# EFFICIENT HEURISTICS FOR SIMULATING POPULATION OVERFLOW IN FEED-FORWARD NETWORKS

Victor F. Nicola
Tatiana S. Zaburnenko

Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, P.O. Box 217
7500 AE Enschede, THE NETHERLANDS

In this paper we propose a state-dependent importance sampling heuristic to estimate the probability of population overflow in feed-forward networks. This heuristic attempts to approximate the "optimal" state-dependent change of measure without the need for difficult analysis or costly optimization involved in other recently proposed adaptive importance sampling algorithms. Preliminary simulation experiments with a 4-node feed-forward network yield asymptotically efficient estimates, with relative error increasing at most linearly in the overflow level, where state-independent importance sampling is demonstrably ineffective.

## 1 INTRODUCTION

Importance sampling is a very effective methodolgry for the efficient simulation of queueing systems and networks involving rare events (see, e.g., Parekh and Walrand 1989, Asmussen and Rubinstein (1995) Heidelberger 1995, Juneja and Nicola 2005). Until recently, only state-independent importance sampling heuristics were developed and considered for analysis. In these heuristics, the change of measure is "static" and independent of the network state (i.e., the number of customers at each node in a Jackson network). A relatively simple (and well known) heuristic change of measure for simulations of population overflow in queueing networks is that proposed in Parekh and Walrand (1989). However, even for the simplest Jackson queueing network (e.g., 2-nodes in series or in parallel), the effectiveness of this heuristic is limited to only some region of the (arrival and service) parameters space (see Glasserman and Kou 1995, de Boer 2004). (We use the term "effectiveness" interchangeably with "asymptotic efficiency," see, e.g., Heidelberger (1995) for a precise definition.)

Recent theoretical and empirical studies (see, e.g., Kroese and Nicola 2002, de Boer and Nicola 2002) reveal that state-dependent change of measures are generally more effective, also where no effective state-independent change of measure exists. In de Boer and Nicola (2002) an adaptive optimization technique based on the method of cross-entropy (Rubinstein 2002) is used to approximate the "optimal" state-dependent change of measure. A drawback of this approach, however, is the excessive computational and storage demands for large state-space models associated with large networks. In Nicola and Zaburnenko (2005a) and Nicola and Zaburnenko (2005b), heuristics are proposed to approximate the "optimal" state-dependent change of measure without the need for a costly optimization. The key observation is that the "optimal" change of measure depends on the network state only along and close to the boundaries (when one or more nodes are empty), and tends to become state-independent in the interior of the state-space. Therefore, if we can determine the change of measure along the boundaries and at the interior of the state-space, then we may be able to combine them appropriately to construct a state-dependent change of measure that approximates the "optimal" one in the entire state-space. The proposed methodology is dubbed "state-dependent heuristic" or SDH in short. The proposed heuristics are effective, easy to implement and could be more efficient than those based on adaptive importance sampling methodologies (e.g., de Boer and Nicola 2002), particularly for large networks. Experimental results for tandem and parallel networks with multiple nodes yield asymptotically efficient estimates, mostly with a bounded relative error (see Nicola and Zaburnenko 2005a, Nicola and Zaburnenko 2006a). In Nicola and Zaburnenko (2005b), a heuristic was also proposed for a feed-forward network; the resulting estimates were correct but not convincingly robust (the relative errors were somewhat high and the effectiveness varied with network parameters).

In this paper we develop a robust state-dependent change of measure for the efficient simulation of rare events in feed-forward networks. The heuristic builds on recently developed and demonstrably effective heuristics for tandem and parallel networks (see Nicola and Zaburnenko 2005a, Nicola and Zaburnenko 2006a).

In Section 2 we introduce the model and notation. In Section 3 we motivate and outline the SDH for feed-forward networks. In Section 4 we present experimental results for the estimation of the probability of network population overflow. Comparisons with the well-known heuristic in Parekh and Walrand (1989) and with the adaptive importance sampling methodology in de Boer and Nicola (2002) are also presented. We conclude in Section refsec:con.

## 2 MODEL AND NOTATION

Consider a Jackson network consisting of $n$ nodes (queues), each having its own buffer of infinite size. Customers arrive at node $i$ $(i = 1, \ldots, n)$ according to a Poisson process with rate $\lambda_i$. The service time of a customer at node $i$ is exponentially distributed with rate $\mu_i$ $(i = 1, \ldots, n)$. Customers that leave node $i$ join node $j$ with probability $p_{ij}$ $(i = 1, \ldots, n; j = 1, \ldots, n)$ or leave the network with probability $p_{ie}$ $(i = 1, \ldots, n)$. We also assume that the queueing network is stable, i.e., $\gamma_i < \mu_i$ for all $i = 1, \ldots, n$, where $\gamma_i$ is the total arrival rate at node $i$, as determined from the traffic equations $\gamma_i = \lambda_i + \sum_{\forall j} \gamma_j \, p_{ji}$.

