Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

EFFICIENT ANALYSIS OF RARE EVENTS ASSOCIATED WITH INDIVIDUAL BUFFERS IN A TANDEM JACKSON NETWORK

Ramya Dhamodaran
Bruce C. Shultes

Department of Mechanical, Industrial, and Nuclear Engineering
University of Cincinnati
PO Box 210072
Cincinnati, OH 45221-0072, U.S.A.

ABSTRACT

Over the last decade, importance sampling has been a popular technique for the efficient estimation of rare event probabilities. This paper presents an approach for applying balanced likelihood ratio importance sampling to the problem of quantifying the probability that the content of the second buffer in a two node tandem Jackson network reaches some high level before it becomes empty. Heuristic importance sampling distributions are derived that can be used to estimate this overflow probability in cases where the first buffer capacity is finite or infinite. The proposed importance sampling distributions differ from previous balanced likelihood ratio methods in that they are specified as functions of the contents of the buffers. Empirical results indicate that the relative errors of these importance sampling estimators is bounded independent of the buffer size when the second server is the bottleneck and is bounded linearly in the buffer size otherwise.

1 INTRODUCTION

The estimation of rare event probabilities has received considerable attention over the last decade. Tandem Jackson networks serve as a simplified model for analyzing rare events in many systems such as switched telecommunication networks, manufacturing systems and computer networks. System performance measures such as the probability that the system size or a specific queue length exceeds a given level are needed to accurately assess system reliability, particularly the time until one of these events occurs.

Importance sampling is gaining popularity as an efficient method for analyzing rare events in queueing and reliability systems (see Asmussen and Rubinstein 1995, Heidelberger 1995). The application of importance sampling involves simulating the model using an auxiliary distribution designed to make the system experience rare events of interest more often. Large deviations theory has been used to derive a heuristic change of measure for estimating the probability that the total system size exceeds a given level before returning to zero in tandem Jackson networks (see Parekh and Walrand 1989). This *exponential twisting* or *tilting* change of measure interchanges the arrival rate and the smallest service rate in the network. This heuristic was later analyzed by Glasserman and Kou (1995) who established necessary and sufficient conditions for the asymptotic efficiency of this importance sampling estimator. More recently, de Boer, Kroese, and Rubinstein (2002) proposed an adaptive importance sampling method that utilizes a minimum cross-entropy optimization approach to estimate the overflow probability in three stages by approximating an optimal tilting parameter.

The balanced likelihood ratio approach to importance sampling (see Alexopoulos and Shultes 1998, 2001) was developed for analyzing system performance in fault-tolerant repairable systems. This approach has been used to derive importance sampling estimators for limiting system unavailability and mean time to system failure that yield bounded relative error. Shultes (2002) applied this approach to estimate the system overflow probability in tandem Jackson networks. This method yields a zero variance importance sampling distribution for a single node system. For systems with more than one node, this method yields asymptotically efficient results with some restrictions on the model parameters.

The rare event studied in this paper is the buffer overflow probability at the second node in a two node tandem Jackson network. An exponential tilting technique was developed by Kroese and Nicola to estimate this overflow probability (see Kroese and Nicola 2002). These authors exponentially tilt a Markov additive process representation of the system to derive an importance sampling estimator. Their distribution is state dependent in that it depends on the contents of the first buffer.

This paper applies the balanced likelihood ratio approach to derive an importance sampling distribution for estimating the overflow probability at the second node in a two node tandem Jackson network. Like all balanced likelihood ratio methods, the distribution has guaranteed variance reduction over standard Monte Carlo methods. Unlike the distributions from Shultes (2002), the proposed distributions depend on the contents of the buffers and can be applied to any set of arrival and service rates. A complete proof of optimality is not available at this time, but numerical results suggest asymptotic efficiency.

Section 2 presents the model studied and provides an overview of importance sampling and the balanced likelihood ratio approach. Section 3.1 and 3.2 provide details of the proposed method for the infinite and finite first buffer cases respectively. Section 4 contains experimental results. Conclusions and future research directions are presented in Section 5.

2 BACKGROUND

Consider a tandem Jackson network with two nodes. Customers arrive at the first queue according to a Poisson process with rate λ.The service time of a customer at the first node is exponential with rate μ_1, independent of the input process and service times at the second node. The output process of the first queue forms the input process of the second queue. The service time at the second node is exponential with rate μ_2, which is also independent of the input process and service times at the first node. Without loss of generality, assume that $\lambda + \mu_1 + \mu_2 = 1$. The queueing system is assumed to be stable, i.e., $\lambda < \min(\mu_1, \mu_2)$.

