Monte Carlo Simulation of Computer System
Availability/Reliability Models

Adrian E. Conway* and Ambuj Goyal**

*McGill University
Montreal, Canada
**IBM Thomas J. Watson Research Center
Yorktown Heights, New York

ABSTRACT

In recent years, there has been an increased interest in evaluating the availability and reliability of computer systems. For this purpose, we have developed a state of the art modeling package called SAVE (System Availability Estimator). This package has refined numerical matrix methods for computing the dependability measures of interest. These methods do, however, have their limitations since the growth of the size of the state space of these models is exponential in the number of component types. In this paper, we investigate the use of Monte Carlo simulation as an alternative for solving models with a large number of components. Since probability of failure is relatively small, it has been observed by many researchers that simulation takes a large amount of computational time. We show that the Importance Sampling variance reduction technique may be applied to reduce the simulation run lengths substantially. We give examples which demonstrate the viability of using simulation for the estimation of large availability/reliability models. Finally, we describe a simulation program which has been developed for the analysis of models that can be generated using the SAVE modeling language. This program makes automatic use of the Importance Sampling technique to reduce the run lengths.

1. Introduction

Availability and reliability prediction is important in the design and configuration of fault-tolerant systems. Care must be taken when modeling component failure/repair behavior, recovery techniques and component interdependencies to reduce the effect of assumptions (or abstractions) on the predictions. The System Availability Estimator (SAVE) package has a modeling language with powerful constructs which aid in considering system details [3,4]. The underlying mathematical models (Markov chains) are automatically constructed from the modeling language. Complex models with a large number of components create large Markov chains which are solved using numerical methods. SAVE takes advantage of the sparsity in the transition rate matrices of Markov chains to solve models with tens of thousands of states. The steady state availability and mean time to failure, together with their sensitivities (i.e. derivatives) with respect to the transition rate parameters, are computed using matrix iterative methods [14]. The interval availability and reliability, together with their sensitivities with respect to the transition rate parameters, are computed using the randomization technique [7,9]. The distribution of availability is computed by a numerical technique described in [6].

Although the above methods are capable of solving models with very large Markov chains, the growth of the size of the state space is exponential in the number of different types of components. This precludes the numerical analysis of general models with many component types. Under special circumstances, the computational effort can be reduced substantially [5]. For example, if the class of models is restricted to those of the product form type, then the known efficient computational algorithms [11] for that class of models may be applied. If the failure and repair characteristics of components are independent, then the combinatorial models [5] may be used. In general, however, the model will require analysis by general purpose matrix methods.

One approach which has been adopted in SAVE for the solution of systems having many types of components is to consider models in which the state space is truncated to a subspace in which there is fewer than some total number \( c \) of failed components. The model is evaluated successively for increasing values of \( c \) until it is observed that there is negligible change in the dependability measure being computed. This method yields an approximation for the measure of interest. Analytical error bounds and convergence results for this technique are currently under investigation [5].

An alternate approach for the solution of large models is Monte Carlo simulation which is the subject of this paper. Simulation is especially useful for those models whose transition rate matrices exceed the available storage. By nature, this approach has the immediate advantage of having relatively small storage requirements. On the other hand, it is apparent that the analysis of large models by simulation will require many regenerative cycles or many long independent replications in order to attain reasonable confidence intervals [1,13]. It is seen, however, that for the types of systems under consideration, it is usual for the repair rates of components to be orders of magnitude larger than the failure rates and in these circumstances highly effective use of the Importance Sampling variance reduction technique [8,10] may be made in order to reduce the simulation run lengths significantly. In this paper, we demonstrate the effectiveness of using Importance Sampling for estimating dependability measures and describe the operation of a simulation program which has been written specifically for models of the SAVE type and which automatically makes use of Importance Sampling.

In the following section, we first illustrate the usefulness of the Importance Sampling technique by means of a simple three state Markov chain example. In Section 3, we briefly discuss the theory of Importance Sampling and show how the various dependability measures can be computed using simulation. In Section 4, we describe the modeling features of SAVE which are included in the simulation program and solve two relatively large examples using the program. Finally, in Section 5, we conclude by describing the directions in which the work is continuing.

