OPTIMAL TEST INTERVAL FOR A MONOTONE SAFETY SYSTEM

TERJE AVEN,* *University of Stavanger*

Abstract

We consider a safety system represented by a monotone (coherent) structure function of n components. The state of the components and the system is only revealed through inspection, which is carried out at intervals of length T. If the inspection shows that the system is in a critical state or has failed, it is overhauled and all components are restored to a good-as-new condition. Costs are associated with tests, system downtime, and repairs. The problem is to find an optimal T minimizing the expected long-run cost per unit of time. The purpose of this paper is to present a formal set-up for this problem and to show how an optimal T can be determined. A special case where the components have three states is given particular attention. It corresponds to a 'delay time type system', where the presence of a fault in a component does not lead to an immediate failure—there will be a 'delay time' between the occurrence of the fault and the failure of the component.

1. Introduction

Multicomponent systems can be represented in various ways. One general formulation is to use the theory of monotone systems; see, e.g. Aven and Jensen (1999). A monotone system is a multicomponent system which has the natural property that the system cannot deteriorate by improving the performance of a component. Our main concern in this paper is safety systems having some degree of redundancy, i.e. the deterioration and failure of one component does not necessarily lead to the failure of the system. A simple example is a binary parallel system, which is functioning if at least one component is functioning. Another example is a system where the component first jumps to a partly defective state before it fails. There will be a time lapse between the occurrence of the fault and the failure of the component—a 'delay time'. This idea is the fundamental concept of the delay time models, which has been thoroughly discussed in the literature. See the reviews by Baker and Christer (1994), Christer and Redmond (1992), and Christer (1999).

The delay time models are used as a framework for determining monitoring strategies for detecting system defects and failures. The state of the system is revealed through inspection. The basic delay time model was introduced for analyzing inspection policies for systems regularly inspected each T time units (see Christer (1982)). If an inspection is carried out during the delay time period, the defect is identified and removed. In most of the models, the objective of the delay time analysis is to determine the optimal inspection interval, that is, the inspection interval that optimizes the (expected) long-run average costs or downtimes.

The framework of the present work incorporates the delay time model concept, as it allows for analysis of systems comprising components having an arbitrary finite number of states, in particular three, as required by the standard delay time model: perfect functioning, partly defective state (delay time), and failure state. The system is inspected every T units of time. If the system is found to be in a critical state or has failed, a complete overhaul (replacement) of the system and its components is performed. The system is in a critical state if further deterioration of a component (component i jumps from state j to state $j - 1$) induces system failure. As the system is a safety system in standby mode, the state of the system and its components is revealed only by testing. The aim of the testing and overhaul is to avoid the system failing and staying in the failure state for a long period. However, this goal has to be balanced against the costs of inspections and overhauls. Too frequent inspections would not be cost optimal.

There exist several papers treating replacement problems related to a monotone system. See Aven and Jensen (1999) and the references therein. In most of the models presented, the state of the system is known and the system is replaced once it has failed. Testing as in the present paper has concentrated on one-unit systems. There is an extensive literature on preventive replacement models. See, e.g. Jensen (1996), Valdez-Flores and Feldman (1989), Nakagawa (2005), Wang and Pham (2006), and Wang (2002).

The problem analyzed in this paper is described in Section 2. In Section 3 we establish formulae for the long-run expected average cost per unit of time, as a function of T. Examples, including the delay-type model, are presented to illustrate the analysis. In Section 4 we provide some final remarks.

2. Model and problem definition

We consider a safety system comprising n components, numbered consecutively from 1 to n. The state of component i at time t, $t \geq 0$, is denoted by $X_i(t)$, $i = 1, 2, \ldots, n$, where $X_i(t)$ can be in one out of $M_i + 1$ states, $0, 1, \ldots, M_i$. The paths $X_i(\cdot)$ are assumed to be right continuous. The states represent different levels of performance, from the worst, 0, to the best, M_i. At time $t = 0$, all components are in the best state, i.e. $X_i(0) = M_i$, $i = 1, 2, \ldots, n$. The random duration time in state M_i is denoted by U_{iM_i}. The component then jumps to state $M_i - 1$ for a random time $U_{i(M_i-1)}$, and so on until the component reaches the absorbing state 0. All sojourn times are positive random variables. The probability distribution of U_{ij} is denoted by F_{ij}. The distributions F_{ij} are assumed to be absolutely continuous, with finite means. The density and 'jump rate' of $F_{ij}(t)$ are denoted by $f_{ij}(t)$ and $r_{ij}(t)$, respectively, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, M_i$. The jump rate $r_{ij}(t)$ is defined as

$$\lim_{h \downarrow 0} \frac{1}{h} P(U_{ij} \leq t + h \mid U_{ij} > t).$$

Hence, $r_{ij}(t)h$ (where h is a small positive number) is approximately equal to the conditional probability that component i makes a jump to state $j - 1$ in the interval $(t, t + h]$, given that the component has stayed in state j during the interval $[0, t]$. The sojourn times $U_{iM_i}, U_{i(M_i-1)}, \ldots, U_{i1}$, $i = 1, 2, \ldots, n$, are assumed to be independent. The distribution of the vector of all the U_{ij}s, $\boldsymbol{U}$, is denoted by F_U.

