FAST SIMULATION OF DEPENDABILITY MODELS WITH GENERAL FAILURE, REPAIR AND MAINTENANCE PROCESSES

by

Victor F. Nicola, Marvin K. Nakayama,
Philip Heidelberger and Ambuj Goyal

TECHNICAL REPORT No. 53

January 1990

Prepared under the Auspices
of
U.S. Army Research Contract
DAAL-03-88-K-0063

DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

FAST SIMULATION OF DEPENDABILITY MODELS WITH GENERAL FAILURE, REPAIR AND MAINTENANCE PROCESSES

Victor F. Nicola *, Marvin K. Nakayama **,
Philip Heidelberger * and Ambuj Goyal *

* IBM Research Division
T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, New York 10598

** Department of Operations Research
Stanford University
Stanford, California 94305

ABSTRACT

The problem of computing dependability measures of repairable systems with general failure, repair and maintenance processes is a hard problem to solve in general either by analytical or by numerical methods. Monte Carlo simulation could be used to solve this problem, however, standard simulation takes a very long time to estimate system reliability and availability with reasonable accuracy because typically the system failure is a rare event. When the failure and repair time distributions are exponential, *importance sampling* has been used successfully in the past to reduce simulation run lengths. In this paper, we extend the applicability of *importance sampling* to non-Markovian models with general failure and repair time distributions. We show that by carefully selecting a heuristic for importance sampling, orders of magnitude reduction in simulation run-lengths can be obtained. We illustrate the effectiveness of the technique by modelling a large repairable computing system. Also, we study the effect of periodic maintenance on systems with components having increasing and decreasing failure rate.

Keywords: *fast simulation, dependability models, importance sampling, non-Markovian models, periodic maintenance*

** *Research supported by the IBM Corporation under SUR-SST Contract* 12480042 *and by the U.S. Army Research Office under contract number DAAL*-03-88-*K*-0063.

1. Introduction

The problem of computing dependability measures of repairable systems with general failure, repair and maintenance processes is a hard problem either by analytical or by numerical methods. Such systems, in general, cannot be modelled by Markov or even semi-Markov processes. HARP [3] solves large models with general failure time distributions by creating a non-homogeneous Markov chain model of the system and then solving the corresponding differential equations numerically. The technique has been applied to non-repairable systems only (transient recoveries are allowed, but they are approximated by instantaneous transitions). Furthermore, only transient measures (e.g., reliability) are estimated. CARE-III [18] uses numerical integration methods to solve similar models.

The goal of this paper is to model systems with general failure, repair and maintenance processes, and solve them for both transient (e.g., reliability and mean time to failure) and stationary (e.g., steady-state availability) measures. In relatively simple cases, one could obtain the Laplace Transform of dependability measures for such models and numerically invert them to obtain the desired results [14]. However, these methods are limited to small models and are prone to unboundable numerical errors.

An alternative approach is to use Monte Carlo simulation. The advantage of this method is that arbitrary system details can be modeled, and furthermore, all the system states need not be generated. The disadvantage of this approach is that standard simulation takes very long time to estimate dependability measures with reasonable accuracy because system failure events are very rare in highly dependable systems [4]. When the failure and repair time distributions are exponential, the *importance sampling* technique has been used successfully in the past to reduce simulation run-lengths significantly [2, 10, 12]. Basically, the system failure events are forced to occur more often by increasing the failure rates; unbiased estimates of dependability measures are obtained by multiplying the value of the measure on a sample path by the likelihood ratio of the sample path. The likelihood ratio for a given sample path is the ratio of the probability of the sample path under the original distributions (e.g., with the original failure and repair rates) over the probability of the same sample path under the new distributions (e.g., with the new failure and repair rates).

In this paper, we extend the applicability of importance sampling to non-Markovian systems with general failure, repair and maintenance processes. For general discrete-event systems, importance sampling has been discussed in [5, 6]. Basically, a Generalized Semi-Markov Processes (GSMPs) formalism is used to represent such systems, and the likelihood ratio of a sample path is written in terms of the various probability distributions (e.g., failure, repair and maintenance distributions) in

the original and the new (simulated) systems. However, in [5, 6] they did not consider the design and implementation of specific importance sampling distributions that are required in order to obtain effective variance reduction in non-Markovian models of highly dependable systems. One possible way to appropriately implement importance sampling, which we propose and use in this paper, is accomplished by canceling and rescheduling previously scheduled events. For example, when one component fails in a system with a redundant component pair, we speed up the failure of the other component so that it fails with high probability before the repair of the first component. This involves cancelling the originally scheduled failure event for the second component and rescheduling it using a new failure distribution with a smaller mean time to failure.

In Sections 2 and 3 we give a concise description of discrete-event systems, which is appropriate for our purpose in this paper; namely, to formally represent the probability of a sample path. This yields a representation for the likelihood ratio which is the key to importance sampling.

In Section 4, we give the basic estimators for some commonly used measures in highly dependable systems, such as reliability, steady-state availability and mean time to failure. A simple example of a two-components system is used to explain these measures as well as the importance sampling technique used to estimate them. In Section 5, we discuss the implementation of these methods in a software tool which we used to generate and simulate large models. This tool is based on the CSIM package [15, 16]. In Section 6, we use three examples to illustrate the effectiveness of the proposed importance sampling techniques. First, we use a small example to experiment with some heuristics for selecting the new probability distributions which make the typically rare system failures occur more often. Second, these heuristics are applied in a large example to show that orders of magnitude reduction in variance can be obtained. We use exponential failure and repair distributions in this example to ascertain the correctness of the results obtained by comparing them against numerical results obtained from the SAVE package [7]. In the third example, we use Weibull failure distribution and periodic maintenance for all individual components in the system. We study the effect of the hazard rate (i.e., increasing, decreasing and constant failure rates) on the optimal maintenance period. Such studies cannot be performed with existing analytical or numerical methods. In Section 7 we give conclusions and some directions for future research.

2. Discrete-Event Systems

In this section we give some notation and basic properties of discrete-event systems, which will assist in representing the probability of a sample path and the likelihood ratio required for importance sampling in simulations of such systems. A precise mathematical framework for the study of discrete-event systems is given by Glynn in [5]: he gives a generalized semi-Markov process

(GSMP) formalism of discrete-event systems. Here, we give an alternative concise description of discrete-event systems, which is appropriate and sufficient for our purpose. In our description we have left out some of the details and generalities which are not needed for the developments in this paper.

A discrete-event system is characterized by a set of events $\mathbf{E}$ which can trigger transitions of its state and a set $\mathcal{Z}$ of integer-valued output state vectors ($\mathcal{Z}$ is possibly a countably infinite set). With each event $e \in \mathbf{E}$ we associate a clock. The reading $c(e)$ is the "remaining lifetime" of clock e, i.e., the time remaining for clock e to expire. $c(e) = \infty$ if clock e is inactive. The choice of the output (observable or measured) state vector in a discrete-event system depends on the application at hand and the desired level of detail.

The internal state of a discrete-event system at a given time is completely determined by its output state and the set of active clocks (i.e., the set of events which can trigger a transition to another internal state) with the associated clock readings. Upon the $i-$th transition, let $\mathcal{Z}_i \in \mathcal{Z}$ be the output state vector and $\mathbf{E}_i \subseteq \mathbf{E}$ be the set of active clocks; $\mathbf{c}_i(\mathbf{E}_i)$ is a vector with the associated clock readings. Then $\mathbf{X}_i = (\mathcal{Z}_i, \mathbf{c}_i(\mathbf{E}_i))$ is the internal state of the discrete-event system upon the $i-$th transition. Notice that the output state and the set of active clocks characterizing the internal state change only in response to transitions (events), while the clock readings are continuously changing at the same rate (in general, different clock rates may be assumed, see e.g., [5]). It is typical in discrete-event systems that the output state does not change between transitions; for example, the number of customers in a queueing system. Therefore, the output state trajectory of a discrete-event system is completely described by the output state at transition times of the internal state. Let t_i, $i \geq 0$, be the time of the $i-$th transition, with $t_0 = 0$. Then $T_i = t_{i+1} - t_i$ is the time between the $i-$th and the $(i+1)-$th transitions. Let $Z(t)$ denote the output state at time t, then $\{Z(t),\ t \geq 0\}$ is the output state trajectory, and $Z(t) = \mathcal{Z}_k$ if $t_k \leq t < t_{k+1}$.

