Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

THE BALANCED LIKELIHOOD RATIO METHOD FOR ESTIMATING PERFORMANCE MEASURES OF HIGHLY RELIABLE SYSTEMS

Christos Alexopoulos

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0205, U.S.A.

Bruce C. Shultes

Naval Postgraduate School
Department of Computer Science
Monterey, CA 93943–5118, U.S.A.

ABSTRACT

Over the past several years importance sampling in conjunction with regenerative simulation has been presented as a promising method for estimating reliability parameters in highly reliable systems. Existing methods fail to provide benefits over crude Monte Carlo for the analysis of systems that contain significant component redundancies. This paper presents refined importance sampling techniques along with a generalized regenerative model. The proposed methods have solid theoretical properties and work well in practice.

1 INTRODUCTION

Consider a system with n_c components that are subject to failure and repair and n_r repairmen. Suppose that each component has a single operating state denoted by 1 and a single failed state denoted by 0. The system is modeled by a network $G = (V, A)$, where $A = \{1, \ldots, a\}$ is the set of links and V is the set of vertices. Link i contains u_i components and $u_1 + \cdots + u_a = n_c$. Let $X_i(t)$ be the number of operating components on link i at time t and let $R_i(t)$ be the number of repairmen working on link i at time t. Assume that only one repairman is needed to repair a failed component. Clearly, $R_i(t)+X_i(t) \leq u_i$ for all $i \in A$ and $R_1(t) + \cdots + R_a(t) \leq n_r$. The state of the system is described by the stochastic process $Y(t) = (X(t), R(t)) = (X_1(t), \ldots, X_a(t), R_1(t), \ldots, R_a(t))$. The state space for $X(t)$ is $\otimes_{i=1}^{a} \mathbb{N}_{u_i}$, where $\mathbb{N}_j = \{0, \ldots, j\}$ and $\otimes$ denotes a Cartesian product. The state space for $R(t)$ is $\{\mathbf{r} \in (\mathbb{N}_{n_r})^a : \sum_{i \in A} r_i \leq n_r\}$ so the state space for $Y(t)$ is $\mathcal{S} = \otimes_{i=1}^{a} \mathbb{N}_{u_i} \times \{\mathbf{r} \in (\mathbb{N}_{n_r})^a : \sum_{i \in A} r_i \leq n_r\}$. The time to failure for each component on link i is an exponential random variable with rate λ_i. Similarly the repair time for a component on link i is an exponential random variable with rate μ_i. All events are assumed to be independent. Consequently, the sojourn time in a generic state $(\mathbf{x}, \mathbf{r}) = (x_1, \ldots, x_a, r_1, \ldots, r_a) \in \mathcal{S}$ is an exponential random variable with rate $q_{(\mathbf{x},\mathbf{r})} = \sum_{i \in A}(x_i\lambda_i + r_i\mu_i)$.

The structure function ϕ is defined by

$$\phi(\mathbf{x}, \mathbf{r}) = \begin{cases} 1 & \text{if the system operates in state } (\mathbf{x}, \mathbf{r}) \\ 0 & \text{if the system is failed in state } (\mathbf{x}, \mathbf{r}). \end{cases}$$

For a review of reliability definitions, see Barlow and Proschan (1981). The definition of operating and failed states depends on the performance measure(s) under consideration.

Assume that $X_i(0) = u_i$ and $R_i(0) = 0$, for $i = 1, \ldots, a$. Let $\mathbf{u} = (u_1, \ldots, u_a)$ and $\mathbf{0} = (0, \ldots, 0)$. Now define the set $\mathcal{F} = \{(\mathbf{x}, \mathbf{r}) \in \mathcal{S} : \phi(\mathbf{x}, \mathbf{r}) = 0\}$ of failed states and the time to failure by $\Gamma = \inf\{t : t > 0, Y(t) \in \mathcal{F}\}$. The limit $U = \lim_{t\to\infty} P[Y(t) \in \mathcal{F}]$ (when it exists) is called the long-run system unavailability whereas $E(\Gamma)$ is the expected time to system failure. Although the system is Markovian, the exact evaluation of these measures and the computation of tight bounds are difficult problems even for moderate-scale systems due to the size of the state space (see Ball et al. 1995). Consequently, computer simulation frequently becomes the most suitable method for their estimation. The demand for modern communications and computer systems to be highly reliable makes the entrance of $Y(t)$ to the set $\mathcal{F}$ a rare event. This property causes crude (standard) Monte Carlo simulation to be inefficient in that it requires prohibitively long runs to produce precise estimates.

