ANALYSING
SCHEDULED MAINTAINANCE POLICIES
OF REPAIRABLE COMPUTER SYSTEMS

E.DE SOUZA E SILVA
H.R. GAIL *

NCE-07/88

Abril/88

Universidade Federal do Rio de Janeiro
Núcleo de Computação Eletrônica
Caixa Postal 2324
20001 - Rio de Janeiro - RJ
BRASIL

* IBM T.J. Watson Research Center

Esta pesquisa contou com o apoio da IBM-Brasil.
Este trabalho foi publicado como relatório técnico do Centro de Pesquisas da IBM T.J. Watson, RC 13379, Dezembro 1987.

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
NÚCLEO DE COMPUTAÇÃO ELETRÔNICA

Resumo

Para sistemas que podem ser reparados, é comum o uso de políticas de manutenção regular como meio de se obter desejada confiabilidade. Neste artigo, desenvolvemos um método de solução para analizar várias políticas de manutenção de sistemas de computação reparáveis. A análise se aplica a sistemas cujo comportamento de falha pode ser modelado por um processo de Markov de tempo contínuo e, por conseguinte, características importantes podem ser incluídas no modelo. Além do mais, não é necessário supor que o sistema é reparado com perfeição ao término de um período de manutenção. Esta suposição não é realística na maioria dos casos. São obtidas tanto medidas transientes quanto medidas em estado estacionário. O método de solução é baseado na técnica de aleatorização e possue vantagens tais como estabilidade numérica e facilidade de implementação.

Abstract

For systems which can undergo repair, it is common to use regular maintainance policies as a means to achieve availability requirements. We develop a solution method to analyze various scheduled maintainance policies of repairable computer systems. The analysis is applicable to systems with failure behavior which can be modeled by a continuous-time Markov process, and thus important characteristics can be included in the model. Furthermore, we do not use the assumption of perfect repair, which is unrealistic for most systems. We obtain both transient and steady-state measures. The solution approach is based on the randomization technique and possesses advantages such as numerical stability and ease of implementation.

1 Introduction

Our interest is in analyzing scheduled maintenance policies of repairable computer systems. The maintenance policies we consider include those for which repair is performed when the system fails, and maintenance/repair is done at regular time intervals. Throughout the paper we use *scheduled repair* and *scheduled maintenance* interchangeably. These indicate a regular maintenance (for example, cleaning contacts, searching for failures, etc.), which may include repairs if the system is not fully operational upon the arrival of the scheduled repairman. The terminology *unscheduled repair* indicates an unscheduled visit of the repairman.

A large amount of literature exists concerning maintenance models and their evaluation [ASCH84]. However, it is usually assumed that the system is restored to full operation after each failure and/or repair times are ignored, which are not very realistic assumptions. In general, the theory of renewal processes is used to attack the problem. Furthermore, in most cases, only simple models of computer structures have been considered [HELV80,ODA81,YAK84]. Makam and Avizienis allow more complex models to be studied [MAKA81], but their approach implicitly assumes an average constant repair time. In their model, no repair is allowed if the system goes down before the next scheduled maintenance, and it is assumed that the system is restored to full configuration at the end of the scheduled repair.

In this paper we evaluate availability measures for models of computer systems in the presence of scheduled maintenance. Specifically, we study the three maintenance policies briefly described below. In the first policy, repair is allowed during fixed intervals of time which are T units apart, and it is not allowed during any other period. This behavior may represent a non-stationary satellite system, where repair can be performed only when the orbiting satellite passes through some points in its orbit. In the second policy, a maintenance period is always scheduled T time units apart. However, if the system goes down before the scheduled service, an extra service is performed to attempt to bring the system to full operation. This policy is typical for large system installations. In the third policy we assume that maintenance is scheduled T units of time apart, similar to the second policy. However, if the system goes down before the regular scheduled repair, an extra service is called for only if the system failure does not occur near the scheduled repair visit. If the extra repair service is performed, the next repair is rescheduled to a later time. If the extra repair is not performed, then the system remains inoperative until the next scheduled repair visit. This behavior is typical of small system facilities.

The basic assumption in all three policies is that the system failure behavior can be represented by a continuous-time Markov process. This is a common assumption, and it allows many important modeling details to be included. For a discussion of details which can be incorporated in a Markov process availability model we refer to [GOYA86]. We discuss cases for which the length of the scheduled maintenance can be dependent on the state of the system upon arrival of the repairman; automated repair may be performed when the system is degraded; complex failure behavior may be modeled when the repair is being made. Imperfect repair models are also considered, i.e., the system may not be left in full operation by the repairman. We allow the system to be repaired to any subset of states in the model. This is an important modeling capability, since in reality it is possible that the repairman may leave without finding all of the malfunctions in the system.

We calculate steady state as well as transient measures. Our solution method requires the evaluation of transient quantities, and as part of the analysis of each policy we use the randomization (or uniformization) technique [CINL75,ROSS83]. Randomization has been used previously to evaluate various transient measures in a Markovian environment [DESO86a,DESO86b,GRAS77,GROS84,MELA84,MILL83]. This technique has proven to be simple and efficient for the quantities we consider. We evaluate measures such as availability and expected number of unscheduled repair visits. However, the approach can be used to obtain other important measures.

The remainder of the paper is organized as follows. In section 2 we review the randomization technique, which is used to analyze the various policies. Sections 3, 4 and 5 describe the scheduled maintenance policies we consider, the assumptions made and the measures to be calculated. In section 6 we present examples which illustrate the approach and the importance of the measures we calculate. Section 7 contains conclusions based on the results of the paper. In appendix A a slightly different approach from that of sections 3, 4 and 5 for calculating transient measures is presented, which may reduce computation for some models.

2 Model Description and Background Material

In this section we first introduce the mathematical model that is used in the analysis of the various scheduled maintenance policies discussed in the previous section. We then give a brief overview of the method of solution, which involves the use of an analytical technique called randomization. We conclude the section by reviewing this technique and emphasizing several important properties of it which we use in our analysis.

The systems considered in this paper are repairable; automatic and/or manual techniques may be used in case of a failure. All failure distributions and repair distributions are assumed to be Markovian, and so general availability Markov models may be analyzed. In addition to unscheduled repairs, which are done when the system goes down, scheduled maintenance is also performed. These scheduled repair periods enable us to identify a sequence of points in time for which there is an imbedded Markov chain. The imbedded points correspond to the beginning of a repair phase in the first model, the arrival of a scheduled repairman in the second model, and the arrival of a repairman (whether scheduled or unscheduled) in the third model. We are able to calculate both transient and steady state availability measures by focusing attention on the imbedded chain.

Specifically, we assume that the behavior of the system of interest evolves according to a stochastic process with finite state space $S = \{a_i : i = 1, \ldots, M\}$. We also assume that there are L operational (up) states and K failed (down) states with $L + K = M$. We use the notation $S_O = \{a_i : i = 1, \ldots, L\}$ for the set of operational states and $S_F = \{a_i : L+1, \ldots, M\}$ for the set of failed states. For each of the models which are analyzed in the paper, we identify an imbedded Markov chain $\mathcal{Y} = \{Y_k : k = 0, 1, \ldots\}$ at time points $\{\tau_k : k = 0, 1, \ldots\}$ ($\tau_0 = 0$). Now suppose we wish to calculate both steady state and transient expected values of a measure $\mathcal{M}$. Such quantities include the fraction of time that the system is operational (the availability of the system) and the number of failures per unit time. We first determine the transition matrix $\mathbf{D}$ of $\mathcal{Y}$ (its ith row is denoted $\mathbf{d}_i$). We next associate a reward $\mathcal{M}_i$ with state a_i of the imbedded chain, which is equal to the value of $\mathcal{M}$ over an interval (τ_{k-1}, τ_k) given that $Y_{k-1} = a_i$ (these rewards are independent of k for the cases considered in the paper). The steady state and transient values of $\mathcal{M}$ are then obtained using results from the theory of Markov chains with rewards.

We first consider steady state results. Let $\mathcal{M}(k)$ be the (unconditional) total reward during $(0, \tau_k)$. From theorem 7.14 of [HEYM82] we conclude that (independent of the state at time 0)

$$\lim_{k\to\infty} \frac{\mathcal{M}(k)}{k} = \sum_{i=1}^{M} E[\mathcal{M}_i]\beta_i = \lim_{k\to\infty} \frac{E[\mathcal{M}(k)]}{k} \tag{1}$$

where the first equality holds with probability 1. Here β_i is the steady state probability of state a_i in the imbedded Markov chain. The vector $\beta = \langle \beta_1, \ldots, \beta_M \rangle$ can be obtained by solving $\beta = \beta\mathbf{D}$.

We next consider transient measures over the interval $(0, \tau_k)$. To calculate $E[\mathcal{M}(k)]$, we first calculate $\mathcal{M}_i$ for all states a_i and then uncondition those results based on the state probability distribution at τ_j ($j = 0, 1, \ldots, k-1$). Let $\beta_i(j) = P(Y_j = a_i)$, and let $\beta(j) = \langle \beta_1(j), \ldots, \beta_M(j) \rangle$ be the corresponding state probability vector. Then $\beta(j) = \beta(j-1)\mathbf{D}$. We have

$$E[\mathcal{M}(k)] = \sum_{i=1}^{M} E[\mathcal{M}_i] \sum_{j=0}^{k-1} \beta_i(j) \tag{2}$$

Once the transition matrix $\mathbf{D}$ and the expected values $\{E[\mathcal{M}_i] : i = 1, \ldots, M\}$ have been calculated, equations 1 and 2 may be used to obtain the steady state and transient measures. As we see below, the evaluation of $\mathbf{D}$ and $E[\mathcal{M}_i]$ involves the calculation of transient quantities in a Markovian environment. This can be done for all three models using a technique called randomization or uniformization. The randomization method is based on transforming a finite state continuous-time Markov process into a discrete-time Markov chain subordinated to a Poisson process. That is, let $\mathcal{X} = \{X(t) : t \geq 0\}$ be a (continuous-time) Markov process with generator $\mathbf{Q}$ and finite state space S. Then we may assume that $X(t) = Z_{N(t)}$ where $\mathcal{Z} = \{Z_n : n = 0, 1, \ldots\}$ is a (discrete-time) Markov chain and $\mathcal{N} = \{N(t) : t \geq 0\}$ is a Poisson process independent of $\mathcal{Z}$. The rate corresponding to $\mathcal{N}$ is $\Lambda \geq \max_{1\leq i\leq M} \lambda_i$ (where λ_i is the rate out of state a_i in $\mathcal{X}$), and the transition matrix for $\mathcal{Z}$ is $\mathbf{P} = \mathbf{Q}/\Lambda + \mathbf{I}$.

In order to use equations 1 and 2 to calculate steady state and transient quantities, we must first obtain values of certain measures over an interval of the form (τ_{k-1}, τ_k). Randomization may be used to calculate various transient quantities related to Markovian models for a given finite time period $(0, t)$. The intervals between the imbedded points τ_k in the first and second models are of constant length, and so randomization may be used in a straightforward manner. However, such intervals are not constant for the third model, and as a consequence the application of randomization in this case is more involved. The quantities of interest for the third model are determined from the corresponding quantities of an auxiliary process, and the latter may be calculated using randomization.

Among the quantities we calculate using randomization, we first consider the state probability vector $P(t) = \langle p_1(t), \ldots, p_M(t) \rangle$ where $p_i(t) = P[X(t) = a_i]$. By conditioning on $N(t)$, we have

$$P(t) = \sum_{n=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^n}{n!} \pi(n) \tag{3}$$

Here $\pi(n) = \langle \pi_1(n), \ldots, \pi_M(n) \rangle$ where $\pi_i(n) = P[Z_n = a_i]$. Note that $\pi(n) = \pi(n-1)\mathbf{P}$.