Let $X(t)$ and $Y(t)$ denote the number of customers at the first and second node at time t, respectively (including customers in service). Let b denote the size of the first buffer, which may be finite or infinite. The quantity of interest is the probability (γ) that the number of customers in the second queue reaches some high level $B \in \mathbb{N}$ before hitting 0. We wish to estimate this probability given that the system starts in state $(X(0) = 0, Y(0) = 0)$ or $(X(0) = 1, Y(0) = 1)$. These probabilities are denoted as γ_0 and γ_1 respectively.

The system can be modeled as a Markov process with system state $Z(t) = (X(t), Y(t))$. Let

$$r(t) = \lambda + 1\,(X(t) > 0)\,\mu_1 + 1\,(Y(t) > 0)\,\mu_2$$

denote the total rate of event transitions out of $Z(t)$. The probability that a buffer overflow is observed depends upon the embedded discrete-time Markov chain whose one-step transition probabilities at time t are: $\lambda/r(t)$ the probability the next event is an arrival, $1\,(X(t) > 0)\,\mu_1/r(t)$ the probability that the next event is a service completion at node one, and $1\,(Y(t) > 0)\,\mu_2/r(t)$ the probability that the next event is a service completion at node two.

2.1 Importance Sampling

Let Ω denote the set of all cycles and for each $\omega \in \Omega$, let $\beta(\omega)$ denote the largest number of customers at the second node within the cycle. The probability $P(\omega)$ of observing the cycle ω is the product of one-step transition probabilities. A new distribution P' is defined using importance sampling, such that $P(\omega) > 0 \implies P'(\omega) > 0$ and

$$\begin{aligned} \gamma &= \sum_{\omega\in\Omega} 1\,(\beta(\omega) = B)\,\frac{P(\omega)}{P'(\omega)}P'(\omega) \\ &= \sum_{\omega\in\Omega} 1\,(\beta(\omega) = B)\,L(\omega)P'(\omega) \end{aligned}$$

where the likelihood ratio $L(\omega)$ is the Radon-Nikodym derivative of P with respect to P'. The likelihood ratio $L(\omega)$ can be decomposed into a product of one-step transition *event likelihood ratios* associated with each individual event within the cycle.

2.1.1 Asymptotic Properties

The asymptotic efficiency of an estimator can be measured using the relative error of the estimated quantity. Relative error is defined as the ratio of the standard deviation of the estimator over its expected value. The estimator yields *bounded relative error* if the relative error remains bounded as the quantity to be estimated approaches zero. This means that, the sample size required to achieve a desired level of accuracy remains bounded in the limit. An estimator is said to be *asymptotically efficient* if the relative error grows at a subexponential rate as the quantity to be estimated approaches zero. For importance sampling estimators, bounded relative error implies asymptotic efficiency.

2.1.2 Variance Reduction Ratio

To compare the performance of two importance sampling estimators, we need to take into account variance reduction and the computational effort required to achieve that reduction. The variance reduction ratio (VRR) measures the trade-off between variance reduction and the associated computational cost. VRRs are computed by multiplying a ratio of the variances of two estimators by a ratio of the corresponding computational effort, i.e., simulation time or number of events sampled to generate that variance. Typically, VRRs are estimated empirically by simulation. If the VRR is less than one, then the approach in the numerator is more efficient and a VRR greater than one implies that the approach in the denominator is more efficient.

2.2 Balanced Likelihood Ratio Approaches

The proposed importance sampling method is based on the balanced likelihood ratio approach. This approach was originally proposed to estimate the reliability of fault-tolerant repairable systems (see Alexopoulos and Shultes 2001) and was later adapted to estimate system overflow probabilities in tandem-Jackson networks (see Shultes 2002). A key feature of this approach is that likelihood ratios associated with regenerative cycles can be bounded from above by controlling event likelihood ratios associated with individual events within cycles.

The application of the balanced likelihood ratio approach to estimate γ_0 and γ_1 proceeds as follows. Classify all system events into 2 classes: events that move the system towards buffer overflow and events that move the system away from buffer overflow. Arrival events and service completion events at the first node belong to the first category and service completion events at the second node fall into the second category. The balanced likelihood ratio method balances the event likelihood ratios associated with events from these two classes.

