Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

AN EFFICIENT IMPORTANCE SAMPLING METHOD FOR RARE EVENT SIMULATION IN LARGE SCALE TANDEM NETWORKS

Lei Wei

School of Electrial Engineering and Computer Science
University of Central Florida
Orlando, FL 32816, U.S.A.

Honghui Qi

Weiston Communications
1400 Pelican Bay TRL
Winter Park, FL 32792, U.S.A.

ABSTRACT

In this paper, we present a variance minimization (VM) procedure for rare event simulation in tandem queueing networks. We prove that the VM method can produce a zero variance. The VM method is suitable to compute optimal importance sampling (IS) parameters for small scale tandem networks. For large scale tandem networks we propose a sub-optimal IS (SOIS) method, which projects the optimal biased transition probabilities of the corresponding small scale system into those of a large scale system. In other words, we establish an efficient IS method for a large scale system by zooming into a small scale system and then projecting our findings into the large scale system. The numerical results show that our SOIS method can produce accurate results with very short CPU time, while many other methods often require much longer.

1 INTRODUCTION

The Monte Carlo (MC) method is commonly used to evaluate the performance of a system or network. MC simulations can produce accurate performance estimates, provided that the number of simulation trials is sufficiently large. However, this condition can be severe. For example, for a 95% confidence interval of $[2P/5, 8P/5]$, where P denotes the probability we want to estimate, the standard MC approach requires at least 10^8 simulations trials for $P = 10^{-7}$ (see Orsak and Aazhang 1984).

A fast simulation technique, known as importance sampling (IS), has been developed to reduce the number of simulation trials, providing substantial run-time savings when determining the performance of communication systems and networks. A comprehensive literature overview of IS in digital communications and a summary of the main techniques can be found in Smith et. al. (1997).

A key paper which popularized the usage of IS in bit error rate estimation is due to Shanmugan and Balaban (1980). This paper described the idea that IS biases the noise density function, so that more samples can be taken from important regions (that is, the regions that cause errors). The method used is to scale the original noise density to increase the probability of simulation samples taken from the important regions.

It is commonly known that most IS techniques come with a price. The greatest difficulties in obtaining an efficient IS method are (a) finding a good way to bias system uncertainties (such as the noise density function); and (b) optimizing system parameters. It has been shown that, in many cases, the computation and experiments required for achieving an efficient IS method may be more complicated than the original problem. Thus, this implies that we have to find simple and efficient IS algorithms.

Analytical methods are often used to simplify simulation tasks. For example, in bit error probability (BEP) estimation we can compute conditional BEP and the true BEP can be then estimated by averaging over the conditional BEP (Wei 1995a). This method has been widely used by engineers in practice without knowing its connection with the IS concept. In Wei (1995b), it was showed that the method is conditional optimal, thus we called it the optimal conditional importance sampling (OCIS) method. OCIS is optimal under the condition that only certain system parameters can be biased. Several applications can be found in Wei (1995a), Wei (1996), and Wei and Schlegel (1995).

The OCIS method cannot be extended to stochastic systems. It is generally a difficult problem to construct an IS method for stochastic systems with a large-scale state space. Sadowsky (1990) proposed an efficient IS method to simulate the Viterbi decoder, and Devetsikiotis and Townsend (1993) presented a system independent method for simulating communication networks. Similarly, in the field of efficient simulation of network and reliability models, several interesting IS-based methods have also been proposed for estimating steady state availability and the mean time to failure (Smith et. al 1997, Glasseman and Kou 1995, Parekh and Walrand 1989, and Heidelberger 1995).

First used by Cottrell (1983) the theory of large deviations techniques (LDT) was developed to design an efficient IS simulation method in the communication field. This work has been developed into a coherent simulation methodology through a series of papers by Sadowsky, Bucklew, and many others.

In a queueing system with finite buffers, some proportion of the customers arriving at a queue is lost due to buffer overflows. While this number will be small in a properly dimensioned system, it is of interest because there is often a large cost associated with such a loss. For example in a computer network, the queue customer may be a packet of data, and the system must arrange for re-transmission of lost packets. The rarity of customer loss makes direct simulation costly interms of computer time. For some simple systems, such as an $M/M/1$ queue, it is possible to calculate analytically the mean time to overflows, so that simulation is unnecessary. However, for some complex systems, the analytical computation is not generally possible, and simulation is often used to fulfill the task.

Based on a heuristic application of LDTs, Parekh and Walrand (1989) established efficient IS estimators for the overflow probabilities in various Jackson networks. By simply changing the arrival rate and a smallest service rate in the system, their estimators can be several orders of magnitude faster than direct simulations. Glasserman and Kou (1995) studied the asymptotic efficiency of these methods and showed that in certain parameter regions the estimator is asymptotic optimal (in term of bounded relative error), while in other regions it is not. Unlike the single queue, the boundaries on the state spaces in queueing networks make it much more difficult to construct an effective IS estimator and to analytically compute the overflow probability.