Based on the above assumptions, the process $Y = (X, R)$ is regenerative with return state $(\mathbf{u}, \mathbf{0})$, and regeneration epochs $0 = T_0 < T_1 < T_2 < \cdots$, where T_i is the time of the ith entry into state $(\mathbf{u}, \mathbf{0})$ and $\lim_{n\to\infty} T_n = \infty$ w.p.1 (for a review of results for regenerative processes see Serfozo 1990, pp. 41-56). Let $W_i = T_i - T_{i-1}$, $i \geq 1$, be the length of cycle i.

Let $1(\cdot)$ denote the identity function, and let $Z_i = \int_{T_{i-1}}^{T_i} 1(Y(t) \in \mathcal{F})\, dt$ denote the time Y spends in the set $\mathcal{F}$ during the cycle $[T_{i-1}, T_i)$. Then the limiting system

unavailability is given by

$$U = \frac{E(Z_1)}{E(W_1)}. \tag{1}$$

In addition, the mean time to failure (MTTF) starting from state $(\mathbf{u}, \mathbf{0})$ can be expressed as a ratio of expectations (see Shahabuddin et al. 1988):

$$E(\Gamma) = \frac{E[\min(W_1, \Gamma)]}{P(\Gamma < W_1)} = \frac{E[\min(W_1, \Gamma)]}{E[1(\Gamma < W_1)]}. \tag{2}$$

Suppose that one simulates Y over n cycles and collects the data (Z_i, W_i), $i = 1, \ldots, n$. Let $\bar{Z}$ and $\bar{W}$ be the respective sample means of Z_i and W_i. Then the classical regenerative estimator of the limiting unavailability is

$$\hat{U} = \bar{Z}/\bar{W}$$

and confidence intervals for U can be computed by a variety of methods (see Iglehart 1975).

The remainder of this paper is organized as follows: Section 2 reviews the main issues related to the importance sampling method and existing methods. Section 3 proposes an alternative process with correlated cycles. Section 4 describes the proposed methods, Section 5 contains experimental results, and Section 6 contains conclusions.

2 ANALYZING HIGHLY RELIABLE SYSTEMS

When the system under study is highly reliable (e.g., the component failure rates are significantly smaller that the respective repair rates) or the system structure does not have small minimum cuts (see Barlow and Proschan 1981), the crude Monte Carlo estimation of U and the MTTF based on regenerative cycles with return state $(\mathbf{u}, \mathbf{0})$ presents problems since cycles containing failures are infrequent.

2.1 Importance Sampling

One way to overcome this problem is to use the *importance sampling* (IS) method. Let $(\Omega, \Im, P)$ be the space of sample paths of the process Y. The IS method attempts to produce an alternative estimator for $E_P(Z_1)$ (the subscript "P" indicates the probability measure) that has smaller mean squared error (MSE). Let P' be another probability measure on $\Im$ such that P is absolutely continuous with respect to P'. Then one can write

$$\begin{aligned} E_P(Z_1) &= \int_\Omega Z_1(\omega) \frac{P(d\omega)}{P'(d\omega)} P'(d\omega) \\ &= \int_\Omega Z_1(\omega) L_1(\omega) P'(d\omega) = E_{P'}(Z_1 L_1), \end{aligned} \tag{3}$$

where the likelihood ratio $L_1(\omega)$ is the Radon-Nikodym derivative of P with respect to P' (for a detailed discussion see Royden 1968, pp. 276–278).

Equation (3) forms the basis of IS. Suppose that one draws n independent samples $Z_i'(\omega)$, $i = 1, \ldots, n$, with probability measure P'. Then the estimator

$$\hat{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i'(\omega) L_i(\omega).$$

is clearly unbiased. If in addition $E_{P'}[(Z_1' L_1)^2] < \infty$, then a confidence interval for $E_P(Z_1)$ can be derived by using the central limit theorem.

In general, one would like to choose P' so that $\mathrm{Var}_{P'}(Z_1 L_1) < \mathrm{Var}_P(Z_1)$ or, equivalently, $E_{P'}[(Z_1 L_1)^2] = E_P(Z_1^2 L_1) < E_P(Z_1^2)$. More specifically, for the problems studied here, one would like to make the probability of system failure (within a cycle) as large as possible while maintaining these properties. This implies that importance sampling probabilities should be as large as possible when $Z_1(\omega)$ is positive forcing the associated likelihood ratios to be small.

2.1.1 The Balanced Failure Biasing (BFB) Method

This method was proposed by Shahabuddin (1994). It models the probability that the next event is a component failure by a single *biasing* parameter. Given that a component failure event occurred, the event is allocated to links uniformly. Unfortunately, the resulting likelihood ratios can become unstable (see Shultes 1997, Chapter 6).

2.1.2 Measure Specific Dynamic Importance Sampling

The computation of confidence intervals for (1) and (2) requires the estimation of covariance terms. Using separate simulation runs for estimating the numerator and the denominator eliminates this problem and allows the use of separate importance sampling distributions in each simulation run.