Every service completion event at the second node must be preceded by an arrival event and a service completion event at the first node. The product of these three event likelihood ratios can be forced to be one for all customers. This assignment causes likelihood ratios associated with cycles to be bounded below one. The proposed method has the following basic balanced likelihood ratio properties established by Shultes (2002).

- Every event that moves the system closer to the rare event (arrival and service completion at the first node) has one corresponding event (service completion at the second node) that effectively cancels out the events that moved the system closer to overflow.
- Events that would complete a cycle before the system experiences a rare event have zero probability in the importance sampling distribution.
- If the events that move the system closer to buffer overflow are forced to be more likely, then the corresponding future event which would move the system away from overflow is forced to be less likely.

3 TANDEM QUEUES

Balanced likelihood ratio methods for estimating the probabilities γ_0 and γ_1 when the first buffer capacity is infinite and finite are described in Sections 3.1 and 3.2 respectively. The importance sampling distribution is the same for estimating both γ_0 and γ_1. However, the method for estimating γ_0 includes cases which do not occur while estimating γ_1, i.e., when the starting state is (1,1). Hence, without loss of generality, the importance sampling distributions are described for the starting state (0,0).

Customer arrival events and service completion events at the first node generate event likelihood ratios denoted by la and ls respectively. These event likelihood ratios are used as multipliers for biasing the probability of service completion at the second node. The importance sampling distribution is formed such that the content of the second buffer reaches the bound B in all cycles. The idea is to avoid paths which fail to experience the rare event within the cycle.

The proposed importance sampling distribution depends on the sample path for the process Z. For simplicity, time and the sample path history are omitted from the following presentation. Let λ' denote the importance sampling probability of an arrival event. Let μ_1' and μ_2' denote the importance sampling probabilities of service completion events at the first and second nodes respectively.

3.1 Infinite First Buffer

The importance sampling approach described in Section 2.2 is directly applied to the infinite first buffer case. Assume the system starts from state $(0, 0)$. There are four cases to consider: (1) The system is empty, (2) All customers are at the first node, (3) All customers are at the second node, and (4) Customers are at both nodes in the system.

Case 1: The system is empty. The next event is a customer arrival with probability one. The event likelihood ratio for this event is replaced by $la' = \lambda/(\lambda + \mu_2)$ in the implementation to ensure that the service completion probability at node two associated with this arrival is reduced. It is easy to show that this deviation from the basic balanced likelihood ratio approach maintains established likelihood ratio properties.

Case 2: All customers in the system are at the first node, i.e., the system state is $(X(t) = x, Y(t) = 0, t \geq 0)$ for some $x \in \mathbb{N}$. In this case, the next event could be either an customer arrival or a service completion at the first node. Deviating from the original balanced likelihood ratio description, the importance sampling probability for a service completion event at the first node is reduced to increase the arrival probability. The importance sampling probabilities in this case are:

$$\mu_1' = la\left(\frac{\mu_1}{\lambda + \mu_1}\right), \text{ and}$$

$$\lambda' = 1 - \mu_1'.$$

Case 3: All customers in the system are at the second node, i.e., the system state is $(X(t) = 0, Y(t) = y, t \geq 0)$ for some $y \in \mathbb{N}$. In this case, the next event could be either a customer arrival or a service completion at the second node. The importance sampling probabilities when $y > 1$ are:

$$\mu_2' = la\,(ls)\left(\frac{\mu_2}{\lambda + \mu_2}\right), \text{ and}$$

$$\lambda' = 1 - \mu_2'.$$

The service completion event is not allowed when $y = 1$ if the rare event has not yet occurred within the cycle. In this latter case, the customer arrival probability is one.

Case 4: Customers in the system are at node one and node two, i.e., the system state is $(X(t) = x, Y(t) = y, t \geq 0)$ for some $x \in \mathbb{N}$, $y \in \mathbb{N}$. The importance sampling probabilities in this case when $y > 1$ derive from:

$$\mu_2' = la\,(ls)\left(\frac{\mu_2}{\lambda + \mu_1 + \mu_2}\right).$$

The remaining probability $\left(1 - \mu_2'\right)$ is split between the customer arrival event and service completion event at the first node based on the number of customers in the system. The importance sampling distribution gives more importance to arrivals when the system size is less than the buffer size B. When the system size is greater than the bound B, the importance sampling probabilities allocated to the arrival event and the service completion at node one are proportional to the respective rates λ and μ_1.

