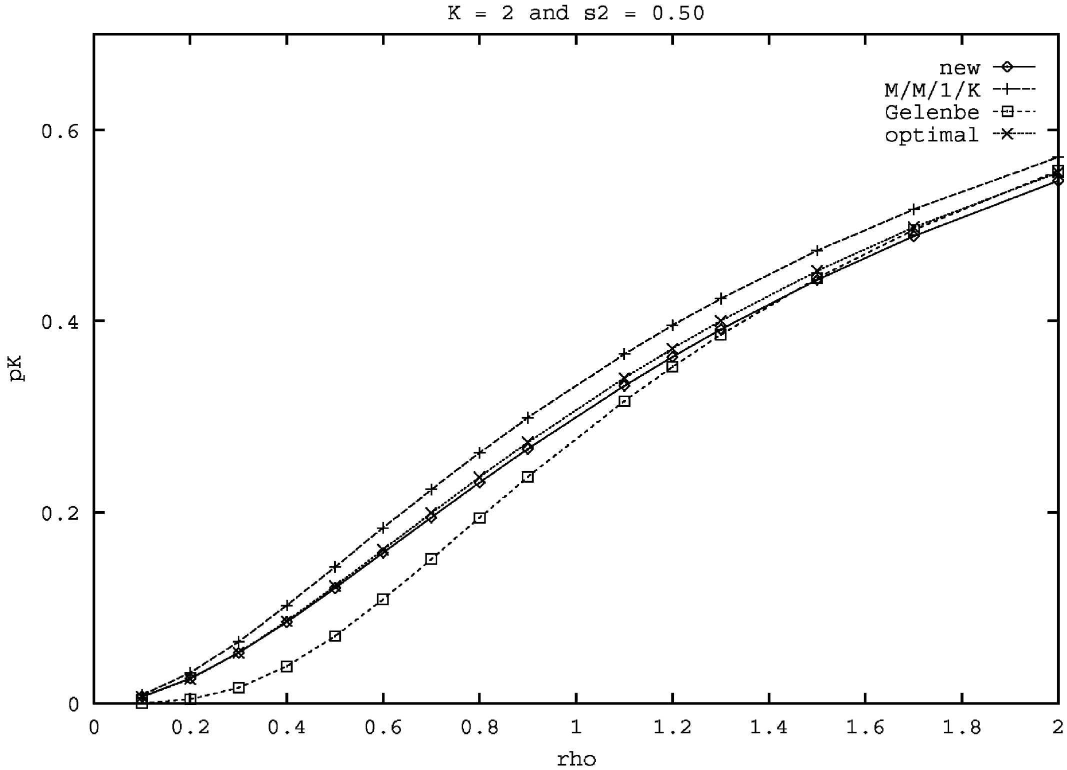


Fig. 8. p_K comparisons for $M/G/1/2$ system, $s^2 = 1/2$.

If one examines the first term in the matrix, we shall see that it is positive everywhere in the range of $0 < \rho < 1$ and $0 < p_K < 1$. A graph of this function is indicated in Fig. 4

For the off-diagonal terms $H_{12} = H_{21}$ we have

$$H_{12} \equiv H_{21} = -\frac{\ln(\rho) p_K \rho - 1 + 2\rho - p_K \rho - \rho^2 + p_K \rho^2}{(1 - \rho + p_K \rho)^2 \ln(\rho)^2 p_K \rho}.$$

This term is negative everywhere in the range $0 < \rho < 1$ and $0 < p_K < 1$. A graph of this function is indicated in Fig. 4.

The second diagonal term H_{22} is given as

$$H_{22} = \frac{(-1 + p_K)^2}{(1 - \rho + p_K \rho)^2 \ln(\rho)} + 2\frac{-1 + p_K}{(1 - \rho + p_K \rho) \ln(\rho)^2 \rho} + 2 \ln\left(\frac{p_K}{1 - \rho + p_K \rho}\right) \ln(\rho)^{-3} \rho^{-2} + \ln\left(\frac{p_K}{1 - \rho + p_K \rho}\right) \ln(\rho)^{-2} \rho^{-2}.$$

This second diagonal term H_{22} unfortunately is not positive everywhere in the range of $0 < \rho < 1$ and $0 < p_K < 1$. A graph of this function is indicated in the range where it is actually positive which is a most practical situation.

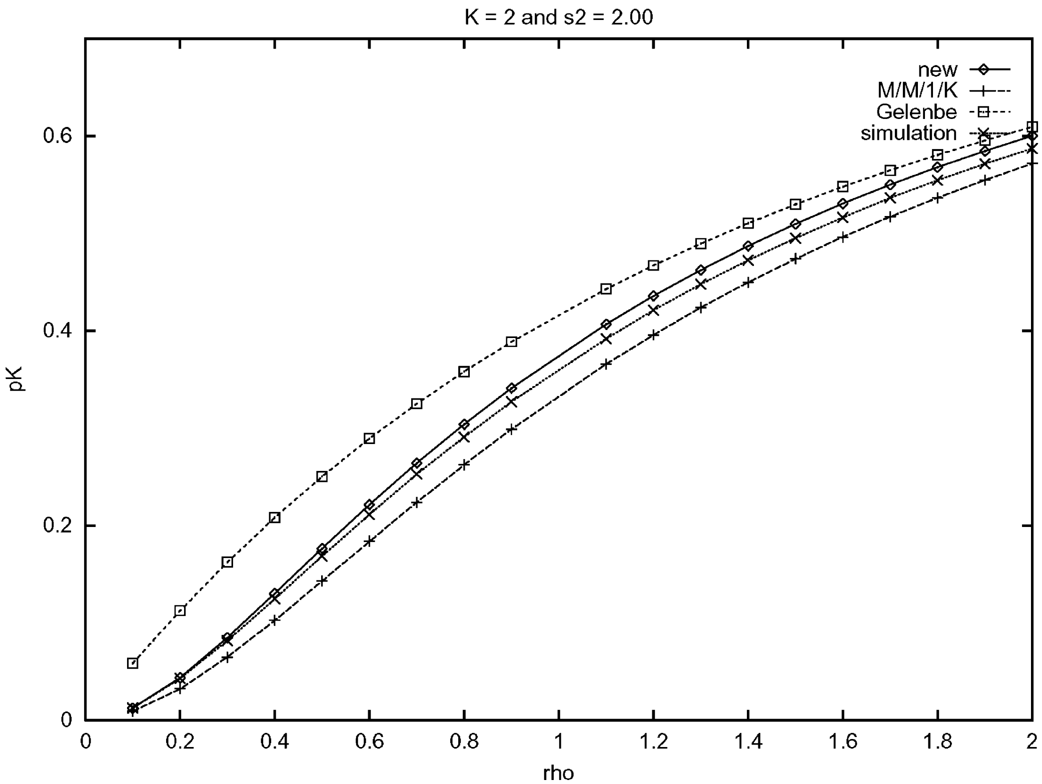


Fig. 9. p_K comparisons for $M/G/1/2$ system, $s^2 = 2$.

Since the first term in the Hessian is positive everywhere, the Hessian will be positive definite if the magnitude of the product of the diagonal terms is strictly larger than the magnitude of the product of the off-diagonal terms, i.e.

$$H_{11} \times H_{22} > H_{12} \times H_{21}.$$

For example, say that $\rho = 0.80$, $p_K = 0.001$, then the Hessian is

$$H = \begin{bmatrix} 4.481 \times 10^6 & -24892.748 \\ -24892.748 & 964.193 \end{bmatrix}.$$

The first two principal minors are 4.481×10^6, 3.701×10^6 and the eigenvalues are 4.481×10^6, 825.894. Thus, one would expect the function to be convex.

Table 1
Optimal buffer capacity for $M/D/1/K$ and $M/M/1/K$ systems

ϵ		ρ								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10^{-2}	Det	2	3	3	4	4	5	6	8	13
	$s = 0$	1.804	2.338	2.839	3.376	4.001	4.830	6.002	7.966	12.392
	Exp	2	3	4	5	6	8	10	14	23
	$s = 1$	1.950	2.724	3.532	4.476	5.658	7.251	9.600	13.600	22.670
10^{-3}	Det	3	4	4	5	6	7	10	13	23
	$s = 0$	2.645	3.449	4.225	5.090	6.150	7.575	9.725	13.582	23.495
	Exp	3	5	6	7	9	12	16	24	44
	$s = 1$	2.954	4.154	5.442	6.981	8.967	11.732	15.998	23.762	43.794
10^{-4}	Det	4	5	5	7	8	10	13	19	35
	$s = 0$	3.487	4.559	5.614	6.808	8.297	10.336	13.475	19.278	34.943
	Exp	4	6	8	10	13	17	23	35	66
	$s = 1$	3.954	5.584	7.354	9.494	12.288	16.237	22.447	34.065	65.571
10^{-5}	Det	4	5	7	8	10	12	16	24	46
	$s = 0$	4.329	5.669	7.002	8.526	10.444	13.097	17.230	24.908	46.426
	Exp	5	8	10	13	16	21	29	45	88
	$s = 1$	4.954	7.014	9.266	12.007	15.610	20.744	28.903	44.382	87.418
10^{-6}	Det	5	6	8	9	12	15	20	29	57
	$s = 0$	5.171	6.780	8.391	10.245	12.592	15.859	20.985	30.684	57.913
	Exp	6	9	12	15	19	26	36	55	110
	$s = 1$	5.954	8.445	11.179	14.520	18.931	25.252	35.358	54.700	109.272
10^{-7}	Det	6	7	9	11	13	17	23	35	68
	$s = 0$	6.013	7.891	9.779	11.963	14.739	18.620	24.740	36.389	69.401
	Exp	7	10	14	18	23	30	42	66	132
	$s = 1$	6.957	9.876	13.091	17.033	22.253	29.759	41.814	65.019	131.126

Table 2
Optimal buffer capacity for $M/E_2/1/K$ system, $s^2 = 1/2$

c	ρ	Method	10^{-10}	10^{-8}	10^{-6}	10^{-4}	10^{-2}
1	0.5	Exact	25	20	15	9	4
		$\tilde{K}$ (continuous)	25.7	20.2	14.8	9.3	3.8
		$\tilde{K}$ (Gelenbe)	21.6	17.0	12.4	7.8	3.2
		K_ϵ (E_2)	25	20	14	9	4
		$\tilde{K}_\epsilon$ (Kimura)	26	21	15	10	4
		K_ϵ^{T}(Tijms)	25	20	15	10	4
1	0.8	Exact	73	57	41	26	10
		$\tilde{K}$ (continuous)	73.7	57.7	41.7	25.7	9.8
		$\tilde{K}$ (Gelenbe)	68.9	53.9	38.9	23.9	9.1
		$K_\epsilon(E_2)$	73	57	41	26	10
		$\tilde{K}_\epsilon$(Kimura)	74	58	42	26	10
		K_ϵ^{T}(Tijms)	73	57	41	26	10

However, in another example, say that $\rho = 0.20$, $p_K = 0.1$, then the Hessian is

$$H = \begin{bmatrix} 62.096 & -17.907 \\ -17.907 & -0.057 \end{bmatrix}.$$

The first two principal minors are 62.096, −324.272 and the eigenvalues are −4.848, 66.887. Thus, we can conclude that the function is not convex.