When simulating highly reliable systems, the primary difficulty with crude Monte Carlo methods is infrequent observations of system failure. Once a system fails, highly reliable systems tend to rapidly return to full functionality. Hence, it is reasonable to apply importance sampling (within a cycle) up to the time of system failure and then utilize crude Monte Carlo sampling for the remainder of the cycle.

These two ideas were introduced by Goyal et al. (1987) and are jointly referred to as *measure specific dynamic importance sampling* (MSDIS). These same ideas are used within the new importance sampling strategies presented here.

3 A SEMI-STATIONARY PROCESS

To overcome problems associated with crude Monte Carlo regenerative simulation, one can consider the stationary portion of Y and "cycles" that start (and end) in a set of states $\mathcal{D}$ such that transitions within $\mathcal{D}$ do not occur. An example is the set of states with k failed components (k is smaller than the size of the smallest minimal cut for the state $(\mathbf{u}, \mathbf{0})$). We redefine T_i to be the time of the ith visit of Y to $\mathcal{D}$ ($W_i = T_i - T_{i-1}$ remains the length of the ith cycle). The process $\tilde{\zeta}_i(t) = Y(T_{i-1} + t)1(t \in [0, W_i))$ contains the information for Y in the interval $[T_{i-1}, T_i)$. Let S_i be the sequence of states visited by Y during $[T_{i-1}, T_i)$, let $H(S_i)$ be the respective sequence of holding times, and let $\tilde{\eta}_i = (S_i, H(S_i))$ be an alternative representation of $\tilde{\zeta}_i$. One can show that the process $\{(W_i, \tilde{\eta}_i) : i \geq 1\}$ is stationary (see Shultes 1997, Lemma 2.2.1), and then $\{(W_i, \tilde{\zeta}_i)\}$ is stationary. If we assume that $\{(W_i, \tilde{\eta}_i)\}$ is ergodic, part (i) of Theorem 2.2 from Serfozo (1972) implies that Y is *semi-stationary* with respect to $\{T_i\}$ and part (iv) of the same theorem with Theorem 3.1 from Serfozo (1972) implies that the limiting system unavailability can be written as

$$U = \lim_{t \to \infty} \frac{1}{t} \int_0^t 1(Y(z) \in \mathcal{F})\, dz = \frac{E(Z_1)}{E(W_1)} \quad \text{w.p.1.}$$

Furthermore, since $\{(W_i, \tilde{\zeta}_i)\}$ is ϕ-mixing (see Shultes 1997, Lemma 2.2.2), a confidence interval for U can be computed by the method of batch means (see Fishman 1996, Chapter 6).

Hordijk et al. (1976) showed that it is sufficient to simulate the embedded Markov chain and replace exponential holding times with the corresponding expected value when utilizing regenerative methods to estimate steady-state quantities. These results remain applicable when one utilizes semi-stationary processes instead of regenerative processes (see Shultes 1997, Theorems 3.3.1 and 3.3.3).

Analyzing the stationary portion of Y in terms of a semi-stationary process and cycles that begin and end in $\mathcal{D}$ is more complicated than regenerative simulation. The limiting hitting distribution for $\mathcal{D}$, known as a *Palm distribution*, must be maintained. Within an importance sampling procedure, the simplest approach is to use a crude Monte Carlo simulation to generate sequential cycle starting points in the set $\mathcal{D}$ and use these choices of starting points for the cycles generated via importance sampling (i.e., at the end of an importance sampling cycle the system jumps to the next crude Monte Carlo starting state). This procedure is loosely related to the procedure for simulating $\mathcal{A}$-cycles presented by Nicola et al. (1993).

4 BALANCED LIKELIHOOD RATIO METHODS

The basic idea behind the proposed methods is that if likelihood ratios associated with individual events within cycles (hereafter called *event likelihood ratios*) are forced to be bounded from above, then the likelihood ratios associated with regenerative and semi-stationary cycles (the product of event likelihood ratios for events that form a cycle) are also bounded from above.

Other authors have considered the problem of bounding likelihood ratios. In particular, Juneja (1993) explored methods of bounding likelihood ratios associated with regenerative cycles. His method built on BFB by providing an algorithm for choosing a biasing parameter that forced the resulting likelihood ratios to be bounded. However, the method does not eliminate the problems component redundancies cause for failure biasing methods.

4.1 Basic Technique

The proposed methods represent a significant departure from the failure biasing methods. The basic technique is based on some simple observations: Since every component repair must be preceded by a component failure, one can force the product of the respective pairs of event likelihood ratios to be one. This assignment causes the likelihood ratio associated with a cycle to be bounded from above by one.