In this paper, we will first show the method derived by Wei and Wei (2000) can produce zero variance estimation. We then propose a sub-optimal IS (SOIS) method to deal with rare-event estimation for a large scale Markov system. The SOIS contains two unique steps: Zoom-in and Projection. In words, we deal with a large scale Markov system by zooming-in into a small size system first and then projecting the results obtained for the small system into the large system for simulation.

This paper is organized as follows. In Section 2, we formulate the problem of estimating a rare event in queueing networks. In Section 3, we revisit the VM method and the key results in Wei and Wei (2001). In Section 4 we introduce near optimal IS methods based on a minimization procedure for two $M/M/1$ queues in tandem. Finally, we show several numerical results in Section 5 and conclude the paper in Section 6.

2 PROBLEM FORMULATION

Consider a discrete time Markov chain $R = (X_t, t = 0, 1, 2, \dots)$ of the queue length at the epoch of arrivals and departures of the queue. Denote F as a set of states that the total buffer size has reached. The set F may contain just one state as in the $M/M/1$ queue, or a number of states as in Jackson networks.

Define a cycle as the duration starting with an empty system and ending at the instant that the system, for the first time, either becomes empty again (type I cycle) or reaches the overflow states F (type II cycle) as shown in Figure 1. Let τ_0 be the first passage time returning to the initial empty state and τ_F be the first passage time to the overflow buffer size, both under the assumption that the system starts empty. Now, our task is to estimate $E_P[\tau_F]$, the expected value of τ_F or the mean time to buffer overflow. Since we are only interested in the first passage to F, we can model F as a set of absorbing states.

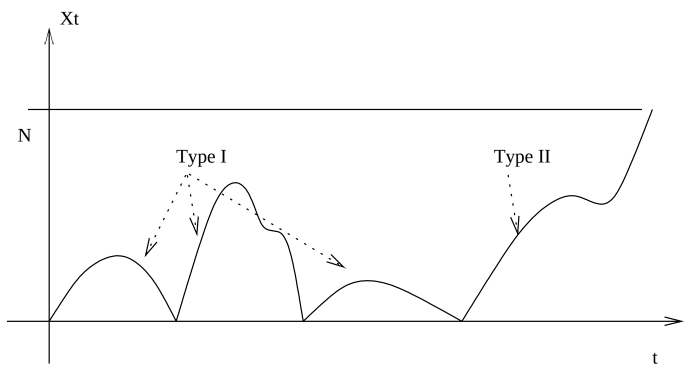


Figure 1: Realization of X_t

Let $\alpha = P(\tau_F < \tau_0)$ be the probability that X_t reaches F before returning to 0 given that the system starts empty. We then have (Parekh and Walrand 1989)

$$E_P[\tau_F] \approx \frac{E_P[\tau_0]}{\alpha}. \tag{1}$$

For stable systems, $E_P[\tau_0]$ can be easily estimated by the Monte Carlo simulations due to short passage time of the type I cycle. The difficult part in estimating $E_P[\tau_F]$ is the estimation of α since the type II cycle is a rare event. So from now on, our primary concern will be on how to estimate α.

3 VARIANCE MINIMIZATION TECHNIQUE

In this section, we will first review the importance sampling principle for the problem shown in the previous section. We then present key results in Wei and Wei (2001) which include the variance minimization (VM) method. Lastly, we prove the VM method produces zero variance.

3.1 Importance Sampling Principle

Define

$$V_k = 1_A(Y_k) \tag{2}$$

where $Y_k = [X_{k,0}, X_{k,1}, \dots]$ is a sample path in cycle k, A is the set of sample paths from initial empty state to the set F, $1_A(.)$ is a zero-one indicator of event A. If event A happens, then the indicator is one; otherwise, it is zero. Denote the likelihood ratio during cycle k by

$$W_k = \prod_{t=0}^{\tau_k - 1} \frac{p(X_{k,t+1}|X_{k,t})}{q_t(X_{k,t+1}|X_{k,t})} \tag{3}$$

where τ_k is the length of path k from empty state to the buffer size state, $p(.)$ and $q_t(.)$ denote the measures induced by the corresponding transition matrices $\mathbf{P}$ and $\mathbf{Q}_t$, respectively. For standard MC simulation, the system is simulated with the original distribution $\mathbf{P}$, i.e., $\mathbf{Q}_t = \mathbf{P}$. For IS simulation, the system is simulated according to $\mathbf{Q}_t$ and the value of α is estimated by

$$\widetilde{\alpha} = \frac{V_1 \cdot W_1 + V_2 \cdot W_2 + \cdots + V_n \cdot W_n}{n} \tag{4}$$

where n denotes the total number of simulation trials. The estimator in (4) is unbiased, if the simulation samples are i.i.d.. The variance of $\widetilde{\alpha}$ can be derived as

$$\sigma^2_{\widetilde{\alpha}} = \frac{E_Q[(V_k \cdot W_k)^2] - \alpha^2}{n}. \tag{5}$$