Let ρ_s and ρ_a denote the fraction of $\left(1 - \mu_2'\right)$ assigned to the service completion at the first node and the arrival event respectively. The importance sampling probabilities for the arrival event and the service completion at node one are:

$$\mu_1' = \rho_s\left(1 - \mu_2'\right), \text{ and}$$

$$\lambda' = \rho_a\left(1 - \mu_2'\right),$$

where

$$\rho_s = \begin{cases} \max\left(0.5, \dfrac{\mu_1}{1 - \mu_2'}\right) & \text{if } x + y \leq B, \\ \dfrac{\mu_1}{\lambda + \mu_1} & \text{if } x + y > B, \end{cases}$$

and

$$\rho_a = 1 - \rho_s.$$

3.1.1 Implementation

Define two stacks: La for storing arrival event likelihood ratios and Ls for storing likelihood ratios for service completion events at the first node. Initially each stack contains one multiplier, $la' = \lambda/(\lambda + \mu_2)$ is on stack La and $ls' = 0$ is on stack Ls where the 0 guarantees that the cycle does not end without observing a buffer overflow event. After each arrival event, the event likelihood ratio (λ/λ') is pushed onto stack La. After each service completion event at the second node, one likelihood ratio from each stack is removed. For each service completion event at the first node, the event likelihood ratio (μ_1/μ_1') is pushed onto stack Ls if the system is in state (x, y) for some $x \in \mathbb{N}$, $y \in \mathbb{N}$ and a likelihood ratio is removed from stack La when the system state is $(x, 0)$ for some $x \in \mathbb{N}$.

3.2 Finite First Buffer

The balanced likelihood ratio method for estimating the probability of buffer overflow in the second node when the first buffer has finite capacity is described below. The approach is similar to the infinite first buffer case.

Assume the system starts from state $(0, 0)$. The same four cases as in the infinite first buffer case are considered. For cases 1, 2 and 3, i.e., when the system is empty and when the system state is $(x, 0)$ and $(0, y)$ for some $x \in \mathbb{N}$, $y \in \mathbb{N}$, the importance sampling distribution is the same as in the infinite first buffer case. When the system is in state (x, y) for some $x \in \mathbb{N}$, $y \in \mathbb{N}$, the importance sampling probabilities derive from the same starting point as before:

$$\mu_2' = la\,(ls)\left(\frac{\mu_2}{\lambda + \mu_1 + \mu_2}\right).$$

As before, the remaining probability $\left(1 - \mu_2'\right)$ is split between the customer arrival event and the service completion event at the first node based on the number of customers in the system. Since the first node has a finite capacity b, the fraction ρ_s of $\left(1 - \mu_2'\right)$ assigned to the service completion at node one is increased, relative to the infinite first buffer case, by a factor c which depends on the number of customers at the first node. However, if $\mu_1 > \mu_2$ then this modification is not necessary, so $c = 0$ in this special case. The importance sampling probabilities for customer arrival events and service completion at node one are:

$$\mu_1' = (\rho_s + c)\left(1 - \mu_2'\right), \text{ and}$$

$$\lambda' = 1 - \mu_1' - \mu_2',$$

where ρ_s is defined as before and

$$c = \frac{x}{b}\left(\frac{\mu_1}{\lambda + \mu_1} - \rho_s\right).$$

The method can be implemented in the same way as that of the infinite first buffer case using two stacks: La for storing arrival event likelihood ratios and Ls for storing likelihood ratios of service completion events at first node.

4 NUMERICAL RESULTS

Experimental results for four, two node tandem Jackson network examples are presented. In the first example, the second server is the bottleneck ($\mu_1 > \mu_2$), in the second and third examples the first server is the bottleneck ($\mu_1 < \mu_2$) and in the fourth example the service rates at the two nodes are equal. Results from experiments that estimate the probability that the contents of the second buffer reach the bound B before reaching zero starting from state $(1, 1)$ and $(0, 0)$ are presented for both finite and infinite first buffer cases. These cases come directly from Kroese and Nicola (2002). The rates in the tables can be normalized so that the normalized rates sum to one.