Next consider $O(t)$, the time that the system is in operational states during $(0, t)$ (the cumulative operational time), and we calculate its expected value using randomization. As before, we condition on $N(t)$ and obtain

$$E[O(t)] = \frac{1}{\Lambda} \sum_{n=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^{n+1}}{(n+1)!} \sum_{m=0}^{n} \|\pi_O(m)\| \tag{4}$$

Here $\pi_O(m) = \langle \pi_1(m), \ldots, \pi_L(m) \rangle$ is that part of the vector $\pi(m)$ which corresponds to the operational states. The vector $\pi_F(m)$ is defined in a similar manner for the failed states, namely, $\pi_F(m) = \langle \pi_{L+1}(m), \ldots, \pi_M(m) \rangle$. Also, for a nonnegative vector $v = \langle v_1, \ldots, v_J \rangle$, the notation $\|v\|$ represents the norm $\|v\| = \sum_{j=1}^{J} v_j$.

A third quantity of interest is the availability $A(t)$, or the fraction of time that the system was operational during $(0, t)$. Noting that $A(t) = O(t)/t$, we may use equation 4 to calculate its expected value.

We end this section by observing that the infinite sums in equations 3 and 4 must be truncated at some index N when these formulas are used for numerical calculations. The error produced in the evaluation of $P(t)$ is

$$e_P(N) \stackrel{\text{def}}{=} \sum_{n=N+1}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^n}{n!} \pi(n) \tag{5}$$

which clearly satisfies

$$\|e_P(N)\| \leq 1 - \sum_{n=0}^{N} e^{-\Lambda t} \frac{(\Lambda t)^n}{n!} \tag{6}$$

Similarly, the error obtained by truncation in equation 4 is

$$e_O(N) \stackrel{\text{def}}{=} \frac{1}{\Lambda} \sum_{n=N+1}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^{n+1}}{(n+1)!} \sum_{m=0}^{n} \|\pi_O(m)\| \tag{7}$$

and it satisfies

$$e_O(N) \leq t \left[1 - \sum_{n=0}^{N} e^{-\Lambda t} \frac{(\Lambda t)^n}{n!} \right] \tag{8}$$

In either case, N can be evaluated in advance for a given error tolerance ϵ.

3 First Model

3.1 Introduction

Consider a satellite in a non-stationary orbit about the earth with components which are subject to failure. During part of each orbit the satellite will pass within range of an earth station, and the station can initiate any needed repairs. If any new failures occur during this time period, these repairs can also be carried out. However, during the second part of each orbit the satellite is out of range of the earth station, and thus repairs cannot be performed. We can view the above situation as the case of a repairman who makes scheduled visits to a site (the time between visits is constant) and remains at the site for a constant length of time. We wish to find both steady state and transient availability measures of the system.

To analyze the above behavior, we introduce the following model. We assume that repair can start only at regular points in time of length T units apart. Furthermore, a repair is allowed only during a fixed period of time R, whether or not the system is brought to full operation. As indicated in Figure 1, the time line is divided into intervals of constant duration T. In each interval, repair is allowed only

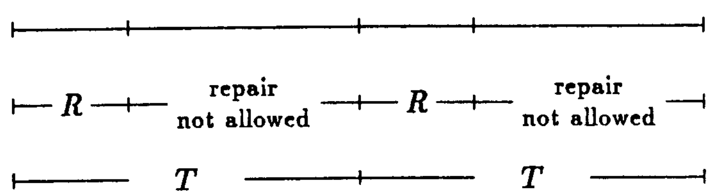

Figure 1: First model.

during the first R units of time (which we call the first phase or repair phase) and disallowed during the remaining $T-R$ units (the second phase or non-repair phase). Distinct generator matrices $\mathbf{Q}_1$, $\mathbf{Q}_2$ corresponding to the two phases are assumed to be given. Note that repair rates appear in $\mathbf{Q}_1$, but they are not present in $\mathbf{Q}_2$. Thus the behavior of the system is governed by a time-nonhomogeneous Markov process $\mathcal{X} = \{X(t), t \geq 0\}$. To achieve our goal of obtaining various availability measures, it is first necessary to examine the system behavior over a single interval of length T.

3.2 Results for Imbedded Points

Let $\tau_k = kT$ $(k = 0, 1, \ldots)$ be the consecutive points in time when repairs can begin to be made. It is easy to see that the behavior of the system during an interval (τ_{k-1}, τ_k) depends only on the state of the system at τ_{k-1}, and is independent of the behavior during other such intervals. Therefore, we can construct an imbedded Markov chain $\mathcal{Y}$ at the points τ_k. Without loss of generality, we begin our analysis of this first model by concentrating on the calculation of various measures for the interval $(\tau_0, \tau_1) = (0, T)$. The quantities of interest include state probability vectors at time T, the cumulative operational time during $(0, T)$ and the system availability. To find these quantities involves the calculation of transient probabilities for an interval of length T of a time-nonhomogeneous Markov process with generator $\mathbf{Q}_1$ during $(0, R)$ and generator $\mathbf{Q}_2$ during (R, T).

As discussed in [GROS84], randomization may be used to obtain transient probabilities for Markov processes with generator changes at discrete time points. Let Λ_1 , Λ_2 be the randomization rates corresponding to $\mathbf{Q}_1$, $\mathbf{Q}_2$ respectively. Also let $\mathbf{P}_1$, $\mathbf{P}_2$ be the transition matrices of the randomized Markov chains $\mathcal{Z}_1$, $\mathcal{Z}_2$ which correspond to the two phases, that is, $\mathbf{P}_1 = \mathbf{Q}_1/\Lambda_1 + \mathbf{I}$ and $\mathbf{P}_2 = \mathbf{Q}_2/\Lambda_2 + \mathbf{I}$. Further, let $\pi(n) = \langle \pi_1(n), \ldots, \pi_M(n) \rangle$ be the vector of state probabilities after n transitions of $\mathcal{Z}_1$, that is, $\pi(n) = \pi(n-1)\mathbf{P}_1$. Similarly, let $\nu(n)$ correspond to $\mathcal{Z}_2$, that is, $\nu(n) = \nu(n-1)\mathbf{P}_2$.

Assume that the state of the system at τ_0 is a_i. We define $\mathbf{c}_i = \langle c_{i1}, \ldots, c_{iM} \rangle$ to be the state probability vector of the system R time units later (at the end of the repair phase). Using randomization on an interval of length R, equation 3 gives

$$\mathbf{c}_i = \sum_{n=0}^{\infty} e^{-\Lambda_1 R} \frac{(\Lambda_1 R)^n}{n!} \pi(n) \tag{9}$$

where $\pi(0) = \mathbf{e}_i$, the unit vector in direction i. Similarly define $\mathbf{d}_i = \langle d_{i1}, \ldots, d_{iM} \rangle$ to be the probability vector corresponding to τ_1, the end of the interval. We now use randomization on an interval of length $T - R$ for a process with generator $\mathbf{Q}_2$ and obtain

$$\mathbf{d}_i = \sum_{n=0}^{\infty} e^{-\Lambda_2 (T-R)} \frac{[\Lambda_2 (T-R)]^n}{n!} \nu(n) \tag{10}$$

where $\nu(0) = \mathbf{c}_i$.

As mentioned in section 2, truncation of the infinite sums in equations 9 and 10 will lead to computational errors. Error bounds for transient probabilities of processes with generator changes are discussed in [GROS84]. As an example, suppose we wish to calculate the vector $\mathbf{d}_i$ within a specified tolerance ϵ. In this case the calculations on the interval $(0, R)$ are done using a tolerance of $(R/T)\epsilon$, and those on the interval (R, T) are done using a tolerance of $(1 - R/T)\epsilon$. The resulting value of $\mathbf{d}_i$ will be within ϵ as desired.

Using these state probabilities we can now find availability measures of system performance. We first consider O_i, the operational time during (τ_0, τ_1) given that the initial state is a_i. We may write

$$E[O_i] = E[O_{i1}] + E[O_{i2}] \tag{11}$$

where O_{i1} corresponds to the first (repair) phase and O_{i2} corresponds to the second (non-repair) phase. Using equation 4 we have

$$E[O_{i1}] = \frac{1}{\Lambda_1} \sum_{n=0}^{\infty} e^{-\Lambda_1 R} \frac{(\Lambda_1 R)^{n+1}}{(n+1)!} \sum_{m=0}^{n} \|\pi_O(m)\| \tag{12}$$

and

$$E[O_{i2}] = \frac{1}{\Lambda_2} \sum_{n=0}^{\infty} e^{-\Lambda_2 (T-R)} \frac{[\Lambda_2 (T-R)]^{n+1}}{(n+1)!} \sum_{m=0}^{n} \|\nu_O(m)\| \tag{13}$$

Finally, the expected availability for (τ_0, τ_1) with initial state a_i is simply

$$E[A_i] = \frac{E[O_i]}{T} \tag{14}$$

3.3 Steady State and Transient Measures

In this section our goal is to calculate steady state and transient availability measures of the system. To this end, we will use equations 1 and 2 applied to the appropriate performance measure. The first measure we consider is the cumulative operational time. Let $O(k)$ be the (unconditional) total operational time during $(0, kT)$. Then equation 1 yields

$$\lim_{k \to \infty} \frac{O(k)}{k} = \sum_{i=1}^{M} E[O_i]\beta_i = \lim_{k \to \infty} \frac{E[O(k)]}{k} \tag{15}$$

where β satisfies $\beta = \beta\mathbf{D}$. Note that the ith row of the matrix $\mathbf{D}$ is simply the vector $\mathbf{d}_i$, which is evaluated using equations 9 and 10. It remains to calculate $\{E[O_i], i = 1, \ldots, M\}$, which is done using equations 11, 12 and 13. We next obtain the limiting availability of the system as follows. Let

$$A(k) = \frac{O(k)}{kT} \tag{16}$$

be the (unconditional) availability of the system over $(0, kT)$. Then we have

$$\lim_{k\to\infty} A(k) = \sum_{i=1}^{M} E[A_i]\beta_i = \lim_{k\to\infty} E[A(k)] \tag{17}$$

where $A_i = O_i/T$.

We now show that transient measures such as the expected availability during an interval $(0, t)$ can be easily obtained for this model. First assume that t is a multiple of T, say $t = kT = \tau_k$. Then we may compute the expected operational time during $(0, t)$ from equation 2 as

$$E[O(t)] = \sum_{i=1}^{M} E[O_i] \sum_{j=0}^{k-1} \beta_i(j) \tag{18}$$

The expected availability over this finite time period is

$$E[A(t)] = \sum_{i=1}^{M} E[A_i] \sum_{j=0}^{k-1} \beta_i(j) \tag{19}$$

The general case of $t = kT + x$ (where $0 < x < T$) follows by first using the above procedure to obtain results for $(0, kT)$, and then calculating the measure over the additional interval of length x using randomization.

4 Second Model

4.1 Introduction

In this section we consider the following maintenance policy. We assume that a repairman makes a visit to inspect, maintain and possibly repair the system at regular time intervals (for example, at the beginning of each month). If the system goes down before the next scheduled visit of the repairman, an extra service is performed in an attempt to bring the system to full operation. This policy is typical for large systems, where extra repair costs are usually small compared to the operating cost of the facility.