The relative error under IS is

$$E(\widetilde{\alpha}) = \sqrt{\frac{E_Q[(V_k \cdot W_k)^2] - \alpha^2}{n\alpha^2}}. \tag{6}$$

The optimal IS (OIS) method proposed in Wei and Wei (2001) can be applied here to estimate α. OIS describes the method to find the biased transition probability $\mathbf{Q}_t$ that minimizes the variance of the estimator, that is, $\min\left(E_Q[(V_k \cdot W_k)^2]\right)$. Then, the standard IS method is applied to estimate the value of α. In Wei and Wei (2001), we see that such a minimization essentially forces the sample trajectories Y_k associated with the likelihood ratio W_k to all belong to the type II cycle. Therefore the variance minimization procedure can be simplified as

$$\min\left(E_Q[(V_k \cdot W_k)^2]\right) = \min\left(E_Q[{W_k}^2]\right). \tag{7}$$

If the minimum variance is zero, then the simulation is not needed. We can compute the value of α following just one path. However, if the variance is not zero, then we still need simulation. We will show in this section that our method will always be a zero-variance estimator. Thus, it is not adequate to call it OIS. From now on, we call it the variance minimization method. Now, let us revisit the key results in Wei and Wei (2001).

3.2 Key Results in Wei and Wei (2001)

For a given observation length L we can model the Jackson network as a trellis diagram. Let x_L denote x given time length L. Clearly, we have $\alpha = \lim_{L\to\infty} \alpha_L$.

Lemma For given nonnegative real $a_1, a_2, ..., a_J$, we have

$$\arg_{x_k} \min_{0\le x_j \le 1} \left\{ \sum_{j=1}^{J-1} \frac{a_j}{x_j} + \frac{a_J}{1 - \sum_{j=1}^{J-1} x_j} \right\} = \frac{\sqrt{a_k}}{\sum_{j=1}^{J} \sqrt{a_j}}, \tag{8}$$

$$\min_{0\le x_j \le 1} \left\{ \sum_{j=1}^{J-1} \frac{a_j}{x_j} + \frac{a_J}{1 - \sum_{j=1}^{J-1} x_j} \right\} = \left(\sum_{j=1}^{J} \sqrt{a_j} \right)^2 \tag{9}$$

for $k = 1, \dots, J$. In this paper we use superscript (o) to denote the optimal value.

Algorithm for Variance Minimization (VM). The optimal IS transition probability matrix $\mathbf{Q}_t^{(o)}$, which minimizes $E_Q[(W_k)^2]|_L$, can be computed by the following procedures:

1. Set $t = L$ and set the vector $\omega_t = [\omega_t^{(0)}, \dots, \omega_t^{(S-1)}]$ in such way that $\omega_t^{(i)} = 1$ if $i \in F$; $\omega_t^{(i)} = 0$ otherwise, where S is the total number of states.
2. Set $t = t - 1$. Update each element in ω_t as follows. Select the $i - th$ element in ω_t, i.e., $\omega_t^{(i)}$. Find out all paths (say a total of J paths) from state i in time t leading to states $j_1, \dots, j_J$ in time $t + 1$. Then update $\omega_t^{(i)}$ as

$$\omega_t^{(i)} = \min_{q_t(j_1|i),\dots,q_t(j_J|i)} \left(\sum_{k=1}^{J} \frac{p^2(j_k|i)\omega_{t+1}^{(j_k)}}{q_t(j_k|i)} \right). \tag{10}$$

According to the previous LEMMA, we have

$$\omega_t^{(i)} = \left(\sum_{k=1}^{J} \sqrt{p^2(j_k|i)\omega_{t+1}^{(j_k)}} \right)^2 \tag{11}$$

and the optimal values of q_t, denoted as $q_t^{(o)}$, are

$$q_t^{(o)}(j_k|i) = \frac{\sqrt{p^2(j_k|i)\omega_{t+1}^{(j_k)}}}{\sum_{k=1}^{J} \sqrt{p^2(j_k|i)\omega_{t+1}^{(j_k)}}}. \tag{12}$$

3 Repeat Step 2 until $t = 0$. $E_Q[(W_k)^2]|_L = \omega_0^{(X_0)}$, where X_0 is the initial empty state.

Other Key Results. We showed that the optimal values for $\mathbf{Q}_t$ are unique. We proved if the system can be modeled as a transient Markov chain, then $\lim_{L\to\infty} q_1^{(o)}(j|i) = p(j|i)$. That is, the optimal biased transition probability matrix will converge to the original transition probability matrix. In other words, the optimal IS estimator is reduced to the Monte Carlo estimator.