Let $X_{i,t}$ $(i = 1, \ldots, n)$ denote the number of customers at node $i$ at time $t \geqslant 0$ (including those in service). Then the vector $\mathbf{X_t} = (X_{1,t}, X_{2,t}, ..., X_{n,t})$ is a Markov process representing the state of the network at time $t$. Denote by $S_t$ the total number of customers in the network (network population) at time $t$, i.e., $S_t = \sum_{i=1}^{n} X_{i,t}$.

Assuming that the initial network state is $\mathbf{X_0}$ (usually, $\mathbf{X_0} = (0, 0, ..., 0)$ corresponding to an empty network), we are interested in the probability that the network population reaches some high level $L \in \mathbb{N}$ before becoming empty. We denote this probability by $\gamma(L)$ and refer to it as the *population overflow probability*, starting from the initial state $\mathbf{X_0}$. Since the associated event is typically rare, importance sampling may be used to efficiently estimate this probability (for a review see, e.g., Heidelberger 1995).

Starting from $\mathbf{X}_0$, define $\tau$ as the first time $S_t$ hits level $L$ or level 0, then

$$\gamma(L) \, = \, \mathbb{E} \, I_{\{S_\tau = L\}} \, = \, \tilde{\mathbb{E}} \, W_\tau \, I_{\{S_\tau = L\}} \,, \tag{1}$$

where $W_\tau$ is the likelihood ratio over the interval $[0, \tau]$; $\mathbb{E}$ and $\tilde{\mathbb{E}}$ are the expectations under the original and the new change of measures, respectively. The relative error is the ratio of the standard deviation of the estimator over its expectation, i.e.,

$$\sqrt{\frac{\tilde{\mathbb{E}}\, {W_\tau}^2\, I_{\{S_\tau = L\}}}{\gamma(L)^2} - 1}\,. \tag{2}$$

The estimator $\tilde{\mathbb{E}}\, W_\tau\, I_{\{S_\tau=L\}}$ is said to be *asymptotically efficient* if its relative error grows at sub-exponential (e.g., polynomial) rate as $L \to \infty$ (i.e., as $\gamma(L) \to 0$). The estimator is said to have *bounded relative error* if its relative error is bounded in $L$ as $\gamma(L) \to 0$. It is important to note that a change of measure may, in general, depend on the state of the system, even if the original underlying distributions do not depend on the system state.

# 3 STATE-DEPENDENT HEURISTIC FOR A FEED-FORWARD NETWORK

Recent theoretical and empirical studies (e.g., Kroese and Nicola 2002, de Boer and Nicola 2002) indicate that the "optimal" change of measure depends on the network state, i.e., the number of customers at each network node. Furthermore, this crucial dependence is strong along the boundaries of the state-space (i.e., when one or more buffers are empty) and diminishes in the interior of the state-space (i.e., when contents of all buffers are sufficiently large). This observation suggests that if we know the "optimal" change of measure along the boundaries and in the interior of the state-space, then we might be able to construct a change of measure that approximates the "optimal" one over the entire state-space. Recently, heuristics based on combining known large deviations results and time-reversal arguments are used to construct such a change of measure for tandem networks (Nicola and Zaburnenko 2005a) and for parallel networks (Nicola and Zaburnenko 2006a). Empirical results show that these heuristics produce asymptotically efficient estimates, mostly with a bounded relative error. These heuristics also form the ingredients for a state-dependent change of measure to efficiently simulate feed-forward networks that is presented next.

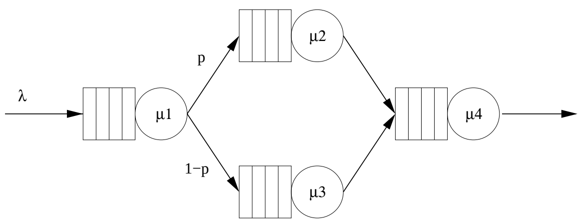


Figure 1: 4-Node Feed-forward Network

To describe our state-dependent heuristic for feed-forward Jackson networks, we use the specific example depicted in Figure 1. Without loss of generality we assume that $\lambda + \sum_{i=1}^{4} \mu_i = 1$. The traffic intensity at Node $i$ is $\rho_i = \gamma_i/\mu_i$, where $\gamma_i$ is the total arrival rate at Node $i$ ($i = 1, 2, 3, 4$). We also assume that $\rho_1 \leqslant \rho_2 \leqslant \rho_3 \leqslant \rho_4$. Let $\mathbf{\Theta}^{\mathrm{T}} = [\lambda, \mu_1, \mu_2, \mu_3, \mu_4, p]$ be a vector with the arrival rate, the service rates at Nodes 1, 2, 3, 4, and the routing probability, respectively. Now consider splitting the parameter vector $\mathbf{\Theta}$ into two vectors $\mathbf{\Theta}^{\mathcal{T}2}$ and $\mathbf{\Theta}^{\mathcal{T}3}$, corresponding to its two input-output paths, $\mathcal{T}2$ and $\mathcal{T}3$ (see Figure 2). Parameters shared between the two paths (namely, the arrival rate $\lambda$, the service rates $\mu_1$ and $\mu_4$, and the routing probability $p$) are allocated to $\mathcal{T}2$