The result from each simulation experiment is based on 1,000,000 cycles. Cycles end when the second node experiences buffer overflow or when the second node empties. Each simulation run provides an estimate for the overflow probability (Mean), a 95% confidence interval halfwidth (Halfwidth) and the relative error, i.e., standard deviation divided by mean (RE). Computation times (CPU) are displayed in terms of average number of events per cycle. The tables include estimates of the overflow probabilities obtained by applying the exponential change of measure technique presented by Kroese and Nicola (2002). The numerical values for these probabilities presented by Kroese and Nicola (2002) are also provided. The numerical values can be obtained by using the algorithm outlined in Garvels and Kroese (1999). The results from the two methods (BLR and exponential change of measure) are compared using Variance Reduction Ratios (VRRs). If the VRR is less than one, then the exponential change of measure method by Kroese and Nicola (K-N method) is more efficient and the BLR method is more efficient if VRR is greater than one. All simulations were implemented in C and run on an HP C3600 workstation.

Tables 1-4 display the results for the estimates of the probability γ_1 for the infinite first buffer cases. Tables 5-8 display the results for the estimates of the probability γ_1 for cases where the first buffer is limited to nine customers. Tables 9 and 10 present the estimates of the probability γ_0 for all the four examples for the infinite and finite first buffer cases respectively.

The relative error of the BLR method is bounded independent of the buffer size when the second server is the bottleneck in both finite and infinite buffer cases. In the other two cases, i.e., when the first server is the bottleneck and when the service rates at both nodes are equal, the relative error is linearly bounded. Based on the numerical results, the BLR method is more efficient than the K-N method when the buffer at the first node is infinite. In contrast, the K-N method is more efficient than the BLR method for B larger than 25 in the finite first buffer cases. This is not surprising given that the BLR relative errors are only linearly bounded in this case while the relative errors for the K-N method are bounded.

The BLR method yields similar results when used to estimate the overflow probabilities γ_0 and γ_1. The K-N method also yields similar results except when the first server is the bottleneck and its capacity is infinite in which case the relative error increases sharply with B. Kroese and Nicola (2002) have suggested that a different change of measure is needed in this case when the starting state is (0,0).

5 CONCLUSIONS

This paper presents a balanced likelihood ratio importance sampling approach for estimating the overflow probability of the second buffer in a two node tandem Jackson network. Numerical results indicate that the relative error is bounded independent of the buffer size except when the first server is the bottleneck in which case the relative error is linearly bounded. Empirical evidence indicates that the BLR method outperforms existing importance sampling distributions when the first node buffer is infinite. More work is needed to determine why the BLR method struggles when the first node buffer is finite. The theoretical properties of the proposed importance sampling distributions including asymptotic characteristics need to be studied in detail. The proposed methods can be readily extended to estimate individual buffer overflow probabilities in tandem Jackson networks with more than two nodes.

REFERENCES

Alexopoulos, C. and B. C. Shultes. 1998. The balanced likelihood ratio method for estimating performance measures of highly reliable systems. In *Proceedings of the 1998 Winter Simulation Conference,* ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 1479-1486. Piscataway, New Jersey: IEEE.

Alexopoulos, C. and B. C. Shultes. 2001. Estimating reliability measures for highly dependable systems, using balanced likelihood ratios. *IEEE Transactions on Reliability* 50 (3): 265-280.

Asmussen, S., and R.Y. Rubinstein. 1995. Steady state rare events simulation in queueing models and its complexity properties. In *Advances in Queueing: Theory, Methods and Open problems,* ed. J. H. Dhashalow, CRC Press, Boca Raton, Florida, 429-462.

De Boer, P. T., D. P. Kroese, and R. Y. Rubinstein. 2002. Estimating buffer overflows in three stages using cross-entropy. In *Proceedings of the 2002 Winter Simulation Conference,* ed. E. Yucesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 301-309.