We denote by $G(t, \boldsymbol{x})$ the distribution of the vector of component states $\boldsymbol{X}(t) = (X_1(t), X_2(t), \ldots, X_n(t))$, i.e.

$$G(t, \boldsymbol{x}) = P(X_1(t) = x_1, X_2(t) = x_2, \ldots, X_n(t) = x_n).$$

Here $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. The state of the system at time t is denoted by $\Phi(t)$ and is a function of the states of the components, i.e.

$$\Phi(t) = \phi(\boldsymbol{X}(t)),$$

where ϕ is the structure function of the system. We assume that Φ and ϕ are binary, equal to 1 if the system is functioning and 0 otherwise. The system is a monotone system, i.e. its structure function ϕ is nondecreasing in each argument, and

$$\phi(0, 0, \ldots, 0) = 0 \quad \text{and} \quad \phi(M_1, M_2, \ldots, M_n) = 1.$$

Since at time $t = 0$ all components are in the best state, $\Phi(0) = 1$. The components deteriorate and at time τ the system fails, i.e.

$$\tau = \inf\{t > 0 \colon \phi(\boldsymbol{X}(t)) = 0\}.$$

The deterioration of the components and the system failure is revealed through inspections. It is assumed that the system is inspected every T units of time. If the system is found to be in the failure state, a complete overhaul is carried out, meaning that all components are repaired to a good-as-new condition. Furthermore, a preventive policy is introduced: if the system is found to be in a *critical* state, a complete overhaul is also conducted. The system is said to be in a critical state if the system is functioning and there exists at least one i such that the system fails if component i jumps to the state $X_i(t) - 1$. Let τ_{c} be the time that the system first becomes critical. Then

$$\tau_{\mathrm{c}} = \inf\{t \geq 0 \colon \phi(\boldsymbol{X}(t)) = 1, \phi((X_i(t) - 1)_i, \boldsymbol{X}(t)) = 0 \text{ for at least one } i\},$$

where $\phi(\cdot_i, \boldsymbol{x}) = \phi(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_n)$. We assume that $\tau_{\mathrm{c}} > 0$, i.e. the system is not critical at time 0.

The distribution of τ_{c} is denoted by $F_{\tau_{\mathrm{c}}}$. The times τ and τ_{c} are functions of the duration times U_{ij}. Let g and g_{c} be defined as

$$\tau = g(\boldsymbol{U}) \quad \text{and} \quad \tau_{\mathrm{c}} = g_{\mathrm{c}}(\boldsymbol{U}).$$

The inspections and overhauls are assumed to take a negligible time.

To further characterize the critical states, we introduce the concept of a critical path vector for the system at level 1.

Definition 1. A state vector $\boldsymbol{x}$ is a critical path vector for the system at level 1 if and only if $\phi(\boldsymbol{x}) = 1$ and $\phi((x_i - 1)_i, \boldsymbol{x}) = 0$ for at least one i.

From this definition we introduce a maximal critical path vector.

Definition 2. A critical path vector $\boldsymbol{x}$ is a maximal critical path vector for the system at level 1 if it cannot be increased without losing its status as a critical path vector.

Note that these concepts are different to the commonly defined path vectors and minimal path vectors in a monotone system (see, e.g. Aven and Jensen (1999, p. 34)).

Based on the maximal critical minimal path vectors we introduce a new structure function, $\phi_{\mathrm{c}}(\boldsymbol{x})$, which is equal to 1 if and only if there exists no maximal critical path vector $\boldsymbol{x}_k$ such that the state $\boldsymbol{x}$ is below or equal to $\boldsymbol{x}_k$, i.e.

$$\phi_{\mathrm{c}}(\boldsymbol{x}) = \prod_k (1 - \mathbf{1}(\boldsymbol{x} \leq \boldsymbol{x}_k)),$$

where k runs through all maximal critical path vectors for the system at level 1 and $\mathbf{1}(\cdot)$ is the indicator function, which is equal to 1 if the argument is true and 0 otherwise. We see that the system ϕ_c fails as soon as a system state becomes critical. As an example, consider a binary parallel system. Then it is seen that the maximal critical path vectors are $(1,0)$ and $(0,1)$, and $\phi_c(\mathbf{x}) = x_1 x_2$, as, if one component fails, the system state becomes critical.