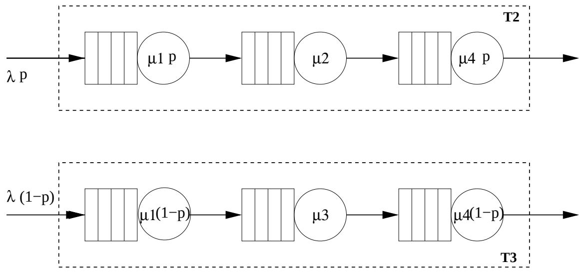


Figure 2: Splitting of the Feed-forward Network into Two Tandem Networks $\mathcal{T}2$ and $\mathcal{T}3$

and $\mathcal{T}3$ proportional to $p$ and $(1-p)$, respectively. Specifically,

$$\mathbf{\Theta}^{\mathcal{T}2} = \left[\lambda p, \mu_1 p, \mu_2, 0, \mu_4 p, p^2\right]$$

is allocated to Path $\mathcal{T}2$ (through Nodes 1, 2 and 4), with the entry corresponding to Node 3 set to zero. Similarly,

$$\mathbf{\Theta}^{\mathcal{T}3} = [\lambda(1-p), \mu_1(1-p), 0, \mu_3, \mu_4(1-p), p(1-p)]$$

is allocated to Path $\mathcal{T}3$ (through Nodes 1, 3 and 4), with the entry corresponding to Node 2 set to zero. (Note that $\mathbf{\Theta}^{\mathcal{T}2} + \mathbf{\Theta}^{\mathcal{T}3} = \mathbf{\Theta}$.)

We denote the state-independent change of measure to overflow Node $i$ ($i = 1, 2, 3, 4$) in the feed-forward network by $\tilde{\mathbf{\Theta}}_i$ (as determined in Juneja and Nicola (2005)). The vector $\tilde{\mathbf{\Theta}}_i$ is also split into two vectors $\tilde{\mathbf{\Theta}}_i^{\mathcal{T}2}$ and $\tilde{\mathbf{\Theta}}_i^{\mathcal{T}3}$, corresponding to the input-output paths, $\mathcal{T}2$ and $\mathcal{T}3$, respectively. Specifically, since $\mu_4 p \leqslant \mu_2 \leqslant \mu_1 p$ (from $\rho_4 \geqslant \rho_2 \geqslant \rho_1$), $\tilde{\mathbf{\Theta}}_i^{\mathcal{T}2}$ simply interchanges $\lambda p$ with the service rate allocated to Path $\mathcal{T}2$ at Node $i$ ($i = 1, 2, 4$), with the entry corresponding to Node 3 set to zero. Similarly, since $\mu_4(1-p) \leqslant \mu_3 \leqslant \mu_1(1-p)$ (from $\rho_4 \geqslant \rho_3 \geqslant \rho_1$), $\tilde{\mathbf{\Theta}}_i^{\mathcal{T}3}$ simply interchanges $\lambda(1-p)$ with the service rate allocated to Path $\mathcal{T}3$ at Node $i$ ($i = 1, 3, 4$). with the entry corresponding to Node 2 set to zero. It follows that the vector $\tilde{\mathbf{\Theta}}_1 = [\mu_1, \lambda, \mu_2, \mu_3, \mu_4, p]$ and is split into:

$$\tilde{\mathbf{\Theta}}_1^{\mathcal{T}2} = \left[\mu_1 p, \lambda p, \mu_2, 0, \mu_4 p, p^2\right],$$

$$\tilde{\mathbf{\Theta}}_1^{\mathcal{T}3} = [\mu_1(1-p), \lambda(1-p), 0, \mu_3, \mu_4(1-p), p(1-p)].$$

The vector $\tilde{\mathbf{\Theta}}_2 = [\mu_2 + \lambda(1-p), \mu_1, \lambda p, \mu_3, \mu_4, \mu_2/(\mu_2 + \lambda(1-p))]$ and is split into:

$$\tilde{\mathbf{\Theta}}_2^{\mathcal{T}2} = [\mu_2, \mu_1 p, \lambda p, 0, \mu_4 p, (\mu_2/(\mu_2 + \lambda(1-p)))p],$$

$$\tilde{\mathbf{\Theta}}_2^{\mathcal{T}3} = [\lambda(1-p), \mu_1(1-p), 0, \mu_3, \mu_4(1-p), (\mu_2/(\mu_2 + \lambda(1-p)))(1-p)].$$