We analyze this policy using the following model. We assume that scheduled repair can start only at intervals of length T units apart. During such a visit the system is shut down until the maintenance is completed. If the state of the system when the repairman arrives is a_i, then we assume that the scheduled repair time is a random variable ξ_i, the distribution of which is exponential truncated at T with parameter θ_i. The truncated exponential is used, since the scheduled repair time until the next such visit can be at most T time units in length. We allow imperfect repairs to occur, i.e., the system may not always be brought to full operation when a scheduled repair is completed. Upon the departure of the scheduled repairman, he leaves the system in a state a_j with a probability c_{ij} which depends on the state of the system a_i that he initially encounters. This modeling aspect is important to consider, since in actual installations some errors may remain unrepaired at the end of the maintenance period. We model the failure behavior of the system during the period after the scheduled repairman leaves until the next scheduled maintenance by a continuous time Markov process $\mathcal{X} = \{X(t) : t \geq 0\}$ with generator $\mathbf{Q}$. During the time between the scheduled maintenance periods, the system may be repaired if it goes down. This unscheduled repair may also be imperfect and dependent on the current down state. The repair performed by the unscheduled repairman is modeled by including transitions out of the failed states in $\mathbf{Q}$. There is a possibility that the unscheduled repair may not be completed at the time of the next scheduled repair. In this case the scheduled repair takes place in the state of the system at that time. Figure 2 illustrates the policy. In that figure the arrows indicate the arrival of the scheduled repairman. The time interval between the arrows has constant length T.

We are interested in calculating both steady state and transient measures of the system. These include operational time and availability quantities similar to those considered in the first model. In addition, we consider the number of unscheduled

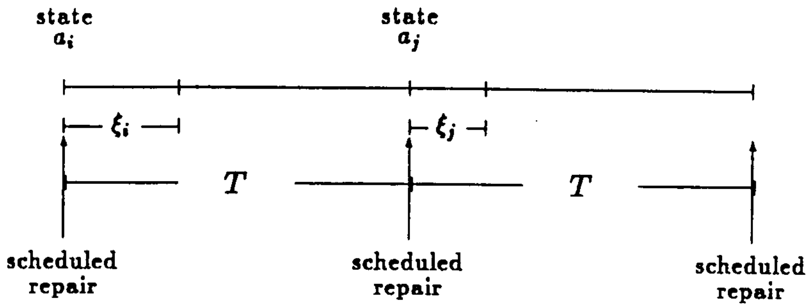

Figure 2: Second model.

repairs, i.e., those repairs that occur due to a failure of the system in the intervals between scheduled repairs, and we also consider the fraction of such intervals in which the unscheduled repairman has to be called. These last measures are important in order to evaluate the effectiveness of the scheduled maintenance contract. If scheduled repairs occur too infrequently, the system probably will suffer several down periods, certainly an undesirable situation. Similar to the previous model, we first examine the system behavior for an interval of length T between two scheduled repairs.

4.2 Results for Imbedded Points

Let τ_k be the beginning of the kth scheduled repair period, that is, $\tau_k = kT$ ($k = 0, 1, \ldots$). As was the case for the first model, the system behavior during any interval (τ_{k-1}, τ_k) depends only on the state of the system at τ_{k-1} and is independent of the behavior during the other intervals. Therefore, we can construct an imbedded Markov chain $\mathcal{Y} = \{Y_k : k = 0, 1, \ldots\}$ at τ_k, the points of arrival of the scheduled repairman. As for the first model, we may begin our analysis by calculating the quantities of interest for the interval $(\tau_0, \tau_1) = (0, T)$. We first need to calculate the transition matrix $\mathbf{D}$ of $\mathcal{Y}$ and the expected operational times $\{E[O_i], i = 1, \ldots, M\}$.

We assume that the state of the system at τ_0 is a_i, and thus the corresponding state probability vector is $\mathbf{e}_i$. As stated above, we assume that the probability occupancy vector for the end of the scheduled repair period and the mean length of the scheduled repair period are supplied as input. That is, the modeler supplies the vector $\mathbf{c}_i = \langle c_{i1}, \ldots, c_{iM} \rangle$ of probabilities that the repairman leaves the system in a certain state at the end of his scheduled visit, given that he found the system in state a_i upon arrival. The modeler also supplies the value of the parameter θ_i of the distribution of the scheduled repair period

$$P[\xi_i \le t] = \begin{cases} 1 - e^{-\theta_i t} & t < T \\ 1 & t \ge T \end{cases} \tag{20}$$

Recall that the behavior of the system after the scheduled repairman leaves is given by the (time-homogeneous) Markov process $\mathcal{X}$ with generator $\mathbf{Q}$. Let Λ be the randomization rate corresponding to $\mathbf{Q}$, and let $\mathbf{P} = \mathbf{Q}/\Lambda + \mathbf{I}$ be the transition matrix of the resulting randomized Markov chain $\mathcal{Z}$. In what follows, we will use randomization to calculate several quantities of interest.

The ith row of $\mathbf{D}$ is the state probability vector at time T, given that the state at the arrival of the scheduled repairman is a_i. This vector, $\mathbf{d}_i = \langle d_{i1}, \ldots, d_{iM} \rangle$, can be obtained from randomization as follows. With probability $e^{-\theta_i T}$ the scheduled repair will not be completed, and the state at τ_1 will again be given by the vector $\mathbf{e}_i$. Otherwise, the repair will finish after say $t < T$, and we use randomization on an interval of length $T - t$. Conditioning on $\xi_i = t$, we have from equation 3

$$\mathbf{d}_i|_{\xi_i = t} = \sum_{n=0}^{\infty} e^{-\Lambda(T-t)} \frac{[\Lambda(T-t)]^n}{n!} \pi(n) \tag{21}$$

where $\pi(0) = \mathbf{c}_i$ and $\pi(n) = \pi(n-1)\mathbf{P}$. Unconditioning gives

$$\mathbf{d}_i = \int_0^T \sum_{n=0}^{\infty} e^{-\Lambda(T-t)} \frac{[\Lambda(T-t)]^n}{n!} \pi(n) \theta_i e^{-\theta_i t}\, dt + e^{-\theta_i T} \mathbf{e}_i \tag{22}$$

We rewrite this as

$$\mathbf{d}_i = \sum_{n=0}^{\infty} I_n(T) \pi(n) + e^{-\theta_i T} \mathbf{e}_i \tag{23}$$

where

$$I_n(T) \stackrel{\text{def}}{=} \int_0^T \theta_i e^{-\theta_i t} e^{-\Lambda(T-t)} \frac{[\Lambda(T-t)]^n}{n!}\, dt \tag{24}$$

We now evaluate $I_n(T)$ and find

$$I_n(T) = \sum_{m=n+1}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T [(\Lambda - \theta_i)T]^{m-(n+1)}}{m!} \tag{25}$$

Substituting this expression for $I_n(T)$ into equation 23, reversing the order of summation, and simplifying yields

$$\mathbf{d}_i = \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T}{(n+1)!} \sum_{m=0}^{n} \left(1 - \frac{\theta_i}{\Lambda}\right)^{n-m} \pi(m) + e^{-\theta_i T} \mathbf{e}_i \tag{26}$$

Similarly, $E[O_i]$ can be obtained as follows. With probability $e^{-\theta_i T}$, the repair will not be completed; in this case $O_i = 0$. Otherwise, the repair will finish after say $t < T$, and we use randomization on an interval of length $T - t$. So conditioning on $\xi_i = t$, using equation 4 we find

$$E[O_i|\xi_i = t] = \frac{1}{\Lambda}\sum_{n=0}^{\infty} e^{-\Lambda(T-t)}\frac{[\Lambda(T-t)]^{n+1}}{(n+1)!}\sum_{l=0}^{n}||\pi_O(l)|| \tag{27}$$

Unconditioning gives

$$E[O_i] = \int_0^T \frac{1}{\Lambda}\sum_{n=0}^{\infty} e^{-\Lambda(T-t)}\frac{[\Lambda(T-t)]^{n+1}}{(n+1)!}\sum_{l=0}^{n}||\pi_O(l)||\theta_i e^{-\theta_i t}\,dt \tag{28}$$

Recognizing $I_{n+1}(T)$ in the above equation, we have

$$E[O_i] = \frac{1}{\Lambda}\sum_{n=0}^{\infty} I_{n+1}(T)\sum_{l=0}^{n}||\pi_O(l)|| \tag{29}$$

Using the value for $I_{n+1}(T)$ from equation 25, we obtain

$$E[O_i] = \frac{1}{\Lambda}\sum_{n=0}^{\infty} e^{-\Lambda T}\frac{(\Lambda T)^{n+1}\theta_i T}{(n+2)!}\sum_{m=0}^{n}\left(1-\frac{\theta_i}{\Lambda}\right)^{n-m}\sum_{l=0}^{m}||\pi_O(l)|| \tag{30}$$

For this second model, we also consider measures involving the number of unscheduled repairs during (τ_0, τ_1). Given that the initial state at τ_0 is a_i, let U_i be a random variable which represents the number of times a repairman had to be called during (ξ_i, τ_1). We wish to calculate the expected number of unscheduled repairs, $E[U_i]$. If the scheduled repair period does not complete (with probability $e^{-\theta_i T}$), then $U_i = 0$. Otherwise we condition on $\xi_i = t$ and use randomization. We mark a transition $r \to s$ in the randomized Markov chain if it causes a visit of the unscheduled repairman. Then we have

$$E[U_i|\xi_i = t] = \sum_{n=1}^{\infty} e^{-\Lambda(T-t)}\frac{[\Lambda(T-t)]^{n}}{n!}\sum_{l=1}^{n}P(l\text{th transition is marked}) \tag{31}$$

Thus we need to find the probability that a particular transition is marked. Let

$$f_{rs} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } r \to s \text{ is marked} \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

The entries f_{rs} are input parameters of the problem and are specified by the analyst. The matrix $\mathbf{U} = [p_{rs} f_{rs}]$ gives the one-step probabilities of being marked. Now define the vectors $\psi(l) = \pi(l-1)\mathbf{U}$ for $l = 1, 2, \ldots$. Then the probability that the lth transition is marked is the sum of the components of $\psi(l)$. Therefore

$$E[U_i|\xi_i = t] = \sum_{n=1}^{\infty} e^{-\Lambda(T-t)} \frac{[\Lambda(T-t)]^n}{n!} \sum_{l=1}^{n} \|\psi(l)\| \quad (3'$$

Unconditioning gives

$$E[U_i] = \sum_{n=1}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T}{(n+1)!} \sum_{m=1}^{n} \left(1 - \frac{\theta_i}{\Lambda}\right)^{n-m} \sum_{l=1}^{m} \|\psi(l)\| \quad (34)$$

We now wish to determine the probability that an unscheduled repair takes place during (τ_0, τ_1). That is, we wish to calculate $P(U_i \geq 1) = 1 - P(U_i = 0)$. We will concentrate on the calculation of $P(U_i = 0)$ instead. If the scheduled repair period does not finish, then $U_i = 0$. Otherwise the scheduled repair is completed at $t < T$, and we use randomization as above to obtain

$$P(U_i = 0|\xi_i = t) = \sum_{n=0}^{\infty} e^{-\Lambda(T-t)} \frac{[\Lambda(T-t)]^n}{n!} P(\text{no marked transitions}) \quad (35)$$

Now define the vectors $\phi(n) = \pi(0)(\mathbf{P} - \mathbf{U})^n$ for $n = 0, 1, \ldots$. Then, given n transitions, the probability that there are no marked transitions is the sum of the components of $\phi(n)$. Therefore

$$P(U_i = 0|\xi_i = t) = \sum_{n=0}^{\infty} e^{-\Lambda(T-t)} \frac{[\Lambda(T-t)]^n}{n!} \|\phi(n)\| \quad (36)$$

Unconditioning gives

$$P(U_i = 0) = \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T}{(n+1)!} \sum_{m=0}^{n} \left(1 - \frac{\theta_i}{\Lambda}\right)^{n-m} \|\phi(m)\| + e^{-\theta_i T} \quad (37)$$

Equations 26, 30, 34 and 37 give formulas for the various quantities of interest which may be evaluated recursively. To illustrate such a recursion, we consider the evaluation of $\mathbf{d}_i$. We first rewrite equation 26 in the form

$$\mathbf{d}_i = \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T}{(n+1)!} f(n) + e^{-\theta_i T} \mathbf{e}_i \quad (38)$$

where

$$f(n) \stackrel{\text{def}}{=} \sum_{m=0}^{n} \left(1 - \frac{\theta_i}{\Lambda}\right)^{n-m} \pi(m) \quad (39)$$

Note that

$$f(n+1) = \left(1 - \frac{\theta_i}{\Lambda}\right) f(n) + \pi(n+1) \tag{40}$$

and thus $f(n)$ may be calculated recursively. As a second example, let us evaluate $E[O_i]$. Equation 30 may be written as

$$E[O_i] = \frac{1}{\Lambda} \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^{n+1} \theta_i T}{(n+2)!} g(n) \tag{41}$$

where

$$g(n) \stackrel{\text{def}}{=} \sum_{m=0}^{n} \left(1 - \frac{\theta_i}{\Lambda}\right)^{n-m} \sum_{l=0}^{m} \|\pi_O(l)\| \tag{42}$$

Note that

$$g(n+1) = \left(1 - \frac{\theta_i}{\Lambda}\right) g(n) + \sum_{l=0}^{n+1} \|\pi_O(l)\| \tag{43}$$

and so $g(n)$ may also be calculated recursively. The expressions given by equations 34 and 37 may be evaluated in a similar manner.

