Fast Simulation of Highly Dependable Systems with General Failure and Repair Processes

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

Abstract—Computing dependability measures of repairable systems with general failure and repair processes is difficult, in general, using either analytical or numerical methods. Monte Carlo simulation could be used to solve this problem; however, in highly dependable systems 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 all failure and repair time distributions are exponential, importance sampling has been used successfully 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 study the effect of periodic maintenance on systems with components having increasing and decreasing failure rate. Also, we study the effect of the component's lifetime distribution on the component's redundancy needed to satisfy a particular system reliability goal. For general failure and repair processes, such studies are not possible with any of the conventional methods known from the literature; simulation is the only alternative and importance sampling makes it feasible.

Index Terms—Dependability evaluation, discrete-event systems, highly reliable systems, importance sampling, rare event simulation, variance reduction.

I. INTRODUCTION

FAULT TOLERANCE and high dependability are essential requirements in many critical systems used in application areas such as aerospace systems, nuclear power plants, communications networks, transaction processing, and automated manufacturing systems. These critical systems are usually characterized by their stringent requirement of high reliability and/or high availability. Therefore, in such systems, the prediction and evaluation of relevant dependability measures is of utmost importance for their efficient design and safe operation. Realistic models of these systems are usually difficult to construct and often hard to analyze using conventional methods.

Manuscript received July 2, 1991; revised January 16, 1992 and July 27, 1992. This work is based on "Fast Simulation of Dependability Models with General Failure, Repair and Maintenance Processes" by V. F. Nicola, M. K. Nakayama, P. Heidelberger, and A. Goyal, which appeared in *Proceedings of the 20th International Symposium on Fault-Tolerant Computing*, Newcastle upon Tyne, U.K., June 1990, pp. 491–498. ©1990 IEEE.

V. F. Nicola is with Tele-Informatics and Open Systems, University of Twente, 7500 AE Enschede, The Netherlands, on leave from IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

M. K. Nakayama, P. Heidelberger, and A. Goyal are with IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598. IEEE Log Number 9208564.

Repairable systems with general failure, repair, and maintenance processes, in general, cannot be modeled by Markov or even semi-Markov processes. However, there has been a significant and continuing effort to develop numerical techniques to deal with special cases. HARP [4] solves large models of nonrepairable systems with general failure time distributions (transient recoveries are approximated by instantaneous transitions). The technique estimates only transient measures (e.g., reliability) by creating a nonhomogeneous Markov chain model of the system and solving the corresponding differential equations numerically. CARE-III [26] uses numerical integration methods to solve similar models. A hierarchical numerical approach to evaluate the reliability of a certain class of systems is described in [15]. In this approach, the failure/repair behavior of components is described by a semi-Markov model. Another approach to model systems with components having decreasing failure rate distributions is provided in [14].

In this paper, we model systems with general failure, repair, and maintenance processes, and solve them for 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 [20]. However, these methods are limited to small models and are prone to unboundable numerical errors.

A common alternative is to use Monte Carlo simulation. The advantage of this approach is that arbitrary system details can be modeled realistically, and, furthermore, not all system states need to be generated. Its disadvantage is that standard simulation takes a very long time to estimate dependability measures with reasonable accuracy because system failure events are very rare in highly dependable systems [6]. When the failure and repair time distributions are exponential, the importance sampling technique has been used successfully to reduce simulation run lengths significantly [2], [12], [16], [19], [24]. 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 a correcting factor known as 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 rates) over the probability of the same sample path under the new distributions (e.g., with the new failure rates).

For various applications, techniques have been developed to speed up simulations involving rare events. An approach, other than importance sampling, called fault-injection simulation, was used in [17] to estimate a stationary measure affected by rare failure events. The theory of large deviations has been used to devise effective importance sampling techniques to estimate buffer overflow probabilities in queueing networks [22]. In the present paper, which is based on our preliminary work in [21], we extend the applicability of importance sampling to highly dependable systems with general failure, repair, and maintenance processes.

For general discrete-event systems, the theory of importance sampling has been discussed in [7] and [8]. However, 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 way to appropriately implement importance sampling, which we propose and use in this paper, is accomplished by scheduling (or canceling and rescheduling previously scheduled) events using distributions other than those in the original system. For example, when one component fails in a system with two redundant components, we speed up the failure of the other component so that it fails with high probability before the repair of the first component, thus causing a system failure. This involves canceling the originally scheduled failure event for the second component and rescheduling it using a new distribution with a smaller mean time to failure. For the same system, when estimating the reliability, we can speed up the first (component) failure in each replication so as to make it occur before the time horizon expires; again, this can be accomplished by initially scheduling the failures of both components using a new distribution with a smaller mean time to failure. Another way to implement importance sampling for estimating reliability in semi-Markov models is described in [5].

In Sections II and III we present a concise description of discrete-event systems; this allows us to formally represent the probability of a sample path and the likelihood ratio, which is the key to importance sampling. A simple example of a two-component system is used to describe failure biasing, which is basic to the implementation of importance sampling in simulations of highly dependable systems. In Section IV, we present the basic estimators for some commonly used measures in highly dependable systems, such as reliability and steady-state availability. Again, the two-component system is used to explain these measures as well as the importance sampling techniques employed to estimate them. Among these techniques are new forcing methods, described in Section IV-A, which are very useful when used in conjunction with failure biasing for reliability estimation (particularly for small time horizons). Failure biasing is also used in the estimation of steady-state availability, which is discussed in Section IV-B. In Section V, we discuss the implementation of these methods in a simulation tool which we used to evaluate large models. In Section VI, we consider several examples to experiment with heuristics and to illustrate the effectiveness of the proposed importance sampling techniques for large models and general failure and repair processes. Assuming

Weibull failure distributions in a system whose components are maintained periodically, we study the effect of the hazard rate (i.e., increasing, decreasing, and constant failure rates) on the optimal maintenance period. By varying the parameters of the Weibull failure distribution, we also study the effect of the component's lifetime distribution on the component's redundancy needed to guarantee a desired level of system reliability. Such studies cannot be performed with existing analytical or numerical methods. We conclude and give directions for future research in Section VII.

II. 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 in [7], where a generalized semi-Markov process (GSMP) formalism of such systems is presented. Here, we present an alternative concise description of discrete-event systems, which is appropriate and sufficient for our purpose. A discrete-event system is characterized by a set of events E that can trigger transitions of its state and a set Z of integer-valued output state vectors (Z is possibly a countably infinite set). With each event $e \in E$, 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, the set of active clocks (i.e., the set of events that can trigger a transition to another internal state) and the associated clock readings. Upon the i th transition, let $Z_i \in Z$ be the output state vector and $E_i \subseteq E$ be the set of active clocks; $c_i(E_i)$ is a vector with the associated clock readings. Then $X_i = (Z_i, c_i(E_i))$ is the internal state of the 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 [7]). 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 $Z(t) = Z_k$ if $t_k \leq t < t_{k+1}$. We only consider the internal state sequence at transition times, since this is sufficient to determine the output state trajectory of the system. A sample path of the discrete-event system up to the n th transition is denoted by the sequence $X_{0,n} = (X_0, X_1, \dots, X_n)$ of internal states at transition times. Let $e_i^* = \arg \min_{e \in E_i} \{c(e)\}$, $i \geq 0$.