Thus for medium to high ρ values and low p_K the function is convex, but for others, it will not be convex. An attempt was made to find the exact region where the Hessian becomes indefinite, but the complexity of the symbolic expression was not readily solvable.

Table 3
Optimal buffer capacity for $M/H_2^g/1/K$ systems $s^2 = 2$

c	ρ	Method	10^{-10}	10^{-8}	10^{-6}	10^{-4}	10^{-2}
1	0.5	Exact	45	35	26	16	7
		$\tilde{K}$ (continuous)	42.2	33.3	24.3	15.3	6.3
		$\tilde{K}$ (Gelenbe)	54.1	42.6	31.1	19.6	8.1
		$K_\epsilon(E_2)$	45	35	26	17	7
		$\tilde{K}_\epsilon$(Kimura)	43	34	24	16	7
		K_ϵ^{T}(Tijms)	46	36	25	17	7
1	0.8	Exact	141	110	80	49	19
		$\tilde{K}$(continuous)	137.5	107.6	77.7	47.8	18.2
		$\tilde{K}$(Gelenbe)	148.3	116.1	83.9	51.7	19.7
		$K_\epsilon(E_2)$	141	110	80	49	19
		$\tilde{K}_\epsilon$(Kimura)	137	109	78	49	19
		K_ϵ^{T}(Tijms)	140	111	80	50	19

Thus, we have the following negative result:

Theorem. *The optimal buffer function for $M/M/1/K$ systems is not a convex function.*

This may seem to be surprising in light of the other well-known results of queueing systems but is specific to this blocking probability function involving the design/performance measure K.

Fig. 5 suffices to indicate the region where the function is not convex. Note the low values of ρ and high values of p_K which makes little sense in a practical situation, but could occur.

3.5. $M/M/c/K$ formulas

Now, let us examine a multi-server Markovian system. We shall limit ourselves to those systems where $(\lambda/c\mu < 1)$. Analytical results from the $M/M/c/K$ model provide the following expression for p_K.

$$p_K = \frac{1}{c^{K-c}c!}\left(\frac{\lambda}{\mu}\right)^K p_0,$$

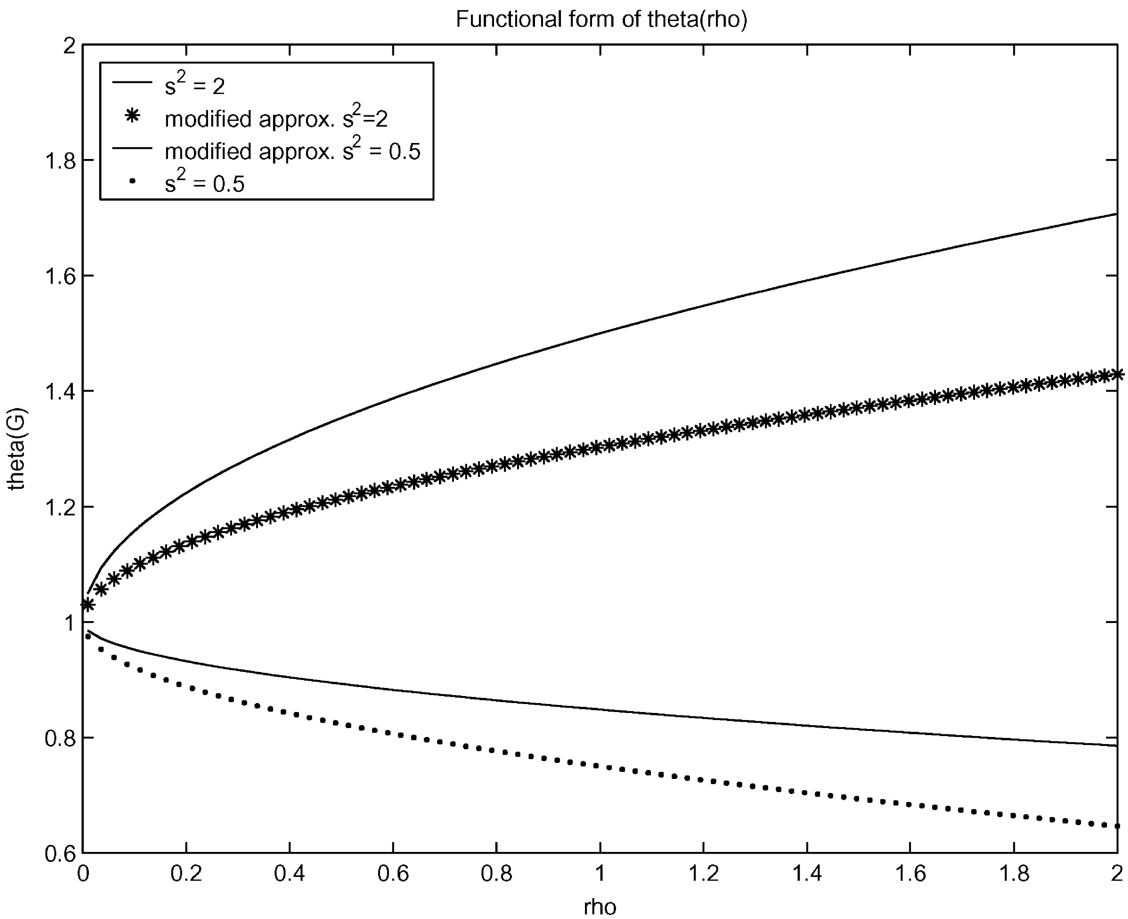


Fig. 10. Modified formula to Kimura's approximation.

where for $(\lambda/c\mu \neq 1)$

$$p_0 = \left[\sum_{n=0}^{c-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{(\lambda/\mu)^c}{c!}\frac{1-(\lambda/c\mu)^{K-c+1}}{1-\lambda/c\mu}\right]^{-1}.$$

$M/M/c/K$ composite formula for the blocking probability is given by

$$p_K = \frac{(\lambda/\mu)^K (c!)^{-1}(c^{K-c})^{-1}}{(\mathrm{e}^{\lambda/\mu}\Gamma(c,\lambda/\mu)(\Gamma(c))^{-1} + (\lambda/\mu)^c(1-(\lambda/c\mu)^{K-c+1})/c!(1-(\lambda/c\mu))}.$$

Since the blocking probability function is very complex, one cannot express the value of K without fixing c. If one fixes $c = 1$, the function becomes the same as the one which was derived from the previous formula for the $M/M/1/K$ system.

3.6. $M/M/2/K$

If one fixes $c = 2$, the following closed-form expression can be developed:

$$K = -\frac{\ln(\frac{1}{2}(p_K(2\mu+\lambda))/(2\mu-\lambda+p_K\lambda))}{\ln(2(\mu/\lambda))}.$$

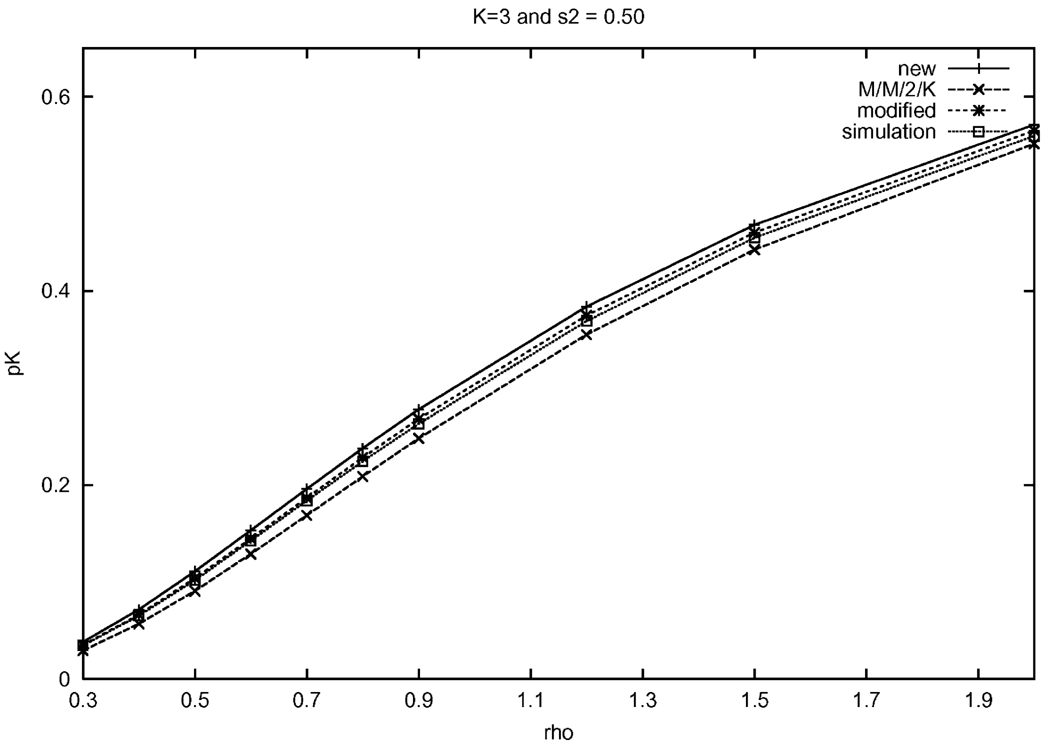


Fig. 11. p_K comparisons for $M/G/2/3$ system, $s^2 = 1/2$.