2. Illustrative Example

The fundamental problem with direct simulation of fault-tolerant systems can be seen by considering a simple system with three states. Consider a 1-of-2 system with perfect coverage and one repairman. The transition state diagram is illustrated in Figure 1. States '0' and '1' are system operational states and state '2' is the unoperational state. The failure and repair rates are \( \lambda \) and \( \mu \), respectively. To obtain the steady-state availability (the probability of being in state '2') by simulation, we can simulate an equivalent discrete time Markov chain, as illustrated in Figure 2. We use regenerative simulation [11] with state
availability is $0.19564 \times 10^{-9}$ and the 95 percent confidence interval is 2000 regenerative cycles (approx. 3000 events) the estimate of unavailability obtained by direct simulation turns out to be zero since no regenerative cycles happen to visit state '2' in the number of events simulated. Using a similar analysis as done in [1, 13], we have found that to obtain an estimate for this case within 10 percent of the mean with 95 percent confidence using direct simulation we would require approximately 1.92 trillion regenerative cycles. Therefore, an excessively large number of cycles will be required to obtain meaningful estimates. For example, if $\lambda = 0.001$ and $\mu = 0.999$, then on the average we exit state '1' towards the unoperational state (state '2') once in one thousand exits. If we could somehow force the simulation to visit the failed state more often, say 50 percent of the time, then we would have more cycles that contribute to the estimate of interest. However, modifying the transition structure of a system introduces bias into the measures. Importance Sampling is a method by which we may change the transition structure, forcing the system towards the failed states, and maintain unbiased estimates.

### Table 1

Unavailability estimates and 99% confidence intervals for three state example with $\mu = 1$.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Importance Sampling</th>
<th>Direct Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-1}$</td>
<td>$0.1644 \times 10^{-1} \pm 0.44 \times 10^{-2}$</td>
<td>$0.1647 \times 10^{-1} \pm 0.68 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>$0.1966 \times 10^{-2} \pm 0.54 \times 10^{-3}$</td>
<td>$0.2018 \times 10^{-2} \pm 0.23 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>$0.1997 \times 10^{-3} \pm 0.55 \times 10^{-4}$</td>
<td>$0.2303 \times 10^{-3} \pm 0.76 \times 10^{-4}$</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>$0.2001 \times 10^{-4} \pm 0.55 \times 10^{-5}$</td>
<td>$0.1538 \times 10^{-4} \pm 0.19 \times 10^{-5}$</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$0.2001 \times 10^{-5} \pm 0.55 \times 10^{-6}$</td>
<td>0.0</td>
</tr>
</tbody>
</table>

### 3. Importance Sampling and Dependability Estimates

The simulation of system availability/reliability models can be performed in two different ways. In the first method, the time to failure and the time to repair of each component are generated at random. The event list is then examined to determine at what time the system changes state. After the occurrence of each state transition, we determine whether the system is in an operational or unoperational state and update the total time which has been spent in both types of states. After a transition, a new set of failure and repair times must be generated at random since a change in system state could mean a change in the failure and repair rates. In the second method, the underlying Markov chain is simulated by generating state transitions at random according to the jump probabilities of the chain. The updating of the total time spent in operational and unoperational states is made after each state transition. For the types of systems we are interested in analyzing, where the number of component types is typically large, the second method is the preferred approach since only two random numbers are used per state transition (one to generate the random jump and the other for the exponential holding time in the state). In the first method, the number of random numbers that must be generated for each transition is, in general, equal to the number of component types. The second method is the approach which has been adopted for the simulation of models of complex systems which arise in the nuclear reactor application [12].

In general, the simulation of highly available systems with many component types usually require extremely long run lengths. Fortunately, since we typically have orders of magnitude difference between the failure and repair rates, we may make effective use of the Importance Sampling variance reduction technique to reduce the lengths of simulations [8, 10]. The effectiveness of using this technique for the analysis of availability/reliability models arising in the nuclear engineering area is described in [12] and [15]. Here we are interested in applying the same variance reduction technique for estimating the dependability measures of computer system models of the SAVE type.