For $\Lambda > \theta_i$, the geometric expressions in equations 26, 30, 34 and 37 involve positive terms which are less than one, and the corresponding recursions are numerically stable. On the other hand, when the maximum transition rate in $\mathbf{Q}$ is at most θ_i, we can always randomize with $\Lambda = \theta_i$. In this case expressions for $\mathbf{d}_i$, $E[O_i]$, $E[U_i]$ and $P(U_i = 0)$ simplify. The value of $\mathbf{d}_i$ becomes

$$\mathbf{d}_i = \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^{n+1}}{(n+1)!} \pi(n) + e^{-\Lambda T} \mathbf{e}_i \tag{44}$$

while the expression for $E[O_i]$ reduces to

$$E[O_i] = \frac{1}{\Lambda} \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^{n+2}}{(n+2)!} \sum_{l=0}^{n} \|\pi_O(l)\| \tag{45}$$

Also, for $\Lambda = \theta_i$, the expression for $E[U_i]$ becomes

$$E[U_i] = \sum_{n=1}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^{n+1}}{(n+1)!} \sum_{l=1}^{n} \|\psi(l)\| \tag{46}$$

and the value of $P(U_i = 0)$ is

$$P(U_i = 0) = \sum_{n=0}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^{n+1}}{(n+1)!} \|\phi(n)\| + e^{-\Lambda T} \tag{47}$$

Error bounds for the truncation of the infinite sum in equations 26, 30, 34 and 37 are similar to those given in equations 6 and 8. As an example, we determine the error in the calculation of $\mathbf{d}_i$ (for $\Lambda > \theta_i$) when the infinite sum in equation 26 is truncated at the index N. In this case, the error is

$$e_d(N) \stackrel{\text{def}}{=} \sum_{n=N+1}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T}{(n+1)!} \sum_{m=0}^{n} \left(1 - \frac{\theta_i}{\Lambda}\right)^{n-m} \pi(m) \tag{48}$$

which satisfies

$$\|e_d(N)\| \leq \sum_{n=N+1}^{\infty} e^{-\Lambda T} \frac{(\Lambda T)^n \theta_i T}{(n+1)!} \left[\frac{1 - (1 - \theta_i/\Lambda)^{n+1}}{1 - (1 - \theta_i/\Lambda)} \right] \tag{49}$$

Thus we have the bound

$$\|e_d(N)\| \leq 1 - \sum_{n=0}^{N+1} e^{-\Lambda T} \frac{(\Lambda T)^n}{n!} \tag{50}$$

Error bounds for other quantities may be obtained in a similar manner.

In many cases, aggregating states which have similar characteristics will considerably reduce the amount of computation necessary for a solution. States of the Markov process $\mathcal{X}$ with the same parameters may be aggregated, if the measure to be calculated is not affected by the aggregation. For example, the cumulative operational time does not change if failed states with the same characteristics are aggregated. Another aggregation which is less apparent involves the imbedded Markov chain $\mathcal{Y}$. States of this chain (which represent the system when the scheduled repairman arrives) may be aggregated if they have the same repair parameter θ and the same probability vector $\mathbf{c}$ for the end of the repair phase. Assuming such states are aggregated into M_A distinct subsets, then the sums in equations 1 and 2 have only M_A terms, and the computation involved in calculating the various measures is reduced. An example which illustrates such an aggregation is presented in section 6.

4.3 Steady State and Transient Measures

We now indicate how steady state and transient availability measures of the system can be calculated. We will apply equations 1 and 2 in a manner identical to that of section 3.3. Limiting expressions involving the cumulative operational time and the system availability are again given by equations 15 and 17. For this second model, the matrix $\mathbf{D}$ and the expected operational time $E[O_i]$ are given by equations 26 and 30, respectively. We can also find the limiting number of unscheduled visits of the repairman between scheduled repair periods. Let $U(k)$ be the (unconditional) number of unscheduled repairs during $(0, kT)$. Applying equation 1, we have

$$\lim_{k\to\infty} \frac{U(k)}{k} = \sum_{i=1}^{M} E[U_i]\beta_i = \lim_{k\to\infty} \frac{E[U(k)]}{k} \tag{51}$$

The quantities $\{E[U_i] : i = 1, \ldots, M\}$ are calculated using equation 34. Finally, we consider the fraction of intervals between scheduled repair periods for which an unscheduled visit is necessary. We concentrate on the equivalent problem of calculating the fraction of intervals with no unscheduled repair. Define a reward V_i which is 1 if no unscheduled repair occurs during an interval and 0 if at least one unscheduled repair does occur, given that the initial state is a_i. Note that $E[V_i] = P(U_i = 0)$. Next let $V(k)$ be the (unconditional) reward during $(0, kT)$, i.e., $V(k)$ is the number of intervals during $(0, kT)$ with no unscheduled repair. We have

$$\lim_{k\to\infty} \frac{V(k)}{k} = \sum_{i=1}^{M} P(U_i = 0)\beta_i = \lim_{k\to\infty} \frac{E[V(k)]}{k} \tag{52}$$

where we use equation 37 to calculate $P(U_i = 0)$.

Transient measures for the above quantities during a time period $(0, t)$ can be obtained for this model in a manner similar to that of the first model. For notational convenience, we consider only the case when $t = kT = \tau_k$ for some k; the case when t is not a multiple of T follows as in section 3.3. The expected operational time during $(0, t)$ and the expected availability during this time period may be calculated using equations 18 and 19. Similarly, the expected number of unscheduled visits during $(0, t)$ is given by

$$E[U(t)] = \sum_{i=1}^{M} E[U_i] \sum_{j=0}^{k-1} \beta_i(j) \tag{53}$$

and the expected number of intervals with no unscheduled repair is

$$E[V(t)] = \sum_{i=1}^{M} P(U_i = 0) \sum_{j=0}^{k-1} \beta_i(j) \tag{54}$$

5 Third Model

5.1 Introduction

In this section we consider the following maintenance policy. As was the case for the second model, we assume that a repairman makes a visit to inspect, maintain and possibly repair the system at regular time intervals. During the scheduled maintenance period, the system is not working and is unavailable to the users. The repairman leaves the system in operation, but it may go down before the next scheduled visit. In this case an extra repair is necessary to attempt to bring the system to full operation. However, unlike the second model, such an unscheduled repair does not always take place. Specifically, if the system failure occurs near to the time of the next scheduled repair visit, an extra repair is not called for. In this case, the system is left down until the scheduled repairman arrives. But if the system goes down with a long period of time remaining until the next scheduled repair, an unscheduled visit is called for. In this case, not only does the unscheduled repairman arrive to repair the system, but also the next scheduled visit is rescheduled to a later time. This policy is typical for small systems, for which it may not always be cost efficient to call for an unscheduled repair.

We analyze this policy using a model, several assumptions of which are identical to those of the second model. For example, we assume that scheduled repair occurs only at intervals of length T time units apart. During the scheduled visit, the system is shut down until the end of the visit. The length of the scheduled visit is a random variable $\xi_{i,S}$, the distribution of which is exponential truncated at T with a parameter $\theta_{i,S}$ which depends on the state of the system a_i at the start of the visit. The scheduled repairman leaves the system in a state a_j with a probability $c_{ij,S}$ which depends on the state a_i at the beginning of the repair period. Unlike the second model, if the system fails after the scheduled repairman leaves, it is possible that an unscheduled repair does not take place. We assume that there is a (constant) value ω such that, if the system failure occurs within ω time units of the next scheduled visit, the unscheduled repairman is not called and the system remains down until the scheduled repairman arrives. However, if the system failure occurs more than ω time units from the next scheduled repair, then an unscheduled repair does take place. In this latter case the next scheduled visit is rescheduled for T time units after the arrival of the unscheduled repairman. The length of the unscheduled repair period is a random variable $\xi_{i,U}$, the distribution of which is exponential truncated at T with a parameter $\theta_{i,U}$ which depends on the state of the system a_i at the time of the unscheduled repair. The system is left in a state a_j at the end of the unscheduled repair with a probability $c_{ij,U}$ which depends on the state a_i at the arrival of the unscheduled repairman. Note that different values of length of repair for scheduled and unscheduled visits are useful for including modeling details such as traveling time for unscheduled repair, etc. We model the system behavior after the (scheduled or unscheduled) repairman leaves by a continuous time Markov process $\mathcal{X} = \{X(t) : t \geq 0\}$ with generator $\mathbf{Q}$. Figure 3 illustrates the policy. In

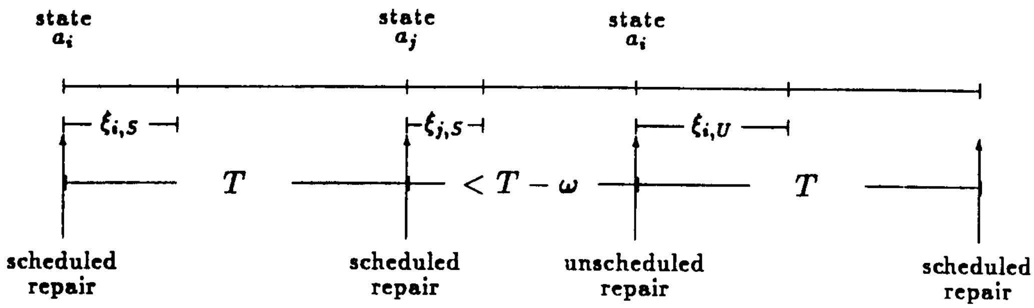

Figure 3: Third model.

that figure the arrows indicate the arrival of a repairman.

We wish to calculate both steady state and transient quantities of system behavior. As before, we are interested in the cumulative operational time and the availability of the system. We also consider the fraction of times that an unscheduled repair is called for before the next scheduled repair.