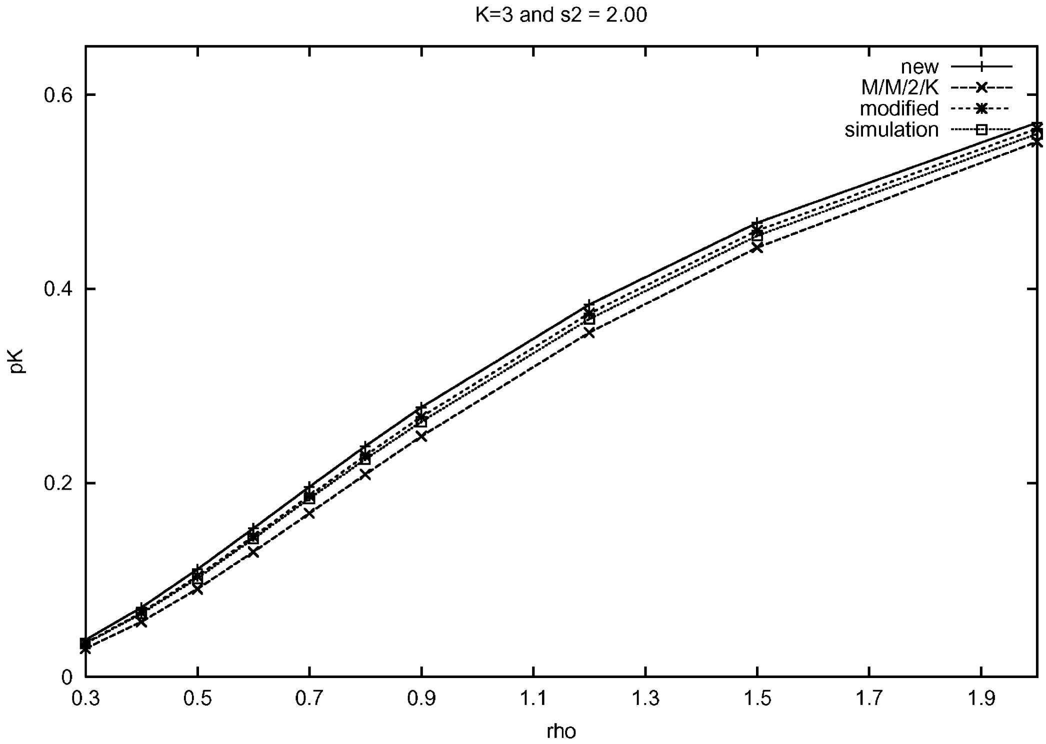


Fig. 12. p_K comparisons for $M/G/2/4$ system, $s^2 = 2$.

Fig. 6 on the left illustrates the monotonic nature of the $c = 2$ expression and also the significantly reduced buffer size as a result of the increased number of servers in the system over the $M/M/1/K$ system compared to the graph on the right of Fig. 3.[1] A similar analysis of the Hessian as was done for the $M/M/1/K$ case of this expression could also be carried out to show the non-convexity of the $c = 2$ function.

3.7. $M/M/3/K$

If one fixes $c = 3$, the following closed-form expression can be developed:

$$K = -\frac{\ln(\frac{1}{9}p_K(6\mu^2 + 4\lambda\mu + \lambda^2)/\mu(3\mu - \lambda + p_K\lambda))}{\ln(3(\mu/\lambda))}.$$

Fig. 6 on the right also illustrates the monotonic nature of the $c = 3$ expression and the marginally reduced buffer size compared with the $M/M/2/K$ case as a result of the increased number of servers in the system. Obviously, we could continue this process, but choose not to at the present stage. Another property to notice is in the denominator in the above expressions for the $M/M/2/K$ and $M/M/3/K$ cases, where the optimal K is reduced at the rate of $\ln(\rho^{-1})$.

[1] The p_K values in Fig. 6 should read 0.001–0.005 instead of 1–5, a small glitch in the output graphics.

4. $M/G/1/K$ two-moment approximations

The blocking probability from Gelenbe's equation with squared arrival and service process coefficient of variation is given by the following equation [4] where a^2 and s^2 are, respectively, the squared coefficient of variations of the arrival and service processes.

Gelenbe's formula is based on approximating the discrete queueing process as a continuous diffusion process

$$p_K = \frac{\lambda(\mu-\lambda)\,\mathrm{e}^{-2((\mu-\lambda)(k-1)/(\lambda a^2+\mu s^2))}}{\mu^2-\lambda^2\,\mathrm{e}^{-2((\mu-\lambda)(k-1)/(\lambda a^2+\mu s^2))}}.$$

The closed-form expression derivable for continuous K from Gelenbe's formula is the following:

$$K = \frac{1}{2}\frac{2\lambda - 2\mu + \ln(p_K\mu^2/\lambda(-\lambda+\mu+p_K\lambda))\lambda a^2 + \ln(p_K\mu^2/\lambda(-\lambda+\mu+p_K\lambda))\mu s^2}{\lambda-\mu}.$$

4.1. $M/G/1/K$ approximations

We will describe in some detail the two-moment approximation schemes of Tijms [25] and Kimura [13,14]. These two approximations were initially developed to model the optimal threshold buffer size,

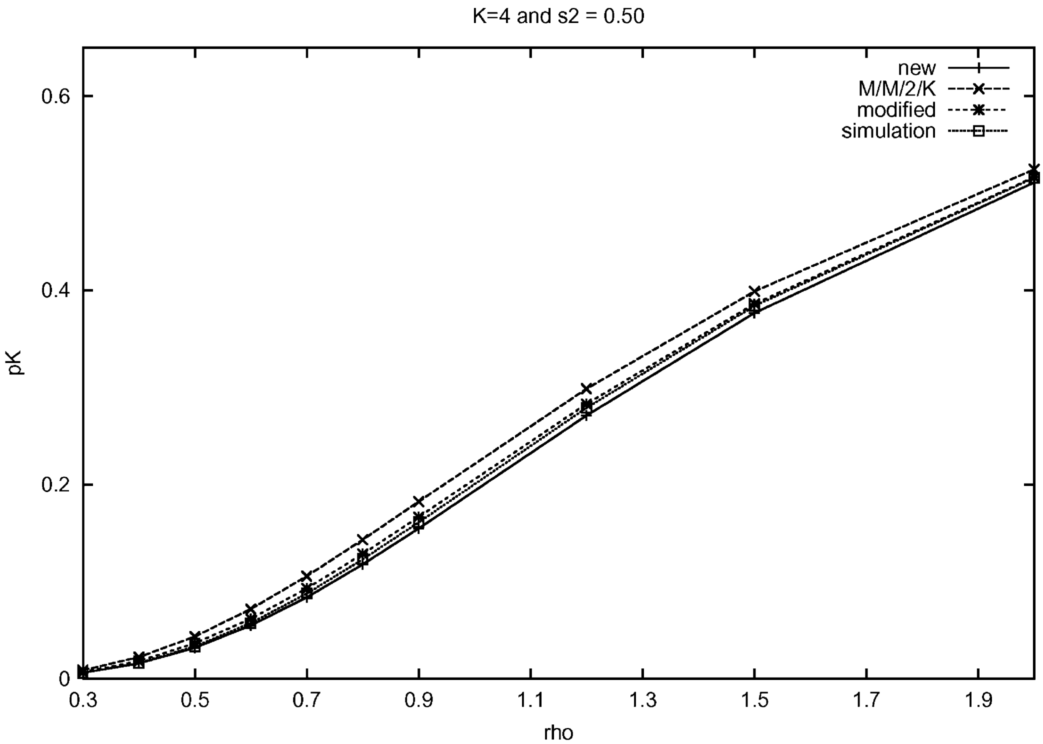


Fig. 13. p_K comparisons for $M/G/2/4$ system, $s^2 = 1/2$.

so that is why they are a bit simpler than the expression for p_K in the infinite models treated by these two authors in the past.

In order to describe Tijms's and Kimura's approximations, an additional bit of notation is needed at this point in the paper. Let us define $B_\epsilon(M)$ as the Markovian expression for the optimal buffer size as a function of the blocking probability and the threshold, $p_K(\epsilon)$. Also, $B_\epsilon(D)$ is the expression of the optimal buffer size as a function of a deterministic service process.

Tijms's two-moment approximation [1,25,27] relies on a weighted combination of an exact (if available) expression of the $M/D/1/K$ blocking probability as well as the blocking probability of the $M/M/1/K$ formula.

$$B_\epsilon^{\mathrm{T}}(s^2) = s^2 B_\epsilon(M) + (1 - s^2) B_\epsilon(D).$$

Of course, if exact expressions are available for both formulas, then Tijms's approximation is exact for the two extreme cases. His approximation has been shown to be very good, and we shall corroborate his results.

Kimura, on the other hand, also has a two-moment approximation that turns out to be a little simpler and is the one we shall build upon since it utilizes Markovian approximations as its basis. His expression is

$$\tilde{B}_\epsilon(s^2) = B_\epsilon(M) + \mathrm{NINT}\left(\tfrac{1}{2}(s^2 - 1)\sqrt{\rho} B_\epsilon(M)\right),$$

where NINT is the nearest integer. As we shall argue in this paper, this weighted combination of the optimal buffer values of Markovian systems is a very effective strategy. The key element in his approximation

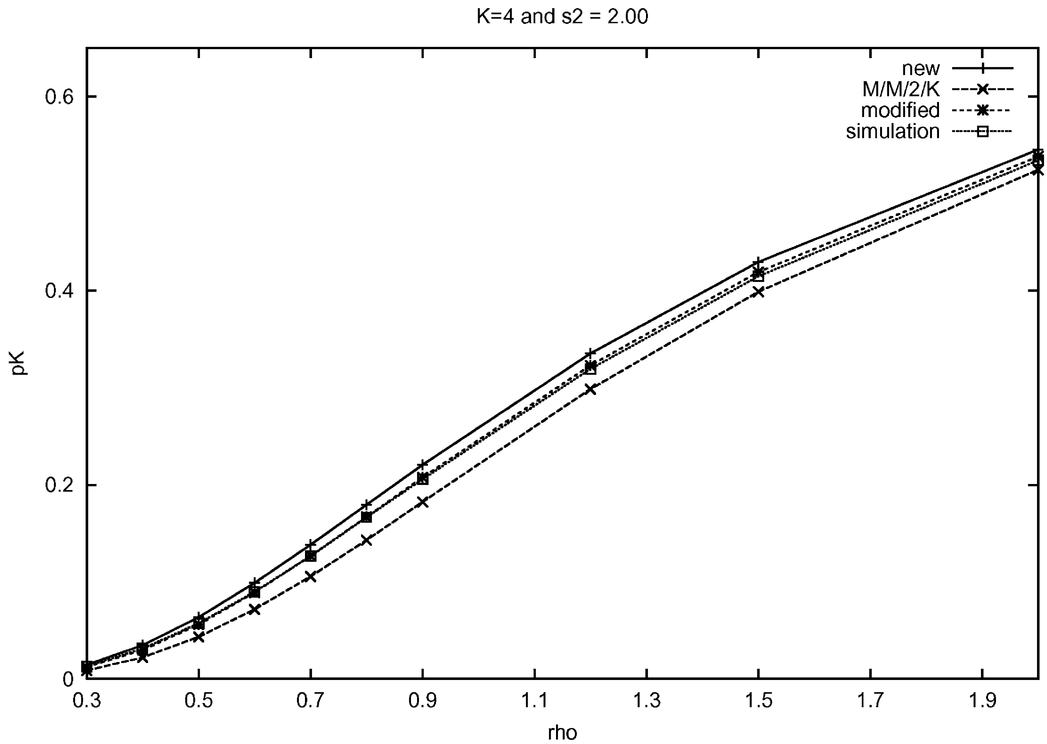


Fig. 14. p_K comparisons for $M/G/2/4$ system, $s^2 = 2$.

is the term

$$\tfrac{1}{2}(s^2 - 1)\sqrt{\rho}$$

which is actually based on the following graph, see Fig. 7.