One might induce that the optimal IS estimator for the overflow probability will also be the Monte Carlo estimator if the above result is true. In fact the induction is not true. In order to estimate the value of α we will focus on type II cycles. Thus, in the trellis the paths back to F will be deleted, but we do not increase the probability of the other path to one. This results in that the chain is not Markovian anymore, since the sum of all transition probabilities is not equal to one. However, we are still able to construct efficient IS methods.

3.3 Proof of Zero Variance

The proof is easy. Let $\pi_t = [\pi_t^{(0)}, \ldots, \pi_t^{(S-1)}]$. For $t = L$ we set it in such way that $\pi_t^{(i)} = 1$ if $i \in F$; $\pi_t^{(i)} = 0$ otherwise. We then recursively update π_t for $t = L-1, L-2, \ldots, 1$ as

$$\pi_t^{(i)} = \sum_{k=1}^{J} p(j_k|i)\pi_{t+1}^{(j_k)}. \tag{13}$$

When $t = 0$, $\pi_0^{(X_0)}$ is actually the sum of probabilities of all possible paths from X_0 to F during the observation period L. Therefore $\pi_0^{(X_0)}$ is the overflow probability from X_0 to F, i.e.,

$$\alpha_L = \pi_0^{(X_0)}. \tag{14}$$

From equations (11) and (13), It is obvious that $\pi_t^{(i)} = \sqrt{\omega_t^{(i)}}$. Therefore

$$\alpha_L^2 = \omega_0^{(X_0)} = E_Q[(W_k)^2]|_L. \tag{15}$$

When $L \to \infty$, we have $\alpha^2 = E_Q[(W_k)^2]$.

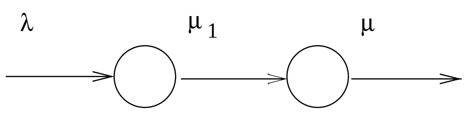
This tells us that the variance is always zero even when L is finite. Since the estimator's variance is zero, α can also be calculated by $\alpha = W_k$ for any path k from an initial empty state to the overflow state F.

4 SOIS FOR TANDEM NETWORKS

It has been shown in Glasseman and Kou (1995) that the PW method may not be asymptotic optimal. In this section, we will introduce suboptimal IS methods to estimate the overflow probability for tandem networks.

4.1 System Description

Figure 2 shows a simple Jackson network in which two $M/M/1$ queues are cascaded together. Let $\lambda+\mu_1+\mu_2 = 1$. Customers arrive to the network at rate λ, and can depart the system only after passing consecutively through two servers operating at rates μ_1 and μ_2, respectively. We assume $\lambda \le \mu_2 \le \mu_1$. If a server is busy, the arriving customer awaits service at the buffer associated with that server. An overflow occurs whenever the total number of customers in the system exceeds the combined buffer capacity, N, of the 2 servers. We estimate the probability of overflow before the system returns to the empty state given that the system is initially empty.



Two queues in tandem

Figure 2: Jackson Networks

Let $s = (s_1, s_2)$ represent the number of customers, including the one in service, at server 1 and server 2, respectively. This queueing system can be modeled as an embedded Markov process with a state space, $s \in R^2$. The overflow condition is represented by states in $F = [s : s_1 + s_2 > N - 1]$. Its transition probabilities are

$$p(1, 0|0, 0) = 1 \tag{16}$$

$$p(s_1 + 1, 0|s_1, 0) = \frac{\lambda}{\lambda + \mu_1}, \tag{17}$$

$$p(s_1 - 1, 1|s_1, 0) = \frac{\mu_1}{\lambda + \mu_1}, \tag{18}$$

$$p(1, s_2|0, s_2) = \frac{\lambda}{\lambda + \mu_2}, \tag{19}$$

$$p(0, s_2 - 1|0, s_2) = \frac{\mu_2}{\lambda + \mu_2}, \tag{20}$$

$$p(s_1 + 1, s_2|s_1, s_2) = \lambda, \tag{21}$$

$$p(s_1 - 1, s_2 + 1|s_1, s_2) = \mu_1, \tag{22}$$

$$p(s_1, s_2 - 1|s_1, s_2) = \mu_2, \tag{23}$$

where $s_1 \geq 1$, $s_2 \geq 1$, and $s_1 + s_2 \leq N$.