A counting process N is introduced which jumps to 1 at the time of the system failure, i.e.

$$N(t) = \mathbf{1}(\tau \le t).$$

Let $V_{ij}(t)$ be the virtual age of component i in state j at time t. Then the intensity $\lambda(t)$ of N is given by

$$\lambda(t) = \sum_{i=1}^{n} \sum_{j=1}^{M_i} r_{ij}(V_{ij}(t)) \, \mathbf{1}(X_i(t) = j)\phi(\mathbf{X}(t))(1 - \phi((j-1)_i, \mathbf{X}(t))),$$

noting that the rate is $r_{ij}(V_{ij}(t))$ at time t for component i to cause system failure by jumping from state j to state $j - 1$. A formal proof can be given following the approach of Aven and Jensen (1999, p. 65). By introducing $\phi_{ij}(\mathbf{x}) = \mathbf{1}(x_i = j)\phi(\mathbf{x})(1 - \phi((j-1)_i, \mathbf{x}))$, the intensity $\lambda(t)$ can be expressed as

$$\lambda(t) = \sum_{i=1}^{n} \sum_{j=1}^{M_i} r_{ij}(V_{ij}(t))\phi_{ij}(\mathbf{X}(t)).$$

Analogously, we define a counting process N_c for the process ϕ_c. This counting process jumps to 1 at the time the system becomes critical, i.e.

$$N_c(t) = \mathbf{1}(\tau_c \le t).$$

The intensity $\lambda_c(t)$ of N_c is given by

$$\lambda_c(t) = \sum_{i=1}^{n} \sum_{j=1}^{M_i} r_{ij}(V_{ij}(t)) \, \mathbf{1}(X_i(t) = j)\phi_c(\mathbf{X}(t))(1 - \phi_c((j-1)_i, \mathbf{X}(t))).$$

Similarly to ϕ_{ij}, we define $\phi_{ijc}(\mathbf{x}) = \mathbf{1}(x_i = j)\phi_c(\mathbf{x})(1 - \phi_c((j-1)_i, \mathbf{x}))$, and, hence, the intensity $\lambda_c(t)$ can be expressed as

$$\lambda_c(t) = \sum_{i=1}^{n} \sum_{j=1}^{M_i} r_{ij}(V_{ij}(t))\phi_{ijc}(\mathbf{X}(t)).$$

In the analysis below we will frequently make use of the following result. If N is a counting process with intensity $\lambda(t)$ then

$$\mathrm{E} \int_0^\infty D(t) \, \mathrm{d}N(t) = \mathrm{E} \int_0^\infty D(t)\lambda(t) \, \mathrm{d}t \tag{1}$$

for all processes $D(t)$ whose values at time t are known just before time t. This result follows from general counting process theory; see, e.g. Aven and Jensen (1999, p. 55). In the general formulation, the requirement is that the process $D(t)$ is a predictable process adapted to a filtration $(\mathcal{F}_t)$.

The following cost structure is assumed. The cost of a complete overhaul is c_p, whereas the cost of each inspection is c_I. If the system is not functioning, a cost c is incurred per unit of time. All costs are positive numbers.

The problem is to find an optimal T minimizing the long-run expected cost per unit of time.

3. Optimization

For a fixed test interval length T, $0 < T < \infty$, the system is overhauled at time τ^T, where τ^T is the time of the first inspection following a critical state, i.e.

$$\tau^T = T\left(\left[\frac{\tau_c}{T}\right]_I + 1\right),$$

where $[x]_I$ equals the integer part of x. This inspection represents a renewal for the cost and time processes, and using Ross (1970, Theorem 3.16), it follows that the long-run (expected) cost per unit time, B^T, can be written as

$$B^T = \frac{E\,C^T}{E\,\tau^T}, \tag{2}$$

where $E\,\tau^T$ expresses the expected length of the first renewal cycle (the time until renewal) and $E\,C^T$ expresses the expected cost associated with this cycle. It is seen that $E\,\tau^T < \infty$ and $E\,C^T < \infty$, observing that $E\,\tau^T \leq \sum_{ij} E\,U_{ij} + T$, and $E\,C^T \leq Tc + c_p + C_I(E\,\tau^T/T + 1)$. Theorem 1, below, establishes an explicit formula for $E\,\tau^T$ and $E\,C^T$, and, hence, for B^T.

Theorem 1. *Under the model assumptions of Section 2, with $\tau = g(U)$ and $\tau_c = g_c(U)$, we have*

$$E\,\tau^T = T \sum_{k=0}^{\infty} (k+1) \int_{u\,:\,kT < g_c(u) \leq (k+1)T} dF_U(u) \tag{3}$$

$$E\,C^T = \sum_{k=0}^{\infty} \int_{u\,:\,kT < g_c(u) \leq (k+1)T} [c_I(k+1) + c_p + c\,\mathbf{1}(g(u) \leq (k+1)T)$$
$$\times \{(k+1)T - g(u)\}]\,dF_U(u). \tag{4}$$

Proof. To establish (3), we write

$$\tau^T = \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \leq (k+1)T)(k+1)T.$$

Then taking expectations we obtain

$$E\,\tau^T = E\sum_{k=0}^{\infty} \mathbf{1}(kT < g_c(U) \leq (k+1)T)(k+1)T$$

$$= T\sum_{k=0}^{\infty} (k+1) \int_{u\,:\,kT < g_c(u) \leq (k+1)T} dF_U(u),$$

which proves (3). To establish (4), we use a similar approach, writing the cost C^T as a function of τ_c and τ:

$$C^T = \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \leq (k+1)T)[c_I(k+1) + c_p + c\,\mathbf{1}(\tau \leq (k+1)T)\{(k+1)T - \tau\}], \tag{5}$$

noting that the system is down a period $(k + 1)T - \tau$ if the system enters a critical state in the interval $(kT, (k + 1)T]$ and the system fails before the inspection at time $(k + 1)T$. Then taking expectations we obtain (4).