Let $p(\mathbf{x}, \mathbf{r})$ denote the transition probability from state $(\mathbf{x}, \mathbf{r})$ to $(\mathbf{x} - e_i, \mathbf{r})$, for some $i \in A$, and let $p'(\mathbf{x}, \mathbf{r})$ be the respective importance sampling probability.

4.1.1 Implementation

Assume that $\mathcal{F}$ does not contain states with a single failed component. As a result, the first two events in any (regenerative or semi-stationary) cycle that includes a visit to $\mathcal{F}$ must be components failing. This guarantees that at least one of the corresponding event likelihood ratios is less than one.

A system is said to be *balanced* if all component failure rates are of the same order of magnitude (i.e., within a factor of ten). Suppose that the set of links is partitioned into sets $A_1, A_2, \ldots$ such that the set A_j contains all links with failure rates of the jth largest order of magnitude. Throughout the simulation of a cycle, we store the event likelihood ratios associated with component failure events from A_j in a stack $\mathcal{L}_j$. Let ℓ_j be the event likelihood ratio on the top of stack $\mathcal{L}_j$. Then $p'(\mathbf{x}, \mathbf{r})$ is set to

$$p'(\mathbf{x}, \mathbf{r}) = 1 - q^{-1}_{(\mathbf{x}, \mathbf{r})} \sum_j \ell_j \sum_{i \in A_j} r_i \mu_i.$$

If the event is a component failure in set A_k, we push the event likelihood ratio containing $p(\mathbf{x}, \mathbf{r})/p'(\mathbf{x}, \mathbf{r})$ onto

stack $\mathcal{L}_k$. Otherwise, we pop the top element of the stack corresponding to the component completing repair and discard it. There are several ways to allocate component failure events to operating components in the network. Some procedures lead to estimates with bounded relative error (BRE).

4.1.2 Bounded Relative Error

Estimates with BRE deserve attention because the amount of computational effort (sample size) required to yield a desired level of accuracy (relative confidence interval half-width) remains bounded as the quantity of interest approaches zero; see Fishman (1996), Nakayama (1996), and Shahabuddin (1994). This is important in the study of highly reliable systems as quantities of interest (e.g., the mean downtime within a cycle) go to zero as component failure rates approach zero.

Theorem 1 *When utilizing BLR estimation procedures for the estimation of $E(Z_1)$ and $P(\Gamma < W_1)$:*

(a) *For balanced systems, the allocation of component failure events to links proportionally to the rate of failure on each link yields BRE.*

(b) *For non-balanced systems, the allocation of component failure events to links proportionally to the number of operating components on each link yields BRE.*

(c) *Any allocation method that is independent of the component failure rates will yield BRE.*

The proof of Theorem 1 appears in Alexopoulos and Shultes (1998). Numerical results for a BLR algorithm utilizing case (a) in Theorem 1 are shown in Section 5.

4.2 Utilizing Structural Information

One can use minimum cuts to identify events on shortest paths to failure. See Ahuja et al. (1996) for a review of network flow properties and Section 5.5.2 of Shultes (1997) for efficient ways to maintain information about minimum cuts. Let $\mathcal{C}_-(\mathbf{x},\mathbf{r})$ be the set of links i on a minimum cut for the network state $(\mathbf{x},\mathbf{r})$, and let $P(\mathcal{C}_-(\mathbf{x},\mathbf{r}))$ be the probability that the next event is a failure in $\mathcal{C}_-(\mathbf{x},\mathbf{r})$. Let $\mathcal{C}_-^c(\mathbf{x},\mathbf{r}) = A - \mathcal{C}_-(\mathbf{x},\mathbf{r})$, and let $P(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))$ be the probability that the next event is a failure in $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$. Similarly, define $\mathcal{C}_+(\mathbf{x},\mathbf{r})$ as the set of links i with components undergoing repair that are on a minimum cut for network state $(\mathbf{x}+e_i,\mathbf{r})$, and let $P(\mathcal{C}_+(\mathbf{x},\mathbf{r}))$ be the probability of a state transition from $(\mathbf{x},\mathbf{r})$ that lengthens the shortest path to failure by one. Let $\mathcal{C}_+^c(\mathbf{x},\mathbf{r})$ be the set of links not in $\mathcal{C}_+(\mathbf{x},\mathbf{r})$ with components completing repair.

4.2.1 Implementation Issues

We store the event likelihood ratios in stacks: Stack $\mathcal{L}_j$ contains event likelihood ratios for components in A_j that were on a minimum cut when they failed (i.e., selected from $\mathcal{C}_-(\mathbf{x},\mathbf{r})$), and stack $\mathcal{L}_j^c$ contains event likelihood ratios for components in A_j that were not on a minimum cut. The *status* of the links in the network depends on the current state of the system. The following heuristic takes this into account: For each component in A_j completing repair, if the component is in $\mathcal{C}_+^c(\mathbf{x},\mathbf{r})$ or stack $\mathcal{L}_j$ is empty, then we multiply the repair probability by the event likelihood at the head of stack $\mathcal{L}_j^c$; otherwise, we multiply the IS repair probability by the event likelihood ratio at the head of stack $\mathcal{L}_j$.