4.2 SOIS Methods

In this subsection, we will first present an example to illustrate how to use the VM method to compute the optimal change of measure for a small N. We then present SOIS for large N.

If N is a small value, we can use the VM method in subsection 3.2 to compute the optimal change of measures. Now we use a simple 2 server tandem network as an example. Similar to the case of $M/M/1$ queues, the $M/M/1$ tandem 2-server queue can be represented by a trellis (see Figure 4). We have excised the path returning to the initial state. Now, according to the VM method, we trace recursively back from any state in F to minimize the cost $\omega_t^{(j)}$ subject to maximizing the one step transition probability $q_t^{(o)}(j|i)$. Figure 4 shows the numerical results for the queue with $\lambda = 0.05$, $\mu_1 = 0.5$, $\mu_2 = 0.45$, $N = 3$, and $L = 9$.

We can easily run the VM method for a large L until all optimal transition probabilities converge. When N is large, the VM method is limited due to the memory requirement to store all $\omega_t^{(i)}$ values. However, we found that for large N the optimal change of measure converges to a stable value rapidly. Therefore, based on the optimal change of measures computed for a reasonable N, we can establish a sub-optimal IS method for a large scale system.

4.2.1 Sub-Optimal IS Method (SOIS)

1. Establish the embedded Markov chain and compute the transition probabilities given in subsection 4.1.
2. Build a Monte Carlo simulation program for simulating the overflow probability based on the embedded Markov model.
3. Compute the optimal change of measures for a reasonably small value of N and run it for a large L until the values converge.
4. Project the optimal transition probabilities with a small N into those with a large N.
5. Run the program in Step 2 with the biased transition probabilities and the likelihood ratio.

Clearly, in SOIS we only need to add in Steps 3 and 4 in the Monte Carlo simulation. Now let us see how to apply SOIS to simulate the overflow probability for a Jackson network with 2 $M/M/1$ queues in tandem with $\lambda = 0.05$, $\mu_1 = 0.5$, $\mu_2 = 0.45$, $N = 30$.

In Step 3, we compute the optimal change of measures based on the same network except $N = 10$. For $L = 50$ the first 5 digits of all values after the decimal point have converged. It only took 0.03 second of CPU time on a 400MHZ PC. It took 19.73 seconds of CPU time using the same PC for the $L = 300$ and $N = 30$ case.

Figure 3 shows how to project the optimal change of measures with a small N (say $N_s = 10$) to those with a large N (say $N = 30$), where $T_s(s_1, s_2)$ denotes the transition probabilities from state (s_1, s_2) for the small system. There are many ways to project the optimal change of measures. Two options, called SOIS1 and SOIS2, are shown in Figure 3, where $K0 = \lfloor (N_s/2) \rfloor$, $K = \lfloor (N_s + 2 - s_2)/2 \rfloor$, $\lfloor x \rfloor$ represents the largest integer smaller than x.

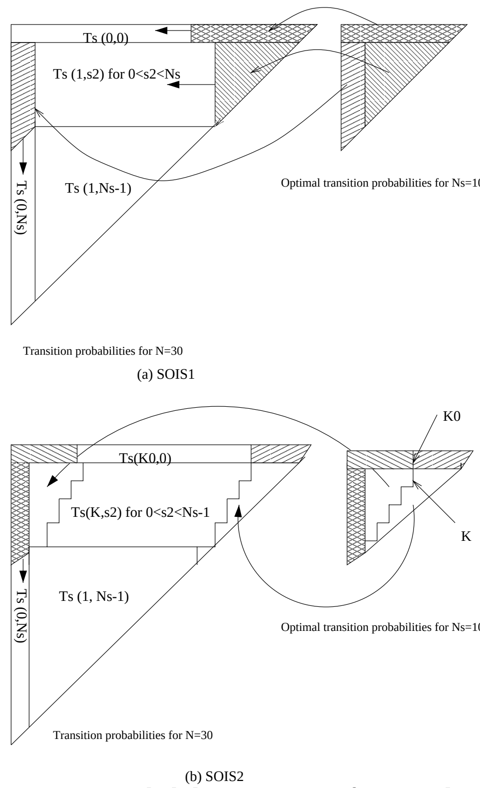


Figure 3: Probability Projection for a Tandem Network in SOIS

In Figure 5, we present several biased transition probabilities of SOIS1 (lines) based on the optimal case with $L = 50$ and $N = 10$. We also present the optimal case for $L = 300$ and $N = 30$ (points) for comparison. We focus on those near the boundary. The others converge to those given in Parekh and Walrand (1989), i.e., swapping the values between λ and μ_2. We found when λ is small, this method provides a good prediction of the optimal change of measures for a large scale system, i.e., SOIS1. However, when λ is large, the better way is to fill those places using the element in the middle of each row, i.e., SOIS2.