The vector $\tilde{\mathbf{\Theta}}_3 = [\mu_3 + \lambda p, \mu_1, \mu_2, \lambda(1-p), \mu_4, \lambda p/(\lambda p + \mu_3)]$ and is split into:

$$\tilde{\mathbf{\Theta}}_3^{\mathcal{T}2} = [\lambda p, \mu_1 p, \mu_2, 0, \mu_4 p, (\lambda p/(\lambda p + \mu_3))p],$$

$$\tilde{\mathbf{\Theta}}_3^{\mathcal{T}3} = [\mu_3, \mu_1(1-p), 0, \lambda(1-p), \mu_4(1-p), (\lambda p/(\lambda p + \mu_3))(1-p)].$$

The vector $\tilde{\mathbf{\Theta}}_4 = [\mu_4, \mu_1, \mu_2, \mu_3, \lambda, p]$ and is split into:

$$\tilde{\mathbf{\Theta}}_4^{\mathcal{T}2} = \left[\mu_4 p, \mu_1 p, \mu_2, 0, \lambda p, p^2\right],$$

$$\tilde{\mathbf{\Theta}}_4^{\mathcal{T}3} = [\mu_4(1-p), \mu_1(1-p), 0, \mu_3, \lambda(1-p), p(1-p)].$$

The state-independent change of measure to simultaneously overflow Nodes 2 and 3 in the feed-forward network is denoted by $\tilde{\mathbf{\Theta}}_{23}$ (as determined in Nicola and Zaburnenko (2006b)). It follows that $\tilde{\mathbf{\Theta}}_{23} = [\mu_2 + \mu_3 + \mu_1, \lambda p, \lambda(1-p), \mu_4, \mu_2/(\mu_2+\mu_3)]$. It is also split into two vectors $\tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}2}$ and $\tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}3}$, corresponding to the input-output paths, $\mathcal{T}2$ and $\mathcal{T}3$, respectively, as follows:

$$\tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}2} = [\mu_2, \mu_1 p, \lambda p, 0, \mu_4 p, (\mu_2/(\mu_2+\mu_3))p]\,,$$

$$\tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}3} = [\mu_3, \mu_1(1-p), 0, \lambda(1-p), \mu_4(1-p), (\mu_2/(\mu_2+\mu_3))(1-p)]\,.$$

With the preceding definitions, the heuristic state-dependent change of measure for the feed-forward network in Figure 1 can now be given in the following proposition.

**Proposition 1 (SDH for a Feed-forward Network)**
Let $\tilde{\mathbf{\Theta}}^{\mathrm{T}}(\mathbf{x}) = \left[\tilde{\lambda}, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3, \tilde{\mu}_4, \tilde{p}\right]$ be a vector with the corresponding state-dependent arrival and service rates at the respective nodes as well as the routing probability under the new change of measure to simulate network population overflow. Then

$$\begin{aligned}
\tilde{\mathbf{\Theta}}(\mathbf{x}) = & \left[\frac{x_4}{b_4}\right]^1 \tilde{\mathbf{\Theta}}_4^{\mathcal{T}2} + \left[\frac{b_4-x_4}{b_4}\right]^+ \times \\
& \{ \left[\frac{x_2}{b_2}\right]^1 \left\{ \left[\frac{x_3}{b_3}\right]^1 \tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}2} + \left[\frac{b_3-x_3}{b_3}\right]^+ \tilde{\mathbf{\Theta}}_2^{\mathcal{T}2} \right\} + \left[\frac{b_2-x_2}{b_2}\right]^+ \times \\
& \{ \left[\frac{x_3}{b_3}\right]^1 \tilde{\mathbf{\Theta}}_3^{\mathcal{T}2} + \left[\frac{b_3-x_3}{b_3}\right]^+ \times \\
& \left[\frac{x_1}{b_1}\right]^1 \tilde{\mathbf{\Theta}}_1^{\mathcal{T}2} + \left[\frac{b_1-x_1}{b_1}\right]^+ \mathbf{\Theta}^{\mathcal{T}2} \} \} \} \\
& + \left[\frac{x_4}{b_4}\right]^1 \tilde{\mathbf{\Theta}}_4^{\mathcal{T}3} + \left[\frac{b_4-x_4}{b_4}\right]^+ \times \\
& \{ \left[\frac{x_3}{b_3}\right]^1 \left\{ \left[\frac{x_2}{b_2}\right]^1 \tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}3} + \left[\frac{b_2-x_2}{b_2}\right]^+ \tilde{\mathbf{\Theta}}_3^{\mathcal{T}3} \right\} + \left[\frac{b_3-x_3}{b_3}\right]^+ \times \\
& \{ \left[\frac{x_2}{b_2}\right]^1 \tilde{\mathbf{\Theta}}_2^{\mathcal{T}3} + \left[\frac{b_2-x_2}{b_2}\right]^+ \times \\
& \left[\frac{x_1}{b_1}\right]^1 \tilde{\mathbf{\Theta}}_1^{\mathcal{T}3} + \left[\frac{b_1-x_1}{b_1}\right]^+ \mathbf{\Theta}^{\mathcal{T}3} \} \} \}\,.
\end{aligned} \tag{3}$$