### 3.1 Importance Sampling

The underlying philosophy of the method of Importance Sampling is described in this section. Consider a random variable $X$ with cumulative distribution function $H(x)$ and suppose we are interested in estimating the quantity $E[f(x)] = \int_{-\infty}^{\infty} f(x) dH(x)$. This can be done by taking $N$ independent samples $\xi$ from the distribution $H(x)$ and computing the estimate $1/N \sum_{i=1}^{N} f(\xi)$. The variance of this estimator is $1/N \sum_{i=1}^{N} (f(\xi) - \theta)^2 dH(x)$ where $\theta = E[f(x)]$. Now assuming $H(x)$ has density $h(x)$, we may write...
\[ E[I(x)] = \int_{-\infty}^{\infty} f(x) \frac{h(x)}{g(x)} dG(x) = \int_{-\infty}^{\infty} f(x) \Lambda(x) dG(x) \]

where \( G(x) \) is a distribution with density \( g(x) \) and \( \Lambda(x) = h(x)/g(x) \) is a likelihood ratio. If we now take \( N \) independent samples \( \xi \) from the distribution \( G(x) \), then

\[
\frac{1}{N} \sum_{\xi=1}^{N} (f(\xi) \Lambda(\xi)) \text{ is an unbiased estimator of } \theta. \text{ The variance of this estimator is }
\]

\[
\frac{1}{N^2} \int_{-\infty}^{\infty} (f(x) \Lambda(x) - \theta)^2 dG(x).
\]

Now if \( g(x) \) is chosen so that \( g(x) = f(x)h(x)\theta^{-1} \), then

\[
(f(x) \Lambda(x) - \theta) = 0 \text{ which implies that we have an estimator with zero variance!}
\]

This estimator, however, requires the knowledge of \( \theta \) which is, by assumption, an unknown quantity. The above considerations indicate to us, however, how we should select \( g(x) \) to obtain estimates with reduced variance. The shape of \( g(x) \) should follow the shape of \( f(x)h(x) \) as closely as possible. In effect, by altering the form of the parent distribution \( H(x) \), we attach more importance in the sampling procedure to those intervals where the contributions to the integral are greatest. By weighting the samples \( f(\xi) \) with the likelihood ratios \( \Lambda(\xi) \), we maintain unbiased estimates. The exact manner in which \( g(x) \) is to be chosen is left open to us. In an application of this technique to a specific problem, when we select the function \( g(x) \), we should make use of any understanding we may have regarding the stochastic structure of the problem.

The above methodology can be applied directly to the simulation of Markov chains arising in SAVE models. In the following, we present the theory of Importance Sampling within this context. For the sake of brevity, we only give a heuristic presentation of results which can be made mathematically rigorous.

Consider an interval \([0,T]\) of interest, and denote by \( S \) the set of all possible sample paths \( \xi \) that the chain may follow in the interval \([0,T]\), given a fixed initial state \( x_0 \). Note that \( S \) is an uncountably infinite set. Also, denote by \( Q \) the transition rate matrix of the chain and consider an objective function \( I(\xi) \). This function could be, for example, the proportion of time in the interval \([0,T]\) during which the system is operational (i.e. interval availability). Now the expected value of the objective function is

\[
E[I(\xi)] = \sum_{\xi \in S} I(\xi) P(\xi). \tag{3.1}
\]

where \( P(\xi) \) is the probability that a sample path \( \xi \) is generated under the transition rate matrix \( Q \). Now suppose that \( Q \) is modified to \( Q' \) so that the probability of realizing path \( \xi \) is \( P'(\xi) \). Multiplying and dividing Equation (3.1) by \( P'(\xi) \) inside the summation sign, we obtain

\[
E[I(\xi)] = \sum_{\xi \in S} I(\xi) \Lambda(\xi) P'(\xi), \tag{3.2}
\]