Then e_i^* is the clock that triggers the $(i+1)$ th transition and $T_i = c(e_i^*)$. The internal state $\mathbf{X}_{i+1} = (\mathbf{Z}_{i+1}, \mathbf{c}_{i+1}(\mathbf{E}_{i+1}))$ 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. Let $\mathbf{A}_i$ be the set of clocks canceled (aborted) upon the $(i+1)$ th transition and $\mathbf{N}_{i+1}$ be the set of new clocks activated upon the $(i+1)$ th transition, then the set of active clocks $\mathbf{E}_{i+1}$ upon the $(i+1)$ th transition is determined by $\mathbf{E}_{i+1} = \mathbf{E}_i - \mathbf{e}_i^* - \mathbf{A}_i + \mathbf{N}_{i+1}$, $i > 0$, where the $+$ and the $-$ are set operations (for example, the set $A - B = \{x: x \in A, x \notin B\}$). The set of clocks $\mathbf{A}_i$ and $\mathbf{N}_{i+1}$ and the output state $\mathbf{Z}_{i+1}$ 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 $\mathbf{Z}_{i+1}$ and the active set of clocks $\mathbf{E}_{i+1}$ with a probability denoted by $p_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*)$. The subscript $i+1$ of p symbolizes the dependence on the sequence $\mathbf{X}_{0,i}$ (routing in queueing networks is an example for the use of these transition probabilities).

We denote by $f_i(t; e)$ (respectively, $\bar{F}_i(t; e)$) the probability density function (respectively, 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 $f_i(t; e) = f(t+a; e)/\bar{F}(a; e)$. Similarly, $\bar{F}_i(t; e) = \bar{F}(t+a; e)/\bar{F}(a; e)$. 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 $\bar{F}_i(t; e) = \bar{F}(t; e)$. Let $\mathbf{O}_{i+1}$ be the set of old clocks that continues 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 $e \in \mathbf{O}_{i+1}$

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

$$\bar{F}_{i+1}(t; e) = \bar{F}_i(t + T_i; e)/\bar{F}_i(T_i; e). \quad (2)$$

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, (1) and (2) 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 $P(\mathbf{X}_{i+1})$, which (by the independence of clocks) can be expressed as

$$P(\mathbf{X}_{i+1}) = f_i(T_i; e_i^*) p_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*) \prod_{e \in \mathbf{E}_i} \bar{F}_i(T_i; e). \quad (3)$$

where $\mathbf{E}_i^* = \mathbf{E}_i - \{e_i^*\}$. It follows that the likelihood of a sample path $\mathbf{X}_{0,n}$, up to the n th transition, is given by

$$P(\mathbf{X}_{0,n}) = \prod_{i=0}^{n-1} f_i(T_i; e_i^*) p_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*) \prod_{e \in \mathbf{E}_i^*} \bar{F}_i(T_i; e). \quad (4)$$

III. IMPORTANCE SAMPLING IN SIMULATIONS OF DISCRETE-EVENT SYSTEMS

The basic idea of importance sampling is to simulate the system under a different probability distribution, 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 (known as the likelihood ratio) is needed to compensate for the resulting bias.

Consider a simulation of a discrete-event system for the purpose of estimating the expected value of a particular performance measure, say M . Let $M(\mathbf{X}_{0,N})$ be the value of the measure on a sample path $\mathbf{X}_{0,N}$, where N is a stopping time relative to the internal state sequence, i.e., $I(N = i)$ is a function of $\mathbf{X}_{0,i}$ ($I(\bullet)$ 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 $\mathbf{X}_{0,N}$ is $P(\mathbf{X}_{0,N})$ as given by (4). Using importance sampling, we simulate the system by generating its sample path using a different likelihood $P'(\bullet)$, for which the following must hold for all $\mathbf{X}_{0,N}$:

$$P'(\mathbf{X}_{0,N}) > 0 \text{ whenever } M(\mathbf{X}_{0,N})P(\mathbf{X}_{0,N}) > 0. \quad (5)$$

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

$$\begin{aligned} E_P(M) &= \sum_{\forall \mathbf{X}_{0,N}} M(\mathbf{X}_{0,N}) P(\mathbf{X}_{0,N}) \\ &= \sum_{\forall \mathbf{X}_{0,N}} M(\mathbf{X}_{0,N}) L(\mathbf{X}_{0,N}) P'(\mathbf{X}_{0,N}) = E_{P'}(ML), \end{aligned} \quad (6)$$

where $E_{P'}(ML)$ is the expectation of ML under $P'(\bullet)$. The likelihood ratio $L(\mathbf{X}_{0,N}) = P(\mathbf{X}_{0,N})/P'(\mathbf{X}_{0,N})$ is the ratio of the sample path likelihoods under the original and the new distributions, P and P' , respectively. In the above equation, the summation should be interpreted as an integral with respect to an appropriate probability measure.

Let $f'_i(t; e)$, $\bar{F}'_i(t; e)$, $e \in \mathbf{E}_i$ and $p'_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*)$ be the new probability distributions, in the simulated system, corresponding to $\bar{F}_i(t; e)$, $e \in \mathbf{E}_i$ and $p_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*)$ in the original system. These distributions should be chosen so as to favorably bias the dynamics of the system while making sure that the condition in (5) is satisfied. The rules for updating $f'_i(\bullet; e)$ and $\bar{F}'_i(\bullet; e)$ at transition times are analogous to those in (1) and (2) for updating $f_i(\bullet; e)$ and $\bar{F}_i(\bullet; e)$. The likelihood ratio associated with the sample path $\mathbf{X}_{0,N}$ is given by

$$\begin{aligned} L(\mathbf{X}_{0,N}) &= \prod_{i=0}^{N-1} \frac{f_i(T_i; e_i^*) p_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*)}{f'_i(T_i; e_i^*) p'_{i+1}(\mathbf{Z}_{i+1}, \mathbf{E}_{i+1}; e_i^*)} \prod_{e \in \mathbf{E}_i} \frac{\bar{F}_i(T_i; e)}{\bar{F}'_i(T_i; e)}. \end{aligned} \quad (7)$$

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

Equation (7) allows us to update the likelihood ratio at transition times in a simple multiplicative manner. Notice 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 canceling 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 (7), 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 (1) and (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 (1) and (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 (product of the above contributions) is $f(T_i + y'; e)/(\bar{F}(T_i; e)f'(y'; e))$. 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 estimating a dependability measure, such as the unreliability. Using standard simulation, a very large number of replications is needed to achieve a reasonably tight confidence interval. Importance sampling is accomplished by

biasing the dynamics of the system so as to make its typical failures occur more frequently [2], [12]. One possible heuristic is failure biasing, which is described as follows: as soon as one of the two components fails, 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 distribution. A heuristic choice for the new distribution is obtained by 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. If the second component fails while the first is in repair, we have a system failure. If the first component is repaired before the second component fails, both components become operational and we reschedule their failures using the original (conditional) distributions.

IV. 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, such as the steady-state availability, are determined by the long-run (or steady-state) behavior in repairable systems and are independent of the initial state. Transient measures, such as the system reliability and the mean time to failure (MTTF), are determined by the transient behavior in repairable and nonrepairable systems; they depend on the initial state, which is usually assumed to be the state with all components fully operational. The distribution and the expectation of interval availability are also examples of transient measures, which are considered in [12] for Markovian models.

In this section we consider the estimation of the system reliability as well as the steady-state availability. (For a discussion of the estimation of MTTF in Markovian and non-Markovian models, the reader is referred to [25] and [21], respectively.) We use the machine-repairman model (described in Section III) as a running example to explain ideas and to numerically illustrate the effectiveness of importance sampling for the estimation of dependability measures. Larger and more complex models are considered in Section VI.