For each set A_j that contains links with failed components, let ℓ_j be the event likelihood ratio on the top of stack $\mathcal{L}_j$, and let ℓ_j^c be the event likelihood ratio on top of stack $\mathcal{L}_j^c$. If $\mathcal{L}_j$ is empty, we set $\ell_j = \ell_j^c$.

4.2.2 Greedy Algorithm

A natural way to define the IS probabilities for allocating component failure events to links is to let

$$P'(\mathcal{C}_-^c(\mathbf{x},\mathbf{r})) = \frac{p(\mathbf{x},\mathbf{r})}{p'(\mathbf{x},\mathbf{r})} P(\mathcal{C}_-^c(\mathbf{x},\mathbf{r})),$$

where $p'(\mathbf{x},\mathbf{r})$ is given by

$$p'(\mathbf{x},\mathbf{r}) = 1 - q_{(\mathbf{x},\mathbf{r})}^{-1} \sum_j \left[\ell_j \sum_{i \in \mathcal{C}_+(\mathbf{x},\mathbf{r})} r_i \mu_i + \ell_j^c \sum_{i \in \mathcal{C}_+^c(\mathbf{x},\mathbf{r})} r_i \mu_i \right].$$

Suppose that component failure events are allocated to links in $\mathcal{C}_-(\mathbf{x},\mathbf{r})$ and $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$ based on probabilities that are proportional to component failure rates. This construction makes the event likelihood ratio for a component failure event in $\mathcal{C}_-(\mathbf{x},\mathbf{r})$

$$\frac{p(\mathbf{x},\mathbf{r})}{p'(\mathbf{x},\mathbf{r})} \frac{P(\mathcal{C}_-(\mathbf{x},\mathbf{r}))}{P'(\mathcal{C}_-(\mathbf{x},\mathbf{r}))} = 1$$

and the event likelihood ratio for a component failure event in $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$ much smaller than one. In fact, it also forces components in $\mathcal{C}_+^c(\mathbf{x},\mathbf{r})$ to have significantly larger repair probabilities than the components from $\mathcal{C}_+(\mathbf{x},\mathbf{r})$.

4.2.3 Expected Downtime Heuristic

The derivation of a near-optimal IS distribution for estimating the expected downtime within a regenerative cycle in a k-out-of-n system suggests another alternative (see Shultes 1997, Chapter 4). For each component failure event that occurs in $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$ the potential number of events

needed to restore the system (i.e., leave the set $\mathcal{F}$) increases by one. Therefore, by keeping track of the number of component failure events that occur in $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$, one can attempt to counter the potential for inflated downtimes.

Suppose that j components, chosen from sets $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$, are in failed states. If all these components are repaired before the system leaves $\mathcal{F}$, then an estimate for the contribution of these j components to the system downtime within a cycle is $j+1$ times the holding time in a state (the one is for the single component from $\mathcal{C}_-(\mathbf{x},\mathbf{r})$ that must be repaired in order for the system to operate). Therefore, one can set $P'(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))$ so that if a component failure event occurs within $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$, then the resulting event likelihood ratio is $(j+1)/(j+2)$ causing the product of the event likelihood ratios associated with component failure events from $\mathcal{C}_-^c(\mathbf{x},\mathbf{r})$ to be $1/(j+2)$.

In some cases, the choice of $p'(\mathbf{x},\mathbf{r})$ might cause $P'(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))$ to be greater than $P(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))$. This can be avoided by using the minimum of the proposed $P'(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))$ and

$$\frac{[p(\mathbf{x},\mathbf{r})/p'(\mathbf{x},\mathbf{r})]\,P(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))}{P(\mathcal{C}_-(\mathbf{x},\mathbf{r})) + [p(\mathbf{x},\mathbf{r})/p'(\mathbf{x},\mathbf{r})]\,P(\mathcal{C}_-^c(\mathbf{x},\mathbf{r}))},$$

a *normalized* form of the importance sampling probability in the Greedy Algorithm.

4.3 Ensuring Bounded Relative Error

Generating shortest paths to failure is an intuitive approach to reducing simulation runtimes. However, the most-likely paths to failure are the most-important paths when estimating a quantity that is nonzero only when the system visits the set $\mathcal{F}$ within a cycle. If all component failure rates in a system are of the same order of magnitude, then it suffices to consider the shortest paths to failure. Otherwise, solely focusing on shortest paths to failure may ignore some of the most-likely paths to failure and does not lead to BRE. Similar results have been proven by Nakayama (1996), who developed a set of path-wise criteria that must be met to ensure BRE.