As indicated above, we only consider the internal state sequence at transition times, since this is sufficient to determine the output state trajectory of the discrete-event system. A sample path of the discrete-event system up to the $n-$th transition is denoted by the sequence $\mathbf{X}_{0,n}$ of internal states at transition times,

$$\mathbf{X}_{0,n} = (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n).$$

Let $e_i^* = \underset{e \in \mathbf{E}_i}{argmin}(c(e))$, $i \geq 0$. Then e_i^* is the clock which triggers the $(i+1)-$th transition and $T_i = c(e_i^*)$. The internal state $\mathbf{X}_{i+1} = (\mathcal{Z}_{i+1}, \mathbf{c}_{i+1}(\mathbf{E}_{i+1}))$ upon the $(i+1)-$th transition is determined by the sequence $\mathbf{X}_{0,i}$ and may depend on the complete history of the system. The set of active clocks $\mathbf{E}_{i+1}$ upon the $(i+1)-$th transition is determined by

$$\mathbf{E}_{i+1} = \mathbf{E}_i - e_i^* - \mathbf{A}_i + \mathbf{N}_{i+1}\,,\ i \geq 0\,,$$

where $\mathbf{A}_i$ is the set of clocks canceled (aborted) upon the $(i+1)-$th transition and $\mathbf{N}_{i+1}$ is the set of new clocks activated upon the $(i+1)-$th transition. The set of clocks $\mathbf{A}_i$ and $\mathbf{N}_{i+1}$ and the output state Z_{i+1} are determined probabilistically, depending on the trigger event e_i^* and the sequence $\mathbf{X}_{0,i}$. Therefore, the $(i+1)-$th transition triggered by e_i^* yields the output state Z_{i+1} and the active set of clocks $\mathbf{E}_{i+1}$ with a probability denoted by $p_{i+1}(Z_{i+1}, \mathbf{E}_{i+1}; e_i^*)$. The subscript $i+1$ of p symbolizes the dependence on the sequence $\mathbf{X}_{0,i}$ (routing in queueing networks is an obvious example for the use of these transition probabilities).

We denote by $f_i(t; e)$ (resp. $\overline{F}_i(t; e)$) the probability density function (resp. the complementary distribution function) of the (conditional) "remaining lifetime" of clock $e \in \mathbf{E}_i$ at the $i-$th transition. The subscript i symbolizes the dependence of this probability density function on the history of the system through its internal state sequence $\mathbf{X}_{0,i}$. For example, if clock e was originally scheduled using a probability density function $f(\bullet; e)$ and if the age of the clock at the $i-$th transition is a, then the density of the remaining lifetime t is $f_i(t; e) = f(t+a; e)/\overline{F}(a; e)$. Similarly, $\overline{F}_i(t; e) = \overline{F}(t+a; e)/\overline{F}(a; e)$. If clock e is newly scheduled at the $i-$th transition, then the age is 0, so that $f_i(t; e) = f(t; e)$ and $\overline{F}_i(t; e) = \overline{F}(t; e)$. Let $\mathbf{O}_{i+1}$ be the set of old clocks which continue to be active upon the $(i+1)-$th transition, i.e., $\mathbf{O}_{i+1} = \mathbf{E}_{i+1} - \mathbf{N}_{i+1}$, $i \geq 0$. The clock reading $c(e)$, $e \in \mathbf{O}_{i+1}$, is updated as follows: $c(e) = c(e) - T_i$. Upon the $(i+1)-$th transition, the probability density function and the complementary distribution function of the remaining time on clock $e \in \mathbf{O}_{i+1}$ are changed to reflect the elapsed time on this clock, i.e., for all $e \in \mathbf{O}_{i+1}$

$$f_{i+1}(t; e) = f_i(t + T_i; e)/\overline{F}_i(T_i; e)\,, \tag{2.1}$$

$$\overline{F}_{i+1}(t; e) = \overline{F}_i(t + T_i; e)/\overline{F}_i(T_i; e)\,. \tag{2.2}$$

Notice that these modified distributions are not needed to determine the clock readings $c(e)$, $e \in \mathbf{O}_{i+1}$, since, as stated above, we can use the remaining lifetime as the updated clock reading for an old clock. However, they are used to describe the probability of a sample path and the likelihood ratio, as we shall see in the following.

Given that the internal state of the discrete-event system is $\mathbf{X}_i$ at the $i-$th transition, we can write the probability density (likelihood) that the next internal state is $\mathbf{X}_{i+1}$ at the $(i+1)-$th transition. We denote this probability by $P(\mathbf{X}_{i,i+1})$, then

$$P(X_{l,l+1}) = f_l(T_l; e_l^*)\, p_{l+1}(Z_{l+1}, E_{l+1}; e_l^*) \prod_{e \in E_l - \{e_l^*\}} \bar{F}_l(T_l; e)\,. \tag{2.3}$$

It follows that the likelihood of a sample path $X_{0,n}$, up to the $n-$th transition, is given by

$$P(X_{0,n}) = \prod_{l=0}^{n-1} f_l(T_l; e_l^*)\, p_{l+1}(Z_{l+1}, E_{l+1}; e_l^*) \prod_{e \in E_l - \{e_l^*\}} \bar{F}_l(T_l; e)\,. \tag{2.4}$$

3. Importance Sampling in Simulations of Discrete-Event Systems

In this section, we discuss importance sampling which can be used to obtain a significant variance reduction over standard simulation when estimating dependability measures. The basic idea of importance sampling is to simulate the system under different probability distributions, so as to appropriately and quickly move the system towards failure. Since the simulated system is dynamically different from the original system, a correction factor is needed to compensate for the resulting bias. This correction factor is called the likelihood ratio and must be used with importance sampling to obtain unbiased estimates.

Consider a simulation of a discrete-event system for the purpose of estimating the expected value of a particular performance measure, say M. Let $M(X_{0,N})$ be the value of the measure on a sample path $X_{0,N}$, where N is a stopping time relative to the internal state sequence, i.e., $I(N = l)$ is a function of $X_{0,l}$ ($I(\bullet)$ is the indicator function which equals one if its argument is true; otherwise it equals zero). In the original system, the likelihood of the sample path $X_{0,N}$ is $P(X_{0,N})$ as given by Equation (2.4). To implement importance sampling, we simulate the system with a different likelihood $P'(\bullet)$ for its sample path. For $P'(\bullet)$, it is necessary that the following must hold for all $X_{0,N}$,

$$P'(X_{0,N}) > 0 \text{ whenever } M(X_{0,N})\, P(X_{0,N}) > 0\,. \tag{3.1}$$

Under $P(\bullet)$, the expectation $E_P(M)$ can be expressed as follows

$$\begin{aligned} E_P(M) &= \sum_{\forall X_{0,N}} M(X_{0,N})\, P(X_{0,N}) \\ &= \sum_{\forall X_{0,N}} M(X_{0,N})\, L(X_{0,N})\, P'(X_{0,N}) \\ &= E_{P'}(M\, L)\,, \end{aligned} \tag{3.2}$$

where $E_{P'}(ML)$ is the expectation of ML under $P'(\bullet)$. The likelihood ratio $L(X_{0,N}) = P(X_{0,N})/P'(X_{0,N})$ is the ratio of the sample path likelihoods under the original and the new distributions, P and P' , respectively. The use of the summation sign in the above equation is not quite precise, since the sample space is uncountable infinite, however, the summation can be interpreted as an integral with respect to an appropriate probability measure to make this fully rigorous.