A. System Reliability

Let the system be initially in the 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.,

$$R(t) = P(T_F > t) = E(I(T_F > t))$$

where $I(\bullet)$ is the indicator function defined in Section III. The method of replications is typically used in simulations for estimating the reliability. In each replication we simulate the system until either a failure occurs or the simulated time exceeds t . Let n_r be the number of replications and T_{F_i} be

the time to failure in the i th replication. The resulting estimate for the unreliability $U(t) (= 1 - R(t))$ is given by

$$\hat{U}(t) = \frac{1}{n_r} \sum_{i=1}^{n_r} I(T_{Fi} \leq t).$$

In a standard simulation of a highly reliable system, 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 to obtain an estimate with a tight confidence interval. Importance sampling can be very effective in improving the efficiency of such simulations. In the i th replication, let N_i be the number of internal state transitions until either a system failure or the first transition to occur after time t ; i.e., N_i is the stopping time in the i th replication. The resulting unbiased estimate is given by

$$\hat{U}_L(t) = \frac{1}{n_r} \sum_{i=1}^{n_r} I(T_{Fi} \leq t) L(X_{0,N_i}), \quad (8)$$

where $L(X_{0,N_i})$ is the likelihood ratio as given by (7).

Forcing ([12], [16]) can be combined with failure biasing (described in Section III) to estimate reliability. This is particularly useful when t is small, in which case even a single component failure is unlikely to occur in the interval $(0, t)$. In the machine-repairman model of Section III, forcing is accomplished in each replication by initially scheduling component failures so that a failure of at least one component is guaranteed to occur before time t . Once a failure occurs, the second component is rescheduled using an accelerated distribution (i.e., failure biasing). In each replication, while forcing can be done every time both components become operational, experimental results show that most effectiveness is already gained by forcing only the first failure.

One way to force the first component failure before time t is to generate a random variate for the (conditioned) minimum lifetime of all components, given that it is less than t . The method can be further explained using the following representation of unreliability

$$U(t) = P(T_F \leq t | \tau_- < t) P(\tau_- < t) \quad (9)$$

where τ_- is the time of the first component failure (minimum lifetime). For a system with n independent components, $P(\tau_- < t)$ can be computed directly from

$$P(\tau_- < t) = 1 - \prod_{i=1}^n P(\tau_i > t)$$

where τ_i is the i th component lifetime. Therefore, in (9), $P(\tau_- < t)$ can be treated as a constant and we are left only with estimating the conditional unreliability $E(I(T_F \leq t | \tau_- < t))$. Effectively, the original sample space is restricted to those sample paths that contain at least one component failure before t ; this can be shown to reduce the variance of the unreliability estimator. We call this method *actual forcing*, which may not be the best way to force the first component failure. Moreover, it requires determining and sampling from the conditional distribution of the minimum lifetime, which may require sophisticated random variate generation techniques.

Once the first failure is scheduled using actual forcing, the failed component is selected with a probability equal to the ratio of its hazard rate (at the time of failure) to the sum of the hazard rates of all components. Therefore, if all components have identical lifetime distributions, then any component may be selected as the failed component with equal probability.

Another practical heuristic to approximately affect forcing in each replication is to initially bias all components towards failure so that the first component failure occurs before t with a given high probability, say p_{af} . This can be accomplished by using accelerated failure distributions to schedule failures of all components in the beginning of each replication. As will be discussed in Section V, these accelerated failure distributions (initial bias level) can be determined from p_{af} . Since we have changed the original component failure distributions, the likelihood ratio must be updated according to (7). We call this method *approximate forcing*, which, according to our experimental results in Section VI, has been shown to be quite robust and very effective (about the same as actual forcing) in reducing the variance of the unreliability estimator. It has the additional advantage of being easier to implement, since we need not determine or sample from the conditional distribution of the minimum lifetime, as we must do in actual forcing.

In each replication, once the first component failure occurs (which was scheduled using actual or approximate forcing), the failure times of all other components are scheduled (rescheduled, in the case of approximate forcing) using accelerated failure distributions. This failure biasing is necessary in order to increase the probability of a system failure before t , and should continue until either the end of the replication (system failure or t expires) or all components become operational.

To illustrate the feasibility and effectiveness of importance sampling (with rescheduling) to estimate system reliability, we consider the machine-repairman example with 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 $q_f = 0.9$, $\lambda_1 = 0.001$ per hour, and $\lambda_2 = 0.01$ per hour. We have selected relatively high failure rates so that standard simulation could provide us with reasonable estimates for the purpose of comparison. The parameters of the repair distribution are $q_r = 0.9$, $\mu_1 = 1$ per hour, and $\mu_2 = 10$ per hour. For importance sampling, we use failure biasing without forcing (actual or approximate). The accelerated failure distribution is the same as the original distribution with its rates scaled up. (While other choices are also possible, determining the optimal accelerated distribution is an open research problem.) The parameters of the accelerated failure distribution are $q'_f = 0.9$, $\lambda'_1 = 0.5$ per hour, and $\lambda'_2 = 5$ per hour.

For the interval between 0 and 10 hours, an accurate estimate of the unreliability $U(10)$ is obtained numerically using the SAVE package ([9], [10]). For the purpose of comparison, we have also used standard simulation as well as importance sampling, each for a total of 128 000 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 failure biasing alone we get more than 10 times reduction in confidence interval width, which is equivalent to more than 100 times reduction in the simulation run length. Further reduction can be obtained by using forcing along with failure biasing, as shown in Section VI.

B. Steady-State Availability

The steady-state availability is defined as the long-run fraction of time the system is available. 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 state.) As a consequence, the steady-state unavailability UA is given by

$$UA = E(D)/E(T),$$

where D and T are the total “down” time and the total “cycle” time between regenerations, respectively [1]. Unfortunately, in non-Markovian models with general failure and repair distributions, a regenerative structure may not exist (for conditions under which a GSMP is regenerative, see [13]). 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 times of individual components are exponentially distributed. Therefore, a regeneration occurs at repair transitions after which all components become operational. In the second case, regenerations occur as a result of a periodic (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 occurs 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 independent of the past.

If the system exhibits a regenerative structure, then regenerative simulation can be used to estimate UA . Let n_c be the number of regenerative cycles used. Then an estimate of UA is given by

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

where D_i and T_i are the total down time and the cycle time in the i th regenerative cycle, respectively. $\hat{D}$ and $\hat{T}$ are estimates of $E(D)$ and $E(T)$, respectively. For highly available systems, standard simulation is very inefficient for estimating the numerator $E(D)$, since only a very small fraction of regenerative cycles will contain system 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. It turns out 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) [11], in which we estimate the numerator using importance sampling, while independently using standard simulation to estimate the denominator. Let n_n and n_d be the number of regenerative cycles used to estimate the numerator and denominator, respectively (their optimal allocation is considered in [12]). For the numerator, estimates for the mean and the variance are given by

$$\begin{aligned} \hat{D}_L &= \frac{1}{n_n} \sum_{i=1}^{n_n} D_i L(\mathbf{X}_{0,N_i}), \\ \hat{\sigma}^2(D_L) &= \frac{1}{n_n} \sum_{i=1}^{n_n} (D_i L(\mathbf{X}_{0,N_i}))^2 - \hat{D}_L^2, \end{aligned} \quad (10)$$

where at the i th regenerative cycle, D_i is the value of the numerator and N_i is the stopping time (i.e., the number of transitions in the i th regenerative cycle). $L(\mathbf{X}_{0,N_i})$ is the associated likelihood ratio as computed from (7). Similar equations hold for the estimators of 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). Let $\beta = n_n/(n_n + n_d)$; it follows that UA has the following estimates for its mean and asymptotic variance [12] (for large n_n and n_d):

$$\begin{aligned} \widehat{UA}_L &= \hat{D}_L / \hat{T}, \\ \hat{\sigma}^2(\widehat{UA}_L) &= \frac{\hat{\sigma}^2(D_L)}{\beta \hat{T}^2} + (\widehat{UA}_L)^2 \frac{\hat{\sigma}^2(T)}{(1 - \beta) \hat{T}^2}. \end{aligned} \quad (11)$$

Let us again consider the machine-repairman example in Section IV-A. To obtain a regenerative system, we change the original failure time distribution to an exponential of a parameter $\lambda = 0.001$ per hour. For importance sampling, we use accelerated failures from an exponential distribution of a parameter $\lambda' = 0.5$ per hour.