5.2 Results for Imbedded Points

Consider the points of arrival of a repairman, both scheduled and unscheduled. Let τ_k $(k = 0, 1, \ldots)$ be the kth such arrival point. Similar to the analysis of the first and second models, we identify an imbedded Markov chain $\mathcal{Y} = \{Y_k : k = 0, 1, \ldots\}$ at these points. We first describe the state space of the imbedded chain $\mathcal{Y}$. Unlike the second model, an interval (τ_{k-1}, τ_k) may begin with either a scheduled or an unscheduled repair. If the initial state is operational, then a scheduled repair must result. However, if the initial state is a failed state, then either type of repair is possible. Since the input parameters (the mean length of the repair period and the state probability vector at the departure of the repairman) for a failed state are different in the two cases, we introduce *two* failed states in the imbedded chain $\mathcal{Y}$ for each failed state in the Markov process $\mathcal{X}$. Specifically, let $S = \{a_i : i = 1, \ldots, M\}$ be the state space of $\mathcal{X}$, with S_O and S_F defined as before. Then the state space

S^* of $\mathcal{Y}$ is a set of pairs which indicate the (original) state and the type of repair. That is, we define

$$\begin{aligned} S_O^* &= \{(a_i, S) : i = 1, \ldots, L\} && \text{(operational states)} \\ S_S^* &= \{(a_i, S) : i = L+1, \ldots, M\} && \text{(failed states, scheduled repair)} \\ S_U^* &= \{(a_i, U) : i = L+1, \ldots, M\} && \text{(failed states, unscheduled repair)} \end{aligned} \tag{55}$$

and we have $S^* = S_O^* \cup S_S^* \cup S_U^* = \{a_\delta^* : \delta = 1, \ldots, M+K\}$ (recall that $L + K = M$).

With this (expanded) definition of state, we see that the behavior of the system during an interval (τ_{k-1}, τ_k) depends only on the state of the system at τ_{k-1}, and is independent of other such intervals. Thus we concentrate on calculating the quantities of interest for the first interval $(0, \tau_1)$. Let a_δ^* be the state of the system at τ_0. The repair period ξ_δ corresponds to a scheduled visit for $\delta = 1, \ldots, M$ and to an unscheduled visit for $\delta = M+1, \ldots, M+K$. As in the second model, we assume that the probability occupancy vector for the end of the repair period and the mean length of the repair period are input parameters. That is, we are given the vector $\mathbf{c}_\delta = \langle c_{\delta,1}, \ldots, c_{\delta,M} \rangle$ of probabilities that the repairman leaves the system in a certain state at the end of the repair period (recall that there are M such states, operational and failed). We are also given that the distribution of the repair period ξ_δ satisfies equation 20 with a parameter θ_δ. Note that ξ_δ, $\mathbf{c}_\delta$, θ_δ correspond to the input parameters $\xi_{i,S}$, $\mathbf{c}_{i,S}$, $\theta_{i,S}$ or $\xi_{i,U}$, $\mathbf{c}_{i,U}$, $\theta_{i,U}$ depending on whether δ corresponds to a scheduled repair or an unscheduled repair.

Note that the interval $(0, \tau_1)$ is not constant, and so randomization cannot be used directly. However, instead of looking at the original process $\mathcal{X}$ on $(0, \tau_1)$, the measures of interest can be calculated by considering an auxiliary process $\mathcal{X}_1$ on $(0, T)$. The generator $\mathbf{Q}_1$ of $\mathcal{X}_1$ is similar to the original generator $\mathbf{Q}$, except that all failed states are absorbing states. We now describe the behavior of the system under the auxiliary process $\mathcal{X}_1$. If the system fails during $(T - \omega, T)$, under $\mathcal{X}_1$ a scheduled repair will occur at time T, and the system will remain in the particular failed state from the time of the failure until time T. In this case the behavior under the auxiliary process $\mathcal{X}_1$ is seen to be identical to that under the original process $\mathcal{X}$. However, if the system fails during $(0, T - \omega)$, under $\mathcal{X}_1$ the system will remain in the particular failed state until time T, and once again an unscheduled repair will occur at time T. This behavior differs from that under the original process $\mathcal{X}$, since in the latter case an unscheduled repair starts immediately. Since additional transitions do not occur once the process $\mathcal{X}_1$ reaches a failed state, the state probability vectors at T and the operational time during $(0, T)$ for the auxiliary process $\mathcal{X}_1$ are the same as the corresponding quantities during the interval $(0, \tau_1)$ for the original process $\mathcal{X}$. We will use randomization to calculate the quantities for the process $\mathcal{X}_1$. To that end, we let Λ_1 be the randomization rate corresponding to the generator $\mathbf{Q}_1$, and we let $\mathbf{P}_1 = \mathbf{Q}_1/\Lambda_1 + \mathbf{I}$ be the transition matrix of the randomized Markov chain $\mathcal{Z}_1$. Note that the state space of $\mathcal{Z}_1$ is simply S.

We now calculate $\mathbf{d}_\delta = \langle d_{\delta,1}, \ldots, d_{\delta,M+K}\rangle$, the state probability vector of the imbedded chain $\mathcal{Y}$ at τ_1. We first consider the failed states which correspond to an unscheduled repair, that is, we determine $\mathbf{d}_U \stackrel{\text{def}}{=} \langle d_{\delta,M+1}, \ldots, d_{\delta,M+K}\rangle$. If the repair period (whether scheduled or unscheduled) lasts at least until $T-\omega$, then these probabilities are zero. Otherwise the repair will finish at $t < T-\omega$, and we use randomization on an interval of length $T-\omega-t$ to find the probabilities of failed states at time $T-\omega$. Similar to the derivation of equation 23, we obtain

$$\mathbf{d}_U = \sum_{n=0}^{\infty} I_n(T-\omega)\pi_F(n) \tag{56}$$

where $\pi(0) = \mathbf{c}_i$ and $\pi(n) = \pi(n-1)\mathbf{P}_1$ (recall that $\pi_F(n) = \langle \pi_{L+1}(n), \ldots, \pi_M(n)\rangle$). This simplifies to

$$\mathbf{d}_U = \sum_{n=0}^{\infty} e^{-\Lambda_1(T-\omega)} \frac{[\Lambda_1(T-\omega)]^n \theta_\delta(T-\omega)}{(n+1)!} \sum_{m=0}^{n} \left(1 - \frac{\theta_\delta}{\Lambda_1}\right)^{n-m} \pi_F(m) \tag{57}$$

We next find $\mathbf{d}_O \stackrel{\text{def}}{=} \langle d_{\delta,1}, \ldots, d_{\delta,L}\rangle$, the probability vector for the operational states (note that these states must correspond to a scheduled repair). With probability $e^{-\theta_\delta T}$, the repair period will last for the interval $(0,T)$, and the operational state probabilities will be given by $\mathbf{e}_O$, the vector $\mathbf{e}_\delta$ restricted to the operational states. Otherwise, the repair period ends at $t < T$, and we use randomization on an interval of length $T-t$ as before. This yields

$$\mathbf{d}_O = \sum_{n=0}^{\infty} I_n(T)\pi_O(n) + e^{-\theta_\delta T}\mathbf{e}_O \tag{58}$$

(recall that $\pi_O(n) = \langle \pi_1(n), \ldots, \pi_L(n)\rangle$). This simplifies to

$$\mathbf{d}_O = \sum_{n=0}^{\infty} e^{-\Lambda_1 T} \frac{(\Lambda_1 T)^n \theta_\delta T}{(n+1)!} \sum_{m=0}^{n} \left(1 - \frac{\theta_\delta}{\Lambda_1}\right)^{n-m} \pi_O(m) + e^{-\theta_\delta T}\mathbf{e}_O \tag{59}$$

Finally, we determine $\mathbf{d}_S \stackrel{\text{def}}{=} \langle d_{\delta,L+1}, \ldots, d_{\delta,M}\rangle$, the probability vector for the failed states which correspond to a scheduled repair. If the repair period does not finish (with probability $e^{-\theta_\delta T}$) and the interval $(0,T)$ began in either of the two failed states $a^*_\delta = (a_\delta, S)$ or $a^*_{\delta+K} = (a_\delta, U)$ ($\delta = L+1, \ldots, M$), then the next interval $(T, 2T)$ will begin with a scheduled repair in the failed state a^*_δ. Otherwise, the repair period ends at $t < T$, and we use randomization to determine the probability of a failed state at T. Note that we must subtract off the probability of a failed state at $T - \omega$, since this was already accounted for in the unscheduled repair calculation of equation 57. Carrying out this procedure, we find

$$\mathbf{d}_S = \sum_{n=0}^{\infty} I_n(T)\pi_F(n) - \mathbf{d}_U + e^{-\theta_\delta T}\{\mathbf{e}_S + \mathbf{e}_U\} \tag{60}$$

or

$$\mathbf{d}_S = \sum_{n=0}^{\infty} e^{-\Lambda_1 T}\frac{(\Lambda_1 T)^n \theta_\delta T}{(n+1)!}\sum_{m=0}^{n}\left(1 - \frac{\theta_\delta}{\Lambda_1}\right)^{n-m}\pi_F(m) - \mathbf{d}_U + e^{-\theta_\delta T}\{\mathbf{e}_S + \mathbf{e}_U\} \tag{61}$$

We now calculate the expected cumulative operational time, $E[O_\delta]$, during the interval $(0, \tau_1)$. Recall that the system is not operational during the repair period (whether scheduled or unscheduled) no matter what the state of the system at the arrival of the repairman. Also recall that failed states are absorbing states in the randomized Markov chain $\mathcal{Z}_1$, so that the cumulative operational time during $(0, T)$ under the auxiliary process $\mathcal{X}_1$ is equal to the cumulative operational time during $(0, \tau_1)$ under the original process $\mathcal{X}$. Thus we may proceed as in the derivation of equation 30 for the second model and obtain

$$E[O_\delta] = \frac{1}{\Lambda_1}\sum_{n=0}^{\infty} e^{-\Lambda_1 T}\frac{(\Lambda_1 T)^{n+1}\theta_\delta T}{(n+2)!}\sum_{m=0}^{n}\left(1 - \frac{\theta_\delta}{\Lambda_1}\right)^{n-m}\sum_{l=0}^{m}\|\pi_O(l)\| \tag{62}$$

In order to obtain the limiting availability of the system, we will need to determine the expected length of time between successive arrivals of a repairman. That is, let C_δ be the value of τ_1 given that the initial state is a_δ^*, and we wish to calculate $E[C_\delta]$. If the repair period ξ_δ satisfies $\xi_\delta \geq T - \omega$, then an unscheduled repair will not be called for, and so $C_\delta = T$. However, if $\xi_\delta < T - \omega$, then C_δ may be thought of as the sum of ξ_δ plus the operational time during $(0, T)$ in a certain time-nonhomogeneous Markov process $\mathcal{X}_2$ with a generator change at $T - \omega$. During $(0, T - \omega)$ the generator of $\mathcal{X}_2$ is identical to $\mathbf{Q}_1$ of the auxiliary process $\mathcal{X}_1$, but during $(T - \omega, T)$ the generator of $\mathcal{X}_2$ corresponds to a process with no state changes. Since $C_\delta = T$ if the system is operational at $T - \omega$, it is clear that the operational time corresponding to $\mathcal{X}_2$ yields the desired quantity. Thus for $\xi_\delta < T - \omega$, we finally have

$$C_\delta = \xi_\delta + O_\delta(T - \omega) + \omega\mathcal{I}_\delta(T - \omega) \tag{63}$$

Here $\mathcal{I}_\delta(u)$ is an indicator random variable which is 1 or 0 depending on whether the system is up or down at time u, and $O_\delta(T - \omega)$ is the cumulative operational time under the auxiliary process $\mathcal{X}_1$ during $(0, T-\omega)$. Accounting for all of the above cases, we obtain the expected value of C_δ as

$$E[C_\delta] = \frac{1}{\theta_\delta}(1 - e^{-\theta_\delta(T-\omega)}) + \omega e^{-\theta_\delta(T-\omega)} + E[O_\delta(T-\omega)] + w\|d_O(T-\omega)\| \quad (64)$$

where

$$E[O_\delta(T-\omega)] =$$
$$\frac{1}{\Lambda_1}\sum_{n=0}^{\infty} e^{-\Lambda_1(T-\omega)}\frac{[\Lambda_1(T-\omega)]^{n+1}\theta_\delta(T-\omega)}{(n+2)!}\sum_{m=0}^{n}\left(1-\frac{\theta_\delta}{\Lambda_1}\right)^{n-m}\sum_{l=0}^{m}\|\pi_O(l)\| \quad (65)$$

and

$$\mathbf{d}_O(T-\omega) = \sum_{n=0}^{\infty} e^{-\Lambda_1(T-\omega)}\frac{[\Lambda_1(T-\omega)]^{n}\theta_\delta(T-\omega)}{(n+1)!}\sum_{m=0}^{n}\left(1-\frac{\theta_\delta}{\Lambda_1}\right)^{n-m}\pi_O(m) \quad (66)$$

Finally, we wish to calculate the probability that an unscheduled repairman had to be called. In the terminology of the second model, this is $P(U_\delta = 1)$ (note that U_δ is either 0 or 1 in this third model). We have

$$P(U_\delta = 1) = \|\mathbf{d}_U\| \quad (67)$$

where $\mathbf{d}_U$, the state probability vector of an unscheduled repair beginning at τ_1, is given by equation 57. Since U_δ is a random variable taking on only the values 0 and 1, we may also note that $E[U_\delta] = P(U_\delta = 1)$.