Let $f'_i(t; e)$, $\bar{F}'_i(t; e)$, $e \in E_i$ and $p'_{i+1}(Z_{i+1}, E_{i+1}; e_i^*)$ be the probability distributions, in the simulated system, corresponding to $f_i(t; e)$, $\bar{F}_i(t; e)$, $e \in E_i$ and $p_{i+1}(Z_{i+1}, E_{i+1}; e_i^*)$ in the original system. These distributions should be chosen appropriately, so as to favorably bias the dynamics of the system while making sure that the condition in Equation (3.1) is satisfied. The rules for updating the new quantities, $f'_i(\bullet\,; e)$ and $\bar{F}'_i(\bullet\,; e)$, at transition times are analogous to those for updating $f_i(\bullet\,; e)$ and $\bar{F}_i(\bullet\,, e)$ given in Equations (2.1) and (2.2). The likelihood ratio associated with the sample path $X_{0,N}$ is given by

$$L(X_{0,N}) = \prod_{i=0}^{N-1} \frac{f_i(T_i; e_i^*)}{f'_i(T_i; e_i^*)} \frac{p_{i+1}(Z_{i+1}, E_{i+1}; e_i^*)}{p'_{i+1}(Z_{i+1}, E_{i+1}; e_i^*)} \prod_{e \in E_i - \{e_i^*\}} \frac{\bar{F}_i(T_i; e)}{\bar{F}'_i(T_i; e)} . \tag{3.3}$$

The above equation is the basis for importance sampling in discrete-event systems. Rather than replicating the r.v. M ($X_{0,N}$) under P to estimate $E_P(M)$, we replicate the r.v. $M(X_{0,N}) L(X_{0,N})$ under P' to estimate $E_{P'}(ML)$, which is equal to $E_P(M)$. When P' is chosen appropriately, significant reduction in the variance of the r.v. ML, under P', can be achieved (compared to the variance of the r.v. M under P). This choice depends on the model at hand and on the measure to be estimated.

Notice that Equation (3.3) allows us to update the likelihood ratio at transition times in simple multiplicative manner. Notice also that at any transition we can actually change the values of any active (old and new) clock according to some chosen, essentially arbitrary, new distribution. This is equivalent to cancelling an active clock and rescheduling (i.e., resampling) its remaining lifetime from the new distribution. We illustrate this by the following: Suppose that clock e is activated at the $i-$th transition and that we assign a value to this clock according to the probability density function $f(\bullet\,; e)$. At the $(i+1)-$th transition, we decide to reschedule clock e, thus we assign to its remaining lifetime a new value y' according to a new probability density function $f'(\bullet\,; e)$. Further, we suppose that clock e continues to run at the $(i+2)-$th transition and it expires at the $(i+3)-$th transition. In effect, clock e has a total lifetime $T_i + y'$. According to Equation (3.3), the contribution of clock e to the likelihood ratio at the $(i+1)-$th transition is

$$\frac{\bar{F}_i(T_i; e)}{\bar{F}'_i(T_i; e)} = \frac{\bar{F}(T_i; e)}{\bar{F}(T_i; e)} = 1 .$$

Using Equations (2.1) and (2.2), the contribution at the $(i+2)-$th transition is

$$\frac{\bar{F}_{i+1}(T_{i+1}; e)}{\bar{F}'_{i+1}(T_{i+1}; e)} = \frac{\bar{F}(T_i + T_{i+1}; e)/\bar{F}(T_i; e)}{\bar{F}'(T_{i+1}; e)} .$$

Using Equations (2.1) and (2.2), the contribution at the $(i+3)-$th transition is

$$\frac{f_{i+2}(T_{i+2}; e)}{f'_{i+2}(T_{i+2}; e)} = \frac{f(T_i + y'; e)/\bar{F}(T_i + T_{i+1}; e)}{f(y'; e)/\bar{F}'(T_{i+1}; e)} .$$

It follows that the overall contribution of clock e to the likelihood ratio between the $i-$th and the $(i+3)-$th transitions is

$$\frac{f(T_i + y'; e)}{\bar{F}(T_i; e) f(y'; e)} .$$

In the above equation, notice that the numerator is simply the likelihood of the total lifetime $(T_i + y')$ of clock e under the original probability density function $f(\bullet; e)$, while the denominator is the likelihood with rescheduling.

As an example to show how importance sampling can be implemented, let us consider a machine-repairman model with two components of the same type and one single server FCFS repair facility. Each component has general failure and repair distributions. The system is initially operational, with all components as good as new, and it continues to be operational as long as at least one component is operational. In a highly dependable system, a component's mean time to failure is usually several orders of magnitude larger than its mean time to repair. Therefore, a system failure is a rare event. Consider the estimation of a dependability measure, such as the unreliability, using the replication method of simulation. Clearly, if we use standard simulation, a very large number of replications is needed to achieve a reasonably tight confidence interval. This implies a very long simulation run. Importance sampling is accomplished by biasing the dynamics of the system so as to make its typical failures occur more frequently. One possible heuristic is what we call dynamic importance sampling (DIS) [2 , 10]; it is described as follows: as soon as one of the two components fail, we accelerate the failure of the second component, either by rescheduling it using a new (accelerated) distribution or by increasing its clock rate. Increasing the failure clock rate is equivalent to rescheduling with a new distribution obtained by scaling the conditional original dis-

tribution. A reasonable heuristic choice for the new distribution is obtained by appropriately scaling the original distribution, such that the new failure "rate" is of the same order of magnitude as the repair "rate" [2]. By rescheduling the failure of the second component, we are also increasing the probability of a system failure (i.e., both components unoperational). If the second component fails while the first is in repair, we have a system failure (this is a stopping time for a replication when estimating the unreliability). If the first component is repaired before the second component fails, both components become operational and we must reschedule their failures using the original distributions. This is crucial in order to appropriately move the system only towards a likely path to failure.

4. Dependability Measures

In this section we discuss the estimation of some measures which are commonly used for the evaluation of highly dependable systems. These measures can be classified as stationary or transient. Stationary measures are determined by the long-run (or steady-state) behavior in repairable systems; they are independent of the initial state. The steady-state availability is a common stationary measure. Transient measures are determined by the transient behavior in repairable and non-repairable systems; they depend on the initial state. It is usually assumed that the system starts with all its components fully operational. The system reliability and the mean time to failure (MTTF) are common transient measures which we consider in this section. Instantaneous availability, distribution and expectation of interval availability are examples of other transient measures; the estimation of these measures in Markovian models is considered in [10].

As a running example, we will consider the machine-repairman model (described in Section 3) to explain our ideas and to numerically illustrate the effectiveness of importance sampling for the estimation of dependability measures.

4.1. System Reliability

In this section we consider the estimation of reliability in non-Markovian discrete-event systems using simulation and importance sampling. No assumptions are made concerning the distributions of time to failure and time to repair of individual system components. The system is initially in a state with all its components operational and as good as new. Let T_F be the time at which the system first enters a failure state. The system reliability $R(t)$ is defined as the probability that the system does not fail in the interval $(0, t)$, i.e.,

$$R(t) = \mathrm{P}(T_F > t) = \mathrm{E}(I(T_F > t)), \tag{4.1}$$

where $I(\bullet)$ is the indicator function. The method of replications is the typical simulation method for estimating the reliability. In each replication we simulate the system until either a failure occurs or the time interval exceeds t. Let n_r be the number of replications. The resulting estimate for the unreliability $U(t)$ is given by

$$\hat{U}(t) = 1 - \hat{R}(t) = \frac{1}{n_r} \sum_{i=1}^{n_r} I(T_{F_i} \leq t), \tag{4.2}$$

where T_{F_i} is the time to failure at the $i-$th replication and $I(T_{F_i} \leq t)$ is the value of the indicator function. Clearly, if we use standard simulation in a highly reliable system, then the value of the indicator function is zero in all but a few replications. A very large number of replications (i.e., a very long simulation) is needed in order to obtain an estimate with a tight confidence interval. Importance sampling, as described in Section 3, is very effective in improving the efficiency of such simulations. Let N_i be the stopping time in the $i-$th replication, i.e., the number of internal state transitions until either a system failure or the first transition to occur after time t. The resulting estimate is unbiased and is given by

$$\hat{U}_I(t) = \frac{1}{n_r} \sum_{i=1}^{n_r} I(T_{F_i} \leq t)\, L(\mathbf{X}_{0,N_i}), \tag{4.3}$$

where $L(\mathbf{X}_{0,N_i})$ is the likelihood ratio as given by Equation (3.3).