An accurate estimate of the unavailability UA is obtained numerically using SAVE. We give estimates using standard simulation and importance sampling, each obtained after a total of 128 000 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: $1.817 \times 10^{-6} \pm 2.61\%$.

Again, we get more than 10 times reduction in confidence interval width by using importance sampling. In Section VI we experiment with a machine-repairman model with periodic maintenance and a large model of a computing system.

V. 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 [23], which is a process-oriented simulation language based on the C programming language. In the following discussion we consider the estimation of steady-state unavailability and system unreliability.

When using importance sampling, we want typical system failures to occur more frequently. One approach to failure biasing is to reschedule events in order to bias the system towards the failed state (as described in Section III). As an alternate approach to failure biasing, one can actually alter the rates at which the clocks associated with the lifetimes of the components advance. This is accomplished by rescaling (i.e., divide by a scaling factor $r \geq 1$) the remaining lifetimes of the operational components at precisely the same instances at which clocks are rescheduled to accelerate failures. The advantage of rescaling clocks is that new random lifetimes do not have to be generated. The resulting likelihood ratio is exactly the same as in the rescheduling case when using scaled conditional distributions for the biased failure distributions. In the experimental results discussed in Section VI, the rescaling technique was used to implement importance sampling in all but the maintenance example of Section VI-C, in which rescheduling was used.

In addition to failure biasing, forcing is also used when estimating system unreliability, particularly for small time horizons. In the method of actual forcing (described in Section IV-A), the first component failure in each replication is scheduled by sampling from the conditional distribution of the minimum lifetime, given that it is less than the time horizon. Inverting this distribution function for the purpose of generating random variates may not be possible except in special cases. However, in general, one can use other known methods [3], e.g., those based on acceptance-rejection schemes. Sampling from the conditioned minimum lifetime is avoided in the method of approximate forcing (described in Section IV-A). In this method, only determining the probability distribution function of the minimum lifetime may be needed in order to compute the biasing level (or scaling factor) necessary to guarantee a component failure in the time horizon with a given high probability, p_{af} . The choice of p_{af} may be optimized with the help of an appropriate analysis of approximate forcing; this is difficult and not within the scope of this paper. However, experiments for a wide range of parameters and time horizons indicate that 0.8 is a good choice for p_{af} .

VI. EXAMPLES AND DISCUSSIONS

In this section we use four examples to illustrate the effectiveness of the proposed importance sampling techniques to estimate steady-state unavailability and system unreliability. First, we use a small model to experiment with some heuristics for selecting the new probability distributions for importance sampling. Second, these heuristics are applied to a model of a fairly complex computing system to demonstrate that orders of magnitude reduction in variance can be obtained in simulations of large models. The components in this system are modeled using exponential repair time distributions

and Weibull failure time distributions with shape parameter α . Recall that when $\alpha = 1$, the component lifetimes are then exponentially distributed; thus, we are able to ascertain the correctness of the simulation estimates by comparing them against numerical (nonsimulation) results obtained using SAVE. We show that by using importance sampling, the relative accuracy of the unavailability estimate is independent of its magnitude. Hence, as system failures become rarer, the amount of improvement gained by using importance sampling increases. Similar results also hold when estimating the system unreliability. The last two examples we consider are systems having components with nonexponential failure and repair time distributions. In particular, the component lifetimes are governed by a Weibull distribution, and we study the impact of changing its hazard rate function (e.g., increasing, decreasing, and constant failure rates). The purpose of these two studies is not so much to analyze the behavior of the systems under consideration but to give examples of the experiments that can be performed using the methodologies developed in this paper. Such studies are virtually impossible to carry out with analytical or numerical methods. Also, without importance sampling, simulation would not be a viable alternative.

A. Experiments with Different Importance Sampling Techniques

When estimating a given measure, different importance sampling techniques may be used to improve the efficiency of simulation (e.g., different scaling factors for failure biasing). When implemented properly, these importance sampling techniques will (at least in theory) yield unbiased estimates for the (same) expected value of the measure. However, the variance of the estimated measure, and hence the efficiency of simulation, depends heavily on the particular importance sampling technique used. In this section we use a small model to investigate the effect of our various importance sampling techniques on the variance of the dependability measure being estimated. First, we examine the effects of the amount of failure biasing on the magnitude and stability of the estimated variance for both the steady-state unavailability and the system unreliability. These experiments provided useful guidelines for developing heuristics to be used when simulating larger and more complex models. In addition, we investigate the effectiveness of the actual forcing and approximate forcing techniques when applied to the estimation of the system unreliability.

The model we consider in this section is a system with two types of components, each having a redundancy of two. The system is considered operational as long as at least one component of each type is operational; otherwise, the system is considered failed. 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 $\mu = 1$. This system can be modeled by a Markov chain which has nine states.

Before analyzing the results, we first define an *up cycle* to be a segment of the sample path between two successive instances when all components in the system become operational. When

TABLE I
ESTIMATED VARIANCE OF DOWN TIME IN A CYCLE FOR THE NINE-STATE
MODEL ($\lambda = 10^{-3}$) USING DIFFERENT SCALING FACTORS r

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}	3.248×10^{-5}	3.617×10^{-5}
4000	2.335×10^{-4}	2.654×10^{-5}	1.643×10^{-5}	1.348×10^{-5}	1.845×10^{-5}
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^{-5}	1.288×10^{-5}
96000	2.094×10^{-4}	2.708×10^{-5}	1.855×10^{-5}	9.191×10^{-5}	2.938×10^{-5}
128000	2.032×10^{-4}	2.840×10^{-5}	1.935×10^{-5}	7.876×10^{-5}	2.212×10^{-5}

TABLE II
ESTIMATED VARIANCE OF DOWN TIME IN A CYCLE FOR THE NINE-STATE
MODEL ($\lambda = 10^{-5}$) 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	—	—	5.232×10^{-9}	1.772×10^{-9}	1.767×10^{-9}	2.971×10^{-9}
4000	—	3.557×10^{-8}	1.463×10^{-8}	2.750×10^{-9}	1.677×10^{-9}	1.513×10^{-9}
16000	—	1.528×10^{-7}	2.020×10^{-8}	2.668×10^{-9}	2.394×10^{-9}	2.112×10^{-9}
32000	—	1.987×10^{-7}	1.887×10^{-8}	2.829×10^{-9}	3.071×10^{-9}	7.025×10^{-9}
64000	9.168×10^{-8}	2.107×10^{-7}	1.918×10^{-8}	2.903×10^{-9}	2.406×10^{-9}	5.537×10^{-9}
96000	6.647×10^{-8}	2.088×10^{-7}	1.914×10^{-8}	2.729×10^{-9}	2.299×10^{-9}	4.936×10^{-9}
128000	4.986×10^{-8}	1.846×10^{-7}	1.960×10^{-8}	2.879×10^{-9}	2.326×10^{-9}	5.495×10^{-9}

using scaling factor r the probability that more than one component fails in an up cycle is $p_r = r(n-1)\lambda/r(n-1)\lambda + \mu$, where n is the total number of components in the system. Note that the situation in which no failure biasing is used is equivalent to using scaling with $r = 1$.

1. *Effect of Failure Biasing on Unavailability Estimate:* We now examine the effect of the scaling factor r on the amount and stability of variance reduction obtained when estimating the steady-state unavailability. 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 (11)). Table I contains estimates of the variance of the total down time in a regenerative cycle after a specified number of simulated events for various values of r when $\lambda = 10^{-3}$. Table II contains similar results when $\lambda = 10^{-5}$. For some small values of r and small number of events, we obtained no results (i.e., no system failures were observed), indicating very inefficient simulation caused by improper choice of r .