Clearly, the advantages of the SOIS methods are (a) it catches the boundary effect easily; (b) it is largely based on the Monte Carlo simulation; (c) it is easy to apply to many different systems. The only thing an engineer needs to do is to implement Step 3. As we will see in the next section, the efficiency of the SOIS methods is much better than the method in Parekh and Walrand (1989). The disadvantages

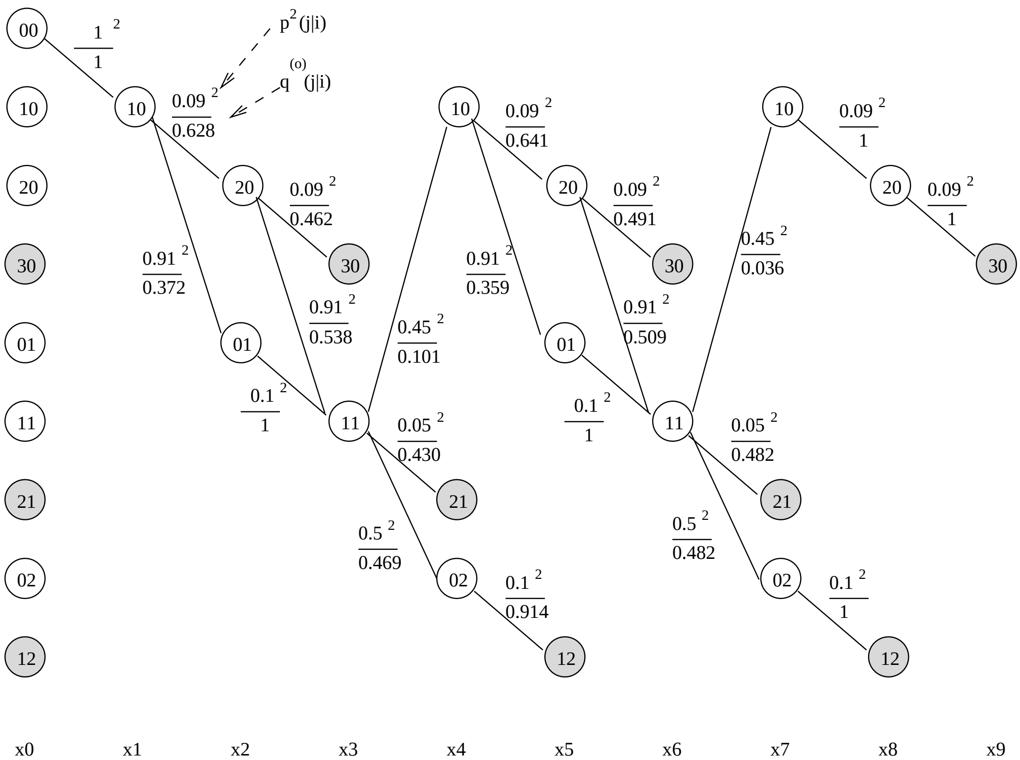


Figure 4: Optimal Change of Measures for an Two $M/M/1$ Tandem Network with $\lambda = 0.05$, $\mu_1 = 0.5$, $\mu_2 = 0.45$, $N = 3$, and $L = 9$

are (a) it is hard to apply for asymptotical analysis, while the LDT method is much easier; (b) it is limited to simulation using the embedded Markov model. The method in Parekh and Walrand (1989) can be easily used in the other types of simulations, for example, the simulation based on Possion exponential distributions. However, the simulation based on the embedded Markov model is much faster and simpler. So the SOIS method can be viewed as a complementary method for the method in Parekh and Walrand (1989).

5 NUMERICAL RESULTS

In this section, we will present numerical results for two types of Jackson networks. In the simulation, the normalized standard deviation is estimated by equation (6), rewritten here

$$E^{\#}(\alpha) = \sqrt{\frac{E_Q^{\#}[(V_k \cdot W_k)^2] - (\alpha^{\#})^2}{n\left(\alpha^{\#}\right)^2}} \tag{24}$$

where

$$\alpha^{\#} = \frac{V_1 \cdot W_1 + V_2 \cdot W_2 + \cdots + V_n \cdot W_n}{n}. \tag{25}$$

$$E_Q^{\#}[(V_k \cdot W_k)^2] = \frac{V_1 \cdot W_1^2 + V_2 \cdot W_2^2 + \cdots + V_n \cdot W_n^2}{n}. \tag{26}$$

For the MC estimator, we have $W_k = 1$. All simulations were produced by an IBM Thinkpad 240X notebook computer (500MHz). Here, MC denotes for the Monte Carlo method, VM denotes our variance minimization calculation, PW represents the method used in Parekh and Walrand (1989), SOIS1 and SOIS2 denote the proposed suboptimal method with different filling procedure, n denotes the total number of simulation trials, n_s denotes the number of trials which hit the target state, and N_s denotes the value of N in the small system.