In the following theorem we establish more explicit formulae for $\mathrm{E}\,\tau^T$ and $\mathrm{E}\,C^T$ by using counting process theory (see (1)). Then we do not need the distribution of $F_U(\boldsymbol{u})$ but the distribution of $X(t)$, $G(t, \boldsymbol{x})$. We consider two special cases.

Case 1. The system is a binary system with binary components, i.e. $M_i = 1$ for $i = 1, 2, \ldots, n$.

Case 2. The rates r_{ij} are independent of t, i.e. the sojourn times are all exponentially distributed.

Theorem 2. *Let*

$$H_{ij}(t, \boldsymbol{x}) = \int_0^t r_{ij}(s)G(s, \boldsymbol{x}) \, \mathrm{d}s.$$

For cases 1 and 2, we then have

$$\mathrm{E}\,\tau^T = \sum_{k=0}^{\infty} T(k + 1) \sum_{i=1}^{n} \sum_{j=1}^{M_i} \sum_{\boldsymbol{x}} \phi_{ij\mathrm{c}}(\boldsymbol{x})[H_{ij}((k + 1)T, \boldsymbol{x}) - H_{ij}(kT, \boldsymbol{x})], \qquad (6)$$

where $\phi_{ij\mathrm{c}}(\boldsymbol{x}) = \mathbf{1}(x_i = j)\phi_{\mathrm{c}}(\boldsymbol{x})(1 - \phi_{\mathrm{c}}((j - 1)_i, \boldsymbol{x}))$. Furthermore, if $G_s(t, \boldsymbol{x} \mid \boldsymbol{x}')$ denotes the conditional distribution of $X(t)$ given $X(s) = \boldsymbol{x}'$ $(t > s)$, we have

$$\mathrm{E}\,C^T = \sum_{k=0}^{\infty} [c_I(k + 1) + c_p] \sum_{i=1}^{n} \sum_{j=1}^{M_i} \sum_{\boldsymbol{x}} \phi_{ij\mathrm{c}}(\boldsymbol{x})[H_{ij}((k + 1)T, \boldsymbol{x}) - H_{ij}(kT, \boldsymbol{x})]$$

$$+ \sum_{k=0}^{\infty} \sum_{\boldsymbol{x}'} \phi_{\mathrm{c}}(\boldsymbol{x}')G(kT, \boldsymbol{x}')$$

$$\times \sum_{i=1}^{n} \sum_{j=1}^{M_i} \sum_{\boldsymbol{x}} \phi_{ij}(\boldsymbol{x}) \int_{kT}^{(k+1)T} c((k + 1)T - t)r_{ij}(t)G_{kT}(t, \boldsymbol{x} \mid \boldsymbol{x}') \, \mathrm{d}t. \qquad (7)$$

Proof. To establish (6), we write

$$\tau^T = \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_{\mathrm{c}} \le (k + 1)T)(k + 1)T = \sum_{k=0}^{\infty} (k + 1)T \int_{kT}^{(k+1)T} \mathrm{d}N_{\mathrm{c}}(t).$$

Then taking expectations, using (1), and noting that we can write $r_{ij}(V_{ij}(t)) = r_{ij}(t)$, we obtain

$$\mathrm{E}\,\tau^T = \mathrm{E} \sum_{k=0}^{\infty} (k + 1)T \int_{kT}^{(k+1)T} \mathrm{d}N_{\mathrm{c}}(t)$$

$$= T \sum_{k=0}^{\infty} (k + 1) \int_{kT}^{(k+1)T} \mathrm{E}\,\lambda_{\mathrm{c}}(t) \, \mathrm{d}t$$

$$= T \sum_{k=0}^{\infty} (k + 1) \int_{kT}^{(k+1)T} \sum_{i=1}^{n} \sum_{j=1}^{M_i} r_{ij}(t) \,\mathrm{E}\,\phi_{ij\mathrm{c}}(X(t)) \, \mathrm{d}t$$

$$= T \sum_{k=0}^{\infty} (k+1) \sum_{i=1}^{n} \sum_{j=1}^{M_i} \sum_{\boldsymbol{x}} \phi_{ijc}(\boldsymbol{x}) \int_{kT}^{(k+1)T} r_{ij}(t) G(t, \boldsymbol{x}) \, dt$$

$$= T \sum_{k=0}^{\infty} (k+1) \sum_{i=1}^{n} \sum_{j=1}^{M_i} \sum_{\boldsymbol{x}} \phi_{ijc}(\boldsymbol{x})[H_{ij}((k+1)T, \boldsymbol{x}) - H_{ij}(kT, \boldsymbol{x})],$$

which proves (6). To establish (7), we rewrite (5) to obtain

$$C^T = \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \leq (k+1)T)[c_I(k+1) + c_p] \tag{8}$$

$$+ \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \leq (k+1)T) c \, \mathbf{1}(\tau \leq (k+1)T)\{(k+1)T - \tau\}. \tag{9}$$