From examining Tables I and II, we see that choosing the scaling factor r such that $(\mu/5(n-1)\lambda) < r \leq (\mu/(n-1)\lambda)$ results in stable estimates of the variance after only a small number of events. This roughly corresponds to selecting r so that $0.1 < p_r \leq 0.5$, which is typically much larger than the probability of two or more component failures in an up cycle under the original distribution. In addition, for r in this range, we obtain the largest amount of variance reduction. Finally, if we choose r too large, the variance actually starts to increase and its estimate becomes less stable; this is caused by more variability in the likelihood ratio.

2. *Effects of Forcing and Failure Biasing on Unreliability Estimate:* We next examine the impact of the scaling factor r on the amount and stability of variance reduction obtained when estimating the system unreliability for different time horizons. Tables III and IV contain estimates of the variance of the unreliability at time horizons $t = 10$ and $t = 1000$, respectively, after a specified number of simulated events and for various values of r when $\lambda = 10^{-4}$. All of the results

TABLE III
ESTIMATED VARIANCE OF UNRELIABILITY AT TIME HORIZON
 $t = 10$ FOR THE NINE-STATE MODEL ($\lambda = 10^{-4}$) USING
ACTUAL FORCING AND DIFFERENT SCALING FACTORS r

Events	$r = 1000$	$r = 1500$	$r = 2000$	$r = 2500$	$r = 3000$
1000	2.013×10^{-12}	2.055×10^{-12}	1.103×10^{-12}	7.334×10^{-13}	5.525×10^{-13}
4000	1.808×10^{-12}	1.766×10^{-12}	1.406×10^{-12}	1.227×10^{-12}	6.407×10^{-13}
16000	1.753×10^{-12}	1.612×10^{-12}	1.389×10^{-12}	1.144×10^{-12}	1.061×10^{-12}
64000	1.774×10^{-12}	1.492×10^{-12}	1.254×10^{-12}	1.257×10^{-12}	1.356×10^{-12}
128000	1.753×10^{-12}	1.537×10^{-12}	1.364×10^{-12}	1.401×10^{-12}	2.087×10^{-12}
256000	1.786×10^{-12}	1.508×10^{-12}	1.299×10^{-12}	1.304×10^{-12}	1.709×10^{-12}

TABLE IV
ESTIMATED VARIANCE OF UNRELIABILITY AT TIME HORIZON
 $t = 1000$ FOR THE NINE-STATE MODEL ($\lambda = 10^{-4}$) USING
ACTUAL FORCING AND DIFFERENT SCALING FACTORS r

Events	$r = 1000$	$r = 1500$	$r = 2000$	$r = 2500$	$r = 3000$
1000	1.581×10^{-8}	1.202×10^{-8}	1.733×10^{-8}	1.492×10^{-8}	2.437×10^{-8}
4000	1.661×10^{-8}	1.811×10^{-8}	1.612×10^{-8}	1.238×10^{-8}	1.473×10^{-8}
16000	1.802×10^{-8}	1.385×10^{-8}	1.416×10^{-8}	1.430×10^{-8}	1.621×10^{-8}
64000	1.885×10^{-8}	1.565×10^{-8}	1.775×10^{-8}	4.225×10^{-8}	2.993×10^{-8}
128000	1.915×10^{-8}	1.772×10^{-8}	1.785×10^{-8}	2.872×10^{-8}	2.551×10^{-8}
256000	1.869×10^{-8}	1.715×10^{-8}	1.707×10^{-8}	2.315×10^{-8}	2.978×10^{-8}

were obtained using actual forcing in conjunction with failure biasing.

Note that the time to first component failure has a mean $1/n\lambda$. Thus, if the time horizon is small (i.e., $t \ll 1/n\lambda$), then the probability that the time to first component failure is less than the time horizon is very small. However, by using (actual) forcing, we guarantee that the first component failure will occur before time t . If the system does not fail in the first up cycle and the time horizon has not expired, then the probability of another component failing in the remaining time is small, since we are no longer using forcing after the first component failure. Thus, when t is small, we want to bias the system enough so that there is a reasonable chance of the system failing during the first up cycle and before the time horizon expires. This suggests that for smaller time horizons, we should choose larger scaling factors, which agrees with our results in Tables III and IV. Note that when $t = 10$, selecting r somewhere between 2000 and 2500 seems to result in the smallest variance. These values of r also give rise to fairly stable variance estimates. However, when $t = 1000$, the value of r resulting in the best variance estimates lies between 1500 and 2000. Also, when using $r = 2500$ at time horizon $t = 1000$, the variance estimates are less stable. Hence, when estimating unreliability, the optimal choice of the scaling factor r depends on the time horizon t . For the model considered, when $t < 1/n\lambda$, it seems that the best choice of r satisfies $(\mu/5(n-1)\lambda) < r \leq (\mu/(n-1)\lambda)$, which roughly corresponds to having $0.1 < p_r \leq 0.5$. As we shall see in the following sections, this heuristic also works well for larger and more complex models with nonexponential distributions. In other situations, the suggested range of scaling factors did not work as well, but it provided a good starting point from where we could search for better values of r .

Table V contains the estimates of the variance of the unreliability at (a large) time horizon $t = 10^5$ when using actual forcing in conjunction with failure biasing for different scaling factors. In this case, $t > 1/n\lambda$. Had we used the heuristic developed for small time horizons, $r = 1000$ would seem an appropriate choice for the scaling factor. However, as

TABLE V
ESTIMATED VARIANCE OF UNRELIABILITY AT TIME HORIZON
 $t = 100,000$ FOR THE NINE-STATE MODEL ($\lambda = 10^{-4}$)
USING ACTUAL FORCING AND DIFFERENT SCALING FACTORS r

Events	$r = 1$	$r = 10$	$r = 50$	$r = 100$	$r = 1000$
1000	—	—	1.056×10^{-4}	5.025×10^{-5}	2.844×10^{-5}
4000	—	4.014×10^{-4}	1.073×10^{-4}	4.681×10^{-5}	6.430×10^{-5}
16000	5.076×10^{-5}	2.690×10^{-4}	1.062×10^{-4}	6.585×10^{-5}	1.592×10^{-4}
64000	3.774×10^{-5}	4.060×10^{-4}	9.435×10^{-5}	7.732×10^{-5}	4.025×10^{-4}
128000	6.285×10^{-5}	3.926×10^{-4}	9.466×10^{-5}	7.237×10^{-5}	3.156×10^{-4}
256000	5.033×10^{-5}	4.412×10^{-4}	9.177×10^{-5}	7.159×10^{-5}	3.406×10^{-4}

seen from Table V, this value does not yield stable estimates for the variance of the unreliability; therefore, we need to develop a different heuristic for large time horizons.

Under the original distribution, when $t > 1/n\lambda$, the probability that the first component failure occurs before the time horizon expires is fairly large. Thus, we do not gain much by forcing in this situation, but failure biasing may still help. Next we consider how to choose the scaling factor. Since the component repair rates are much larger than the failure rates, an up cycle typically consists of one component failing and then getting repaired. Hence, the average length of an up cycle (without importance sampling) is $1/n\lambda + 1/\mu \approx 1/n\lambda$, and so the expected number (without importance sampling) of up cycles completed before the time horizon expires is approximately $tn\lambda$. Furthermore, when using failure biasing (without forcing), the expected number of up cycles completed before the time horizon expires is about the same, since failures are not biased when all components are operational. Thus, similar to our heuristic for small t , we want to select the scaling factor so that the probability of more than one failure occurring in at least one of the up cycles during the interval $(0, t)$ is between 0.1 and 0.5. Using a rough approximation, since the expected number of up cycles in $(0, t)$ is approximately $tn\lambda$, we should choose r so that p_r is between $0.1/tn\lambda$ and $0.5/tn\lambda$. This implies $(\mu/5tn(n-1)\lambda^2) < r \leq \mu/tn(n-1)\lambda^2$. If r turns out to be close to (or less than) 1.0, then standard simulation should be used. Following this heuristic, we should select r between 17 and 83 for our model. Upon examining the results in Table V, we see that the best value of r seems to lie between 50 and 100, which overlaps with our suggested range.