The various remarks at the end of section 4.2 also apply to this third model. For example, if the maximum transition rate in $\mathbf{Q}_1$ is at most θ_δ, we can randomize with $\Lambda_1 = \theta_\delta$ and obtain simplified expressions for the measures of interest. These expressions may be calculated using numerically stable recursions as before. We also observe that error bounds for the quantities of interest may be easily obtained in a manner similar to that for the second model. Finally, states with similar characteristics may be aggregated in the imbedded Markov chain $\mathcal{Y}$ to give computational savings when calculating various measures.

5.3 Steady State and Transient Measures

In this section we calculate steady state and transient availability measures of the system. As for the previous two models, equation 15 again holds, where $O(k)$ is the cumulative operational time during $(0, \tau_k)$. We next let $A(k)$ be the availability of the system during $(0, \tau_k)$. However, unlike the first two models, the value of τ_k in this third model is not constant. In this case, we have

$$A(k) = \frac{O(k)}{\tau_k} \tag{68}$$

Dividing numerator and denominator of equation 68 by k, we may apply theorem 7.14 of [HEYM82] to obtain (independent of the state at time 0)

$$\lim_{k\to\infty} A(k) = \frac{\sum_{\delta=1}^{M+K} E[O_\delta]\beta_\delta}{\sum_{\delta=1}^{M+K} E[C_\delta]\beta_\delta} \tag{69}$$

with probability 1.

To find the limiting fraction of times that an unscheduled repair was called for before the next scheduled repair, let $U(k)$ be the number of unscheduled repairs during $(0, \tau_k)$. Recalling that $E[U_\delta] = P(U_\delta = 1)$, we have

$$\lim_{k\to\infty} \frac{U(k)}{k} = \sum_{\delta=1}^{M+K} P(U_\delta = 1)\beta_\delta = \lim_{k\to\infty} \frac{E[U(k)]}{k} \tag{70}$$

where $P(U_\delta = 1)$ is given by equation 67.

We next calculate transient availability measures of the system. Since the length of an interval (τ_{k-1}, τ_k) is not constant in this model, we cannot obtain transient measures for a time period $(0, t)$ of fixed length. However, expressions similar to those of equations 18 and 53 may be obtained for the interval $(0, \tau_k)$. As an example, we may determine the expected number of the first k intervals in which a repairman had to be called before the next scheduled visit. Proceeding as before, we have

$$E[U(k)] = \sum_{\delta=1}^{M+K} P(U_\delta = 1) \sum_{j=0}^{k-1} \beta_\delta(j) \tag{71}$$

The expected length of an interval (τ_{k-1}, τ_k) may then be used to find an approximate result for $(0, t)$, the period of interest.

6 Examples

In this section we present examples which illustrate the applicability of the approach developed in this paper. We have chosen to use the second scheduled maintenance policy, which is typical of large computer installations. Examples illustrating the other policies will appear in future reports.

Consider a computer system with redundancy. The system has two processors and three memory units connected to a bus, as illustrated in Figure 4. One of

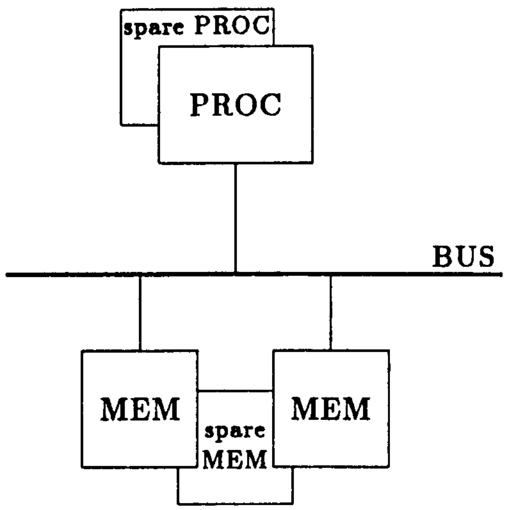

Figure 4: Example of a system with redundancy.

the two processors and one of the three memory units are spares. We assume that the failure of a component in the system is independent of the failure of any other component. Furthermore, the spares are in a cold standby mode, i.e., they are deactivated and cannot fail. We also assume that once an operating component fails, a spare (if there is any left) is immediately switched to full operation. We emphasize that the above assumptions are not modeling restrictions, but are used to simplify the model, since our goal is to illustrate the approach we use. An operating processor may fail in two different modes, *soft* fail and *hard* fail. If an operating processor fails in soft mode, the system can be restarted automatically (if there is a spare processor, the spare substitutes for the failed unit during restart). The restart process takes an average of 1 hour. On the other hand, a hard fail requires the intervention of a repairman. The rate at which processor soft failures occur is 1 per 14 days, and hard failures occur on an average of 1 per 3 months. A working memory unit fails at a rate of 1 per 2 months. The bus fails at a rate of 1 per 4 months. The system is considered up if at least one processor, two memory units and the bus are operational. In this example we assume that no failures can occur if the system is down.

We now consider the repair schedule. We assume that a repairman inspects the system once every T units of time. Upon arrival, the repairman may find the system either operational or inoperative. In the first case, the repairman runs certain diagnostic programs which will indicate if any unit is failed. He then replaces the failed units and brings the system to operation with all units repaired. This process takes only 1 hour on the average, during which the system is unavailable for use. If the system is found inoperative upon the scheduled visit of the repairman, then it is assumed that more time is needed to repair the system than in the previous case. We assume that it takes 4 hours for the system to be repaired on the average, and that all the units are brought to full operation.

If the system fails during the period between scheduled maintenance times, an unscheduled repairman is called. Upon arrival, the repairman performs the necessary repairs and brings the system to full operation. It is reasonable to assume that the average interval from the time that an unscheduled repairman is called for until the system is brought to full operation is considerably larger than the time required to fix the system by a scheduled repairman, since in the former, there are usually extra delays (such as travel delays) which occur due to the unexpected failure. In this example we assume that the average elapsed time between a failure and the full repair of the system is 24 hours.

The state of the system between scheduled visits of the repairman is represented by a quadruple $(d_{ps}, d_{ph}, d_m, d_b)$, where d_{ps} (d_{ph}) is the number of processors in a soft (hard) fail mode, d_m is the number of failed memory units and d_b indicates whether or not the bus is operational. The total number of states in the model is 21, and 6 of them represent an operational system. An unscheduled visit of a repairman is represented by transitions from the failed states to the state $(0,0,0,0)$ for which all units are operational. For the measures we consider, the failed states of the Markov process representing the system behavior between scheduled repairs can be aggregated to a single state, since down states are indistinguishable by the unscheduled repairman. However, we emphasize that this aggregation may not hold in general. Furthermore, we note that more complex cases can be easily represented. For instance, if the probability that an unscheduled repairman leaves the system in full operation is less than 1, then the Markov process would have non-zero transitions from the failed states to states other than the $(0,0,0,0)$ state.

Since the scheduled repairman always leaves the system in full operation and there are only two different means for the scheduled repair times, then the imbedded

Markov chain $\mathcal{Y}$ used to calculate the measures (matrix $\mathbf{D}$) has only two states, one represents an operational system upon arrival of the scheduled repairman and the other represents a down system.

Figure 5 shows the availability of the system as a function of the interval T between scheduled repairs. The horizontal line indicates the availability when no

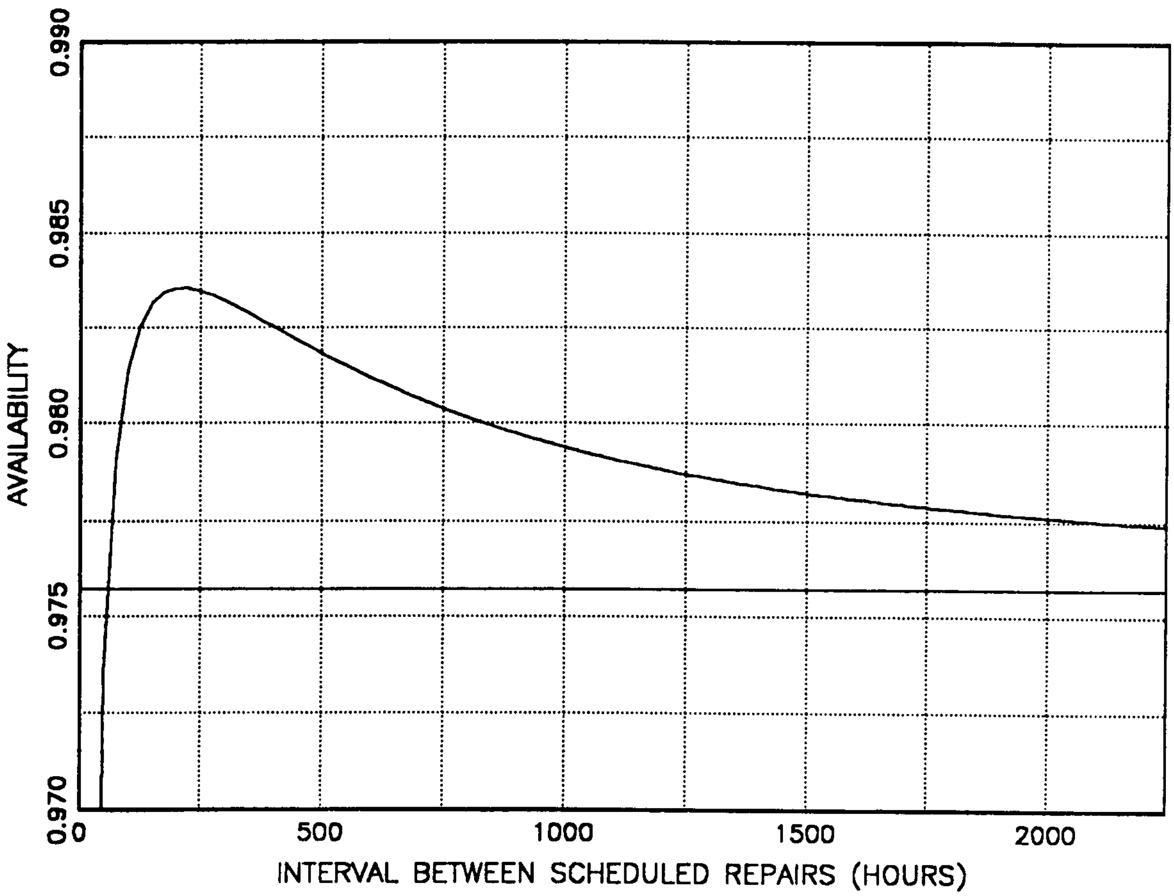

Figure 5: Availability versus the interval between scheduled repairs.

scheduled repair takes place, i.e., the repairman comes only when the system fails. The shape of the curve in Figure 5 can be explained as follows: if the scheduled repair interval is small, then the availability is also small, since the system is shut down for the scheduled maintenance for a greater proportion of time. As the scheduled repair interval increases, the availability increases to a maximum. Then the availability decreases since, if the time between repair periods increases, the benefit gained with a scheduled maintenance (say, the ability to repair failed units before the system fails) is lost. In the limit as the interval between scheduled repairs goes to infinity, only unscheduled repairs take place. From the figure, we see that for this example, the highest availability is achieved if the interval between scheduled repairs is 9 days (216 hours).