Glasserman, P. and S.-G. Kou. 1995. Analysis of an importance sampling estimator for tandem queues. *ACM*

Table 1: Estimates of γ_1 in Example 1 $(\lambda, \mu_1, \mu_2 = 1, 4, 2)$ with $b = \infty$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	1.43e-06	BLR	1.43e-06	1.26e-09	0.45e-03	85	3.8
		K-N	1.43e-06	3.20e-09	1.13e-03	51	—
25	4.47e-08	BLR	4.47e-08	3.95e-11	0.45e-03	110	3.8
		K-N	4.51e-08	1.00e-10	1.13e-03	65	—
50	1.33e-15	BLR	1.33e-15	1.18e-18	0.45e-03	235	3.8
		K-N	1.35-15	3.01e-18	1.13e-03	136	—
60	1.30e-18	BLR	1.30e-18	1.15e-21	0.45e-03	285	3.8
		K-N	1.33e-18	2.95e-21	1.13e-03	164	—
100	1.18e-30	BLR	1.18e-30	1.05e-33	0.45e-03	485	3.8
		K-N	1.22e-30	2.72e-33	1.13e-03	276	—

Table 2: Estimates of γ_1 in Example 2 $(\lambda, \mu_1, \mu_2 = 1, 2, 3)$ with $b = \infty$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	2.05e-11	BLR	2.05e-11	5.24e-14	1.30e-03	70	9.2
		K-N	2.05e-11	1.97-13	4.89e-03	46	—
25	4.61e-14	BLR	4.61e-14	1.35e-16	1.49e-03	89	9.1
		K-N	4.63e-14	5.07e-16	5.59e-03	57	—
50	4.31e-27	BLR	4.30e-27	1.93e-29	2.29e-03	186	8.5
		K-N	4.28e-27	7.27e-29	8.66e-03	112	—
60	2.96e-32	BLR	2.96e-32	1.49e-34	2.57e-03	224	8.4
		K-N	2.94e-32	5.62e-34	9.76e-03	133	—
100	8.60e-53	BLR	8.58e-53	6.02e-55	3.58e-03	378	8.4
		K-N	8.49e-53	2.32e-54	13.8e-03	218	—

Table 3: Estimates of γ_1 in Example 3 $(\lambda, \mu_1, \mu_2 = 3, 4, 6)$ with $b = \infty$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	1.35e-08	BLR	1.35e-08	3.54e-13	1.34e-03	97	6.6
		K-N	1.35e-08	1.38e-20	5.20e-03	42	—
25	1.97e-10	BLR	1.97e-10	5.95e-13	1.54e-03	125	6.4
		K-N	1.98e-10	2.33e-12	5.99e-03	52	—
50	2.20e-19	BLR	2.20e-19	1.03e-21	2.39e-03	264	6.1
		K-N	2.22e-19	4.13e-21	9.49e-03	101	—
60	6.54e-23	BLR	6.53e-23	3.46e-25	2.70e-03	320	6.0
		K-N	6.68e-23	1.39e-24	10.7e-03	120	—
100	6.79e-37	BLR	6.79e-37	5.08e-39	3.80e-03	541	5.8
		K-N	6.96e-37	2.05e-38	15.2e-03	194	—

Table 4: Estimates of γ_1 in Example 4 $(\lambda, \mu_1, \mu_2 = 1, 2, 2)$ with $b = \infty$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	2.79e-07	BLR	2.79e-07	6.11e-10	1.11e-03	95	3.1
		K-N	2.78e-07	1.59e-09	2.90e-03	43	—
25	7.66e-09	BLR	7.68e-09	1.84e-11	1.22e-03	122	2.9
		K-N	7.67e-09	4.67e-11	3.10e-03	54	—
50	1.56e-16	BLR	1.56e-16	4.83e-19	1.58e-03	257	2.4
		K-N	1.56e-16	1.16e-18	3.79e-03	107	—
60	1.38e-19	BLR	1.38e-19	4.54e-22	1.67e-03	310	2.3
		K-N	1.39e-19	1.08e-21	3.99e-03	127	—
100	9.62e-32	BLR	9.61e-32	3.72e-34	1.98e-03	520	2.2
		K-N	9.58e-32	8.63e-34	4.59e-03	208	—

Table 5: Estimates of γ_1 in Example 1 $(\lambda, \mu_1, \mu_2 = 1, 4, 2)$ with $b = 9$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	1.43e-06	BLR	1.43e-06	1.27e-09	0.45e-03	85	3.8
		K-N	1.43e-06	3.20e-09	1.13e-03	51	—
25	4.45e-08	BLR	4.45e-08	3.99e-11	0.45e-03	110	3.7
		K-N	4.48e-08	9.99e-11	1.13e-03	65	—
50	1.30e-15	BLR	1.30e-15	1.21e-18	0.47e-03	235	3.5
		K-N	1.32e-15	2.99e-18	1.13e-03	136	—
60	1.26e-18	BLR	1.26e-18	1.19e-21	0.48e-03	285	3.5
		K-N	1.29e-18	2.92e-21	1.13e-03	164	—
100	1.12e-30	BLR	1.11e-30	1.10e-33	0.49e-03	485	3.3
		K-N	1.12e-30	2.66e-33	1.17e-03	277	—