The $\theta(g)$ on the y-axis in Fig. 7 is used by Kimura to describe how the variability of the service time cdf of G affects the buffer size. Thus, a square root function of ρ is a very reasonable functional approximation for this problem.

4.2. M/G/1/K expression

Recall the formula for the optimal buffer size for the $M/M/1/K$ formula

$$K = \frac{\ln(p_K/(1-\rho+p_K\rho))}{\ln(\rho)}.$$

Here, we add Kimura's expression for the approximation of the optimal buffer size based on his two-moment approximation formula. It is important to note here that we subtract the space for the server in order to estimate the true buffer space in the queue

$$\left(\frac{\ln(p_K/(1-\rho+p_K\rho))}{\ln(\rho)} - 1\right) + \frac{s^2-1}{2}\sqrt{\rho}\left(\frac{\ln(p_K/(1-\rho+p_K\rho))}{\ln(\rho)} - 1\right).$$

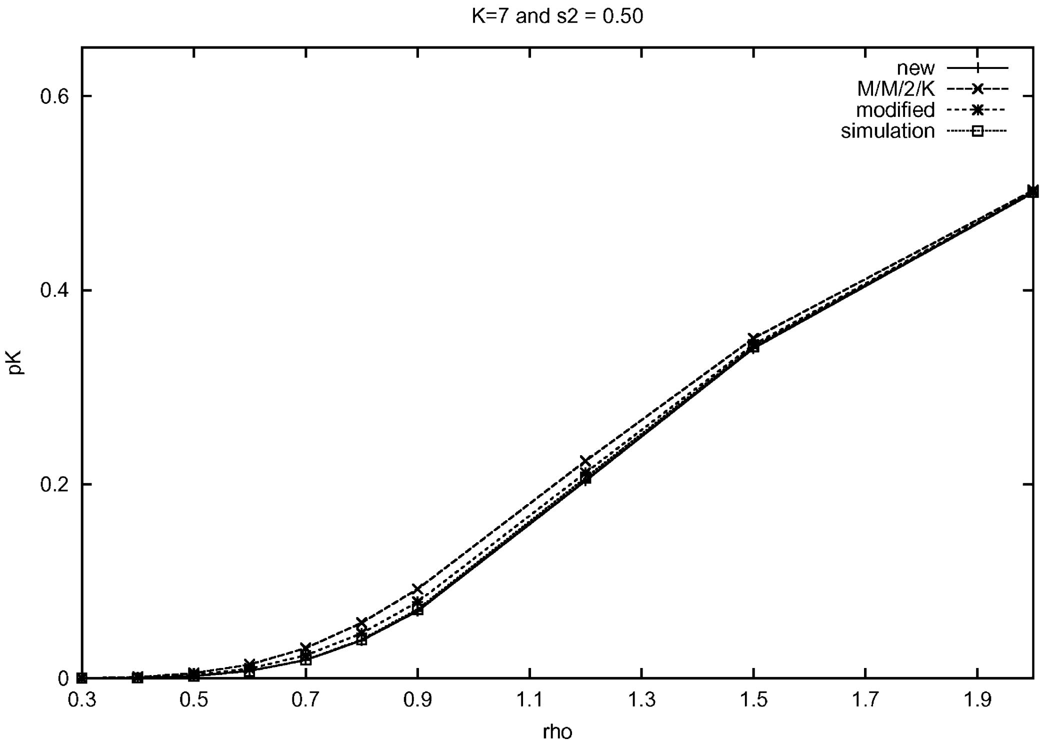


Fig. 15. p_K comparisons for $M/G/2/7$ system, $s^2 = 1/2$.

Now we factor the terms of the above expression to give the following simplified expression for an approximation to the optimal buffer size B in $M/G/1/K$ formulas:

$$\tilde{B} = -\frac{(-\ln(p_K/(1-\rho+p_K\rho)) + \ln(\rho))(2+\sqrt{\rho}s^2-\sqrt{\rho})}{2\ln(\rho)}$$

Notice in the above formula, that when $s^2 = 1$, we achieve the optimal buffer size of the $M/M/1/K$ queue with the server removed.

4.3. $M/G/1/K$ blocking probability

By taking the previous formula and inverting it, we can obtain the following expression for the $M/G/1/K$ blocking probability:

$$p_K = \frac{\rho^{(2+\sqrt{\rho}s^2-\sqrt{\rho}+2B)/(2+\sqrt{\rho}s^2-\sqrt{\rho})}(-1+\rho)}{\rho^{2((2+\sqrt{\rho}s^2-\sqrt{\rho}+B)/(2+\sqrt{\rho}s^2-\sqrt{\rho}))}-1}.$$

One caveat of our approach, however, is that we do not have an explicit formula for the case when $\rho = 1$. This is because the original blocking probability formula for the $M/M/1/K$ formula does not include ρ in its calculation for the case when $\rho = 1$. Thus, we must linearly interpolate the value when $\rho = 1$, albeit, a crude approximation, but not extraordinary difficult.

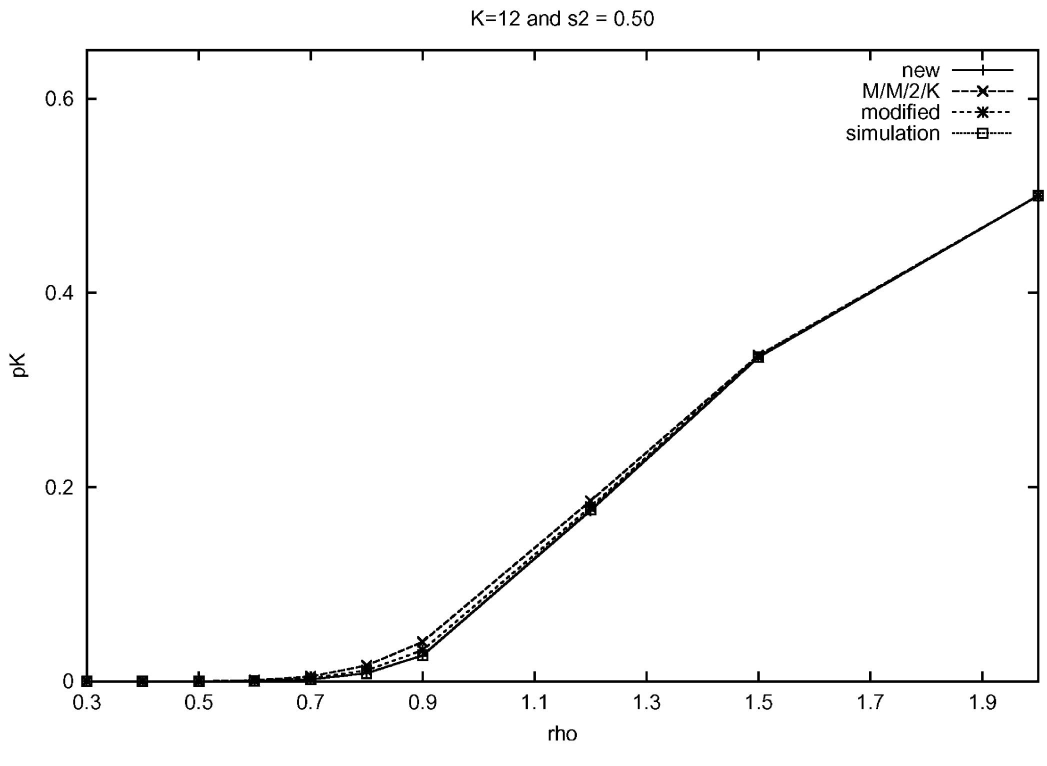


Fig. 16. p_K comparisons for $M/G/2/12$ system, $s^2 = 1/2$.

Table 4
$M/G/1/K$ to $M/G/5/K$ blocking probability formulas

$M/G/1/K$	$p_K = \dfrac{\rho^{(2+\sqrt{\rho}s^2-\sqrt{\rho}+2B)/(2+\sqrt{\rho}s^2-\sqrt{\rho})}(-1+\rho)}{\rho^{(2(2+\sqrt{\rho}s^2-\sqrt{\rho}+B)/(2+\sqrt{\rho}s^2-\sqrt{\rho}))}-1}$
$M/G/2/K$	$p_K = \dfrac{2\rho^{2((2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}+B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}(2\mu-\lambda)}{-2\rho^{(2(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}+B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}+\lambda+2\mu+\lambda}$
$M/G/3/K$	$p_K = \dfrac{-9\rho^{(6+3\sqrt{\rho/e}s^2-3\sqrt{\rho/e}+2B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e})}\mu(-3\mu+\lambda)}{-9\rho^{(6+3\sqrt{\rho/e}s^2-3\sqrt{\rho/e}+2B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e})}\mu\lambda+6\mu^2+4\lambda\mu+\lambda^2}$
$M/G/4/K$	$p_K = \dfrac{64\rho^{(2(4+2\sqrt{\rho/e}s^2-2\sqrt{\rho/e}+B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}\mu^2(4\mu-\lambda)}{-64\rho^{(2(4+2\sqrt{\rho/e}s^2-2\sqrt{\rho}+B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}\mu^2\lambda+24\mu^3+18\mu^2\lambda+6\lambda^2\mu+1\lambda^3}$
$M/G/5/K$	$p_K = \dfrac{-625\rho^{(-1(10+5\sqrt{\rho/e}s^2-5\sqrt{\rho/e}+2B)/(2+\sqrt{\rho/e}s^2-1\sqrt{\rho/e}))}\mu^3(-5\mu-\lambda)}{-625\rho^{(-1(10+5\sqrt{\rho/e}s^2-5\sqrt{\rho/e}+2B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}\mu^3\lambda+120\mu^4+36\lambda^2\mu^2+8\lambda^3\mu+\lambda^4+96\lambda\mu^3}$