where \( \Lambda(\xi) = P(\xi)/P'(\xi) \) is the likelihood ratio. The only conditions placed on the modified transition rate matrix \( Q' \) are that \( P'(\xi) \neq 0 \), if \( I(\xi) P(\xi) \neq 0 \), and that \( P'(\xi) = 0 \), if \( I(\xi) P(\xi) = 0 \) for all sample paths \( \xi \in S \). Hence, if we carry out \( R \) independent replications \([11]\), an unbiased estimator of \( E[I(\xi)] \) is

\[
\frac{1}{R} \sum_{r=1}^{R} I(\xi_r) \Lambda(\xi_r), \tag{3.3}
\]

where \( \xi_r \) is the sample path realized on the \( r^{th} \) replication. The exact manner of selecting \( Q' \) is, of course, left open to us. The theory of Importance Sampling suggests that in order to reduce the variance of the estimator we should choose \( Q' \) in such a way as to increase the likelihood of generating sample paths \( \xi \) which have high values of \( I(\xi) \) and which would be unlikely to be generated under \( Q \).

In the estimation of availability and reliability measures of SAVE models, where the failure rates are typically orders of magnitude smaller than the repair rates, the likelihood of generating sample paths that pass through the system unoperational states is extremely small. To reduce the variance of the estimates, we bias the transition rates so as to increase the probability of component failures and consequently increase the likelihood that the sample paths pass through the system unoperational states. The details of the manner in which this is done is described in the following section.

For the estimation of objective functions which are simply a function of the equilibrium probabilities, such as steady-state availability, it is preferable to simulate an associated semi-Markov process rather than the original continuous time Markov chain. In the semi-Markov process, the holding times are deterministic and equal to the mean of the holding times in the original continuous time Markov chain. This semi-Markov process has the same equilibrium distribution as that of the original Markov chain. By simulating the associated semi-Markov chain, however, we obtain estimates with reduced variance since there is no variation in the holding times to contribute variance to the estimates of interest.

Let \( P = [P_{ij}] \) denote the transition matrix of the imbedded discrete time chain of the semi-Markov process and let the mean holding time in state \( i \) be denoted by \( \bar{h}_i \). Also, let \( \pi = [\pi_i] \) be the equilibrium distribution and suppose that we are interested in estimating a dependability measure of the form \( \omega(\pi) = \sum \omega_{ij} \pi_j \), where \( B \) is the state space of the semi-Markov process and \( \omega_{ij} \) is a set of weights. The quantity \( \omega(\pi) \) can be estimated using the method of regenerative simulation \([11]\). An estimate of \( \omega(\pi) \) can be found by simulating the discrete time chain for \( R \) regenerations starting with the initial state \( x_0 \) and terminating when \( x_0 \) is 'hit' for the first time after leaving \( x_0 \). Assume that we simulate the semi-Markov process under a modified \( P' \), say, \( P' \). The estimate of \( \omega(\pi) \) is

\[
\frac{1}{R \sum_{r=1}^{R} \left( \sum_{j=0}^{I-1} \omega_{ij}(s_{j+1}) \lambda_{ij}(s_j) \right)} \lambda_{ij}(s_j) \left( \sum_{j=0}^{I-1} \omega_{ij}(s_{j+1}) \lambda_{ij}(s_j) \right), \tag{3.4}
\]

where \( J \) is the number of jumps realized in the \( r^{th} \) regenerative cycle and \( s \) is the sequence of states (or the sample path) realized in the \( r^{th} \) regenerative cycle. If the sequence of states is \( s = (s_0, s_1, ..., s_{I-1}, s_0) \), then the likelihood ratio is simply

\[
\lambda_{ij}(s_j) = \frac{P_{ij} \pi_j}{P'_{ii} \pi_i}.
\]

Here again, the selection of \( P' \) is left open to us and the theory of Importance Sampling suggests to us that we should select \( P' \) so as to increase the likelihood of generating regenerative cycles which have high values of \( \sum_{j=0}^{I-1} \omega_{ij}(s_{j+1}) \lambda_{ij}(s_j) \) and which would be unlikely to be generated under \( P \).