Similarly to the above analysis for $\mathrm{E}\,\tau^T$, it is seen that term (8) equals

$$\sum_{k=0}^{\infty} [c_I(k+1) + c_p] \sum_{i=1}^{n} \sum_{j=1}^{M_i} \phi_{ijc}(\boldsymbol{x})[H_{ij}((k+1)T, \boldsymbol{x}) - H_{ij}(kT, \boldsymbol{x})].$$

Hence, it remains to establish the desired expression for the downtime costs. Term (9) can be expressed as

$$\sum_{k=0}^{\infty} \phi_c(\boldsymbol{X}(kT)) \int_{kT}^{(k+1)T} c((k+1)T - t) \, dN(t),$$

as $\phi_c(\boldsymbol{X}(t))$ is 1 as long as $t < \tau_c$. Then again taking expectations and using (1), we find that this expected cost term equals

$$\mathrm{E} \sum_{k=0}^{\infty} \phi_c(\boldsymbol{X}(kT)) \int_{kT}^{(k+1)T} c((k+1)T - t) \, dN(t)$$

$$= \mathrm{E} \sum_{k=0}^{\infty} \phi_c(\boldsymbol{X}(kT)) \int_{kT}^{(k+1)T} c((k+1)T - t)\lambda(t) \, dt$$

$$= \mathrm{E} \sum_{k=0}^{\infty} \phi_c(\boldsymbol{X}(kT)) \int_{kT}^{(k+1)T} c((k+1)T - t) \sum_{i=1}^{n} \sum_{j=1}^{M_i} \phi_{ij}(\boldsymbol{X}(t)) r_{ij}(t) \, dt$$

$$= \sum_{k=0}^{\infty} \sum_{\boldsymbol{x}'} \phi_c(\boldsymbol{x}') G(kT, \boldsymbol{x}') \sum_{i=1}^{n} \sum_{j=1}^{M_i} \sum_{\boldsymbol{x}} \phi_{ij}(\boldsymbol{x})$$

$$\times \int_{kT}^{(k+1)T} c((k+1)T - t) r_{ij}(t) G_{kT}(t, \boldsymbol{x} \mid \boldsymbol{x}') \, dt.$$

Equation (7) follows, and the theorem is proved.

We seek an optimal T_{opt} minimizing B^T given by (2) and the expressions for $\mathrm{E}\,C^T$ and $\mathrm{E}\,\tau^T$ in Theorems 1 and 2. Such a minimum always exists if we include the 'perform no testing and overhaul' policy $T = \infty$ as B^T is a continuous function and $\lim_{T \to 0} B^T = \infty$. We have

$B^\infty = \lim_{T \to \infty} B^T = c$. The expected average long-run cost per unit of time when there is no testing and overhaul equals c. If we perform very frequent testing, the long-run expected average cost will be very high due to a large number of inspections.

To find T_{opt}, it is convenient to search for Ts minimizing the functions

$$B^T(\delta) = \mathrm{E}\, C^T - \delta\, \mathrm{E}\, \tau^T.$$

If T_δ minimizes $B^T(\delta)$ and $B^{T_\delta}(\delta) = 0$, then T_δ minimizes B^T, i.e. T_δ is optimal, and $\delta = B^{T_\delta} = \inf_{0 < T \le \infty} B^T$. This result is well known from the literature; see Aven and Bergman (1986).

3.1. Special case: a parallel system of two components

Assume that $\phi(x) = 1 - (1 - x_1)(1 - x_2)$, i.e. the system is a parallel system comprising two components. The time that the system first becomes critical, τ_c, can then be expressed as

$$\tau_c = \min\{U_{11}, U_{21}\},$$

noting that if a component fails, the system is functioning if and only if the other component is functioning. Furthermore, the time to system failure, τ, equals the maximum component lifetime, i.e.

$$\tau = \max\{U_{11}, U_{21}\}.$$

It follows that

$$
\begin{aligned}
\mathrm{E}\,\tau^T &= T \sum_{k=0}^{\infty} (k+1)[F_{\tau_c}((k+1)T) - F_{\tau_c}(kT)] \\
&= \sum_{k=0}^{\infty} (k+1)[\bar{F}_{11}(kT)\bar{F}_{21}(kT) - \bar{F}_{11}((k+1)T)\bar{F}_{21}((k+1)T)],
\end{aligned}
$$

where $\bar{F} = 1 - F$. By similar arguments, first considering the costs c_I and c_p, and then the cost c, conditioning on $U_{11} = u_1$ and $U_{21} = u_2$, we obtain