It is important to mention that forcing [12] can be combined with failure biasing (described in Section 3) to estimate reliability. This is particularly useful when t is small so that a failure is unlikely to occur in the interval $(0, t)$. In the machine-repairman model of Section 3, forcing is accomplished by scheduling component failures using the original distribution conditioned so that a failure is guaranteed to occur before time t. Once a failure occurs, the second component is rescheduled using an accelerated distribution. In each replication, forcing should be done every time both components become operational.

In Markovian models, conditioning out the holding time in the initial (fully operational) state [10] has also proven quite effective in improving the efficiency of simulations to estimate transient measures. However, extending this technique to systems with general failure time distributions is difficult.

To illustrate the feasibility and effectiveness of importance sampling (with rescheduling) to estimate system reliability, we consider the machine-repairman example (of Section 3) with a two-stage hyperexponential failure and repair distributions. A two-stage hyperexponential failure time is

generated from an exponential of parameter λ_1 with a probability q_f and from an exponential of parameter λ_2 with a probability $1 - q_f$. The parameters of the failure distribution are $q_f = .9$, $\lambda_1 = .001$ per hour and $\lambda_2 = .01$ per hour. We have selected relatively high failure rates so that tandard simulation could provide us with reasonable estimates for the purpose of comparison. The parameters of the repair distribution are $q_r = .9$, $\mu_1 = 1$ per hour and $\mu_2 = 10$ per hour. For importance sampling, we use an accelerated failure distribution which is the same as the original distribution with its rates scaled up (while other choices are also possible, determining the optimal is an open research problem). The parameters of the accelerated failure distribution are $q'_f = .9$, $\lambda'_1 = .5$ per hour and $\lambda'_2 = 5$ per hour.

For the interval between 0 and 10 hours, a very accurate estimate of the unreliability $U(10)$ is obtained numerically using the SAVE package [8]. For the purpose of comparison, we have also used standard simulation as well as importance sampling, each for a total of 128000 simulated events. The numerical and simulation estimates are as follows (with the 90% half-width confidence interval as a percentage of the point estimate):

Numerical: 5.775×10^{-5}

Standard simulation: $4.737 \times 10^{-5} \pm 94.98\%$

Importance sampling: $5.560 \times 10^{-5} \pm 7.89\%$

Notice that by using importance sampling we get more than 10 times improvement in the confidence interval, which is equivalent to more than 100 times reduction in the simulation run-length.

4.2. Steady-State Availability

The steady-state availability is defined as the long-run fraction of time the system is available. It is typically used as a metric for evaluating repairable systems. In Markovian models, regenerative simulations are typically used to estimate the steady-state availability [2] (the state in which all components of the system are operational is usually chosen as a regeneration point). As a consequence, a simple estimator for the steady-state unavailability UA follows from a basic result of renewal theory [1]

$$UA = \frac{\mathrm{E}(D)}{\mathrm{E}(T)} , \tag{4.4}$$

where D and T are the total "down" time and the total "cycle" time between regenerations, respectively.

Unfortunately, in non-Markovian models with general failure and repair distributions, a regenerative structure may not be present (for conditions under which a discrete-event system or a GSMP

is regenerative the reader is referred to [11]). Let us again consider the machine-repairman model with general failure and repair distributions. We consider two cases in which a regenerative structure can be recognized. In the first case, we assume that the failure time of individual components is exponentially distributed. Therefore, a regeneration point is readily identified at repair transitions after which all components become operational. In the second case, regenerations occur as a result of a periodic (and deterministic) maintenance on all components. This is true for general failure and repair distributions, since after maintenance a component is as good as new. In this case, a regeneration point is identified at the lowest common multiple of all maintenance periods, provided that no component has failed since its last maintenance. At these points, all components are operational and the conditional distribution of the time to failure of each individual component is the same for all regenerations and is conditionally independent of the past.

If a regenerative structure can be recognized in a discrete-event system, then regenerative simulation can be used to estimate the steady-state unavailability by using Equation (4.4). Let n_c be the number of regeneration cycles used. Then an estimate of UA is given by

$$\hat{UA} = \frac{\hat{D}}{\hat{T}} = \frac{\frac{1}{n_c}\sum_{i=1}^{n_c} D_i}{\frac{1}{n_c}\sum_{i=1}^{n_c} T_i}, \tag{4.5}$$

where D_i and T_i are the total "down" time and the "cycle" time in the $i-$th regeneration cycle, respectively. $\hat{D}$ and $\hat{T}$ are estimates of E(D) and E(T), respectively. For highly available systems, system failure is a rare event. Therefore, standard simulation is very inefficient for estimating the numerator E(D), since only a very small fraction of regeneration cycles will contain failures. Again, importance sampling provides an efficient solution by biasing the dynamics of the system appropriately, so that a likely path to failure is encountered more often. Notice that the denominator can be estimated efficiently using standard simulation; in fact, using importance sampling to estimate the denominator E(T) may increase its variance. Therefore, a better estimate for the steady-state unavailability can be obtained by using measure specific dynamic importance sampling (MSDIS) [9], in which we estimate the numerator E(D) using importance sampling as described in Section 3, while independenttly using standard simulation to estimate the denominator E(T). The optimal allocation of the simulation run lengths for estimating the numerator and the denominator is considered in [10]. Notice that here, regeneration times are used as stopping times. Let n_n and n_d be the number of regeneration cycles used to estimate the numerator and denominator, respectively. For the numerator, estimates for the mean and the variance are given by

$$\hat{D}_L = \frac{1}{n_n} \sum_{i=1}^{n_n} D_i \, L(\mathbf{X}_{0,N_i}),$$
$$\hat{\sigma}^2(D_L) = \frac{1}{n_n} \sum_{i=1}^{n_n} (D_i \, L(\mathbf{X}_{0,N_i}))^2 - \hat{D}_L^{\,2}, \tag{4.6}$$

where at the $i-$th replication, D_i is the value of the numerator and N_i is the stopping time. $L(\mathbf{X}_{0,N_i})$ is the associated likelihood ratio as computed from Equation (3.3). Similar equations hold for the mean and the variance of the denominator, $\hat{T}$ and $\hat{\sigma}^2(T)$, respectively, except here the likelihood ratio is identical to one (since we are using standard simulation). It follows that UA has the following estimates for its mean and asymptotic variance [10] (for large n_n and n_d):

$$\hat{UA}_L = \frac{\hat{D}_L}{\hat{T}},$$
$$\hat{\sigma}^2(\hat{UA}_L) = \frac{\hat{\sigma}^2(D_L)}{\beta \, \hat{T}^2} + (\hat{UA}_L)^2 \frac{\hat{\sigma}^2(T)}{(1-\beta)\,\hat{T}^2}, \tag{4.7}$$

with $\beta = n_n/(n_n + n_d)$.

Let us again consider the machine-repairman example in Section 4.1 to illustrate the effectiveness of importance sampling to estimate the steady-state unavailability. We change the failure time distribution to an exponential of a parameter $\lambda = .001$ per hour; this is done to obtain a regenerative system for which we can use a regenerative simulation. For importance sampling, we use accelerated failures from an exponential distribution of a parameter $\lambda' = .5$ per hour.