Theorem 2 *For balanced systems, the Greedy Algorithm or the Expected Downtime Heuristic with allocation of component failure events to links proportionally to the component failure rates yield estimates for* $E(Z_1)$ *and* $P(\Gamma < W_1)$ *with BRE.*

The proof of Theorem 2 is in Alexopoulos and Shultes (1998). Numerical results for an algorithm (BLRC) utilizing Theorem 2 are shown in Section 5.

5 NUMERICAL RESULTS

Consider the communications network depicted in Figure 1. This network contains 21 links, each containing two identical components (e.g., dedicated communication lines). For simplicity, let each component represent one unit of capacity between respective nodes. There are numerous factors that could cause a component to fail (e.g., hardware failure, software failure, or other external factors). Components fail at a rate of one every $33\frac{1}{3}$ hours. There are 4 repairmen that repair components, as good as new, at rate of one every $2\frac{1}{2}$ hours. Upon completing a repair, a repairman selects the next component to repair uniformly over the links in the network that contain failed components.

A variety of reliability parameters can be defined. Assume that the network is functional if nodes 1 and 10 communicate (via a path with an operating component on each link). Table 1 compares 90% confidence intervals for the limiting network unavailability from the crude Monte Carlo and BFB methods in regenerative simulations. The crude Monte Carlo simulation utilized 10,000,000 cycles to establish a benchmark. The BFB algorithm used MSDIS with 10,000,000 cycles to estimate the mean downtime $E(Z_1)$ within a cycle and 100,000 cycles to estimate the mean cycle time $E(W_1)$. The runtime required for the BFB estimate is significantly less than the runtime for the crude Monte Carlo estimate, but the half-width for the crude Monte Carlo confidence interval is significantly smaller than the corresponding BFB half-width. Based on the variance reduction time ratio (VRTR), i.e., the product of the ratio of variances and the ratio of runtimes, BFB is dramatically less efficient than crude Monte Carlo in this instance.

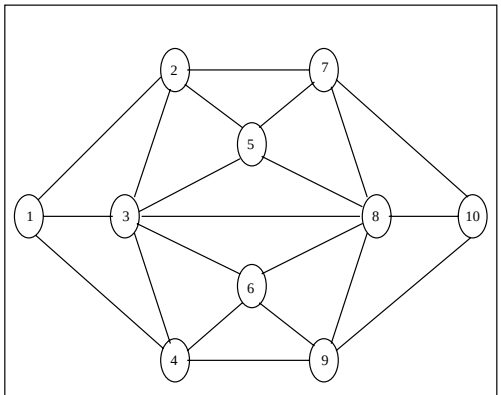

Figure 1: Communications Network

Table 1: Estimation of U Using Crude Monte Carlo and Balanced Failure Biasing

Quantity	Crude*	IS Algorithm BFB**	Reduction
Estimate	1.377E-05	1.502E-05	—
Half-width	7.060E-07	2.785E-06	0.25
Runtime	63081.00	49133.67	1.28
VRTR	—	0.08	—

* Estimates based on 10,000,000 cycles
** Estimates based on 10,000,000 cycles for $E(Z_1)$ and 100,000 for $E(W_1)$

Table 3: Estimation of U Using BLRC with Semi-stationary Cycles Starting with 3 Failed Components

Quantity	Crude(R)*	IS Algorithm BLRC**	Reduction
Estimate	1.377E-05	1.408E-05	—
Half-width	7.060E-07	1.951E-07	3.62
Runtime	63081.00	25727.06	2.45
VRTR	—	32.12	—

* Estimates based on 10,000,000 cycles
** Estimates based on 1,000,000 cycles for $E(Z_1)$ and 100,000 for $E(W_1)$

Results from applying BLR (Theorem 1, case (a)) and BLRC (Theorem 2) within regenerative simulations are shown in Table 2. Each method was implemented using MSDIS. The BLRC method yielded the tightest 90% confidence interval half-width, but an examination of the VRTRs shows that the runtime associated with maintaining information about minimum cuts in the network outweighed the improvement in half-width over the BLR method. Both methods illustrate modest improvement over the crude Monte Carlo method.

The use of semi-stationary cycles dramatically improves the performance of the BLRC method. Table 3 displays numerical results from applying BLRC and MSDIS with $\mathcal{D}$ being the set of states with 3 failed components. The confidence intervals were computed by the method of batch means (Fishman 1996) with 30 batches. The crude Monte Carlo results based on regenerative simulation are redisplayed. Notice that semi-stationary cycles dramatically reduce the runtime for BLRC over the corresponding runtime with regenerative cycles in Table 2. The BLRC algorithm utilized event likelihood ratios associated with component failure events within the importance sampling distribution, so the method's ability to quickly force systems to fail depends on the magnitude of these event likelihood ratios. For states in $\mathcal{D}$ the repair probability is much greater than the repair probability for the regenerative state $(\mathbf{u}, \mathbf{0})$, which leads to smaller event likelihood ratios.