For the VM method, we need to run it several times with different L values to make sure it has converged. Typically, we select $L = 10 \times N$, $L = 50 \times N$, and $L = 250 \times N$. For others, we set the required normalized standard deviation be 5%. The value of $E^{\#}(\alpha)$, is calculated every 500 simulation trials for MC and PW, every 100 simulation trials for SOIS. We select a small value for SOIS, since $E^{\#}(\alpha)$ for SOIS is much less sensitive to the seed values of the random number generator. The simulation is terminated for MC, PW, and SOIS methods, if the estimated value of $E^{\#}(\alpha)$ is below 5% for three random seeds or the total number of simulation trails reaches 10^7.

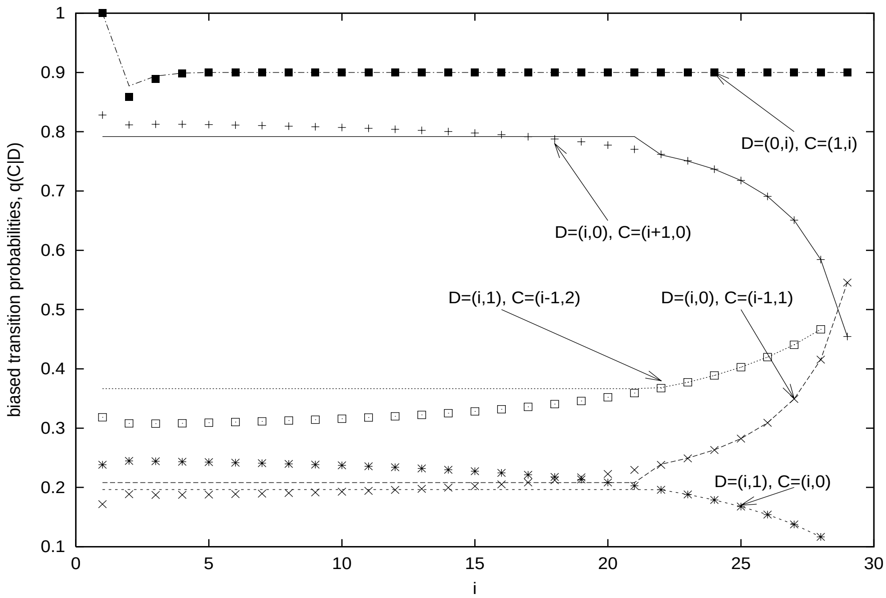


Figure 5: Transition Probability Comparison Between SOIS1 and Optimal Ones

In Table 1 we compare several simulation methods. We will focus on those cases which the PW method may not be able to deliver an efficient estimation result.

Firstly, let us study a tandem network with a small λ. In Table 1, we show the SOIS method can produce an accurate estimation within 3 seconds of CPU time. We also find the estimated standard deviation ($E^{\#}(\alpha)$) is very sensitive to the seed values used in the random number generator.

Table 1: Simulation Efficiency Comparison

Parameters	$\lambda = 0.05,\ \mu_1 = 0.5,\ \mu_2 = 0.45,$ $N = 30,\ N_s = 10$		
methods	PW	SOIS2	VM
n	1654500	1000	2
n_s	1556593	all	all
L	--	2500	300,1500
$E^{\#}(\alpha)$	4.7%	2.6%	0
CPU times	414.9s	0.91s	7.2s
α	1.72e-27	1.816e-27	1.806e-27
Parameters	$\lambda = 0.05,\ \mu_1 = 0.5,\ \mu_2 = 0.45,$ $N = 100,\ N_s = 10$		
methods	PW	SOIS2	VM
n	74000	2300	1
n_s	69616	all	all
L	--	2500	1000
$E^{\#}(\alpha)$	3.6%	3.6%	0
CPU times	64.7s	2.85s	563.3s
α	2.37e-94	3.17e-94	3.01e-94

We studied the above two difficult systems with very tight requirement (i.e., $E^{\#}(\alpha) < 5\%$). It is often unnecessary to have such a tight requirement. In Table 2, we show the results for simulations with a fixed and small value of n. We also show the results for several large scale systems with near one million states.