An accurate estimate of the unavailability UA is obtained numerically using the SAVE package [8]. We give estimates using standard simulation and importance sampling, each for a total of 128000 simulated events. The results are as follows (with the 90% half-width confidence interval as a percentage of the point estimate):

Numerical: 1.799×10^{-6}

Standard simulation: $1.623 \times 10^{-6} \pm 29.70\%$

Importance sampling: $1.617 \times 10^{-6} \pm 2.61\%$

Again, we get more than 10 times improvement in the confidence interval by using importance sampling. In Section 6 we present experimentation results for estimating the steady-state unavailability in a machine-repairman model with periodic maintenance and in a large model of a computing system. For some experiments, we select typical failure rates in the range of 10^{-5} to 10^{-6}. In this range, standard simulation produces meaningless results, while the estimates obtained using importance sampling converge as quickly as those in the above example.

4.3. Mean Time to Failure (MTTF)

MTTF is typically thought of as a transient measure, since it depends on the initial state of the system. Assuming that the system is initially in a state with all its components operational and as good as new, the MTTF is defined as the expected time the system first enters a failure state. The replication method of simulation is typically used to estimate the MTTF. Again, standard simulation of highly dependable systems means very long replications and, hence, excessively long simulation runs. When the replication method of simulation is used, importance sampling may actually increase the variance of the MTTF estimate; this is because a likely sample path to failure in the biased system is, roughly, much shorter (in terms of simulated time) than a likely sample path in the original system.

If the initial state of the system is a regeneration point, then a ratio representation for the MTTF is possible [17],

$$MTTF = \frac{E(\tau)}{P(T_f < T)} , \tag{4.8}$$

where $\tau\ (= \min(T_F, T))$ is the minimum of the time to system failure (T_F) and the cycle time (T). $P(T_f < T)$ is the probability that a system failure occurs before a regeneration. In the above ratio representation, both the numerator and the denominator can be estimated using regenerative simulations. The numerator $E(\tau)$ can be estimated efficiently using standard simulation. However, in highly dependable systems, the denominator $P(T_f < T)$ is a very small quantity; hence, it can be estimated much more efficiently using importance sampling. Here also, MSDIS is recommended for estimating the MTTF, in which the numerator and the denominator are simulated independently.

Unfortunately, a regenerative structure may not be exhibited in a general discrete-event system; this limits the validity of the ratio representation for the MTTF, and hence the use of importance sampling, to only those systems in which the initial state is a regeneration point.

Let us again consider the machine-repairman model with general failure and repair distributions. In Section 4.2 we have recognized two cases in which the system exhibits a regenerative structure. In particular, if the time to failure of individual components is exponentially distributed, then the initial state, with all components operational, is a regeneration point. In this case, the ratio representation of the MTTF is valid and importance sampling can be used to estimate $P(T_F < T)$.

Again, the heuristic for importance sampling is as described in Section 3, except that here, the stopping time is either the regeneration time or the time to system failure, whichever occurs first.

Let n_d be the number of cycles used, and N_i be the stopping time in the $i-$th regeneration cycle. The resulting estimate $\hat{PF}_L$ of $P(T_F < T)$ is given by

$$\hat{PF}_L = \frac{1}{n_d} \sum_{i=1}^{n_d} I(T_{F_i} < T_i)\, L(\mathbf{X}_{0,N_i}), \tag{4.9}$$

where $I(T_{F_i} < T_i)$ and $L(\mathbf{X}_{0,N_i})$ are the indicator function and the likelihood ratio (from Equation (3.3)), respectively, evaluated in the $i-$th regeneration cycle. The estimate $\hat{\tau}$ of $E(\tau)$ is obtained independently using standard simulation. The resulting estimates for the mean and variance of the MTTF are computed from equations similar to Equations (4.6) and (4.7) for the steady-state unavailability UA.

Here also we consider the machine-repairman example in Section 4.2 to illustrate the effectiveness of importance sampling to estimate the MTTF. Notice that the failure time distribution is assumed to be exponential with a parameter $\lambda = .001$ per hour. We obtain a regenerative system for which the ratio representation is valid; thus we can use regenerative simulation and importance sampling. Again, we use accelerated failures from an exponential distribution of a parameter $\lambda' = .5$ per hour.

An accurate estimate of the MTTF is obtained numerically using the SAVE package [8]. In the following we also give estimates using standard simulation and importance sampling, each for a total of 128000 simulated events (with the 90% half-width confidence interval as a percentage of the point estimate):

Numerical: 5.510×10^5

Standard simulation: $6.039 \times 10^5 \pm 22.60\%$

Importance sampling: $5.450 \times 10^5 \pm 1.96\%$

We obtain more than 10 times improvement in the confidence interval by using importance sampling.

5. Implementation Issues

In this section we consider the implementation of the variance reduction techniques described in the previous sections. We have implemented these techniques using CSIM [15,16], which is a process-oriented simulation language based on the C programming language. In a process-oriented simulation, a model is defined as a collection of interacting processes. Each process is an independent program which runs in parallel with the other processes, with a main program synchronizing all of the processes and controlling the interactions between them. For example, in the reliability system simulations which we consider here, a separate process is created for each indi-

vidual component of the system. Each process simulates the failures and repairs of its respective component. In our models solved with CSIM, we only consider steady state unavailability, which we estimate using regenerative simulation.

We define an *up cycle* to be a segment of the sample path between two successive times when a component comes out of repair or scheduled maintenance and finds all other components operational. As we will see later, there may be more than one up cycle in a regenerative cycle.

In models of highly reliable systems, the repair rates of the components are typically orders of magnitude larger than the failure rates. A method of implementing importance sampling is to reschedule events in order to bias the system towards the failed state. This is called *failure biasing*.

When using importance sampling, we want to cause the system to fail using the most likely path to failure. This suggests using the following strategy for implementing failure biasing. After the first component failure in an up cycle, we reschedule all of the other components' failure times by generating new remaining lifetimes using specified biased distributions. The biased distributions are selected so that the probability that some operating component fails before the component in repair completes service is in the range of .1 to .5, thus greatly increasing the probability of a system failure. Until either a system failure occurs or we reach the end of an up cycle, we continue to schedule all failure lifetimes using the biased failure distributions. Once we reach the end of an up cycle, we reschedule the remaining lifetimes of all components using the original failure distributions and repeat the entire process. However, if we reach the failed state during the time failure biasing is activated, we immediately reschedule all of the remaining lifetimes of the operational components using the original failure distributions and do not use failure biasing for the rest of the regenerative cycle. By doing this, we ensure that the probability of two system failures occurring in one regenerative cycle remains small. For continuous time Markov chains the discrete time conversion of the above strategy was shown to be an effective technique in [2, 9].

As an alternate approach to rescheduling failures, one can actually alter the rates at which the clocks associated with the lifetimes of the components advance. In order to implement importance sampling in a manner similar to rescheduling, we rescale, i.e., divide by a scaling factor r, the remaining lifetimes of the operational components at precisely the same instances at which clocks were rescheduled when using the rescheduling technique. The advantage of rescaling clocks is that new random lifetimes do not have to be generated. In our implementation, we actually altered the repair clock rate instead of altering all of the failure clock rates. Since we are assuming that there is only one repairman, this allows us to reschedule only one event, hence saving computational effort. In order to avoid numerical problems with the likelihood ratio, we pretended that we actually changed the failure clock rates and did nothing to the repair clock rate. The resulting likelihood ratio is

exactly the same as in the rescheduling case when using scaled conditional distributions for the biased distributions.

In the experimental results discussed below, the rescaling technique was used to implement importance sampling in all of the models except for the maintenance model, in which rescheduling was used. It should be noted that the amount of CPU time needed to simulate a fixed number of events using importance sampling took an extra 10% to 150% over standard simulation, depending on the size of the model solved and the importance sampling implementation used. However, the extra computation time needed is due to special "tricks" we had to use in CSIM in order to adjust the event list of the simulator, and in a different implementation where we are able to directly access the event list, there would be minimal extra cost. This observation is supported by experiments in [10] when using importance sampling for simulating Markovian models.