$$
\begin{aligned}
C^T &= \mathrm{E} \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \le (k+1)T)[c_I(k+1) + c_p + c\,\mathbf{1}(\tau \le (k+1)T)\{(k+1)T - \tau\}] \\
&= \mathrm{E} \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \le (k+1)T)[c_I(k+1) + c_p] \\
&\quad + \mathrm{E} \sum_{k=0}^{\infty} \mathbf{1}(kT < \tau_c \le (k+1)T)[c\,\mathbf{1}(\tau \le (k+1)T)\{(k+1)T - \tau\}] \\
&= \sum_{k=0}^{\infty} [\bar{F}_{11}(kT)\bar{F}_{21}(kT) - \bar{F}_{11}((k+1)T)\bar{F}_{21}((k+1)T)][c_I(k+1) + c_p] \\
&\quad + \sum_{k=0}^{\infty} \int_0^{\infty} \int_0^{\infty} \mathbf{1}(kT < \min\{u_1, u_2\} \le (k+1)T)[c\,\mathbf{1}(\max\{u_1, u_2\} \le (k+1)T) \\
&\qquad\qquad \times \{(k+1)T - \max\{u_1, u_2\}\}]\,\mathrm{d}F_{21}(u_1)\,\mathrm{d}F_{11}(u_2).
\end{aligned}
$$

The last term due to the system downtime can be simplified to

$$\sum_{k=0}^{\infty} \int_{kT}^{(k+1)T} c\{(k+1)T - u_1\}[F_{21}(u_1) - F_{21}(kT)]\,\mathrm{d}F_{11}(u_1)$$
$$+ \sum_{k=0}^{\infty} \int_{kT}^{(k+1)T} c\{(k+1)T - u_2\}[F_{11}(u_2) - F_{11}(kT)]\,\mathrm{d}F_{21}(u_2).$$

These results are presented in Proposition 1, below.

Proposition 1. *For a parallel system of two binary components, the expected renewal cycle and the expected associated costs are given by*

$$\mathrm{E}\,\tau^T = \sum_{k=0}^{\infty}(k+1)[\bar{F}_{11}(kT)\bar{F}_{21}(kT) - \bar{F}_{11}((k+1)T)\bar{F}_{21}((k+1)T)],$$

$$C^T = \sum_{k=0}^{\infty}[\bar{F}_{11}(kT)\bar{F}_{21}(kT) - \bar{F}_{11}((k+1)T)\bar{F}_{21}((k+1)T)][c_I(k+1) + c_p]$$
$$+ \sum_{k=0}^{\infty} \int_{kT}^{(k+1)T} c\{(k+1)T - u_1\}[F_{21}(u_1) - F_{21}(kT)]\,\mathrm{d}F_{11}(u_1)$$
$$+ \sum_{k=0}^{\infty} \int_{kT}^{(k+1)T} c\{(k+1)T - u_2\}[F_{11}(u_2) - F_{11}(kT)]\,\mathrm{d}F_{21}(u_2).$$

An optimal T can then be determined.

Similar expressions can be easily derived based on Theorem 2.

3.2. Special case: the delay time model with three components

We consider a system comprising $n = 3$ components, with $M_i = 2$, i.e. each component has three states. State 2 is a perfect functioning state, whereas state 1 is a 'partly defective' state, as a result of a 'fault'. There will be a time lapse between the occurrence of the fault and the failure of the component—a 'delay time'. To simplify the mathematical analysis, we assume that all sojourn times U_{ij} are exponentially distributed. The constant rates are denoted by r_{ij}. Components 1 and 2 are assumed to have the same rates. The rates for different arrival states are assumed to be different, i.e. $r_{i2} \neq r_{i1}$ for $i = 1, 2, 3$.

The state of the system is given by the structure function

$$\phi(\boldsymbol{x}) = \mathbf{1}(x_1 + x_2 \geq 1)\,\mathbf{1}(x_3 \geq 1).$$

Hence, the system is functioning if either component 1 or 2 is in state 1 or better, and component 3 is in state 1 or better. We may think of the system as a parallel system comprising components 1 and 2, in series with component 3, with each component having a delay time before failure occurs.

The maximal critical path vectors for level 1 are $(0,1,2)$, $(1,0,2)$, and $(2,2,1)$, and this defines $\phi_c(\boldsymbol{x})$ and $\phi_{ijc}(\boldsymbol{x})$. We see that $\phi_c(\boldsymbol{x}) = 1$ for $\boldsymbol{x} = (1, 1, 2)$ and $\boldsymbol{x} > (1, 1, 2)$, as well as for $\boldsymbol{x} = (0, 2, 2)$ and $\boldsymbol{x} = (2, 0, 2)$.