To emphasize the power of the balanced likelihood ratio methods, a sequence of simulation experiments was performed by varying the failure rate of the components in the computer network and checking the performance of the regenerative method using BFB with MSDIS, the regenerative method using BLRC with MSDIS, and the semi-stationary method using BLRC with MSDIS. For each experiment utilizing semi-stationary methods, the set $\mathcal{D}$ of states with k failed components was chosen so that the component failure rate and the component repair rate out of the set were approximately equal. The first column of Table 4 displays the sequence of multipliers for the component failure rates used to create the sequence of simulation experiments. Subsequent columns display the observed VRTRs. These results are plotted in Figure 2. The solid line at one represents crude Monte Carlo. Clearly, BLRC dominates BFB and crude Monte Carlo. The performance of BFB remains approximately constant while the BLRC method exhibits better performance as component failure rates approach zero (systems exhibit greater reliability).

Table 4: Variance Reduction Time Ratios for the Network in Figure 2. All Estimates are Based on 1,000,000 Cycles for $E(Z_1)$ and 100,000 Cycles for $E(W_1)$

Multiplier	VRTR BFB	BLRC	BLRC-SS
1.331	8.540E-03	4.388E-01	2.327E+00
1.210	3.866E-02	9.010E-01	4.467E+00
1.100	7.505E-04	1.629E+00	2.003E+01
1.000	8.251E-02	2.733E+00	3.212E+01
0.909	1.509E+00	1.120E+01	7.772E+01
0.826	3.284E+00	3.459E+01	1.168E+02
0.751	4.912E+00	1.545E+02	3.354E+02
0.683	5.622E+00	3.660E+01	9.316E+02
0.621	5.345E+00	4.467E+02	7.333E+02
0.564	5.754E+00	6.627E+02	8.802E+02
0.513	4.897E+00	1.182E+03	2.791E+03

The choice of k varies with the multiplier:
$k = 1$ for the multiplier 0.513,
$k = 2$ for the multipliers in the range $[0.564, 0.751]$,
$k = 3$ for the multipliers in the range $[0.826, 1.210]$,
$k = 4$ for the multiplier 1.331

Table 2: Estimation of U Using BLR Methods

		IS Algorithm		IS Algorithm	
Quantity	Crude*	BLR**	Reduction	BLRC**	Reduction
Estimate	1.377E-05	1.409E-05	—	1.410E-05	—
Half-width	7.060E-07	4.407E-07	1.60	3.338E-07	2.12
Runtime	63081.00	42553.34	1.48	103223.77	0.61
VRTR	—	3.79	—	2.74	—

* Estimates based on 10,000,000 cycles

** Estimates based on 1,000,000 cycles for $E(Z_1)$ and 100,000 for $E(W_1)$

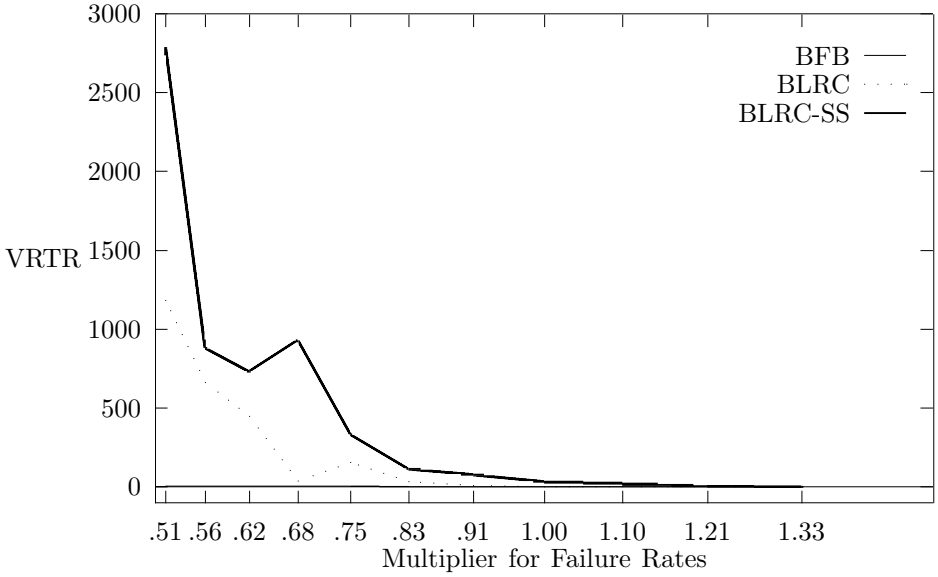

Figure 2: Impact of Varying Failure Rates on Algorithm Performance

6 CONCLUSIONS

Importance sampling can be a powerful tool for the estimation of reliability measures of highly dependable systems with repairs. The proper selection of an importance sampling distribution can "make or break" an analysis procedure. The proposed balanced likelihood ratio methods in conjunction with semi-stationary models provide some level of assurance for performance improvement over the crude Monte Carlo method. The ideas motivating BLR methods provide hope that packaged routines for importance sampling are within reach.