We next compare the efficiency of the various simulation techniques proposed in this paper. Table VI contains the results from estimating the system unreliability at different time horizons using standard simulation, only failure biasing, approximate forcing (with $p_{af} = 0.8$, in conjunction with failure biasing), and actual forcing (combined with failure biasing). When using failure biasing, we ran several pilot runs to determine the best value of the scaling factor. All simulation results were obtained from running 1 024 000 events. The numerical results were obtained using SAVE.

The first thing to note is that as the time horizon shrinks, standard simulation becomes less effective (as measured by the width of the 99% confidence interval). In fact, when $t = 10$, no system failures were observed after simulating the given number of events, and so the point estimate is 0.0 in this case but we have no confidence interval. If failure biasing alone is used, we can improve the quality of our estimates for all

TABLE VI
ESTIMATES OF UNRELIABILITY AND 99% CONFIDENCE INTERVALS FOR
THE NINE-STATE MODEL ($\lambda = 10^{-4}$) USING DIFFERENT METHODS
(SIMULATION RESULTS OBTAINED AFTER 1 024 000 EVENTS)

Time Horizon	Numerical Results	Standard Simulation	Failure Biasing Only	Approx. Forcing and Biasing	Actual Forcing and Biasing
10	3.600×10^{-7}	—	3.847×10^{-7} $\pm 13.3\%$	3.589×10^{-7} $\pm 1.6\%$	3.585×10^{-7} $\pm 1.5\%$
100	3.960×10^{-8}	6.328×10^{-6} $\pm 106.1\%$	3.967×10^{-8} $\pm 4.6\%$	3.985×10^{-8} $\pm 1.7\%$	3.944×10^{-8} $\pm 1.5\%$
1000	3.996×10^{-8}	4.746×10^{-5} $\pm 50.0\%$	3.967×10^{-8} $\pm 2.0\%$	3.995×10^{-8} $\pm 1.8\%$	4.000×10^{-8} $\pm 1.6\%$

time horizons. However, the degradation in the estimates as t diminishes is again present. Finally, we obtain very good estimates for all time horizons when using failure biasing along with either approximate or actual forcing. It is interesting to note that the accuracy of the estimates is the same for all time horizons (similar conclusions were drawn in [12]). Also, the effectiveness of approximate and actual forcing are about equal. This suggests using approximate forcing in practice, since it is easier to implement.

B. A Large Model

In this section, we provide simulation results to show that the methods described in this paper are also feasible and effective for large and complex systems. We demonstrate that, as long as a system failure is a rare event, the relative width of the confidence intervals is independent of the magnitude of the unavailability (or the unreliability) estimate.

The system we examine is based on a model of a fairly complex computing system (also considered in [18]), with its block diagram shown in Fig. 1. The computing system is composed of two sets of processors with two processors per set, two sets of controllers with two controllers per set, and six clusters of disks, each consisting of four disk units. In a disk cluster, data are replicated so that one disk can fail without affecting the system. All repair time distributions are exponential; all of the component lifetimes (failure times) follow Weibull distributions having the same shape parameter α . Recall that when $\alpha = 1$, the failure times of the components are exponentially distributed. In this case, the system can be modeled as a continuous-time Markov chain, and we can validate our simulation results with numerical (nonsimulation) results for the unavailability and the unreliability obtained using SAVE. However, when $\alpha \neq 1$, numerical results for these dependability measures cannot be obtained using any known method.

When estimating the steady-state unavailability, we fix $\alpha = 1$ in order to maintain the regenerative structure of the system. For this dependability measure, we examine the model under two different sets of failure rates, in order to show that the relative width of the confidence interval obtained using importance sampling is insensitive to the magnitude of the steady-state unavailability. In the first (large) set, the failure rates of the processors, controllers and disks are assumed to be 1/2000, 1/2000, and 1/6000 per hour, respectively. We chose these relatively large failure rates so that stable estimates of the unavailability and its variance could be obtained in a reasonable amount of time using standard simulation. In

Fig. 1. A block diagram of the computing system modeled.

the second (small) set, we divide all of the failure rates by 100, thus obtaining more realistic values. The repair rate for all components is 1 per hour. Components are repaired by a single repairman according to a FCFS discipline. The system is defined to be operational if all data are accessible to both processor types, which means that at least one processor of each type, one controller in each set, and three out of four disk units in each of the six disk clusters are operational. Operational components continue to fail at the given rates when the system is failed.

When estimating the system unreliability, we need not have a regenerative structure in the system. Therefore, we can experiment with general failure processes by varying the shape parameter α of the failure time distributions. In these experiments, we adjust the scale parameters of the Weibull failure distributions so that the mean lifetimes of the processors, controllers, and disks are fixed at 200 000, 200 000, and 600 000 hours, respectively; i.e., we use the more realistic (small) set of failure parameters. All other aspects of the model remain the same.

1. Unavailability Results: In Table VII, we display the estimates of unavailability and the width of their 90% confidence intervals for the different sets of failure rates when using standard simulation and failure biasing after 1 024 000 simulated events. When using failure biasing, the scaling factor r was selected according to the heuristic derived from the results of the small model example given in Section VI-A-1. The first (second) row of the table contains the results from using the large (small) set of failure rates. For the large parameter set, the width of the confidence interval is reduced by a factor of 3.6 by using failure biasing rather than standard simulation, which translates into a 13-fold potential reduction in run length. For the small parameter set, the results from standard simulation are meaningless because the variance estimate had not yet stabilized by the end of the simulation. However, the results from using importance sampling are quite accurate, with the width of the relative 90% confidence interval being the same as that for the large set of failure rates. Thus, our failure biasing technique is relatively independent of the magnitude of unavailability. In other words, as system failure becomes rarer, the improvement gained by using importance sampling increases.

2. Unreliability Results: Table VIII contains the results from estimating the system unreliability at time $t = 100$ after

TABLE VII
ESTIMATES OF UNAVAILABILITY AND 90% CONFIDENCE INTERVALS FOR A LARGE MODEL (SIMULATION RESULTS OBTAINED AFTER 1 024 000 EVENTS)

Parameter Set	Numerical Result	Standard Simulation	Failure Biasing	Scaling Factor r
Large	4.055×10^{-6}	4.121×10^{-6} $\pm 13.6\%$	4.124×10^{-6} $\pm 3.8\%$	10^2
Small	4.000×10^{-10}	6.165×10^{-10} $\pm 159.5\%$	4.027×10^{-10} $\pm 3.7\%$	10^4

TABLE VIII
ESTIMATES OF UNRELIABILITY AT TIME HORIZON $t = 100$ AND 99% CONFIDENCE INTERVALS FOR A LARGE MODEL (SIMULATION RESULTS OBTAINED AFTER 512 000 EVENTS)

Shape Parameter α	Numerical Result	Standard Simulation	Approx. Forcing and Failure Biasing	Actual Forcing and Failure Biasing	Scaling Factor r
0.75	—	—	3.287×10^{-6} $\pm 6.2\%$	3.315×10^{-6} $\pm 6.1\%$	2000
1.00	3.960×10^{-8}	—	3.869×10^{-8} $\pm 6.4\%$	4.067×10^{-8} $\pm 5.6\%$	5000
1.25	—	—	5.967×10^{-10} $\pm 7.2\%$	6.177×10^{-10} $\pm 5.7\%$	10000

simulating 512 000 events. We used three different values (0.75, 1.0, and 1.25) for the shape parameter α of the failure time distributions. Recall that if $\alpha > 1$ (respectively, $\alpha < 1$), the distribution has an increasing (respectively, decreasing) failure rate. The importance sampling technique used in these experiments was failure biasing in conjunction with either approximate or actual forcing. The heuristic (given in Section VI-A-2) for estimating the unreliability at small time horizons can also be used here by replacing the failure rates with the inverses of the component mean lifetimes, thus giving us the range $2000 < r \leq 10\,000$ for the scaling factor used in failure biasing. For each of the different values of α , we varied the scaling factor in this interval to find the one yielding the best results. Thus, we selected $r = 2000$ when $\alpha = 0.75$, $r = 5000$ when $\alpha = 1.0$, and $r = 10\,000$ when $\alpha = 1.25$.