4.4. $M/G/1/K$ experimental comparison

In order to test the efficacy of the blocking probability formula a small set of experiments are performed comparing our approximation with Gelenbe's and the $M/M/1/K$ model and optimal results from Seelen et al. [18] for an $M/G/1/K$ model with $s^2 = 1/2$, see Fig. 8. The results of our new formula when compared to Gelenbe's are surprisingly accurate, especially in the $\rho < 1$ values and in relation to the growth of the blocking probability values over the range of the parameters. Gelenbe's formula does better in $\rho \geq 1.50$, however, this is beyond the range of ρ we feel is most appropriate to consider. An additional

Table 5
Comparison of $M/G/5/K$ system

s^2		$\rho = 0.50$			$\rho = 0.8$			$\rho = 1.5$	
		$K = 6$	$K = 7$	$K = 10$	$K = 6$	$K = 7$	$K = 10$	$K = 6$	$K = 10$
0	App	0.0286	0.0105	0.00036	0.1221	0.0742	0.0179	0.3858	0.3348
	Exa	0.0287	0.0108	0.00035	0.1224	0.0746	0.0175	0.3870	0.3352
	Mod	0.0277	0.0113	0.00079	0.1211	0.0791	0.0265	0.3870	0.3369
	New	0.0228	0.0077	0.00031	0.1051	0.0616	0.0156	0.3674	0.3338
1/2	App	0.0311	0.0135	0.0010	0.1306	0.0882	0.0308	0.3975	0.3395
	Exa	0.0314	0.0137	0.0010	0.1318	0.0895	0.0314	0.4000	0.3400
	Mod	0.0309	0.0140	0.0013	0.1302	0.0899	0.0347	0.3971	0.3408
	New	0.0289	0.0123	0.0009	0.1245	0.0831	0.0294	0.3900	0.3381
1	App								
	Exa	0.0337	0.0166	0.0020	0.1374	0.0990	0.0425	0.0405	0.3453
	Mod								
	New								
2	App	0.0370	0.0211	0.0046	0.1450	0.1123	0.0603	0.4114	0.3555
	Exa	0.0366	0.0206	0.0044	0.1435	0.1103	0.0587	0.4092	0.3537
	Mod	0.0382	0.0213	0.0038	0.1481	0.1134	0.0569	0.4149	0.3547
	New	0.0406	0.0240	0.0051	0.1532	0.1207	0.0645	0.4195	0.3604

experimental comparison when $s^2 = 2$ appears in Fig. 9 where simulation is used to compare the s^2 model. A gamma distribution with appropriate parameters was used in an Arena simulation model of 200,000 time units to compare the results.

Additional experiments and an application of the $M/G/1/K$ blocking probability formula to a buffer allocation problem in open finite queueing networks appear in [21].

Finally, it would be nice to be able to compute other performance measures with our approach, but this is also not possible, since our focus is on calculating p_K rather than expected length, mean waiting time, and the entire probability distribution for the $M/G/1/K$ system.

4.5. $M/G/1/K$ threshold experiments

Now let us in the first instance compare numerically the quality of these approximations with some known optimal results (from Tijms's [25] for single-server systems, see page 308). Table 1 compares the approximation with known results for $M/D/1/K$ and $M/M/1/K$ systems. In this table, we have included the number in service with the estimate of those in the buffer to be consistent with the reported results of Tijms's. The "det" row is Tijms's approximation, the "exp" is the exponential model results while the real number values are the approximation results. The approximation with K rounded to the largest integer is remarkably accurate for the $M/D/1/K$ systems and of course exact for the $M/M/1/K$ systems.

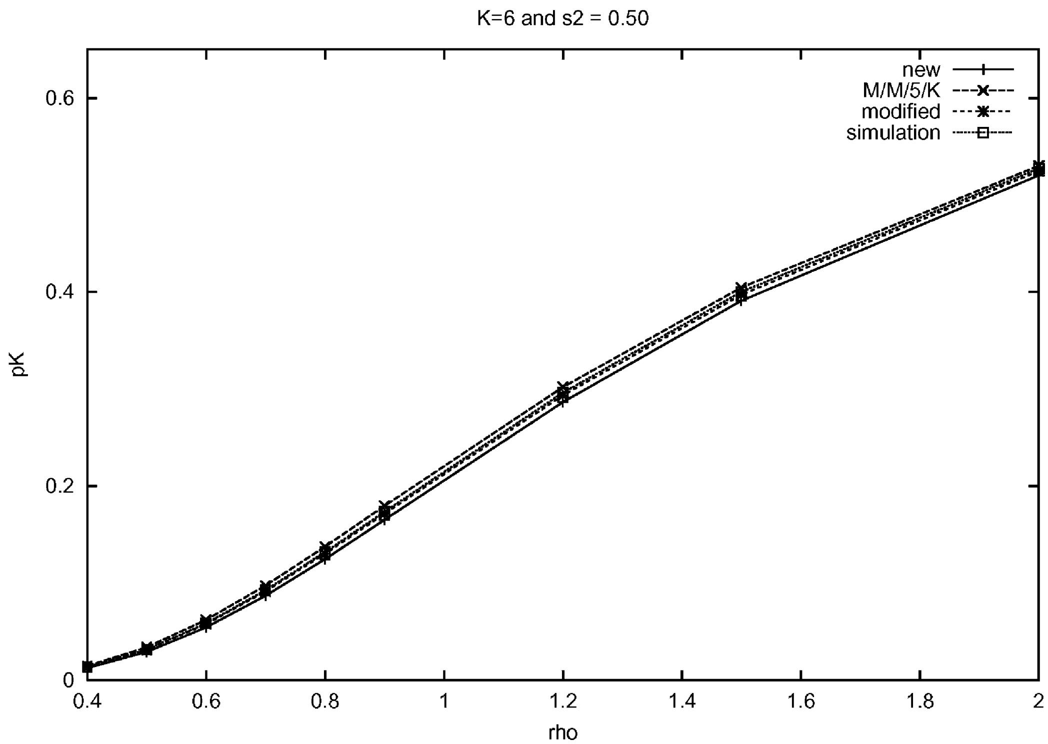


Fig. 17. p_K comparisons for $M/G/5/6$ system, $s^2 = 1/2$.

Tables 2 and 3 compare the approximation with the exact buffer values and those approximate values of Tijms and Kimura for the $M/E_2/1/K$ and $M/H_2^g/1/K$ systems, respectively, for various threshold levels from 10^{-2} to 10^{-10}.

Thus, as one can see, Gelenbe's approximation does not do as well in predicting the optimal buffer size, it underestimates the buffer in low variability situations and overestimates it in high variability situations. The closed form approximation developed in this present paper based on Kimura's expression, does very well for the lower variability queues, see Table 2, although not as well on the higher variability queues with very small ϵ, see Table 3. Our expression actually does a little better than Kimura's in two of the cases in Table 3. The fact that the expression for the optimal buffer size is a closed-form expression, means that it could be very helpful in a simulation model or optimization search process for networks of finite queues, let alone single queues.

There are two papers where we have demonstrated the efficacy of our approximation, one in open queueing networks [21] and the other more recently in closed queueing networks [22].

5. $M/G/c/K$ blocking probability formula

In this section of the paper, we extend our methodology based on the $M/G/1/K$ system to $M/C/c/K$ systems. As in the single-server systems, we develop the optimal continuous buffer size for the $M/M/c/K$

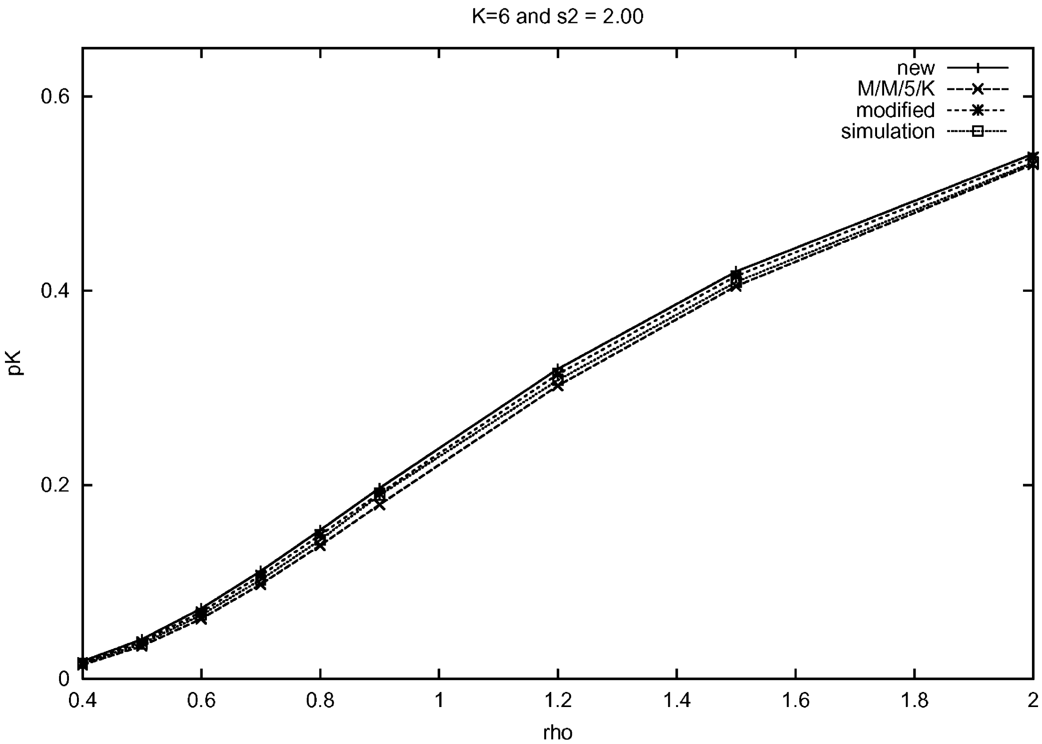


Fig. 18. p_K comparisons for $M/G/5/6$ system, $s^2 = 2$.

system, transform it to the $M/G/c/K$ system buffer size, then invert the formula to get the blocking probability p_K.