Table 6: Estimates of γ_1 in Example 2 $(\lambda, \mu_1, \mu_2 = 1, 2, 3)$ with $b = 9$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	1.89e-11	BLR	1.88e-11	4.15e-14	1.12e-03	68	2.8
		K-N	1.87e-11	8.69e-14	2.37e-03	43	—
25	3.76e-14	BLR	3.76e-14	1.00e-16	1.37e-03	87	1.8
		K-N	3.76e-14	1.75e-16	2.37e-03	53	—
50	1.25e-27	BLR	1.25e-27	6.90e-30	2.83e-03	182	0.4
		K-N	1.25e-27	5.80e-30	2.37e-03	107	—
60	5.06e-33	BLR	5.00-33	7.48e-35	9.62e-03	221	0.1
		K-N	5.06e-33	2.35e-35	2.37e-03	128	—
100	1.38e-54	BLR	1.39-54	2.48e-56	9.12e-03	371	0.04
		K-N	1.37e-54	6.39e-57	2.37e-03	214	—

Table 7: Estimates of γ_1 in Example 3 $(\lambda, \mu_1, \mu_2 = 3, 4, 6)$ with $b = 9$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	1.15e-08	BLR	1.15e-08	2.74e-11	1.21e-03	91	1.6
		K-N	1.15e-08	5.12e-11	2.27e-03	41	—
25	1.41e-10	BLR	1.40e-10	3.89e-13	1.41e-03	117	1.2
		K-N	1.41e-10	6.25e-13	2.27e-03	52	—
50	3.89e-20	BLR	3.88e-20	1.93e-22	2.54e-03	246	0.3
		K-N	3.88e-20	1.73e-22	2.27e-03	103	—
60	5.84e-24	BLR	5.85e-24	3.57e-26	3.11e-03	299	0.2
		K-N	5.89e-24	2.61e-26	2.27e-03	124	—
100	2.98e-39	BLR	2.99e-39	3.87e-41	6.61e-03	506	0.1
		K-N	2.98e-39	1.33e-41	2.27e-03	207	—

Table 8: Estimates of γ_1 in Example 4 $(\lambda, \mu_1, \mu_2 = 1, 2, 2)$ with $b = 9$

Buffersize	Numerical	Method	Mean	Halfwidth	RE	CPU	VRR
20	2.56e-07	BLR	2.56e-07	3.29e-10	0.65e-03	96	3.8
		K-N	2.55e-07	9.56e-10	1.91e-03	43	—
25	6.40e-09	BLR	6.40e-09	9.98e-12	0.79e-03	125	2.5
		K-N	6.42e-09	2.40e-11	1.91e-03	54	—
50	6.34e-17	BLR	6.34e-17	1.84e-19	1.48e-03	268	0.7
		K-N	6.33e-17	2.37e-19	1.91e-03	110	—
60	3.99e-20	BLR	3.99e-20	1.40e-22	1.79e-03	324	0.5
		K-N	3.99e-20	1.49e-22	1.91e-03	132	—
100	6.24e-33	BLR	6.25e-33	3.92e-35	3.20e-03	552	0.1
		K-N	6.21e-33	2.33e-35	1.91e-03	221	—