Note that all vectors on the r.h.s. of Equation 3 are state-independent. However, the new parameters to simulate the network under importance sampling ($\tilde{\lambda}(\mathbf{x})$, $\tilde{\mu}_i(\mathbf{x}), i = 1,2,3,4$, and $\tilde{p}(\mathbf{x})$) are state-dependent. Moreover, the equality $\sum_{i=1}^{n} (\tilde{\lambda}_i(\mathbf{x}) + \tilde{\mu}_{\mathbf{i}}(\mathbf{x})) = 1$ still holds under the above change of measure.

According to the above change of measure, the two tandem networks $\mathcal{T}2$ and $\mathcal{T}3$ are overloaded simultaneously, depending on the buffer contents at their respective nodes. Similar to a 3-node tandem network (see Nicola and Zaburnenko 2005a), the state-independent change of measure $\tilde{\mathbf{\Theta}}_i^{\mathcal{T}2}$ (resp. $\tilde{\mathbf{\Theta}}_i^{\mathcal{T}3}$) is nested in the same order of Node $i$'s utilization in $\mathcal{T}2$ (resp. $\mathcal{T}3$). Since $\rho_1 \leqslant \rho_2 \leqslant \rho_3 \leqslant \rho_4$, dependence on $x_4$ supersedes dependence on $x_3$ and $x_2$ which supersedes dependence on $x_1$. The above change of measure implies that, upon arrival to an empty network, Node 1 is gradually overloaded according to $[x_1/b_1]^1\, \tilde{\mathbf{\Theta}}_1$ (i.e., gradually interchange $\lambda$ and $\mu_1$ depending on $x_1$). As $x_2$ (resp. $x_3$) increases, Node 1 is gradually downloaded while Node 2 (resp. Node 3) is gradually overloaded according to $[x_2/b_2]^1\, \tilde{\mathbf{\Theta}}_2^{\mathcal{T}2}$ (resp. $[x_3/b_3]^1\, \tilde{\mathbf{\Theta}}_3^{\mathcal{T}3}$). As $x_4$ increases, Node 2 (resp. Node 3) is gradually downloaded while Node 4 is gradually overloaded according to $[x_4/b_4]^1\, \tilde{\mathbf{\Theta}}_4$. The choice of the variables $b_i$ (the dependence range at Node $i$, $i = 1, \ldots, n$) is crucial for the effectiveness of the heuristic.

# 4 EXPERIMENTAL RESULTS

Importance sampling to estimate the probability of population overflow ($\gamma(L)$) involves generating, say, $N$, independent and identically distributed (i.i.d.) busy cycles (i.e., starting with an arrival to an empty network). Starting a cycle at time 0, define $\tau_L$ as the instant when the network population reaches level $L$ for the first time. Similarly, define $\tau_0$ as the instant when the network population returns to 0 for the first time. The indicator function $I_i(\tau_L < \tau_0)$ takes the value 1 if the population overflow (level $L$) is reached in cycle $i$, otherwise it takes the value 0.

In each cycle, the change of measure is applied until either the population overflow event is reached or the network population returns to 0. Let $W_i$ be the likelihood ratio associated with cycle $i$, then an unbiased estimator $\tilde{\gamma}$ of $\gamma(L)$ is given by $\tilde{\gamma} = (1/N)\sum_{i=1}^{i=N} I_i\, W_i$. The second moment of $IW$ is estimated by $\tilde{\gamma^2} = (1/N)\sum_{i=1}^{i=N} I_i\, W_i{}^2$. The variance and the relative error of the importance sampling estimator $\tilde{\gamma}$ are given by $\mathrm{VAR}(\tilde{\gamma}) = (\tilde{\gamma^2} - (\tilde{\gamma})^2)\,/\,(N-1)$ and $\mathrm{RE}(\tilde{\gamma}) = \sqrt{\mathrm{VAR}(\tilde{\gamma})}\,/\,\tilde{\gamma}$, respectively. Another useful measure for comparing the efficiency of different estimators is the "relative time variance" (RTV) product, which is defined as the simulation time (in seconds) multiplied by the squared relative error of the estimator. As the estimate becomes more stable, its RTV tends to a constant value, which is smaller for a more efficient estimator. For example, if $\mathrm{RTV}_2$ (for Estimator 2) is larger than $\mathrm{RTV}_1$ (for Estimator 1), then it will take Estimator 2 a longer simulation time to reach the same accuracy. For efficiency comparisons we use the variance reduction ratio, $\mathrm{VRR} = \mathrm{RTV}_2\,/\,\mathrm{RTV}_1$, which represents the efficiency gain when using Estimator 1 relative to that when using Estimator 2.