### 3.2 Dependability Measures

The dependability measures we are interested in estimating are the availability, the interval availability over an interval \([0,T]\), the distribution of availability over an interval \([0,T]\), the mean time to failure (MTTF) and the reliability at time \( T \). The steady state availability is obtained using the regenerative method. The estimator for the steady state availability is given by Equation (3.4) with

\[
\omega_{ij} = \mathcal{A}(s_j) = \begin{cases} 1, & \text{if state } x_i \text{ is an up state} \\ 0, & \text{otherwise} \end{cases} \tag{3.5}
\]

The regenerative state \( x_0 \) is selected to be the state in which all components of all types are operational. The interval availability is estimated using the method of independent replications. The estimator for interval availability is given by Equation (3.3) with \( I(\xi) = \mathcal{A}(s) = \begin{cases} 1, & \text{if the system is operational} \\ 0, & \text{otherwise} \end{cases} \).
respectively. The estimator of the probability that the interval availability is less than or equal to \( x \) for \( x \leq T \) (i.e. the distribution of availability) is also computed using Equation (3.3) with \( I(\varepsilon) \) given by

\[
I(\varepsilon) = \left\{ \begin{array}{ll}
1, & \text{if } I(\varepsilon) \leq x, \\
0, & \text{otherwise.}
\end{array} \right.
\]

(3.7)

The MTTF and reliability are also estimated using the method of independent replications. The estimator for the MTTF is given by Equation (3.3) with

\[
I(\varepsilon) = \left\{ \begin{array}{ll}
0, & \text{if } L(\varepsilon) < x, \\
1, & \text{otherwise.}
\end{array} \right.
\]

(3.8)

with

\[
I(\varepsilon) = \left\{ \begin{array}{ll}
1, & \text{if no unoperational state is encountered} \\
0, & \text{otherwise.}
\end{array} \right.
\]

(3.9)

The manner in which we have chosen to modify the transition probabilities to achieve variance reduction is as follows. If the system is in a state such that there are both outgoing failure and repair transitions, then we scale the failure transitions so that the total probability of a failure occurring is \( BIAS \) and the total probability of a repair occurring is \( 1 - BIAS \). In all other situations, we do not alter the transition probabilities of the system. The above described method of altering the transition matrix is known as 'failure biasing'. Typically, \( BIAS \) is chosen to be approximately 0.5. We have experimented with various choices of \( BIAS \) and have found that this choice works well. The results of Table 2 illustrate how the confidence intervals depend upon the choice of \( BIAS \) for the three state example of Section 2.

<table>
<thead>
<tr>
<th>BIAS</th>
<th>Unavailability</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.2034 \times 10^{-3}</td>
<td>±2.14 \times 10^{-5}</td>
</tr>
<tr>
<td>0.10</td>
<td>0.1991 \times 10^{-3}</td>
<td>±0.78 \times 10^{-5}</td>
</tr>
<tr>
<td>0.30</td>
<td>0.1960 \times 10^{-3}</td>
<td>±0.54 \times 10^{-5}</td>
</tr>
<tr>
<td>0.50</td>
<td>0.1966 \times 10^{-3}</td>
<td>±0.54 \times 10^{-5}</td>
</tr>
<tr>
<td>0.70</td>
<td>0.1951 \times 10^{-3}</td>
<td>±0.61 \times 10^{-5}</td>
</tr>
<tr>
<td>0.90</td>
<td>0.1929 \times 10^{-3}</td>
<td>±0.96 \times 10^{-5}</td>
</tr>
<tr>
<td>0.99</td>
<td>0.1940 \times 10^{-3}</td>
<td>±3.00 \times 10^{-5}</td>
</tr>
</tbody>
</table>

In the systems we are interested in simulating, it is typical for the repair rates to be much larger than the failure rates. This causes two practical problems when failure biasing is employed. The first is that the likelihood ratios will most likely underflow and the second is that the regenerative cycles may become excessively long. The first problem can be solved by "turning off" the failure biasing when the likelihood ratio \( \Lambda(\epsilon) \) falls below a certain threshold, say \( 10^{-6} \). The
second problem can be solved by "turning off" the biasing after a certain number MXDOWN of system unoperational states have been encountered. A typical choice (by experimentation) is MXDOWN=1. When a regenerative simulation is begun in a state \( x_0 \) where all components are operational, the failure biasing mechanism is "turned on". After MXDOWN unoperational states have been encountered the failure biasing is "turned off". Since the repair rates are much larger than the failure rates the system returns quickly to the regenerative state \( x_0 \).