It is usually costly if the system fails unexpectedly and a repairman has to be called. Ideally, the interval between scheduled repairs should be small enough so that the probability that the repairman has to be called between scheduled maintenance periods is insignificant. However, frequent scheduled visits may be extremely costly. To study this relationship, we have assigned a cost of $\mathcal{S}$ units for each visit of a scheduled repairman, and a cost of $\mathcal{U}$ units for each unscheduled visit. It is reasonable to assume that $\mathcal{S} < \mathcal{U}$. Figure 6 shows the total expected cost per unit of time versus the interval T between scheduled repairs, when $\mathcal{S} = 10$ and $\mathcal{U} = 1000$. With this cost for the system, a high price is paid if the interval

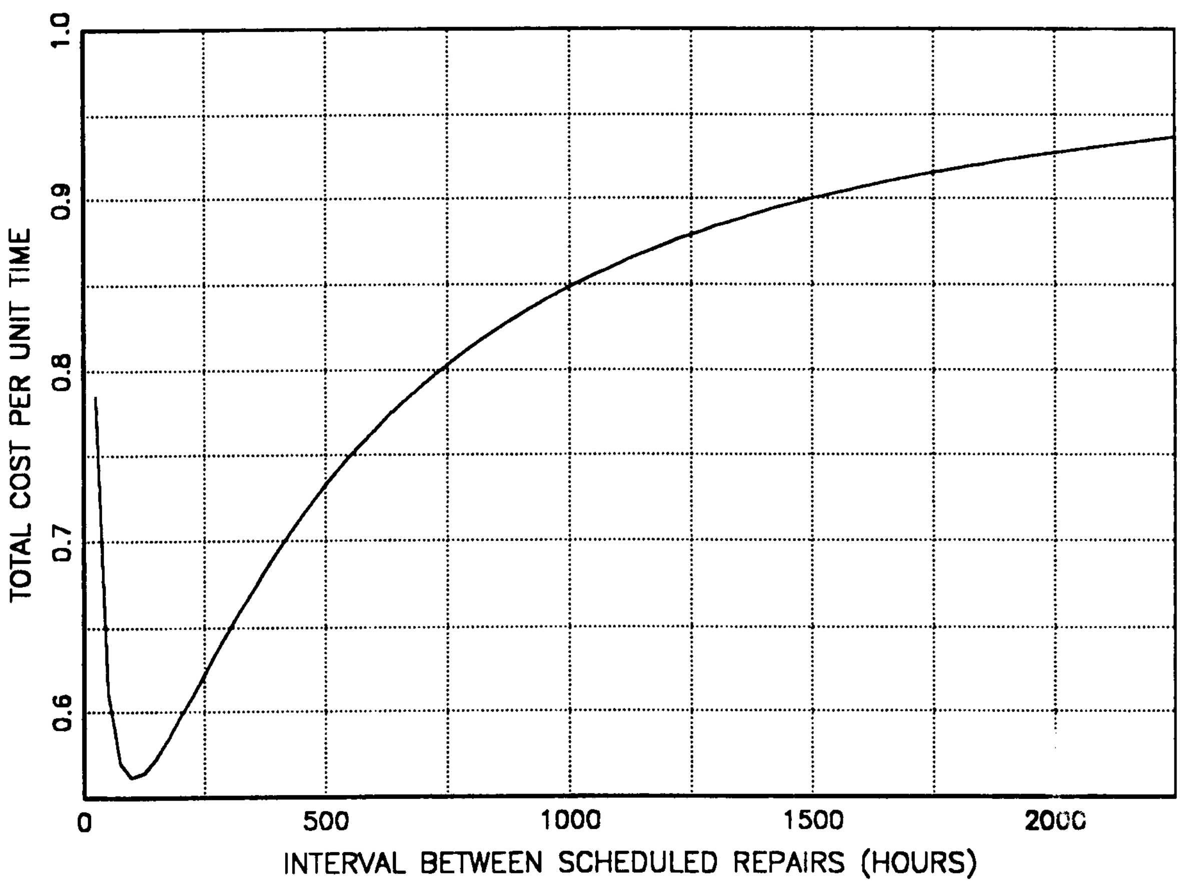

Figure 6: Total cost per unit of time: $\mathcal{S} = 10$, $\mathcal{U} = 1000$.

between scheduled repairs is too small. The cost decreases to a minimum, and then it increases due to the cost of unscheduled visits of the repairman. From the figure we see that the cost is minimized if the interval between repairs is 4 days.

Figure 7 shows the cost per unit of time for the same system when we increase the cost of the scheduled repair visits to $\mathcal{S} = 100$. We see that now the cost favors

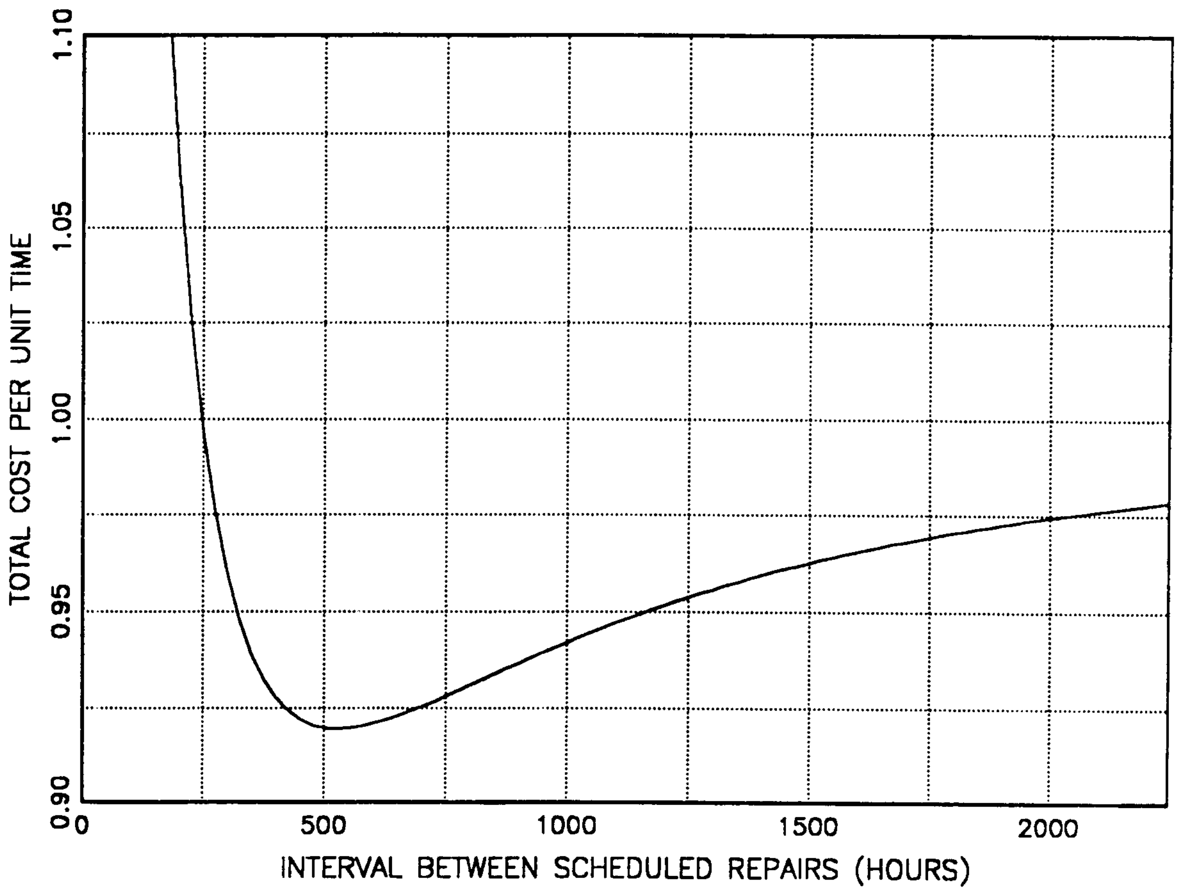

Figure 7: Total cost per unit of time: $\mathcal{S} = 100$, $\mathcal{U} = 1000$.

less scheduled visits in comparison to the previous value of $\mathcal{S}$. From the figure the minimum cost is achieved when the interval between scheduled repairs is 21 days (504 hours).

From the figures we also note that there is a tradeoff between cost and availability, namely, we would like to minimize cost and maximize availability. This can be studied by constructing tradeoff functions based on assigning weights to the availability and the cost.

7 Conclusions

In this paper we obtain transient as well as steady state measures of models of repairable computer systems with different scheduled maintenance policies. Among the contributions of the paper we mention the ability to model complex system failure and repair behavior and the calculation of transient measures, unlike previous work. Our solution method allows the modeling of important features such as imperfect repair and repair time dependent on the state of the system upon arrival of the repairman. Since the behavior of the system to be studied is modeled by a continuous-time Markov process, modeling details such as Coxian failure distributions, coverage, spares and failure dependencies can be included. Furthermore, the transient and steady state quantities we calculate can be compared with the corresponding ones obtained for the same model of a system, but without scheduled repair.

In addition to availability, we calculate quantities related to the number of unscheduled visits of the repairman. These quantities are important in evaluating the scheduled maintenance policy being considered. As we have shown in the examples, our solution method allows us to perform tradeoff studies between achieved availability and cost. Furthermore, the results can be easily extended to evaluate the distribution of unscheduled visits of the repairman. Finally, we have shown that the expressions obtained for the measures of interest can be easily evaluated using numerically stable recursions.

Appendix A

A Transient Solutions by Interval

The technique for calculating transient measures which was presented in the previous sections involves the following state by state approach. Given the initial state of the system, quantities of interest are calculated over a single interval, and the (unconditional) transient measure is then obtained from the conditional quantities. In this section we describe another method of solution for calculating transient measures over a time period $(0,t)$ for the above three models. This method involves calculating the state probabilities of the imbedded Markov chain at the points τ_k up to time t and evaluating the quantities of interest over each interval (τ_{k-1},τ_k). This process requires approximately the same amount of computation as evaluating a transient measure of a time-homogeneous Markov process using randomization. As is the case with other calculations of transient quantities using randomization, error bounds can be easily computed. For notational convenience, we assume that t is a multiple of T, say $t = \kappa T$.