In this section we present experimental results on the feed-forward network depicted in Figure 1. These results are obtained using the following somewhat simpler and less refined version of the heuristic given in Section 3 (here the r.h.s. does not include $\tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}2}$ and $\tilde{\mathbf{\Theta}}_{23}^{\mathcal{T}3}$, which constitute a change of measure that simultaneously overload Nodes 2 and 3):

$$
\begin{aligned}
\tilde{\mathbf{\Theta}}(\mathbf{x}) = {} & \left[\frac{x_4}{b_4}\right]^1 \tilde{\mathbf{\Theta}}_4^{\mathcal{T}2} + \left[\frac{b_4 - x_4}{b_4}\right]^+ \times \\
& \{ \left[\frac{x_2}{b_2}\right]^1 \tilde{\mathbf{\Theta}}_2^{\mathcal{T}2} + \left[\frac{b_2 - x_2}{b_2}\right]^+ \times \\
& \{ \left[\frac{x_1}{b_1}\right]^1 \tilde{\mathbf{\Theta}}_1^{\mathcal{T}2} + \left[\frac{b_1 - x_1}{b_1}\right]^+ \mathbf{\Theta}^{\mathcal{T}2} \} \} \\
& + \left[\frac{x_4}{b_4}\right]^1 \tilde{\mathbf{\Theta}}_4^{\mathcal{T}3} + \left[\frac{b_4 - x_4}{b_4}\right]^+ \times \\
& \{ \left[\frac{x_3}{b_3}\right]^1 \tilde{\mathbf{\Theta}}_3^{\mathcal{T}3} + \left[\frac{b_3 - x_3}{b_3}\right]^+ \times \\
& \{ \left[\frac{x_1}{b_1}\right]^1 \tilde{\mathbf{\Theta}}_1^{\mathcal{T}3} + \left[\frac{b_1 - x_1}{b_1}\right]^+ \mathbf{\Theta}^{\mathcal{T}3} \} \}\,.
\end{aligned}
\tag{4}
$$

We consider Four sets of network parameters for which the well-known heuristic in Parekh and Walrand (1989)is shown to be ineffective (this is verified empirically by showing that PW yields wrong or unstable estimates). In each experiment (for a given set of network parameters), the above variant of the heuristic (termed SDH) is used to obtain estimates of the population overflow probability $\gamma(L)$ for different overflow levels. For the purpose of comparison and/or verification, we include estimates obtained using the adaptive importance sampling methodology in de Boer and Nicola (2002) (termed SDA) as well as the known heuristic in Parekh and Walrand (1989) (termed PW), although the latter (PW) estimates are not necessarily accurate or stable. Each estimate is obtained from one simulation run of $10^6$ replications (same for all heuristics and all parameter points). These estimates and the associated relative errors (in percentage) are displayed in tables; one table for each experiment. For the SDH estimates, we also include the ratio VRR (relative to SDA); hence, VRR $> 1$ implies efficiency gain of SDH over SDA. In general, numerical results are difficult to obtain for larger and/or higher overflow levels (i.e., for larger sate-space). In the absence of numerical results (as marked by "$*$" in the corresponding table entry), agreement of the SDH and SDA estimators may be an indication of correctness.

Both SDH and SDA assume state-dependence only over a (small) number of boundary layers (say, $b_i$ along the state variable $x_i, i = 1, \ldots, n$) which must be properly determined to ensure the effectiveness and efficiency of these methods. Unless stated otherwise, all $b_i$s along the different boundaries are set equal to some $b$. All results presented in this section are obtained using the corresponding "optimal" $b$, which may differ for SDH and SDA estimates. For consistency with the assumption made in Section 3, we also choose the network parameters such that $\rho_1 \leqslant \rho_2 \leqslant \rho_3 \leqslant \rho_4$.
In the first set, $\lambda = 0.0455, \mu_1 = 0.7272, \mu_2 = 0.0455, \mu_3 = \mu_4 = 0.0909, p = 0.1$. The corresponding node utilizations are $\rho_1 = 0.06, \rho_2 = 0.1, \rho_3 = 0.45, \rho_4 = 0.5$.
In the second set, $\lambda = 0.064, \mu_1 = 0.564, \mu_2 = 0.039, \mu_3 = 0.192, \mu_4 = 0.141, p = 0.1$. The corresponding node utilizations are $\rho_1 = 0.11, \rho_2 = 0.16, \rho_3 = 0.3, \rho_4 = 0.45$.
In the third set, $\lambda = 0.069, \mu_1 = 0.571, \mu_2 = 0.022, \mu_3 = 0.198, \mu_4 = 0.14, p = 0.1$. The corresponding node utilizations are $\rho_1 = 0.12, \rho_2 = \rho_3 = 0.31, \rho_4 = 0.49$ (the two parallel nodes have equal utilizations).
In the fourth set, $\lambda = 0.074, \mu_1 = 0.617, \mu_2 = 0.024, \mu_3 = 0.135, \mu_4 = 0.15, p = 0.1$. The corresponding node utilizations are $\rho_1 = 0.12, \rho_2 = 0.31, \rho_3 = \rho_4 = 0.49$ (Nodes 3 and 4 have equal utilizations).