For two distribution functions F_1 and F_2, let $\bar{F}_1 * F_2(t) = \int_0^t \bar{F}_1(t-s)\,\mathrm{d}F_2(s)$. Then the distribution $G(t, \boldsymbol{x})$ can be expressed as

$$G(t, (2, 2, 2)) = \bar{F}_{12}(t)\bar{F}_{22}(t)\bar{F}_{32}(t) = \exp\left\{-t\sum_{i=1}^{3} r_{i2}\right\},$$

$$G(t, (1, 2, 2)) = [\bar{F}_{11} * F_{12}(t)]\bar{F}_{22}(t)\bar{F}_{32}(t)$$

$$= \frac{r_{12}}{r_{12} - r_{11}}(\exp\{-r_{11}t\} - \exp\{-r_{12}t\})\exp\left\{-t\sum_{i=2}^{3} r_{i2}\right\},$$

$$G(t, (1, 1, 2)) = [\bar{F}_{11} * F_{12}(t)][\bar{F}_{21} * F_{22}(t)]\bar{F}_{32}(t)$$

$$= \frac{r_{12}}{r_{12} - r_{11}}(\exp\{-r_{11}t\} - \exp\{-r_{12}t\})$$

$$\times \frac{r_{22}}{r_{22} - r_{21}}(\exp\{-r_{21}t\} - \exp\{-r_{22}t\})\exp\{-tr_{32}\},$$

$$G(t, (0, 2, 2)) = F_{12} * F_{11}(t)\bar{F}_{22}(t)\bar{F}_{32}(t)$$

$$= \left\{1 - \exp\{-r_{12}t\} - \frac{r_{12}}{r_{12} - r_{11}}[\exp\{-r_{11}t\} - \exp\{-r_{12}t\}]\right\}$$

$$\times \exp\left\{-t\sum_{i=2}^{3} r_{i2}\right\}.$$

From these expressions, compact formulae can be derived for $H_{ij}(t, \boldsymbol{x}) = r_{ij}\int_0^t G(s, \boldsymbol{x})\,\mathrm{d}s$.

Similar equations can be established for $G_s(t, \boldsymbol{x} \mid \boldsymbol{x}')$, the conditional distribution of $X(t)$ given $X(s) = \boldsymbol{x}'$. We need to compute the conditional distribution of $\mathrm{P}(X_i(t) = j_2 \mid X_i(s) = j_1)$ for $j_2 \leq j_1$, $j_1 = 1, 2$, $i = 1, 2$. We see that $\mathrm{P}(X_i(t) = 2 \mid X_i(s) = 2) = \bar{F}_{i2}(t-s)$, $\mathrm{P}(X_i(t) = 1 \mid X_i(s) = 2) = \bar{F}_{i1} * F_{i2}(t-s)$, and $\mathrm{P}(X_i(t) = 1 \mid X_i(s) = 1) = \bar{F}_{i1}(t-s)$. Furthermore, $\mathrm{P}(X_i(t) = 0 \mid X_i(s) = 1) = F_{i1}(t-s)$ and $\mathrm{P}(X_i(t) = 0 \mid X_i(s) = 2) = F_{i2} * F_{i1}(t-s)$. From these formulae we see for example that

$$G_s(t, (2, 2, 2) \mid (2, 2, 2)) = \bar{F}_{12}(t-s)\bar{F}_{22}(t-s)\bar{F}_{32}(t-s)$$

$$= \exp\left\{-(t-s)\sum_{i=1}^{3} r_{i2}\right\},$$

$$G_s(t, (1, 2, 2) \mid (2, 2, 2)) = [\bar{F}_{11} * F_{12}(t-s)]\bar{F}_{22}(t-s)\bar{F}_{32}(t-s)$$

$$= \frac{r_{12}}{r_{12} - r_{11}}(\exp\{-r_{11}(t-s)\} - \exp\{-r_{12}(t-s)\})$$

$$\times \exp\left\{-(t-s)\sum_{i=2}^{3} r_{i2}\right\}.$$

In this way, all terms in $\mathrm{E}\,\tau^T$ and $\mathrm{E}\,C^T$ can be derived and an optimal T determined.

3.2.1. *Numerical example.* We assume that the failure rates are as follows: $r_{12} = r_{22} = 0.5$, $r_{11} = r_{21} = 1.0$, and $r_{32} = \frac{1}{3}$, $r_{31} = \frac{1}{2}$. Hence, the expected times to failure for the three components are $2 + 1 = 3$, $2 + 1 = 3$, and $3 + 2 = 5$, respectively. The following costs are assumed: $c = 100$, $c_I = 1$, and $c_p = 5$, i.e, the cost of an overhaul is five times the inspection
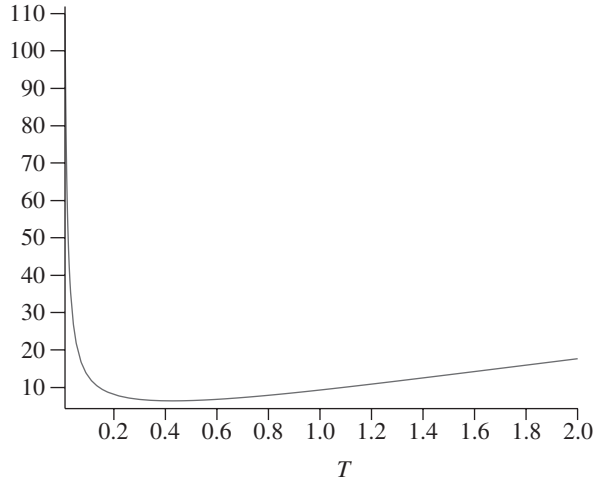
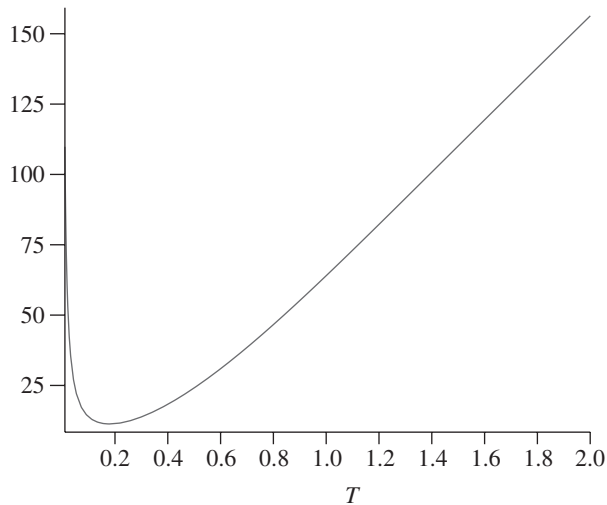FIGURE 1: The B^T function for the base case example with $c = 100$.