6. Examples and Discussions

In this section we use three examples to illustrate the effectiveness of the proposed importance sampling techniques. First, we use a small example to experiment with some heuristics for selecting the new probability distributions which make the typically rare system failures occur more often. Second, these heuristics are applied to a model of a fairly complex computing system to demonstrate that the methods described in this paper are effective and that orders of magnitude reduction in variance can be obtained in simulations of large models. We also show that the relative accuracy of our estimate of unavailability when using our importance sampling technique is independent of the magnitude of the unavailability. We use exponential failure and repair distributions in this example to ascertain the correctness of the results obtained by comparing them against numerical results obtained from the SAVE package [7]. In the third example, we use Weibull failure distribution and periodic maintenance for all individual components in the system. We study the effect of the hazard rate (i.e., increasing, decreasing and constant failure rates) on the optimal maintenance period. Such studies cannot be performed with existing analytical or numerical methods.

6.1. Effects of Different Biased Failure Distributions

In this section, we use a small model to analyze the behavior of the variance when using our importance sampling technique. In particular, we examine the effect on the stability and magnitude of the estimated variance from how much we bias the system towards failure when using importance sampling.

The model consists of two types of components, each having a redundancy of two. The failure distributions of the components are exponential, with the failure rate denoted by λ. There is one repairman, who services failed components in a FCFS fashion, with repair times being exponentially distributed and repair rate $\mu = 1$.

We now examine the effect on the amount of variance reduction gained and the stability of the variance by choosing different scaling factors r. Table 1 contains the values of the variance of the amount of down time in a regenerative cycle after a specified number of simulated events for various values of r when $\lambda = 10^{-3}$, and Table 2 contains similar results when $\lambda = 10^{-5}$. If the probability of system failure is small, then the variance of the down time is the dominating term in the expression for the variance of steady state unavailability when estimated as a ratio (see Equation (4.7)). By choosing the scaling factor r such that $\mu/10 < r(n-1)\lambda \leq \mu$, where n is the total number of components in the system, we obtain stable estimates of the variance quickly. Also, for r in this range, we obtain the largest amount of variance reduction. It is also interesting to point out that if we choose r too large, the variance actually starts to increase and becomes less stable. The increase is caused by the added variability in the likelihood ratio. The above experiment was a useful guide in selecting the scaling factor for larger models.

6.2. A Large Model

In this section, we provide empirical results from a large model, showing that the methods described in this paper are also feasible and effective for larger systems than the ones described above. Also, we demonstrate that the relative size of the confidence intervals when using importance sampling is independent of the magnitude of the unavailability of the system, as long as a system failure is still a rare event. The system we will examine is based on a model of a fairly complex computing system (also considered in [13]), with its block diagram shown in Figure 1. The computing system is composed of two sets of processors with 2 processors per set, two sets of controllers with 2 controllers per set, and 6 clusters of disks, each consisting of 4 disk units. In a disk cluster, data is replicated so that one disk can fail without affecting the system. The "primary" data on a disk is replicated such that one third is on each of the other three disks in the same cluster. Thus, one disk in each cluster can be inaccessible without losing access to the data. The connectivity of the system is shown in Figure 1. All failure time distributions and repair time distributions are exponential. We examine the model under two different sets of failure rates in order to show that the relative width of the confidence interval is insensitive to the magnitude of the unavailability. In the first set, the failure rates of processors, controllers and disks are assumed to be 1/2000, 1/2000 and 1/6000 per hour, respectively. These rates are much larger than one typically would find in the real world,

but we chose these values so that we could obtain stable estimates of both the unavailability and its variance using standard simulation in a reasonable amount of time. In the second set, we divide all of the failure rates by 100, thus creating more realistic failure rates and causing the unavailability to be even smaller. The repair rates for all components is 1 per hour. Components are repaired by a single repairman who repairs the components in a FCFS discipline. The system is defined to be operational if all data is accessible to both processor types, which means that at least one processor of each type, one controller in each set, and 3 out of 4 disk units in each of the 6 disk clusters are operational. We also assume that operational components continue to fail at the given rates when the system is failed.

Since all failure and repair time distributions are exponential, the resulting system is a continuous time Markov chain. We designed the system in this manner so that we could obtain numerical (non-simulation) results for the unavailability using the SAVE package [7, 8]. Since the system has a few hundred thousand states, only bounds could be computed [13]. These bounds are very tight and typically do not differ from the exact results significantly.

In Table 3, we have the estimates of unavailability and their 90% confidence intervals for the different sets of failure rates when using standard simulation and importance sampling after 1,024,000 simulated events. When using importance sampling, the scaling factor r was selected in a manner analogous to the results from the small model example given in Section 6.1. The first row of the table contains the results from using the first set of failure rates. The width of the confidence interval is reduced by a factor of 3.6 by using importance sampling over standard simulation, which translates into a 13-fold improvement in run length. The results from using the second set of failure rates are given in the second row of the table. The results from standard simulation are meaningless because the variance had not yet stabilized by the end of the simulation. However, the results from using importance sampling are quite accurate, with the size of the relative 90% confidence interval being the same as that with the first set of failure rates when using importance sampling. Thus, our importance sampling technique is relatively independent of the magnitude of unavailability, and as the occurrence of a system failure becomes rarer, the amount of improvement gained by using our importance sampling technique over standard simulation increases, which is a favorable conclusion.

6.3. A Study of Effects of Maintenance Policies

We now demonstrate the types of studies that can be made with the aid of the importance sampling schemes described in this paper. We examine a non-Markovian model with scheduled periodic (deterministic) maintenances and determine the effect of varying the length of time between main-

tenances when component lifetime distributions have increasing failure rate (IFR), constant failure rate, and decreasing failure rate (DFR). Because of the complexity of the model, analytic results are extremely difficult to obtain. Also, as we will see, since system failures occur very rarely, standard simulation is very inefficient, and importance sampling is the only practical alternative.

We consider a simple maintenance model consisting of one type of component with a redundancy of two. The distribution of the lifetime of each component is Weibull, with shape parameter α and scale parameter β. Recall that if $\alpha = 1$, the component lifetime distributions are exponential. Also, if $\alpha > 1$, the distribution has an increasing failure (hazard) rate, and we have a decreasing failure rate distribution if $\alpha < 1$. In our experiments, we fixed $\beta = 10^{-5}$ and varied α. There is one repairman who fixes failed components in FCFS fashion. The length of the repair times is the sum of a constant c plus an exponentially distributed random quantity with rate μ. The constant c corresponds to the travel time of the repair man. In all of our simulations, c was 2.0 hours and the repair rate μ was 0.5 per hour. In addition, each component has a periodic scheduled maintenance every d hours, where d is deterministic. One component has its first scheduled maintenance at the beginning of the simulation, and the other component has its first scheduled maintenance after $d/2$ simulated hours have passed. Thus, the maintenance cycles of the two components are staggered. All scheduled maintenances take 0.5 hours. Also, after a component comes out of repair from a failure, the next scheduled maintenance is skipped, and a maintenance is performed on a component only if the other component is operational. A component is considered to be as good as new immediately after completing a scheduled maintenance. There is a single repairman, different from the one who repairs failed components, who performs scheduled maintenances. The system is considered operational if at least one component is operational, i.e., not failed or in scheduled maintenance.

Figure 2 shows a plot of the unavailability versus the time between maintenances (d) for the different values of α. The graph was constructed by running simulations using the different parameter values, plotting the point estimates, and using linear interpolation between the points. We ran all the experiments long enough so that the relative half-width of the 90% confidence interval was less than 10%. It is interesting to note how smooth the curves are for each of the value of α, thus demonstrating the effectiveness of our importance sampling technique. Also note that as $d \to \infty$, the system becomes equivalent to one without scheduled, periodic maintenances. This is demonstrated by observing that the curve for $\alpha = 1.0$ is beginning to flatten out for $d > 1000$.