Experimental results in Tables 1, 2, 3 and 4 show that SDH (using the above less refined variant of the heuristic proposed in Section 3) works very well and yields stable estimates with small (and seemingly bounded) relative errors. (The results using the more refined heuristic in Section 3 were not completed at the time of writing this paper.) Exact values of the true probabilities (being estimated) are not feasible with the numerical algorithms available to us. However, correctness is verified by agreement of the SDH and SDA estimates, both are typically accurate with small relative errors. VRR ratios are mostly higher (or much higher) than one, indicating that SDH could be more efficient than SDA. Estimates using PW do not always agree with those using SDH and SDA. In fact, for an increasing number of replications (beyond $10^6$ used in all simulation runs), PW estimates eventually exhibit unstable behaviour symptomatic of infinite variance.

Generally, the key to any effective heuristic is its ability to induce a sample path behaviour sufficiently close to those most likely to hit the rare event - the closer the better! The heuristic in this paper seems to do just that for the relatively simple feed-forward network in Figure 1. Also, it seems to be robust despite the convenient choice of equal dependence range ($b$) at different nodes. Other heuristics may be developed that could also prove effective for this and other feed-forward networks of various topologies.

# 5 CONCLUSIONS

In this paper we have proposed and experimented with a heuristic state-dependent change of measure to estimate (via importance sampling) the probability of population overflow in feed-forward networks. Preliminary experimental results indicate that the heuristic yields asymptotically efficient estimates, with relative error growing at most (sub-)linearly with the overflow level. The efficiency of the obtained changes of measure compares well with those determined using adaptive importance sampling methodologies. Yet, our approach does not require costly pre-computation and avoids complicated (and often intractable) mathematical analyses. Moreover, its effectiveness is not diminished for larger networks. The applicability of the methodology to other networks of more complex topologies awaits further testing.

# References

[1] 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. Dshalalow, 429–461. CRC Press, New York.

[2] de Boer, P.T., and V.F. Nicola. 2002. Adaptive state-dependent importance sampling simulation of Markovian queueing networks. *European Trans. on Telecommunications* 13 (4): 303–315.

[3] de Boer, P.T. 2004. Analysis of sate-independent IS measures for the two-node tandem queue. *International Workshop on Rare Event Simulation (RESIM'04),* Budapest, Hungary.

[4] Glasserman, P., and S-G. Kou. 1995. Analysis of an importance sampling estimator for tandem queues. *ACM Trans. on Modeling and Computer Simulation* 5 (1): 22–42.

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

[6] Juneja, S.K., and V.F. Nicola. 2005. Efficient simulation of buffer overflow probabilities in Jackson networks with feedback. *ACM Trans. on Modeling and Computer Simulation* 15 (4): 281–315.

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

[8] Nicola, V.F., and T.S. Zaburnenko. 2005a. Importance sampling simulation of population overflow in two-node tandem networks. In *Proc. of the 2nd International Conference on the Quantitative Evaluation of Systems (QEST'05),* Torino, Italy, 220–229.

[9] Nicola, V.F., and T.S. Zaburnenko. 2005b. Efficient Importance Sampling Heuristics for the Simulation of Population Overflow in Jackson Networks. In *Proc. of the 2005 Winter Simulation Conference (WSC'05),* Orlando, Florida, pp. 538-546

[10] Nicola, V.F., and T.S. Zaburnenko. 2006a. Efficient Heuristics for the Simulation of Buffer Overflow in Series and Parallel Queueing Networks. In *Proc. of the First International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS'2006),* Pisa, Italy.

[11] Nicola, V.F., and T.S. Zaburnenko. 2006b. Efficient Simulation of Population Overflow in Parallel Queues. In *Proc. of the 2006 Winter Simulation Conference (WSC'06),* Monterey, California.

[12] Parekh, S., and J. Walrand. 1989. A quick simulation method for excessive backlogs in networks of queues. *IEEE Trans. on Automatic Control* 34: 54–66.