FIGURE 2: The B^T function for $c = 1000$.

cost and the unit downtime cost is 100 times the inspection cost. Then we can compute the B^T function and determine an optimal inspection time. Figure 1 shows the B^T function as a function of T, computed using MAPLE® 10. By inspection, an optimal value is obtained for $T = 0.43$.

A number of sensitivity analyses should be performed to see the effect of changes in the input data. Figure 2 shows an example where the unit downtime cost is increased by a factor 10, from 100 to 1000, to reflect the serious safety risk caused by downtime. The optimal inspection interval is then reduced to 0.18.

4. Final remarks

The analytical optimization of B^T needs to be carried out using numerical methods. For the numerical example considered in the previous section, the optimization criterion is of the standard form seen for many maintenance models (nonincreasing up to a minimum value and then nondecreasing). In general, this is however not the case for the model studied in this paper. Examples can be constructed where the optimization function has several local minimum values, in line with the examples for a one-component system in Aven and Castro (2009).

The model can be extended in many ways, for example, by allowing a more general cost structure. As an example, we may distinguish between the cost of an overhaul when the system is in a critical state and when it has failed. The calculations of $\mathrm{E}\,C^T$ in (5) then need to be modified, by considering a cost term $c_p + c_p' \mathbf{1}(\tau < (k+1)T)$, where c_p' is the additional overhaul cost if the system has failed compared to being in a critical state. This analysis is analogous to that carried out for $\mathrm{E}\,C^T$ in Section 3. The next step would be to allow the overhaul cost to depend on the state vector. The analysis would then become more complicated, but still within the framework and approach presented.

The analysis is based on independent components. If such independence cannot be justified, the theory and results in this paper do not apply. Attempts may be made to explicitly model the dependencies, but the arbitrariness in such modeling would typically be large. It is difficult to provide sufficient support for one particular such dependency model. In some cases it is possible to separate out the dependencies (common cause failures) and add a component in series with the rest of the system to reflect the common cause. Then we would be back to the framework considered in this paper.

Acknowledgements

References

AVEN, T. AND BERGMAN, B. (1986). Optimal replacement times—a general set-up. *J. Appl. Prob.* **23,** 432–442.

AVEN, T. AND CASTRO, I. T. (2009). A delay time model with safety constraint. *Reliab. Eng. System Safety* **94,** 261–267.

AVEN, T. AND JENSEN, U. (1999). *Stochastic Models in Reliability* (Appl. Math. **41**). Springer, New York.

BAKER, R. D. AND CHRISTER, A. H. (1994). Review of delay-time OR modelling of engineering aspects of maintenance. *Europ. J. Operat. Res.* **73,** 407–422.

CHRISTER, A. H. (1982). Modelling inspection policies for building maintenance. *J. Oper. Res. Soc.* **33,** 723–732.

CHRISTER, A. H. (1999). Developments in delay time analysis for modelling plant maintenance. *J. Oper. Res. Soc.* **50,** 1120–1137.

CHRISTER, A. H. AND REDMOND, D. F. (1992). Revising models of maintenance and inspection. *Internat. J. Production Econom.* **24,** 227–234.

JENSEN, U. (1996). Stochastic models of reliability and maintenance: an overview. In *Reliability and Maintenance of Complex Systems*, ed. S. Ozekici, Springer, Berlin, pp.3–36.

NAKAGAWA, T. (2005). *Maintenance Theory of Reliability*. Springer, New York.

ROSS, S. M. (1970). *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco, CA.

VALDEZ-FLORES, C. AND FELDMAN, R. M. (1989). A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Res. Logistics* **36,** 419–446.

WANG, H. (2002). A survey of maintenance policies of deteriorating systems. *Europ. J. Operat. Res.* **139,** 469–489.

WANG, H. AND PHAM, H. (2006). *Reliability and Optimal Maintenance*. Springer, New York.