The curves show that when component lifetimes have exponential or DFR distributions, performing scheduled maintenances actually increases the unavailability of the system. When $\alpha = 1.0$, the component lifetime distributions have constant failure rate, which means that the conditional dis-

tribution of a failure given that it is greater than t does not depend on t. Thus, a component's reliability does not improve by performing a maintenance on it. Actually, performing scheduled maintenances increases the system's unavailability, which can be explained as follows. Since a scheduled maintenance for a component takes a deterministic amount of time, the conditional probability of the other component failing during the maintenance time given that it has lived, say s units of time already, is the same for all values of s. Thus, by decreasing the time between maintenances, we are increasing the frequency with which the system can fail by having a maintenance and then a failure occurring. This, in turn, leads to the higher unavailability. We see similar results when $\alpha = 0.75$. However, since the component lifetimes now have DFR distributions, the effect is more pronounced. This is because the conditional probability that a component fails given that it has already lived t units of time is a decreasing function of t. Hence, by decreasing the time between scheduled maintenances, we not only increase the frequency with which the system can fail by having one component in maintenance and the other failing during the maintenance, but also the conditional probability of the operational component failing during the maintenance of the other component also increases. Thus, one should not perform scheduled maintenances on systems of components with DFR distributions. When $\alpha = 1.25$, the components have IFR lifetime distributions. In this case, the unavailability is large for small values of d, attains its minimum around $d = 500$, and then increases. When $\alpha = 1.5$, the unavailability behaves in a similar manner, with its minimum being attained around $d = 100$. Hence, in a maintainable system composed of components having IFR lifetime distributions, scheduled maintenances should be performed more frequently at higher component failure rates.

7. Summary

In this paper we have described an approach for simulating models of highly dependable systems with general failure and repair time distributions. The approach combines importance sampling with event rescheduling in order to obtain variance reduction in such rare event simulations. The approach is general in nature and allows us to effectively simulate a variety of features commonly arising in dependability modeling. For example, in this paper we have shown how the technique can be applied to systems with periodic maintenance. We have explored how the steady-state availability is affected by the maintenance period and by different failure time distributions.

We described some of the trade-offs involved in the design of specific rescheduling rules, and demonstrated their potential effectiveness in simulations of systems with both exponential, and non-exponential failure and repair time distributions. We found that an effective method for selecting the rescheduling distribution is by making the probability of a failure transition in the range from

0.1 to 0.5. In addition, we used a rescaling of clock values as an inexpensive way to implement rescheduling. While this can be effective when the clocks have nearly constant hazard rates, different rescheduling algorithms may be required when the clock densities are more general. The use of importance sampling for estimating steady-state availability and MTTF requires that the underlying model of the system has a regenerative structure. This requires either exponential failure distributions or general failure distributions with periodic (deterministic) maintenance. On the other hand, the use of importance sampling for estimating transient measures, such as reliability, is completely general and does not require any assumption on the failure and the repair processes.

We are currently in the process of implementing importance sampling for estimating reliability, MTTF and interval availability in large models (here, we have only experimented with the estimation of these measures in small models). We are also working on the problem of estimating the gradient of dependability measures in non-Markovian models using importance sampling.

References

[1] Cinlar, E. (1975). *An Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, New Jersey.

[2] Conway, A.E. and Goyal, A. (1987). Monte Carlo Simulation of Computer System Availability/Reliability Models. *Proceedings of the Seventeenth Symposium on Fault-Tolerant Computing*, Pittsburgh, Pennsylvania, 230-235.

[3] Dugan, J.B., Trivedi, K.S., Smotherman, M.K. and Geist, R.M. (1986). The Hybrid Automated Reliability Predictor. *Journal of Guidance, Control, and Dynamics*. **9**, 3, 319-331.

[4] Geist, R.M. and Trivedi, K.S. (1983). Ultra-High Reliability Prediction for Fault-Tolerant Computer Systems. *IEEE Transactions on Computers* **C-32**, 1118-1127.

[5] Glynn, P.W. (1989). A GSMP Formalism for Discrete Event Systems. *Proceedings of the IEEE* **77**, 1, 14-23.

[6] Glynn, P.W. and Iglehart, D.L. (1989). Importance sampling for Stochastic Simulations. To appear in *Management Science*.

[7] Goyal, A., Carter, W.C., de Souza e Silva, E., Lavenberg, S.S., and Trivedi, K.S. (1986). The System Availability Estimator. *Proceedings of the Sixteenth Symposium on Fault-Tolerant Computing*. Vienna, Austria, 84-89.

[8] Goyal, A. and Lavenberg, S.S. (1987). Modeling and Analysis of Computer System Availability. *IBM Journal of Research and Development*, **31**, 6, 651-664.

[9] Goyal, A., Heidelberger, P. and Shahabuddin, P. (1987). Measure Specific Dynamic Importance Sampling for Availability Simulations. *1987 Winter Simulation Conference Proceedings*. A. Thesen, H. Grant and W.D. Kelton (eds.). IEEE Press, 351-357.

[10] Goyal, A., Shahabuddin, P., Heidelberger, P., Nicola, V.F. and Glynn, P. (1988). A Unified Framework for Simulating Markovian Models of Highly Dependable Systems. IBM Research Report RC 14772, Yorktown Heights, New York.

[11] Haas, P.J. and Shedler, G.S. (1987). Regenerative Generalized Semi-Markov Processes. *Commun. Statist.-Stochastic Models* **3**, 3, 409-438.

[12] Lewis, E.E. and Böhm, F. (1984). Monte Carlo Simulation of Markov Unreliability Models. *Nuclear Engineering and Design* 77, 49-62.

[13] Muntz, R.R., de Souza e Silva, E. and Goyal, A. (1989). Bounding Availability of Repairable Computer Systems. *Proceedings of the ACM Sigmetrics and Performance'89*. Berkeley, California.

[14] Nicola, V.F., Bobbio, A. and Trivedi, K.S. (1989). A Unified Performance Reliability Analysis of a System with a Cumulative Down Time Constraint. IBM Research Report, Yorktown Heights, New York.

[15] Schwetman, H. (1986). CSIM: A C-Based, Process-Oriented Simulation Language. *1986 Winter Simulation Conference Proceedings*. J. Wilson, J. Henriksen, and S. Roberts (eds.). IEEE Press, 387-396.

[16] Schwetman, H. (1988). Using CSIM to Model Complex Systems. *1988 Winter Simulation Conference Proceedings*. M.A. Abrams, P.L. Haigh and J.C. Comfort (eds.). IEEE Press, 491-499.

[17] Shahabuddin, P., Nicola, V.F., Heidelberger, P., Goyal A. and Glynn P.W. (1988). Variance Reduction in Mean Time to Failure Simulations. *1988 Winter Simulation Conference Proceedings*. M.A. Abrams, P.L. Haigh and J.C. Comfort (eds.). IEEE Press, 491-499.

[18] Stiffler, J. and Bryant, L. (1982). CARE III Phase III Report-Mathematical Description. NASA Contractor Report 3566.