5.1. $M/G/c/K$ expressions

Unfortunately, when applying Kimura's formula to $M/G/c/K$ systems, the blocking probabilities did not turn out quite as accurate as for the $M/G/1/K$ system.

Another functional form was found which is a modified square root function and this yields a better approximation when the blocking probabilities are large. This is felt to be important for the buffer allocation problem because in the optimization process, surprisingly the blocking probabilities tend not to be very small.

The "modified" approximation is

$$\frac{s^2-1}{2}\sqrt{\frac{\rho}{e}}.$$

This modified approximation is the one with asterisks in Fig. 10 compared with the curve for $s^2 = 2$ for the previous terms.

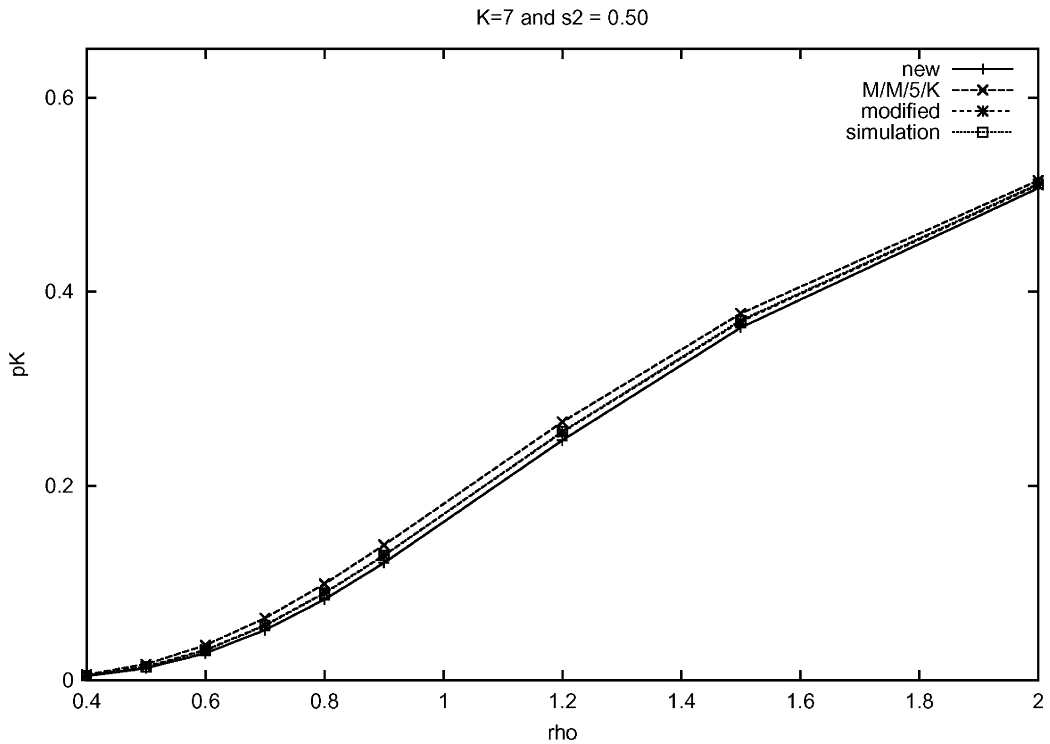


Fig. 19. p_K comparisons for $M/G/5/7$ system, $s^2 = 1/2$.

5.2. $M/G/2/K$ blocking probability

If we add the modified square root function, we achieve the following approximation for the buffer size B of an $M/G/2/K$ system

$$\left(\ln\left(\frac{1}{2}\frac{p_K(2\mu+\lambda)}{2\mu-\lambda+p_K\lambda}\right)(\ln(\rho))^{-1}-2\right) + \left(\frac{s^2-1}{2}\right)\sqrt{\frac{\rho}{e}}\left(\ln\left(\frac{1}{2}\frac{p_K(2\mu+\lambda)}{2\mu-\lambda+p_K\lambda}\right)(\ln(\rho))^{-1}-2\right).$$

Now, if we invert this function and solve for p_K we get

$$p_K = \frac{2\rho^{2((2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}+B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}(2\mu-\lambda)}{-2\rho^{(2(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}+B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e}))}\lambda+2\mu+\lambda}.$$

In Fig. 11 we see the tabular and close graphical comparison of the modified approximation as compared with a simulation model of the same system. Here we see that the modified approximation is well within the envelope of the $M/M/1/K$ bound and the previous formula for p_K. Now let us compare the blocking probability with the previous system with $s^2 = 2$.

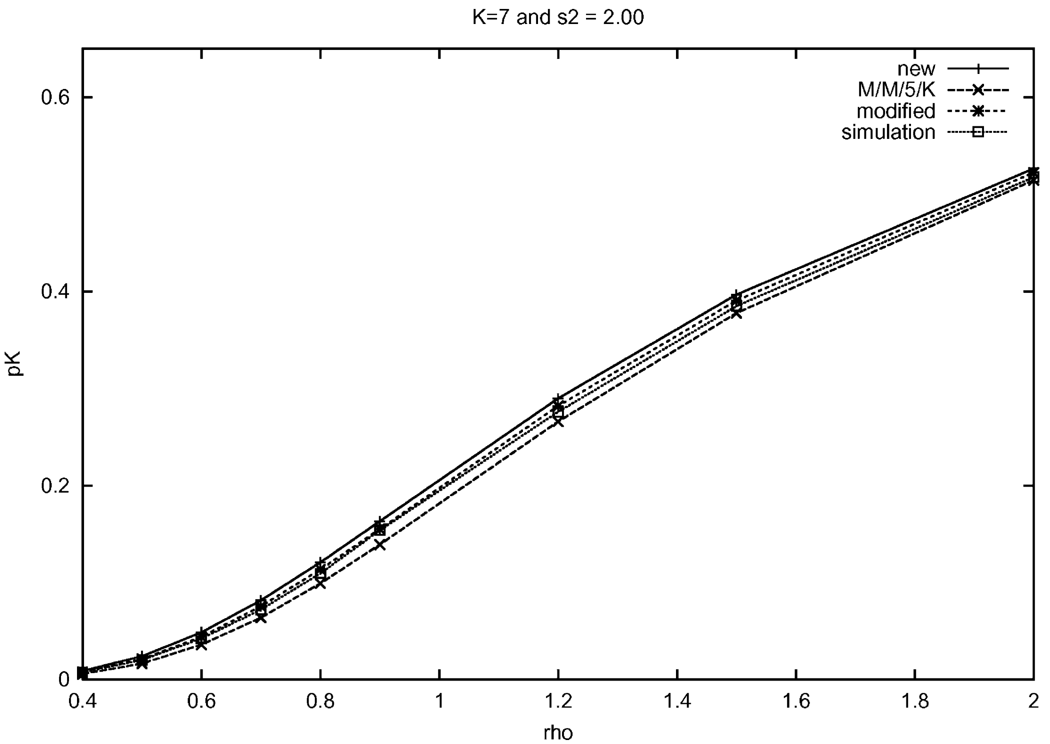


Fig. 20. p_K comparisons for $M/G/5/7$ system, $s^2 = 2$.

As can be seen in the table and graph of Fig. 12 the modified formula does a better job of not overestimating the blocking probability in this highly variable system whereas the $M/M/2/K$ is way off until $\rho \gg 1$. These results hold true also for the $M/G/2/4$ system as depicted in Figs. 13 and 14 for $s^2 = 1/2$ and 2.

Now let us increase the buffer and examine the $M/G/2/7$ and $M/G/2/12$ systems. These results appear in Figs. 15 and 16. As can be seen in Fig. 15, the modified formula is not as accurate when the blocking probabilities are low, but is more accurate than the "new" formula when they are high. This is also apparent in Fig. 16.

5.3. $M/G/3, 4/K$ expressions

If one fixes $c = 3$ in the $M/G/c/K$ formula, the following closed-form expression for the optimal buffer size can be developed

$$K = -\frac{\ln(\frac{1}{9}p_K(6\mu^2 + 4\lambda\mu + \lambda^2)/(3\mu - \lambda + p_K\lambda))}{\ln(3(\mu/\lambda))}.$$

The blocking probability found by using the modified square root function and inverting the previous formula is given as

$$p_K = \frac{-9\rho^{(6+3\sqrt{\rho/e}s^2-3\sqrt{\rho/e}+2B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e})}\mu(-3\mu+\lambda)}{-9\rho^{(6+3\sqrt{\rho/e}s^2-3\sqrt{\rho/e}+2B)/(2+\sqrt{\rho/e}s^2-\sqrt{\rho/e})}\mu\lambda + 6\mu^2 + 4\lambda\mu + \lambda^2}.$$

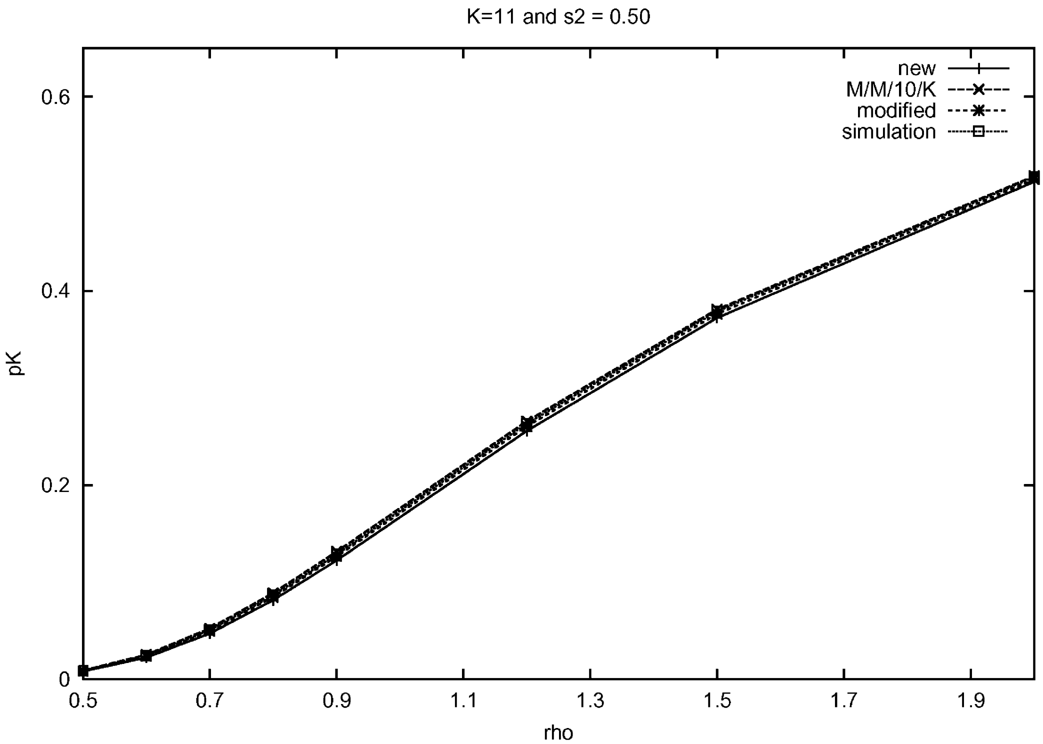


Fig. 21. p_K comparisons for $M/G/10/11$ system, $s^2 = 1/2$.