[13] Rubinstein, R.Y. 2002. The cross-entropy method and rare events for maximal cut and bipartition problems. *ACM Trans. on Modeling and Computer Simulation* 12 (1): 27–53.

[14] Zaburnenko, T.S., and V.F. Nicola. 2005. Efficient heuristics for simulating population overflow in tandem networks. In *Proc. of the 5th St. Petersburg Workshop on Simulation (SPWS'05),* ed. S.M. Ermakov, V.B. Melas, and A.N. Pepelyshev, 755–764. St. Petersburg University Publishers.

Table 1: 4-Node Feed-forward Network ($\lambda = 0.0455, \mu_1 = 0.7272, \mu_2 = 0.0455, \mu_3 = 0.0909, \mu_4 = 0.0909, p = 0.1$) ($\rho_1 = 0.06, \rho_2 = 0.1, \rho_3 = 0.45, \rho_4 = 0.5$)

| $L$ | **Numerical** | **PW** | **SDA** | | **SDH** | | |
|---|---|---|---|---|---|---|---|
| | $\gamma(L)$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | VRR |
| 25 | * | 3.9476e-07 ± 2.05 | 3 | 4.0064e-07 ± 0.07 | 6 | 4.0027e-07 ± 0.27 | 2.80 |
| 50 | * | 1.2825e-14 ± 3.44 | 4 | 1.3298e-14 ± 0.05 | 6 | 1.3291e-14 ± 0.25 | 4.56 |
| 100 | * | 1.1920e-29 ± 3.50 | 4 | 1.2568e-29 ± 0.06 | 6 | 1.2540e-29 ± 0.26 | 2.13 |

Table 2: 4-Node Feed-forward Network ($\lambda = 0.064, \mu_1 = 0.564, \mu_2 = 0.039, \mu_3 = 0.192, \mu_4 = 0.141, p = 0.1$) ($\rho_1 = 0.11, \rho_2 = 0.16, \rho_3 = 0.3, \rho_4 = 0.45$)

| $L$ | **Numerical** | **PW** | **SDA** | | **SDH** | | |
|---|---|---|---|---|---|---|---|
| | $\gamma(L)$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | VRR |
| 25 | * | 2.0830e-08 ± 5.89 | 3 | 1.9674e-08 ± 0.03 | 6 | 1.9602e-08 ± 0.25 | 1.23 |
| 50 | * | 5.2291e-17 ± 0.73 | 4 | 5.2223e-17 ± 0.02 | 6 | 5.2090e-17 ± 0.26 | 0.82 |
| 100 | * | 3.6552e-34 ± 0.67 | 4 | 3.6807e-34 ± 0.02 | 6 | 3.6921e-34 ± 0.18 | 0.43 |

Table 3: 4-Node Feed-forward Network ($\lambda = 0.069, \mu_1 = 0.571, \mu_2 = 0.022, \mu_3 = 0.198, \mu_4 = 0.14, p = 0.1$) ($\rho_1 = 0.12, \rho_2 = \rho_3 = 0.31, \rho_4 = 0.49$)

| $L$ | **Numerical** | **PW** | **SDA** | | **SDH** | | |
|---|---|---|---|---|---|---|---|
| | $\gamma(L)$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | VRR |
| 25 | * | 1.9267e-07 ± 3.75 | 5 | 2.1408e-07 ± 0.04 | 11 | 2.1278e-07 ± 0.54 | 1.88 |
| 50 | * | 3.9177e-15 ± 2.82 | 4 | 4.4567e-15 ± 0.07 | 11 | 4.4220e-15 ± 0.56 | 1.80 |
| 100 | * | 1.9168e-30 ± 12.0 | 4 | 1.9285e-30 ± 0.10 | 10 | 1.9101e-30 ± 0.61 | 0.74 |

Table 4: 4-Node Feed-forward Network ($\lambda = 0.074, \mu_1 = 0.617, \mu_2 = 0.024, \mu_3 = 0.135, \mu_4 = 0.15, p = 0.1$) ($\rho_1 = 0.12, \rho_2 = 0.31, \rho_3 = \rho_4 = 0.49$)

| $L$ | **Numerical** | **PW** | **SDA** | | **SDH** | | |
|---|---|---|---|---|---|---|---|
| | $\gamma(L)$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | $b$ | $\tilde{\gamma}(L) \pm$ RE% | VRR |
| 25 | * | 1.6065e-06 ± 8.08 | 4 | 1.6963e-06 ± 0.11 | 9 | 1.7018e-06 ± 0.47 | 9.27 |
| 50 | * | 5.7736e-14 ± 31.6 | 4 | 7.7208e-14 ± 0.40 | 8 | 7.6820e-14 ± 0.48 | 116. |
| 100 | * | 2.2486e-29 ± 14.9 | 4 | 7.2017e-29 ± 0.56 | 10 | 7.0982e-29 ± 0.50 | 127. |