Table 9: Estimates of γ_0 with $b = \infty$

Buffersize	Example	Method	Mean	Halfwidth	RE	CPU	VRR
50	1	BLR	7.40e-16	1.40e-18	0.97e-03	230	1.0
		K-N	7.40e-16	2.05e-18	1.41e-03	112	—
	2	BLR	1.03e-26	9.80e-29	4.86e-03	188	18.4
		K-N	9.06e-27	5.54e-28	31.1e-03	108	—
	3	BLR	3.86e-19	3.10e-21	4.10e-03	267	10.4
		K-N	3.82e-19	1.65e-20	22.1e-03	98	—
	4	BLR	1.54e-16	4.87e-19	1.62e-03	257	6.0
		K-N	1.54e-16	1.93e-18	6.39e-03	98	—
100	1	BLR	6.56e-31	1.23e-33	0.96e-03	480	1.0
		K-N	6.56e-31	1.82e-33	1.41e-03	224	—
	2	BLR	2.39e-52	5.76e-54	12.3e-03	380	17.8
		K-N	2.22e-52	3.29e-53	75.7e-03	207	—
	3	BLR	1.23e-36	1.87e-38	7.72e-03	543	11.8
		K-N	1.23e-36	1.09e-37	45.4e-03	187	—
	4	BLR	9.53e-32	3.76e-34	2.01e-03	520	3.6
		K-N	9.59e-32	1.18e-33	6.28e-03	187	—

Table 10: Estimates of γ_0 with $b = 9$

Buffersize	Example	Method	Mean	Halfwidth	RE	CPU	VRR
50	1	BLR	7.23e-16	1.40e-18	0.99e-03	233	1.0
		K-N	7.35e-16	2.06e-18	1.43e-03	111	—
	2	BLR	1.96e-27	1.16e-29	3.02e-03	185	1.9
		K-N	1.96e-27	2.15e-29	5.61e-03	103	—
	3	BLR	5.59e-20	3.05e-22	2.79e-03	251	0.8
		K-N	5.64e-20	4.30e-22	3.89e-03	100	—
	4	BLR	5.89e-17	1.78e-19	1.54e-03	269	1.0
		K-N	5.86e-17	2.89e-19	2.52e-03	101	—
100	1	BLR	6.21e-31	1.23e-33	1.00e-03	483	1.0
		K-N	6.45e-31	1.83e-33	1.45e-03	225	—
	2	BLR	2.16e-54	3.93e-56	9.27e-03	374	0.2
		K-N	2.18e-54	2.41e-56	5.65e-03	204	—
	3	BLR	4.30e-39	6.10e-41	7.23e-03	512	0.1
		K-N	4.33e-39	3.29e-41	3.88e-03	199	—
	4	BLR	5.83e-33	3.86e-35	3.38e-03	554	0.3
		K-N	5.78e-33	2.85e-35	2.51e-03	200	—

Transactions on Modeling and Computer Simulation 5 (1): 22-42.

Garvels, M. J. J., and D. P. Kroese. 1999. On the entrance distribution in RESTART simulation. In *Proceedings of the Second Workshop on Rare Event Simulation (RESIM' 99)*, Enschede, The Netherlands, 65-88.

Heidelberger, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation* 5 (1): 43-85.

Kroese, D. P. and V. F. Nicola. 2002. Efficient simulation of a tandem Jackson network. *ACM Transactions on Modeling and Computer Simulation* 12 (2): 119-141.

Parekh, S. and J. Walrand. 1989. A quick simulation method for excessive backlogs in networks of queues. *IEEE Transactions on Automatic Control* 34 (1): 54-66.

Shultes, B. C. 2002. A balanced likelihood ratio approach for analyzing rare events in a tandem Jackson network. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yucesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 424-432.

AUTHOR BIOGRAPHIES

RAMYA DHAMODARAN is a graduate student in the Department of Mechanical, Industrial, and Nuclear Engineering at the University of Cincinnati. She received her B.S. in Industrial Engineering from Anna University, College of Engineering, India. Her research interests are in simulation and optimization. Recent research efforts focus on applying generalized regenerative analysis to derive closed-form expressions for reliability, and efficient importance sampling techniques. Her email address is <dhamodr@email.uc.edu>.

BRUCE C. SHULTES is an Assistant Professor of Industrial Engineering at the University of Cincinnati. Previously, he spent two years at the Naval Postgraduate School as an Assistant Research Professor and a National Research Council Research Associate. He received his B.S. in Applied Mathematics from Carnegie Mellon University, his M.S. in Management Science from Case Western Reserve University, and his Ph.D. in Industrial Engineering (Stochastic Systems) from Georgia Institute of Technology. His research interests are in the areas of simulation, analysis of stochastic systems, and software engineering. His recent research efforts include: analysis of rare events, probabilistic risk analysis, estimating ANSI form tolerances, and quickly assessing scheduling feasibility. He is a member of IIE, INFORMS, and the INFORMS College on Simulation. His email address is `<Bruce.Shultes@uc.edu>`.