For the sake of the argument, the previous blocking probability formula without the modified square root expression is given below:

$$p_K = \frac{-9\rho^{(6+3\sqrt{\rho}s^2-3\sqrt{\rho}+2B)/(2+\sqrt{\rho}s^2-1\sqrt{\rho})}\mu(-3\mu+\lambda)}{-9\rho^{(6+3\sqrt{\rho}s^2-3\sqrt{\rho}+2B)/(2+\sqrt{\rho}s^2-1\sqrt{\rho})}\mu\lambda+6\mu^2+4\lambda\mu+\lambda^2}.$$

For the $M/G/4/K$ system, the blocking probability with the modified square root function is given in the table since it is quite cumbersome.

We will not present results for the comparison of the previous and modified systems for the sake of the argument, but the results are similar to before for both the $s^2 = 1/2, 2$ situations.

5.4. $M/G/5/K$ blocking probability

For the $c = 5$ multi-server case, we have the following expression for the blocking probability which appears in Table 4 because of its length.

Before presenting the numerical results based upon our perturbation experiments, we will present a published table of results from Tijms's (page 360) to indicate the accuracy of our results. In Table 5, the "app" represents Tijms's approximation, "exa" represents an exact solution, "new" is the approximation based on the square root function and "mod" is the modified square root function.

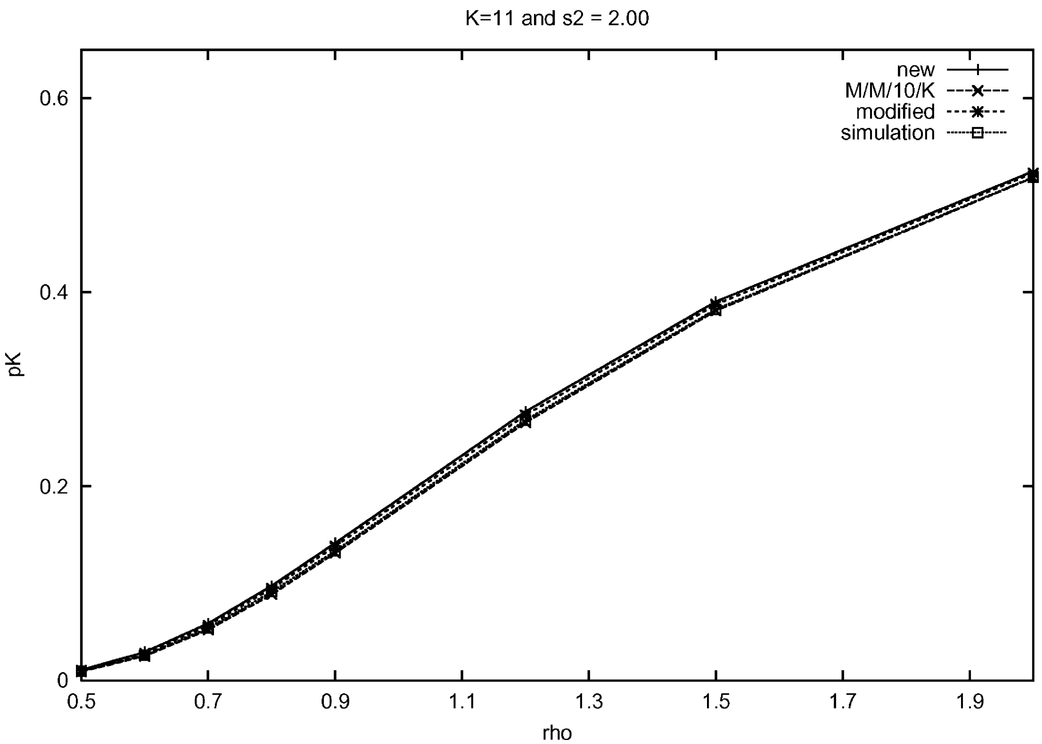


Fig. 22. p_K comparisons for $M/G/10/11$ system, $s^2 = 2$.

As can be seen in Table 5 the "modified" probability formula works very well in most all cases over the previous "new" one. Still, the "new" formula is useful when the blocking probabilities are relatively small as they are when computing the threshold buffer values.

Figs. 17 and 18 illustrate our numerical results for the $M/G/5/6$ queues with $s^2 = 1/2$ and 2. Additional results for an $M/G/5/7$ queues also appear in Figs. 19 and 20. We will not present any more perturbation experiments for the $M/G/5/K$ system but lastly address the $M/G/10/K$ system.

5.5. $M/G/10/K$ blocking probability

Figs. 21–24 illustrate the performance of the approximation for these multi-server systems. As $c = 10$, it is apparent that the Markovian $M/M/10/K$ approximation is very effective as p_K gets larger.

5.6. $M/G/10/K$ threshold systems

The only experimental results we have available at this point of time is to compare our approximation approach for these multi-server systems is through the experiments of Tijms and Kimura for $c = 10$ servers.

For the $M/G/10/K$ system, our approximation formula becomes extremely long and complicated and is essentially similar to the $c = 2, 3$ cases. In principle, we could extend our approximation approach to

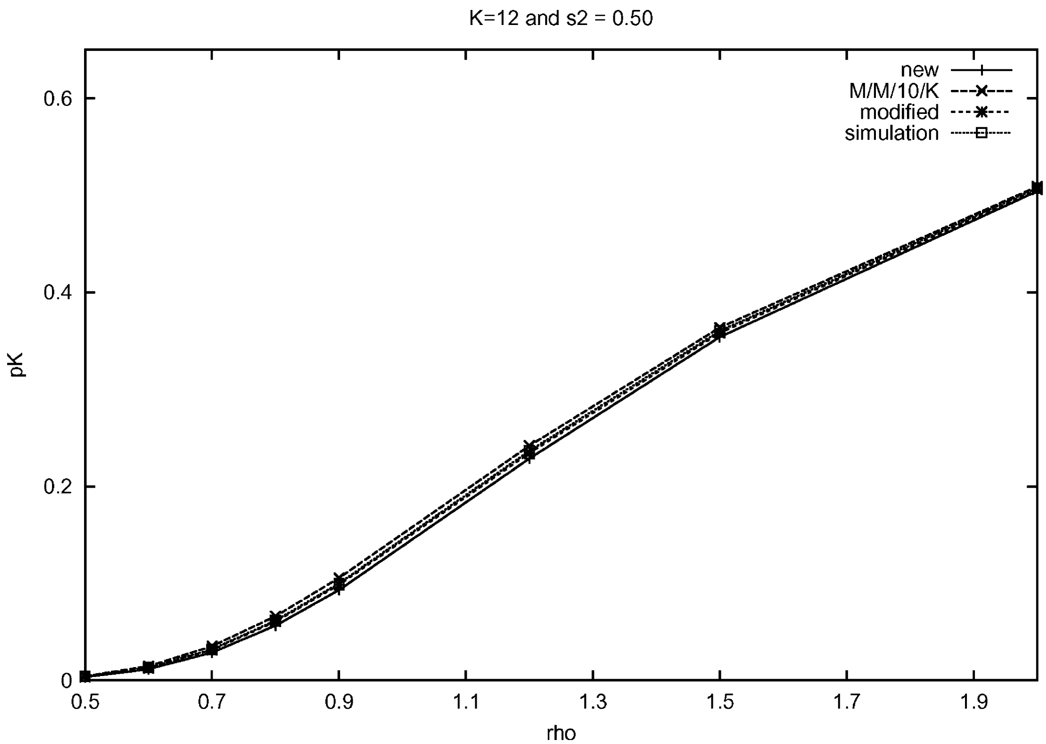


Fig. 23. p_K comparisons for $M/G/10/12$ system, $s^2 = 1/2$.

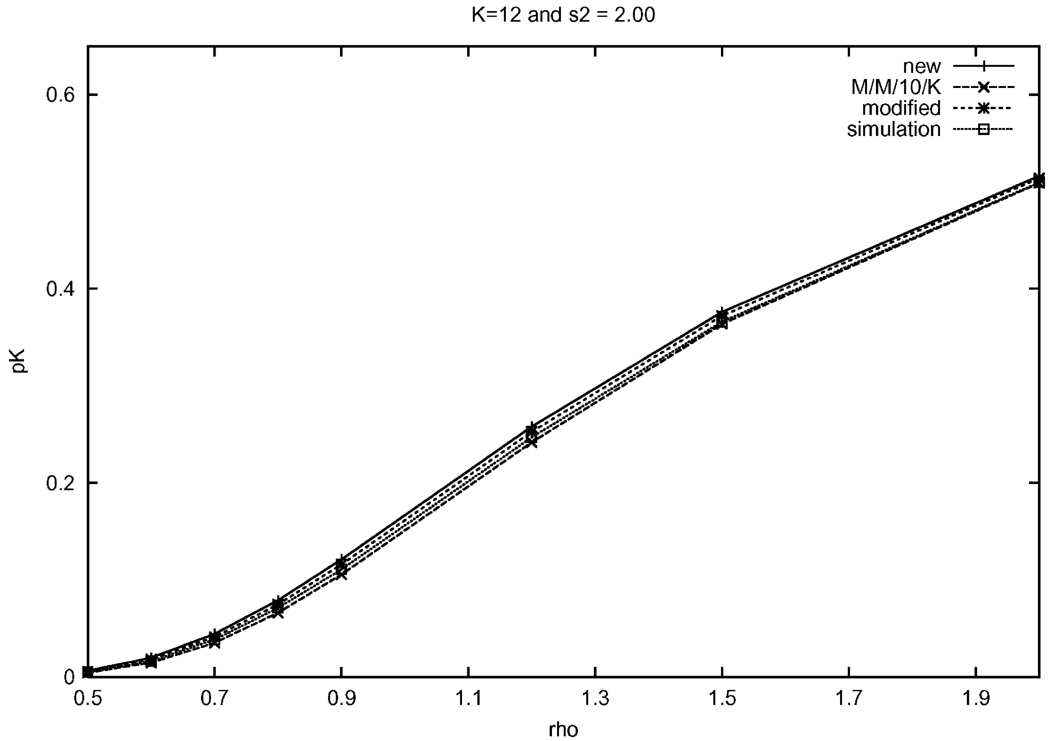


Fig. 24. p_K comparison for $M/G/10/12$ system, $s^2 = 2$.