Using standard simulation, we did not observe any system failures before the designated number of events, and so the point estimate in this case is 0.0 with no confidence interval. However, as seen from Table VIII, by using forcing along with failure biasing we are able to obtain almost equally accurate estimates regardless of the value of the shape parameter. Thus, our importance sampling techniques, as well as the suggested heuristics, seem to work for large and non-Markovian models. Note that the accuracy of the estimates obtained using actual and approximate forcing are about the same.

C. A Study of the Effect of Maintenance on Unavailability

In this section, we examine a non-Markovian model with scheduled periodic (deterministic) maintenances and determine the effect on the steady-state unavailability of varying the length of time between maintenances when component lifetime distributions have increasing failure rate (IFR), constant failure rate, and decreasing failure rate (DFR). Because the model is non-Markovian, analytic results are extremely difficult to obtain. Also, 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 n . For this study, we

Fig. 2. Effect of maintenance on system unavailability.

fixed $n = 2$. The distribution of the lifetime of each component is Weibull, with shape parameter α and scale parameter β . 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 ($\alpha = 2.0$ hours) plus an exponentially distributed random variable with a mean $1/\mu$ ($= 2.0$ hours). In addition, each component has a repairman who performs periodic (deterministic) scheduled maintenance every d hours. A scheduled maintenance takes 0.5 hours ($d > 0.5$). 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. Thus, the maintenance cycles of the two components are staggered. A maintenance is performed on a component only if the other component is operational (i.e., not failed or in scheduled maintenance). Thus, $d \geq 1.0$ hour in this example. Also, after a component comes out of repair from a failure, the next scheduled maintenance is skipped. A component is considered to be as good as new immediately after completing a scheduled maintenance or a repair. The system is considered operational if at least one component is operational.

Fig. 2 shows a plot of the unavailability versus the time between maintenances (d) for the different values of α . Notice that, for a fixed β , the mean lifetime of individual components decreases for increasing values of α . This causes the unavailability to increase with α . We ran all the experiments long enough so that the relative half-width of the 90% confidence interval was less than 10% of the point estimate. Note the smooth curves for all value of α , indicating the high accuracy obtained using our importance sampling technique. Also note that as $d \rightarrow \infty$, the system becomes equivalent to one without 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. For $\alpha = 1.0$, the components lifetime distributions have constant failure rates. Thus, a component's reliability does not improve by performing a maintenance on it. Actually,

performing scheduled maintenances increases the system's unavailability, since by decreasing the time between maintenances, we are increasing the frequency of system failures caused by a failure of one component during the maintenance of the other. For $\alpha = 0.75$, the component lifetimes have DFR distributions; i.e., a component's reliability decreases by performing a maintenance on it. Therefore, by decreasing the time between scheduled maintenances, we are increasing the frequency of components failures, and hence increasing the system unavailability. For $\alpha = 1.25$, 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. For $\alpha = 1.5$, 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 often for components having higher failure frequencies.

D. A Study of the Effect of Redundancy on Unreliability

In this section we examine the same non-Markovian model studied in Section VI-C but without scheduled maintenances. We estimated the unreliability at time horizon $t = 100$ and $\beta = 10^{-5}$ for different values of the shape parameter α and different component redundancies n . Our goal is to determine the number of components n needed in order to achieve a certain level of unreliability U . We selected as our goals $U \leq 10^{-4}$, $U \leq 10^{-7}$, and $U \leq 10^{-10}$. Fig. 3 is a graph showing the redundancies needed for the different levels of unreliability as the value of the shape parameter increases. Thus, for $\alpha = 1.5$, in order to ensure that the unreliability of the system is no more than 10^{-4} , we must have a redundancy of at least two components. Similarly, for the same value of α , there must be at least three (respectively, four) components to guarantee that the system unreliability is less than 10^{-7} (respectively, 10^{-10}). Recall that as α increases (with β fixed), the failure frequencies of the individual components also become larger. Thus, as one would expect, as the value of α grows, the redundancy needed to ensure a given level of unreliability increases monotonically.

The results in this section were obtained from simulations using the heuristics developed in Section VI-A-2. However, it should be mentioned that the effectiveness of our importance sampling techniques diminishes for systems with high levels of redundancy (e.g., more than 4). For these systems, other importance sampling techniques are more effective [24].

VII. CONCLUSIONS

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 reductions in such rare event simulations. The approach is general in nature and allows us to simulate effectively 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 redundant components and/or

Fig. 3. Effect of failure time distribution on the redundancy needed to satisfy different reliability goals.

periodic maintenance. For different failure time distributions, we have explored how the steady-state availability is affected by the maintenance period. We have also determined the amount of component redundancy needed to achieve a certain reliability level. Because of the general non-Markovian nature of these systems, such studies are not possible using known analytical or numerical methods. Since standard simulation is very inefficient, importance sampling is the only practical alternative.

We described some of the trade-offs involved in the design of specific importance sampling techniques, such as failure biasing and forcing, and demonstrated their potential effectiveness in simulations of systems with non-exponential failure and repair time distributions. We have also used rescaling of clock values as an inexpensive way to implement rescheduling. While rescaling can be effective for some failure distributions, such as the exponential and Weibull distributions, its applicability for other failure distributions needs to be examined.

In addition to failure biasing, we have shown that forcing is very effective in estimating system unreliability, particularly for small time horizons. We described and experimented with two methods for forcing, namely, actual forcing and approximate forcing. While both methods are about equal in efficiency, approximate forcing is somewhat more robust and easier to implement in general.

The use of importance sampling for estimating steady-state availability requires a regenerative structure of the underlying model. This holds for 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 a regenerative structure or any assumption on the failure and repair processes.

One direction for future research is to investigate further the basis of some of the heuristics presented in this paper, so that additional refinements or generalizations can be made. While our importance sampling techniques are very effective for systems with highly reliable components, they are less so

for systems with high component redundancy. Such systems are considered in [24], but more research is needed to develop efficient methods for their simulation. Another research direction is in the development of theory and methods to estimate the sensitivity of dependability measures with respect to system parameters; this is very useful in systems design and optimization. While such methods have been presented in [19] for Markovian models, extensions to the non-Markovian case are of much interest.