A.1 First Model

We wish to calculate transient measures over $(0,t)$ for the first model introduced above. Let $\mathbf{c}(k)$, $\mathbf{d}(k)$ be the state probability vectors at the end of the kth repair phase (the point $\tau_{k-1}+R = (k-1)T+R$) and the end of the kth interval (the point $\tau_k = kT$) respectively. We assume that the state probability vector at time 0, $\mathbf{d}(0)$, is given. The expected cumulative operational time over the interval $(0,t)$ can be easily obtained as follows. The operational time during the kth interval (τ_{k-1},τ_k) is $\Theta(k) = \Theta_1(k)+\Theta_2(k)$, where $\Theta_1(k)$ and $\Theta_2(k)$ are the operational time during the repair and non-repair phases respectively. The cumulative operational time during $(0,\tau_k)$ is simply $O(k) = \Theta(1)+\cdots+\Theta(k)$. Assuming that quantities for the $(k-1)$st interval have been calculated, we calculate them for the kth interval as follows. We clearly have from equations 9 and 12

$$\mathbf{c}(k) = \sum_{n=0}^{\infty} e^{-\Lambda_1 R}\frac{(\Lambda_1 R)^n}{n!}\pi(n) \tag{72}$$

and

$$E[\Theta_1(k)] = \frac{1}{\Lambda_1}\sum_{n=0}^{\infty} e^{-\Lambda_1 R}\frac{(\Lambda_1 R)^{n+1}}{(n+1)!}\sum_{m=0}^{n}||\pi_O(m)|| \tag{73}$$

where $\pi(0) = \mathbf{d}(k-1)$. We also have from equations 10 and 13

$$\mathbf{d}(k) = \sum_{n=0}^{\infty} e^{-\Lambda_2(T-R)} \frac{[\Lambda_2(T-R)]^n}{n!} \nu(n) \tag{74}$$

and

$$E[\Theta_2(k)] = \frac{1}{\Lambda_2} \sum_{n=0}^{\infty} e^{-\Lambda_2(T-R)} \frac{[\Lambda_2(T-R)]^{n+1}}{(n+1)!} \sum_{m=0}^{n} \|\nu_O(m)\| \tag{75}$$

where $\nu(0) = \mathbf{c}(k)$. Using these expressions we can calculate the expected operational time during $(0, t)$ as

$$E[O(t)] = \sum_{k=1}^{\kappa} E[\Theta(k)] \tag{76}$$

and the expected availability during $(0, t)$ as

$$E[A(t)] = \frac{E[O(t)]}{t} \tag{77}$$

Note that the above procedure can be viewed as calculating transient availability measures for a time-nonhomogeneous Markov process with $2\kappa - 1$ generator changes. That is, the process has generator $\mathbf{Q}_1$ during the intervals $((k-1)T, (k-1)T + R)$ and generator $\mathbf{Q}_2$ during the intervals $((k-1)T + R, kT)$ for $k = 1, \ldots, \kappa$. Error bounds for the various measures may be obtained using a straightforward extension of the method that was discussed in section 3.2 for a process with a single generator change.

The choice of method for calculating $E[A(t)]$ depends on the tradeoff between the number of intervals to be computed (the value of κ), the number of states of S (the value of M), the sparseness of the matrices P_1 and P_2, and the value of Λt. The tradeoff can be easily seen by evaluating the number of matrix operations for each method, since the cost of these operations will dominate the total cost in both cases. For instance, if M is much larger than κ, then the interval by interval method introduced in this section should be used.

A.2 Second Model

In this section we discuss an alternate method of solution for calculating transient measures for the second model. Instead of using the original process $\mathcal{X}$, this method involves an interval by interval approach using a Markov process $\mathcal{X}_E$ which has an extended state space S_E with $2M$ states. Specifically, S_E has M (transient) states α_i ($i = 1, \ldots, M$) which explicitly represent the scheduled repair during an interval (τ_{k-1}, τ_k), while the time after the scheduled repair is modeled as before using the M states a_i ($i = 1, \ldots, M$). The generator of $\mathcal{X}_E$ is

$$\mathbf{Q}_E = \begin{bmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \tag{78}$$

where $\mathbf{Q}$ is the generator of $\mathcal{X}$, $\mathbf{R}$ is a diagonal matrix with entries $r_{ii} = -\theta_i$, and $\mathbf{S}$ is a matrix with entries $s_{ij} = \theta_i c_{ij}$. Let Λ_E be the randomization rate corresponding to $\mathbf{Q}_E$, and let $\mathbf{P}_E = \mathbf{Q}_E/\Lambda_E + \mathbf{I}$ be the transition matrix of the randomized Markov chain $\mathcal{Z}_E$. The state probability vector after n transitions of $\mathcal{Z}_E$, $\gamma(n) = \langle \gamma_1(n), \ldots, \gamma_{2M}(n) \rangle$, is given by $\gamma(n) = \gamma(n-1)\mathbf{P}_E$. We define $\nu(n) = \langle \gamma_1(n), \ldots, \gamma_M(n) \rangle$ to be the vector corresponding to the scheduled repair states α_i and $\pi(n) = \langle \gamma_{M+1}(n), \ldots, \gamma_{2M}(n) \rangle$ to correspond to the states a_i.

We identify an imbedded Markov chain $\mathcal{Y}_E$ at the points $\tau_k = kT$ which represent the beginning of the kth scheduled repair period. Observe that there are only M states of $\mathcal{Y}_E$, namely, α_i ($i = 1, \ldots, M$). We define $\mathbf{d}(k) = \langle d_1(k), \ldots, d_M(k) \rangle$ to be the state probability vector of $\mathcal{Y}_E$ at the point τ_k. We assume that $\mathbf{d}(0)$ is given, and we wish to calculate the vectors $\mathbf{d}(k)$ ($k = 1, \ldots, \kappa$) and the expected operational time during $(0, t)$. Although $\mathbf{d}(k)$ has length M, it is determined using the randomized Markov chain $\mathcal{Z}_E$ which has $2M$ states. Note that $\mathcal{Z}_E$ must be restarted in the states α_i ($i = 1, \ldots, M$) at the beginning of each interval (τ_{k-1}, τ_k). The probability of state α_i at the end of such an interval is the sum of the probabilities of the two states α_i and a_i at time T in the randomized chain $\mathcal{Z}_E$. In addition, since we have assumed that the system is not working during the scheduled repair period, the states α_i are classified as down states when determining the operational time. Thus the only operational states are the states of S_O.

Using the above observations, we may easily calculate quantities for the kth interval (τ_{k-1}, τ_k), assuming that they have been calculated for the $(k-1)$st interval. We have

$$\mathbf{d}(k) = \sum_{n=0}^{\infty} e^{-\Lambda_E T} \frac{(\Lambda_E T)^n}{n!} \{\nu(n) + \pi(n)\} \tag{79}$$

where $\gamma(0) = \langle \mathbf{d}(k-1), \mathbf{0} \rangle$. We also have

$$E[\Theta(k)] = \frac{1}{\Lambda_E} \sum_{n=0}^{\infty} e^{-\Lambda_E T} \frac{(\Lambda_E T)^{n+1}}{(n+1)!} \sum_{m=0}^{n} \|\pi_O(m)\| \tag{80}$$

Measures involving the number of unscheduled repairs can also be calculated using randomization on an interval by interval basis. Let $\Upsilon(k)$ be the (unconditional) number of unscheduled repairs during the interval (τ_{k-1}, τ_k). The number of unscheduled repairs during $(0, \tau_k)$ is $U(k) = \Upsilon(1) + \cdots + \Upsilon(k)$. Since unscheduled visits can occur only after the scheduled repair period is completed, marked transitions (transitions which cause an unscheduled repair) must involve the states a_i and not the states α_i. Therefore, the marking matrix $\mathbf{U}_E$ for the chain $\mathcal{Z}_E$ is

$$\mathbf{U}_E = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U} \end{bmatrix} \tag{81}$$

where $\mathbf{U}$ is defined as in section 4.2. Setting $\psi_E(l) = \gamma(l-1)\mathbf{U}_E$ for $l = 1, 2, \ldots$, then $\psi_E(l) = \langle \mathbf{0}, \psi(l) \rangle$ where $\psi(l) = \pi(l-1)\mathbf{U}$. Thus we have

$$E[\Upsilon(k)] = \sum_{n=1}^{\infty} e^{-\Lambda_B T} \frac{(\Lambda_E T)^n \theta_\delta T}{(n+1)!} \sum_{m=1}^{n} \left(1 - \frac{\theta_\delta}{\Lambda_E}\right)^{n-m} \sum_{l=1}^{m} \|\psi(l)\| \tag{82}$$

Note that $\psi(1) = \mathbf{0}$, because $\pi(0) = \mathbf{0}$ (the first transition in the randomized Markov chain $\mathcal{Z}_E$ always involves a scheduled repair state and thus cannot be marked).

The choice of the method for calculating measures of interest depends on the number of states in the imbedded Markov chain $\mathcal{Y}$ (call this number A) and the number of states in the Markov process $\mathcal{X}_E$. Note that, as mentioned in section 4.2, it is usually possible to aggregate states of the imbedded Markov chain $\mathcal{Y}$. Therefore, the number of states of $\mathcal{Y}$ is usually much less than M. If $\mathcal{Y}$ has less than M states, then $\mathcal{X}_E$ has less than $2M$ states ($\mathcal{X}_E$ has $M + A$ states). The choice of the method depends roughly on the value of A and the number of intervals to be computed. For instance, if $\kappa > A$ then the state by state method of section 4.3 should be used.

A.3 Third Model

The interval by interval method for calculating transient quantities for the third model also makes use of an extended Markov chain. In this case, (transient) states for both scheduled and unscheduled repair are added ($M + K$ extra states). Thus the randomized Markov chain for this third model has $2M + K$ states. The interval by interval method is then similar to that introduced in section A.2 for the second model. The choice of solution method depends roughly on the tradeoff between the number of states in the imbedded Markov chain $\mathcal{Y}$, and the number of intervals to be computed. For example, large values of κ compared with A favor the use of the method presented in section 5.3.

Acknowledgment

We wish to thank W. C. Carter for helpful discussions.

References

[ASCH84] Ascher and H. Feingold, "Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes," Marcel Dekker, 1984.

[CINL75] E. Cinlar, "Introduction to Stochastic Processes," Prentice-Hall, 1975.

[DESO86a] E. de Souza e Silva and H.R. Gail, "Calculating Cumulative Operational Time Distributions of Repairable Computer Systems," *IEEE Transactions on Computers*, C-35, no. 4, pp. 322-332, April 1986.

[DESO86b] E. de Souza e Silva and H.R. Gail, "Calculating Availability and Performability Measures of Repairable Computer Systems Using Randomization," IBM Research Report RC12386, December 1986.

[GOYA86] A. Goyal, W.C. Carter, E. de Souza e Silva, S.S. Lavenberg and K.S. Trivedi, "The System Availability Estimator," *Proceedings of FTCS-16*, pp. 84-89, June 1986.

[GRAS77] W.K. Grassmann, "Transient Solutions in Markovian Queueing Systems," *Computers and Operations Research* 4, pp. 47-53, 1977.

[GROS84] D. Gross and D.R. Miller, "The Randomization Technique as a Modeling Tool and Solution Procedure for Transient Markov Processes," *Operations Research* 32, no. 2, pp. 343-361, March-April 1984.

[HELV80] B.E. Helvik, "Periodic Maintenance, On the Effect of Imperfectness," *Proceedings of FTCS-10*, pp. 204-206, 1980.

[HEYM82] D.P. Heyman and M.J. Sobel, "Stochastic Models in Operations Research," Volume I, McGraw-Hill, 1982.

[MAKA81] S.V. Makam and A. Avizienis, "Modeling and Analysis of Periodically Renewed Closed Fault-Tolerant Systems," *Proceedings of FTCS-11*, pp. 134-141, June 1981.

[MELA84] B. Melamed and M. Yadin, "Randomization Procedures in the Computation of Cumulative-Time Distributions over Discrete State Markov Processes," *Operations Research* 32, no. 4, pp. 926-944, July-August 1984.

[MILL83] D.R. Miller, "Reliability Calculation using Randomization for Markovian Fault-Tolerant Computing Systems," *Proceedings of FTCS-13*, pp. 284-289, June 1983.

[ODA81] Y. Oda, Y. Tohma and K. Furaya, "Reliability and Performance Evaluation of Self-Reconfigurable Systems with Periodic Maintenance," *Proceedings of FTCS-11*, pp. 142-147, June 1981.

[ROSS83] S.M. Ross, "Stochastic Processes," John Wiley & Sons, 1983.

[YAK84] Y.W. Yak, T.S. Dillon and K.E. Forward, "The Effect of Imperfect Periodic Maintenance on Fault Tolerant Computer Systems," *Proceedings of FTCS-14*, pp. 66-70, June 1984.