Table 2: Simulation Efficiency Comparison for Large State Space Cases

Parameters	$\lambda = 0.30,\ \mu_1 = 0.35$ $\mu_2 = 0.35$ $N = 300,\ N_s = 10$	
methods	PW	SOIS2
n	10000	300
n_s	4649	all
L	--	2500
$E^{\#}(\alpha)$	27.6%	37.8%
CPU times	114.6s	8.3s
α	1.48e-19	2.13e-19
Parameters	$\lambda = 0.05,\ \mu_1 = 0.5$ $\mu_2 = 0.45$ $N = 300,\ N_s = 10$	
methods	PW	SOIS2
n	10000	300
n_s		all
L	--	2500
$E^{\#}(\alpha)$	71.6%	4.7%
CPU times	28.7s	1.1s
α	1.12e-285	3.96e-285

6 CONCLUSION AND DISCUSSION

In this paper, we presented a suboptimal IS method (called SOIS) based on this variance minimization procedure. The key concept is to compute the optimal change of measures for a system with a small state space and then project these optimal values into a system with a large state space.

For all systems we examined, SOIS performs much better than all other methods in terms of delivering an accurate estimation using moderate computation effort. Currently, we are extending this method for othertypes of applications, such as QoS in inventory systems, IS for light-intensity analysis in physics, and software tests in software development.

ACKNOWLEDGMENTS

The authors would like to thank STRICOM for partly supporting this work through the grant no. 16-23-504.

REFERENCES

Cottrell, M. 1983. Large deviations and rare events in the study of stochastic algorithms. *IEEE Trans Automatic. Control*, AC-28 (9): 907-920.

Devetsikiotis, M., and J. K. Townsend. 1993. Statistical optimization of dynamic importance sampling parameters for efficient simulation of communication networks *IEEE/ACM Trans on Networking* 1 (3): 293–305.

Glasseman, P., and S. Kou. 1995. Analysis of an Importance Sampling Estimator for tandem Queue. *IEEE/ACM Trans. On Modeling and Computer Simulation* 3 (1): 22-42.

Heidelberger, P. 1995. Fast Simulation of rare event in queuing and Reality Models. *IEEE/ACM Trans. on Modeling and Computer Simulation* 5 (1): 43-85.

Orsak, G., and B. Aazhang. 1984. Techniques for estimating the bit error rate in the simulation of digital communication systems *IEEE JSAC*, 2, (1): 153–170.

Parekh, S., and J. Walrand. 1989. A Quick Simulation Method for Excessive Backlogs in Networks of Queues. *IEEE Trans. on Autom. Control* 34 (1): 54-66.

Sadowsky, J. S. 1990. A new method for Veberti decoder simulation using importance sampling. *IEEE Trans. on Communications* COM-38, (9): 1341–1351.

Shanmugam, K. S., and P. Balaban. 1980. A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems. *IEEE Trans. on Communications*, COM-28 (11): 1916-1924.

Smith, P. J., H. Shafi, and H. Gao. 1997. Quick simulation: a review of importance sampling techniques in communications systems *IEEE JSAC*, 15, (5): 597–613.

Wei, L. 1995a. Estimated bit error probability of DS-SSMA/MPSK with coherent detector on satellite mobile channel. *IEEE JSAC*, 13 (2): 250–263.

Wei, L. 1995b. Efficiency of optimal conditional importance sampling. *Electronics Letters*, (August): 1322-1323.

Wei, L. and C. Schlegel. 1995. Synchronization requirements for multi-user OFDM on satellite mobile and two-path rayleigh channels. *IEEE Trans. on Communications*, COM-43, (Feb./Mar./Apr.): 887–895.

Wei L. 1996. Estimated bit error probability of DS-SSMA/MDPSK with differential phase detector on satellite mobile channel. *IEE Proceeding I, (communications)*, 143 (10): 297–304.

Wei, Y. C., and L. Wei. 2001 Construction of fast recovery codes using a new optimal importance sampling method, *IEEE Trans. Information Theory*, 47 (7): 3006-3019.

AUTHOR BIOGRAPHIES

LEI WEI is an Associate Professor at School of Electrical Engineering and Computer Science, University of Central Florida. He is currently serving as an Editor for wireless CDMA systems for IEEE Transaction on Communications. His research interests include wireless communications, error control coding, importance sampling and fast simulation techniques, and ultra wide band systems. He is a member of IEEE. His e-mail address is `<lei@ee.ucf.edu>`.

HONGHUI QI is a Visiting Assistant Professor at SEECS, University of Central Florida. She was an assistant professor at the Australian National University and University of Wollongong. Her research interests include multiple access for wireless systems, Markov analysis for system/network modeling, fast simulation, and Internet.