REFERENCES

- [1] E. Cinlar, *An Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
- [2] A. E. Conway and A. Goyal, "Monte Carlo simulation of computer system availability/reliability models," in *Proc. 17th Symp. Fault-Tolerant Computing*, Pittsburgh, PA, 1987, pp. 230-235.
- [3] L. Devroye, *Non-uniform Random Variate Generation*. New York: Springer-Verlag, 1986.
- [4] J. B. Dugan, K. S. Trivedi, M. K. Smotherman, and R. M. Geist, "The hybrid automated reliability predictor," *J. Guidance, Control, and Dynamics*, vol. 3, no. 9, pp. 319-331, 1986.
- [5] R. M. Geist and M. K. Smotherman, "Ultrahigh reliability estimates through simulation," in *Proc. Annual Reliability and Maintainability Symp.*, Atlanta, GA, 1989, pp. 350-355.
- [6] R. M. Geist and K. S. Trivedi, "Ultra-high reliability prediction for fault-tolerant computer systems," *IEEE Trans. Comput.*, vol. C-32, pp. 1118-1127, 1983.
- [7] P. W. Glynn, "A GSMP formalism for discrete event systems," *Proc. IEEE*, vol. 77, no. 1, pp. 14-23, 1989.
- [8] P. W. Glynn and D. L. Iglehart, "Importance sampling for stochastic simulations," *Management Sci.*, vol. 35, pp. 1367-1392, 1989.
- [9] A. Goyal, W. C. Carter, E. de Souza e Silva, S. S. Lavenberg, and K. S. Trivedi, "The system availability estimator," in *Proc. 16th Symp. Fault-Tolerant Computing*, Vienna, Austria, 1986, pp. 84-89.
- [10] A. Goyal and S. S. Lavenberg, "Modeling and analysis of computer system availability," *IBM J. Res. Develop.*, vol. 31, no. 6, pp. 651-664, 1987.
- [11] A. Goyal, P. Heidelberger, and P. Shahabuddin, "Measure specific dynamic importance sampling for availability simulations," in *1987 Winter Simulation Conf. Proc.*, pp. 351-357.
- [12] A. Goyal, P. Shahabuddin, P. Heidelberger, V. F. Nicola, and P. Glynn, "A unified framework for simulating Markovian models of highly dependable systems," *IEEE Trans. Comput.*, vol. 41, pp. 36-51, 1992.
- [13] P. J. Haas and G. S. Shedler, "Regenerative generalized semi-Markov processes," *Commun. Statist. Stochastic Models*, vol. 3, no. 3, pp. 409-438, 1987.
- [14] J. C. Laprie, K. Kanoun, C. Beoune, and M. Kaaniche, "The KAT (knowledge-action-transformation) approach to the modeling and evaluation of reliability and availability growth," *IEEE Trans. Software Eng.*, vol. 17, no. 4, pp. 370-382, 1991.
- [15] D. Lee, J. Abraham, D. Rennels, and G. Gilley, "A numerical technique for the hierarchical evaluation of large, closed fault-tolerant systems," in *Proc. 2nd Int. Conf. Dependable Computing for Critical Applications*, Tucson, AZ, 1991, pp. 49-56.
- [16] E. E. Lewis and F. Böhm, "Monte Carlo simulation of Markov unreliability models," *Nucl. Eng. Des.*, vol. 77, pp. 49-62, 1984.
- [17] A. P. A. van Moorsel, B. R. Haverkort, and I. G. Niemegeers, "Fault injection simulation: A variance reduction technique for systems with rare events," in *Proc. 2nd Int. Conf. Dependable Computing for Critical Applications*, Tucson, AZ, 1991, pp. 57-64.
- [18] R. R. Muntz, E. de Souza e Silva, and A. Goyal, "Bounding availability of repairable computer systems," *IEEE Trans. Comput.*, vol. 38, pp. 1714-1723, 1989.
- [19] M. K. Nakayama, "Asymptotics for likelihood ratio derivative estimators in simulations of highly reliable Markovian systems," IBM, Yorktown Heights, NY, Res. Rep. RC 17357, 1991.
- [20] V. F. Nicola, A. Bobbio, and K. S. Trivedi, "A unified performance reliability analysis of a system with a cumulative down time constraint," *Microelectron. Rel.*, vol. 32, no. 1-2, pp. 49-65, 1992.
- [21] V. F. Nicola, M. K. Nakayama, P. Heidelberger, and A. Goyal, "Fast simulation of dependability models with general failure, repair and maintenance processes," in *Proc. 20th Symp. Fault-Tolerant Computing*, Newcastle Upon Tyne, England, 1990, pp. 491-498.

- [22] S. Parekh and J. Walrand, "A quick simulation method for excessive backlogs in networks of queues," *IEEE Trans. Automatic Control*, vol. 34, no. 1, pp. 54-66, 1989.
- [23] H. Schwetman, "Using CSIM to model complex systems," in *1988 Winter Simulation Conf. Proc.*, pp. 491-499.
- [24] P. Shahabuddin, "Simulation and analysis of highly reliable systems," Ph.D. dissertation, Dept. Operations Res., Stanford Univ., Stanford, CA, 1990.
- [25] P. Shahabuddin, V. F. Nicola, P. Heidelberger, A. Goyal, and P. W. Glynn, "Variance reduction in mean time to failure simulations," in *1988 Winter Simulation Conference Proc.*, pp. 491-499.
- [26] J. Stiffler and L. Bryant, "CAREIII phase III report-mathematical description," NASA Contractor Rep. 3566, 1982.

Victor F. Nicola (M'89) received the B.S. and M.S. degrees in electrical engineering from Cairo University, Egypt, and Eindhoven University of Technology, The Netherlands, respectively, and the Ph.D. degree in computer science from Duke University, Durham, NC.

In 1979 he held scientific and research staff positions at Eindhoven University and Duke University. Since 1987 he has been a Research Staff Member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. He is currently on leave

from IBM and holds the position of a Visiting Associate Professor in the group of Tele-Informatics and Open Systems at the University of Twente, The Netherlands. His research interests include performance and reliability modeling of computer and communication systems, queueing theory, fault tolerance, and simulation.

Dr. Nicola is a member of the IEEE Computer Society.

Marvin K. Nakayama received the B.A. degree in mathematics/computer science from the University of California, San Diego, in 1986 and the M.S. and Ph.D. degrees in operations research from Stanford University, Stanford, CA, in 1988 and 1991, respectively.

He is currently a post-doctoral fellow at the IBM Thomas J. Watson Research Center in Yorktown Heights, NY. His research interests include discrete event simulation, reliability theory, and applied probability.

Dr. Nakayama is a member of the Operations Research Society of America.

Philip Heidelberger (M'82-SM'91) received the B.A. degree in mathematics from Oberlin College, Oberlin, OH, in 1974 and the Ph.D. degree in operations research from Stanford University, Stanford, CA, in 1978.

He has been a Research Staff Member at the IBM Thomas J. Watson Research Center in Yorktown Heights, NY, since 1978. His research interests include modeling and analysis of computer performance, probabilistic aspects of discrete-event simulations, and parallel simulation.

Dr. Heidelberger is currently an Area Editor for the ACM's *Transactions on Modeling and Computer Simulation*, and was an Associate Editor of *Operations Research*. He was the Program Chairman of the 1989 Winter Simulation Conference, and the Program Co-Chairman of the ACM Sigmetrics/Performance '92 Conference. He is a member of the Operations Research Society of America and the ACM.

Ambuj Goyal (S'81-M'82) received the B.Tech. degree from the Indian Institute of Technology, Kanpur, in 1978, and the M.S. and Ph.D. degrees from the University of Texas at Austin, in 1979 and 1982, respectively.

Since 1982 he has been at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, as a Research Staff Member. Currently he is the Senior Manager for the Parallel and Fault-Tolerant Computer Department. He has worked in the areas of high-speed networking, computer architecture,

performance and reliability modeling, and parallel and distributed operating systems and database management systems. He has authored over 50 articles in these areas.

Dr. Goyal has authored a tutorial on *Effectiveness Evaluation of Multiprocessor Systems*, and has served as the Program Committee Chairman of the *Workshop on Reliability/Availability Modeling Tools and Their Applications*. Also, he is the Program Committee Chairman for the *Twenty-Third Fault-Tolerant Computing Symposium*. He has received several IBM and external honors including Best Paper Awards, Invention Achievement and Outstanding Innovation Awards, and Merit Scholarships. He is a member of the IEEE Computer Society and the IFIP Working Group 10.4 on Reliable Computing.