ACKNOWLEDGMENTS

This research was supported in part by Air Force Office of Scientific Research Grant No. F49620-93-1-43.

REFERENCES

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. 1993. *Network Flows: Theory Algorithms and Applications*. Prentice Hall, Englewood Cliffs, New Jersey.

Alexopoulos, C., and B. C. Shultes. 1998. Balanced likelihood ratio methods for estimating network reliability. Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Ball, M. O., C. J. Colbourn, and J. S. Provan. 1995. Network reliability. In *Handbooks in Operations Research and Management Science: Network Models*, eds. M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, 7:673–762. Elsevier Science B.V., Amsterdam

Barlow, R. E. and F. Proschan. 1981. *Statistical Theory of Reliability and Life Testing*, 2nd edition. TO BEGIN WITH, Silver Spring, Maryland.

Fishman, G. S. 1996. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag, New York.

Goyal, A., P. Heidelberger, and P. Shahabuddin. 1987. Measure specific dynamic importance sampling for availability simulations. In *Proceedings of the 1987 Winter Simulation Conference*, eds. A. Thesen, H. Grant, and W. D. Kelton, 491–499. IEEE, Piscataway, New Jersey.

Hordijk, A., D. L. Iglehart, and R. Schassberger. 1976. Discrete time methods for simulating continuous time Markov chains. *Advances in Applied Probability* 8:772–788.

Iglehart, D. L. 1975. Simulating stable stochastic systems, V: Comparison of ratio estimators. *Naval Research Logistics Quarterly* 22:553–565.

Juneja, S. 1993. Efficient rare event simulation of stochastic systems. Ph.D. Thesis, Department of Operations Research, Stanford University, Palo Alto, California.

Nakayama, M. K. 1996. General conditions for bounded relative error in simulations of highly reliable Markovian systems. *Advances in Applied Probability* 28:687–727.

Nicola, V. F., P. Shahabuddin, P. Heidelberger, and P. W. Glynn. 1993. Fast simulation of steady-state availability in non-Markovian highly dependable systems. In *Proceedings of the 23rd Symposium on Fault Tolerant Computing*, 38–47. IEEE, New York.

Serfozo, R. F. 1972. Semi-Stationary Processes, *Z. Wahrscheinlichkeitstheorie verw. Geb.* 23:125–132.

Serfozo, R. F. 1990. Point processes, In *Handbooks in Operations Research and Management Science: Stochastic Models*, eds. D. P. Heyman and M. J. Sobel, Chapter 1. North-Holland, New York.

Shahabuddin, P., V. F. Nicola, P. Heidelberger, A. Goyal, and P. W. Glynn. 1988. Variance reduction in mean time to failure simulations. In *Proceedings of the 1988 Winter Simulation Conference*, eds. M. A. Abrams, P. l. Haigh, and J. C. Comfort, 491–499. IEEE, Piscataway, New Jersey.

Shahabuddin, P. 1994. Importance sampling for the simulation of highly reliable Markovian systems, *Management Science* 40(3):333–352.

Shultes, B. C. 1997. Regenerative techniques for estimating performance measures of highly dependable systems with repairs. Ph.D. Thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

AUTHOR BIOGRAPHIES

CHRISTOS ALEXOPOULOS is an Associate Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. in Operations Research from the University of North Carolina at Chapel Hill. His research interests are in the areas of applied probability, statistics, and optimization of stochastic systems. His recent work involves problems related to the optimal design of telecommunications and transportation systems. Dr. Alexopoulos is a member of INFORMS and the INFORMS College of Simulation. He served as a Co-Editor for the *Proceedings of the 1995 Winter Simulation Conference*.

BRUCE C. SHULTES is a National Research Council Research Associate in the Department of Computer Science (Software Engineering Group) at the Naval Postgraduate School. 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 the Georgia Institute of Technology. His research interests are in the areas of applied probability, simulation, analysis of stochastic systems, algorithms, and software engineering. His recent research efforts include efficient algorithms for analyzing rare events, methods for transitioning "state-of-the-art" theory into practice, and methods for enabling distributed computing. He is a member of INFORMS and the INFORMS College of Simulation.