$c \gg 10$ but have not done so. Obviously, the formulas should be extremely complex, but they should be directly translatable into numerical formulas which should be useful. This will be done at a future date.

As can be seen in Tables 6 and 7, the results for the approximation perform as well as for the $M/G/1/K$ expression and roughly parallel Kimura's results.

Table 6
Optimal buffer capacity for $M/E_2/10/k$ system, $s^2 = 1/2$

c	ρ	Method	10^{-10}	10^{-8}	10^{-6}	10^{-4}	10^{-2}
10	0.5	Exact	23	17	12	7	1
		$\tilde{K}$ (continuous)	22.6	17.1	11.6	6.2	0.7
		$K_\epsilon(E_2)$	22	17	11	6	1
		$\tilde{K}_\epsilon$ (Kimura)	23	17	12	7	1
		K_ϵ^{T} (Tijms)	23	17	13	7	1
10	0.8	Exact	71	55	39	24	8
		$\tilde{K}$ (continuous)	71.4	55.4	39.4	23.4	7.5
		$K_\epsilon(E_2)$	70	55	39	23	8
		$\tilde{K}_\epsilon$ (Kimura)	71	56	40	24	8
		K_ϵ^{T} (Tijms)	71	55	39	24	8

Table 7
Optimal buffer capacity for $M/H_2^g/10/k$ system, $s^2 = 2$

c	ρ	Method	10^{-10}	10^{-8}	10^{-6}	10^{-4}	10^{-2}
10	0.5	Exact	39	29	20	10	1
		$\tilde{K}$ (continuous)	37.1	28.1	19.1	10.2	1.2
		$K_\epsilon(H_2)$	40	30	21	11	2
		$\tilde{K}_\epsilon$ (Kimura)	38	28	20	11	1
		K_ϵ^{T} (Tijms)	39	29	20	10	1
10	0.8	Exact	135	105	74	44	14
		$\tilde{K}$ (continuous)	133.1	103.2	73.4	43.5	13.9
		$K_\epsilon(H_2)$	136	106	75	45	15
		$\tilde{K}_\epsilon$ (Kimura)	133	104	74	45	14
		K_ϵ^{T} (Tijms)	135	106	75	45	14

5.7. *Final coda*

As one final demonstration of the use of our approximations of p_K let us illustrate how the p_K formula can be used directly to optimize a system problem.

Let us say that we are interested in the optimal throughput of a single queue where we wish to maximize the throughput of the system varying the service rate μ and the buffer size B. We will demonstrate this process for a single-server system, although we could do it in principle for the other $M/G/c/K$ systems as well. Since λ will be given, we would like to know what values of both μ and B should be to optimize the system, where c_1 is the profit per unit throughput and say $c_2(\mu)$ is the cost of the service, $c_3(B)$ is the cost of the buffers.

Our optimization problem is

$$\text{Maximize } Z(\mu, B) = c_1\theta - c_2(\mu) - c_3(B).$$

If we use the p_K formula in the objective function, where the throughput $\theta = \lambda(1 - p_K)$ then we have more explicitly

$$c_1\,\lambda\left(1.0 - \frac{(\lambda/\mu)^{(2+\sqrt{\lambda/\mu}s^2-\sqrt{\lambda/\mu}+2B)/(2+\sqrt{\lambda/\mu}s^2-\sqrt{\lambda/\mu})}(-1+(\lambda/\mu))}{(\lambda/\mu)^{2(2+\sqrt{\lambda/\mu}s^2-\sqrt{\lambda/\mu}+2B)/(2+\sqrt{\lambda/\mu}s^2-\sqrt{\lambda/\mu})}}\right) - c_2\,\mu - c_3\,B.$$

This is a non-trivial optimization problem because the decision variable μ in the first term of our objective is under the radical sign, so a closed form solution is not easily derivable. Numerical solutions, however, are possible and we shall demonstrate a couple of instances with a non-linear programming subroutine based on sequential quadratic programming in the IMSL program library programming language. In these examples, Table 8, $\lambda = 0.50$ for all cases

The run times on a Dell PC-Dimension 266 running Windows NT 4.0 were very fast. It is interesting to see in the above table the implicit value tradeoffs in the optimal decision variables that occur for these different experiments. This type of optimization process with the p_K formulas will be applied in our future research to open and closed networks of queues.

Table 8
Example optimization experiments

	Case i		Case ii		Case iii	
	$s^2 = 1/2$	$s^2 = 2$	$s^2 = 1/2$	$s^2 = 2$	$s^2 = 1/2$	$s^2 = 2$
c_1	100.0	–	50.0	–	1000.0	–
c_2	1.0	–	5.0	–	1.0	–
c_3	5.0	–	1.0	–	10.0	–
μ	0.90	1.06	1.85	2.20	0.91	1.08
B	3.02	3.48	1.00	1.00	6.11	7.42

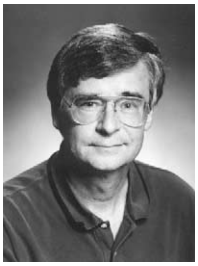
6. Summary and conclusions

We have presented an overview of results for estimating an approximation to the blocking probability of $M/G/c/K$ systems and some of their associated properties. It is felt that the methodology is a unique and new approach to the development of the blocking probabilities of these systems, and the approximation formulas based on a weighted combination of the formulas for the $M/M/c/K$ optimal buffer formula do quite well in comparison with simulation and well-known established results.

References

[1] A.G. De Kok, H. Tijms, A two-moment approximation for a buffer design problem requiring a small rejection probability, Perform. Eval. 5 (1985) 77–84.
[2] D. Gaver, G.S. Shedler, Processor utilization in multi-programming systems via diffusion approximations, Oper. Res. 21 (1973) 569–576.
[3] D. Gaver, G.S. Shedler, Approximate models for processor utilization in multiprogrammed computer systems, SIAM J. Comput. 2–3 (1973) 183–192.
[4] E. Gelenbe, On approximate computer system models, JACM 22 (2) (1975) 261–269.
[5] D. Gross, C. Harris, Fundamentals of Queueing Theory, Wiley, New York, 1985.
[6] E. Gelenbe, X. Mang, R. Onvural, Diffusion based call admission control in ATM, Perform. Eval. 27–28 (1996) 411–436.
[7] E. Gelenbe, X. Mang, Y. Feng, Diffusion cell loss estimates for ATM with multiclass bursty traffic, Comput. Syst.—Sci. Eng. 11 (6) (1996) 325–334.
[8] A. Harel, P. Zipkin, Strong convexity results for queueing systems, Oper. Res. 35 (3) (1987) 405–418.
[9] A. Harel, Sharp bounds and simple approximations for the Erlang delay and loss formulas, Manage. Sci. 34 (8) (1988) 959–972.
[10] A. Harel, The convexity of the Erlang Delay formula with respect to the number of servers, Working Paper #90-24, Graduate School of Management, Rutgers, November 1990.
[11] D. Heyman, Comments on a queueing inequality, Manage. Sci. 26 (2) (1908) 956–959.
[12] T. Kimura, Refining diffusion approximations for GI/G/1 queues: a tight discretization method, Teletraffic Congr. (1985) 317–323.
[13] T. Kimura, A transform-free approximation for the finite capacity $M/G/s$ queue, Oper. Res. 44 (6) (1996) 984–988.
[14] T. Kimura, Optimal buffer design of an $M/G/s$ queue with finite capacity, Commun. Statist.—Stochastic Mod. 12 (1) (1996) 165–180.
[15] L. Liyanage, G. Shantikumar, Second-order properties of single-stage queueing systems, in: U.N. Bhat, I.V. Basawa (Eds.), Queueing and Related Models, Clarendon Press, Oxford, 1992, pp. 129–160.
[16] A. Pacheco, Second-order properties of the loss probability in $M/M/s/s+c$ systems, Questa 15 (1994) 289–308.

[17] A. Pacheco, Some properties of the delay probability in $M/M/s/s+c$ systems, Questa 15 (1994) 309–324.
[18] L.P. Seelen, H. Tijms, M.H. VanHoorn, Tables for Multi-server Queues, North-Holland, Amsterdam, 1985.
[19] H. Sakasegawa, M. Miyazawa, G. Yamazaki, Evaluating the overflow probability using the infinite queue, Oper. Res. 39 (1993) 1238–1245.
[20] P. Schweitzer, A. Konheim, Buffer overflow calculations using an infinite-capacity model, Stochastic Proc. Appl. 6 (1978) 267–276.
[21] J.M. Smith, F.R. Cruz, The buffer allocation problem for general finite buffer queueing networks, submitted for publication.
[22] J.M. Smith, L. Kerbache, Buffer space allocation in general closed finite queueing networks, submitted for publication.
[23] M. Sobel, Simple inequalities for multiserver queues, Manage. Sci. 26 (2) (1980) 951–956.
[24] M. Springer, P. Makens, Queueing models for performance analysis and selection of single station models, EJOR 58 (1991) 123–145.
[25] H. Tijms, Stochastic Modeling and Analysis, Wiley, New York, 1986.
[26] H. Tijms, Heuristics for finite-buffer queues, Prob. Eng. Inform. Sci. 6 (1992) 277–285.
[27] H. Tijms, Stochastic Models: An Algorithmic Approach, Wiley, New York, 1994.
[28] D. Yao, J.A. Buzacott, Queueing models for a flexible machining station. Part 1. The diffusion approximation, Eur. J. Oper. Res. 19 (1985) 233–240.



J. MacGregor Smith received his Bachelor's and Master's of Architecture from the University of California at Berkeley in 1969 and 1970, and his Ph.D. in Mechanical and Industrial Engineering from the University of Illinois at Champaign-Urbana in 1978. His academic and professional interests include the design of algorithms for combinatorial optimization and stochastic network problems, the optimal design of facilities, networks, and the design and analysis of manufacturing systems. Dr. Smith has carried our sponsored research for the National Science Foundation (NSF), the Digital Equipment Corporation (DEC), IBM and the US Army Natick Research and Development Laboratories. He is a member of the INFORMS, ACM, SIAM, The Mathematical Programming Society, and $\Sigma\Xi$.