Events	$r = 10$	$r = 10^2$	$r = 3.3 \times 10^2$	$r = 10^3$	$r = 10^4$
1000	5.419×10^{-4}	1.764×10^{-5}	1.745×10^{-5}	3.218×10^{-5}	3.617×10^{-6}
4000	2.335×10^{-4}	2.654×10^{-5}	1.643×10^{-5}	1.318×10^{-5}	1.845×10^{-6}
16000	2.138×10^{-4}	2.877×10^{-5}	1.466×10^{-5}	2.380×10^{-5}	1.608×10^{-5}
32000	2.114×10^{-4}	2.888×10^{-5}	1.600×10^{-5}	7.072×10^{-5}	1.388×10^{-5}
64000	2.093×10^{-4}	2.896×10^{-5}	1.579×10^{-5}	1.189×10^{-4}	1.288×10^{-5}
96000	2.094×10^{-4}	2.708×10^{-5}	1.855×10^{-5}	9.191×10^{-5}	2.938×10^{-4}
128000	2.032×10^{-4}	2.840×10^{-5}	1.935×10^{-5}	7.876×10^{-5}	2.212×10^{-4}

Table 1: Estimated variance of down time in a cycle for 9-state model ($\lambda = 10^{-3}$) using different scaling factors r

Events	$r = 10$	$r = 10^2$	$r = 10^3$	$r = 10^4$	$r = 3.3 \times 10^4$	$r = 10^5$
1000	N/A	N/A	5.232×10^{-9}	1.772×10^{-9}	1.767×10^{-9}	2.971×10^{-9}
4000	N/A	3.557×10^{-8}	1.463×10^{-8}	2.750×10^{-9}	1.677×10^{-9}	1.513×10^{-9}
16000	N/A	1.528×10^{-7}	2.020×10^{-8}	2.668×10^{-9}	2.391×10^{-9}	2.112×10^{-9}
32000	N/A	1.987×10^{-7}	1.887×10^{-8}	2.829×10^{-9}	3.071×10^{-9}	7.025×10^{-9}
64000	9.168×10^{-6}	2.107×10^{-7}	1.918×10^{-8}	2.903×10^{-9}	2.406×10^{-9}	5.537×10^{-9}
96000	6.647×10^{-6}	2.088×10^{-7}	1.914×10^{-8}	2.729×10^{-9}	2.299×10^{-9}	4.936×10^{-9}
128000	4.986×10^{-6}	1.846×10^{-7}	1.960×10^{-8}	2.879×10^{-9}	2.326×10^{-9}	5.495×10^{-9}

Table 2: Estimated variance of down time in a cycle for 9-state model ($\lambda = 10^{-5}$) using different scaling factors r

Numerical Result	Standard Simulation	Importance Sampling	Scaling Factor r
4.055×10^{-6}	4.121×10^{-6} ±13.6%	4.124×10^{-6} ±3.8%	10^2
4.000×10^{-10}	6.165×10^{-10} ±159.5%	4.027×10^{-10} ±3.7%	10^1

Table 3: Estimates of unavailability and 90% confidence intervals for a large model (1,024,000 events)

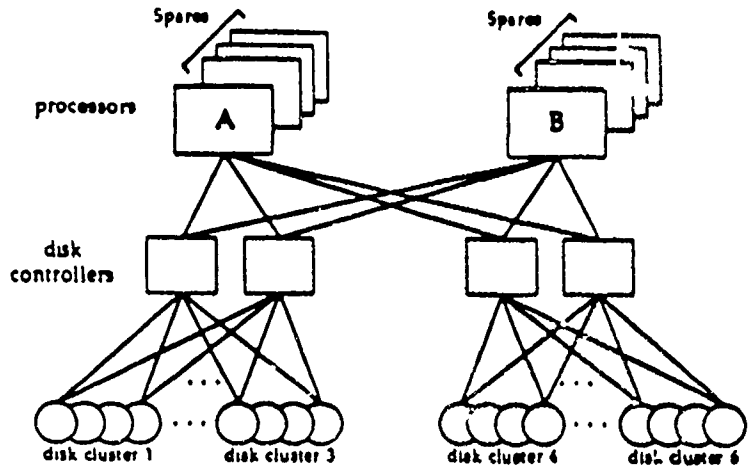

Figure 1. A block diagram of the computing system modeled

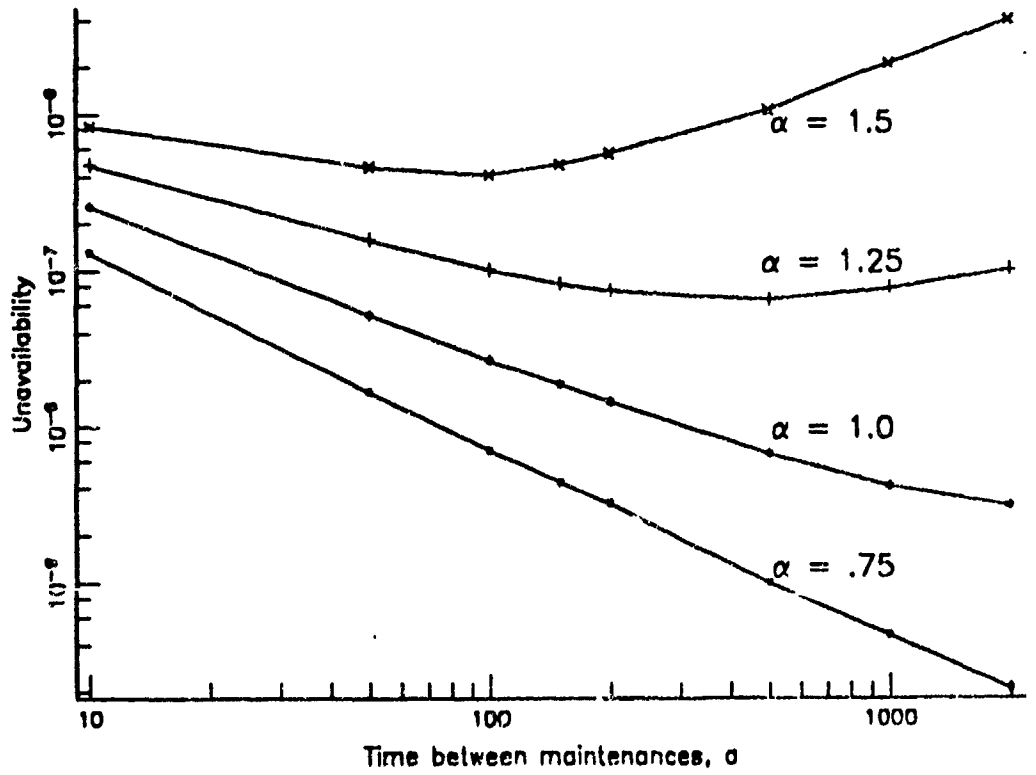

Figure 2. Effect of maintenance on system unavailability.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

MASTER COPY - FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS:
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): Technical Report No. 53	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARO 25839.29-MA
6a. NAME OF PERFORMING ORGANIZATION: Dept. of Operations Research; 6b. OFFICE SYMBOL (If applicable):	7a. NAME OF MONITORING ORGANIZATION: U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code): Stanford, CA 94305-4022	7b. ADDRESS (City, State, and ZIP Code): P. O. Box 12211, Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U. S. Army Research Office; 8b. OFFICE SYMBOL (If applicable):	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAAL03-88-K-0063
8c. ADDRESS (City, State, and ZIP Code): P. O. Box 12211, Research Triangle Park, NC 27709-2211	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Fast simulation of dependability models with general failure, repair and maintenance processes

12. PERSONAL AUTHOR(S)
Victor F. Nicola, Marvin K. Nakayama, Philip Heidelberger, Ambuj Goyal

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	January 1990	24

16. SUPPLEMENTARY NOTATION
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
fast simulation, dependability models, importance sampling, non-Markovian models, periodic maintenance

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

(please see next page)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

FAST SIMULATION OF DEPENDABILITY MODELS WITH GENERAL FAILURE, REPAIR AND MAINTENANCE PROCESSES

Victor F. Nicola *, Marvin K. Nakayama **,
Philip Heidelberger * and Ambuj Goyal *

* IBM Research Division
T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, New York 10598

** Department of Operations Research
Stanford University
Stanford, California 94305

ABSTRACT

The problem of computing dependability measures of repairable systems with general failure, repair and maintenance processes is a hard problem to solve in general either by analytical or by numerical methods. Monte Carlo simulation could be used to solve this problem, however, standard simulation takes a very long time to estimate system reliability and availability with reasonable accuracy because typically the system failure is a rare event. When the failure and repair time distributions are exponential, *importance sampling* has been used successfully in the past to reduce simulation run lengths. In this paper, we extend the applicability of *importance sampling* to non-Markovian models with general failure and repair time distributions. We show that by carefully selecting a heuristic for importance sampling, orders of magnitude reduction in simulation run-lengths can be obtained. We illustrate the effectiveness of the technique by modelling a large repairable computing system. Also, we study the effect of periodic maintenance on systems with components having increasing and decreasing failure rate.