### 4.3 Fault-Tolerant Database Example

We illustrate the capabilities of the simulation method by using a hypothetical example of a fault-tolerant database system as shown in Figure 3. The system has two front-end systems, two databases, and two processing subsystems each of which contains a switch, a memory, and two processors. A processing subsystem is considered operational if the memory, the switch, and one of the two processors in that subsystem is functioning. The entire system is operational if a database, a front-end, and at least one of the two processing subsystems is operational.

![Figure 3](image_url)

A Fault-Tolerant Database System

We assume that the repair and failure time distributions of all components are exponentially distributed with means 1 and 2,400 hours, respectively, except for the processors which have means 1 and 120 hours, respectively. We further assume that when a processor fails it contaminates (or fails) the database with probability \( (1-c) \), where \( c = 0.99 \) is the coverage probability. There is a single repairman in the system with highest repair priority given to the databases and the front-ends, the next highest priority to the memory and the switch elements, and the lowest priority to the processors. Components at the same priority level are selected at random for repair. The simulation results for this example are given in Table 3. We see that failure biasing gives us tighter confidence intervals than those obtained with no biasing. The latter method would require approximately 100 times longer simulation runs to obtain confidence intervals similar to those obtained by the former method.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias on</th>
<th>Bias off</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unavailability</td>
<td>( 0.352 \times 10^{-5} )</td>
<td>( 0.51 \times 10^{-5} )</td>
<td>( 0.344 \times 10^{-5} )</td>
</tr>
</tbody>
</table>

### 4.4 A Large Example

In our experiments, we have analyzed many large models. As an illustration, we provide the simulation results for a simple system that has many component types and an extremely large state space. The system is described as follows. There are 70 components and there is a single repairman that operates on a random order service (ROS) basis. Although, the simulation considers each component of being a different type, the failure and repair rates of all components are assumed to be 1 and 10^5 per year, respectively. This is done so that the exact results computed numerically could be compared with the simulation results. The condition under which the system is considered operational is that at least one of the components of type 1 or 2 is operational and that at least one of the components of type 3, 4 or 5 is operational. The state of the other components do not effect the system, except through the congestion they cause at the repair queue. The results of this experiment are given in Table 4. The BIAS was set to 0.5 and MXDOWN was set to 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias on</th>
<th>Bias off</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unavailability</td>
<td>( 0.198 \times 10^{-9} )</td>
<td>( 0.50 \times 10^{-11} )</td>
<td>( 0.200 \times 10^{-9} )</td>
</tr>
</tbody>
</table>

### 5. Summary and Future Directions

In this paper, we have described a method for simulating large reliability and availability models. These models are intractable using the numerical techniques described in the literature. The variance reduction scheme described in this paper is very useful in reducing the simulation run-lengths by orders of magnitude. A typical simulation takes a few minutes on an IBM 3090 computer.

There are several directions in which the work is continuing. We intend to carry out an extensive simulation study to determine the ranges of the parameter values over which the simulation method is a viable alternative to the matrix methods currently employed in SAVE. This would provide a set of guidelines so that a user of SAVE could select an appropriate method for a particular modeling situation.
We are also working on estimating the sensitivity of the availability measure with respect to parameter values. The method we are experimenting with uses the likelihood ratio gradient estimation theory described in [2] in conjunction with the Importance Sampling method described in this paper. The preliminary results have been encouraging. We are also working on an Adaptive Importance Sampling method in which the bias values are adjusted as the simulation progresses. We adjust the bias based on the gradient of the variance of the dependability measure with respect to the bias value as estimated after each completed replication or regenerative cycle.

Acknowledgments

We would like to thank Philip Heidelberger, Stephen Lavenberg and Thomas Rosse for helpful suggestions and discussions relating to this work.

References


