A correlated M/G/1-type queue with randomized server repair and maintenance modes

David Perry\textsuperscript{a}, M.J.M. Posner\textsuperscript{b, *}

\textsuperscript{a}Department of Statistics, The University of Haifa, Haifa, 31905, Israel
\textsuperscript{b}Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, ONT, Canada M5S 3G8

Received 1 June 1998; received in revised form 1 August 1999

Abstract

We consider a single machine, subject to breakdown, that produces items to inventory, continuously and uniformly. While the machine is working (an ON period), there is a deterministic net flow into the buffer. There are also two different types of OFF periods; one is a repair operation initiated after a breakdown, and the other is a preventive maintenance operation. The length of an OFF period depends on the length of the preceding ON period. During OFF periods there is a uniform outflow from the inventory buffer. The buffer content process can then be described as a fluid model. The main tool employed in determining its steady-state law exploits an equivalence relationship between the buffer content process and the virtual waiting time process of a special single-server queue in which the interarrival times and the service requirements depend on each other. We present an appropriate cost function and provide an optimization scheme illustrated by some numerical results. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Production/inventory; Correlated fluid queues; Repair; Maintenance; On/off unreliable machine; Level crossings

1. Introduction

In this paper, we study a special type of manufacturing problem incorporating machine reliability and maintenance. Items are produced continuously and uniformly by a single machine that is subject to breakdown. During a machine ON time there is a deterministic net flow into the inventory buffer at rate \( \lambda > 0 \), where \( \lambda \) is the production rate minus the demand rate. If a failure now occurs before some time \( \tilde{a} \) has elapsed, then a machine repair operation will start; this is an OFF time of type 1. However, if a failure does not occur before time \( \tilde{a} \), then the controller stops production to initiate a preventive maintenance (PM) action; this is an OFF
time of type 2. During an OFF time of either type 1 or type 2, there is a deterministic demand generating a linear outflow from the inventory buffer at rate \( \beta \), whenever the buffer level is positive. However, negative inventory is not allowed, so that during OFF times there will be neither inflow nor outflow whenever the system is empty.

In the context of such a model framework, we attempt to maximize the net revenue accumulation rate with respect to the decision variable \( \tilde{a} \), namely, how long we should permit the machine to operate before initiating PM. The net revenue function will incorporate revenue from production, holding costs, penalty cost for unsatisfied demands, and repair and maintenance costs. Each component is determined in accordance with the steady-state distribution of the buffer contents.

Our model can be interpreted either as a certain queueing model with dependence between an interarrival time and the preceding service request, or as a fluid/inventory model with a two-state random environment. The two models are directly related in the sense that the steady-state law of the buffer contents in the fluid interpretation can be expressed in terms of the workload in the queueing interpretation. As a stochastic model, the fluid interpretation may be more natural than the queueing one; however, the queueing interpretation enables us to locate the problem in the general setting of queueing models and to use well-established tools and results from queueing theory for the solution. The analysis is based on the fact (cf. [13]) that the conditional steady-state buffer content distribution, given that it is positive, is independent of the inflow rate \( \alpha \). Thus, by setting \( x = \alpha \), we generate a sample path in which the ON periods are deleted and are represented by upward jumps, while the OFF periods are then glued together. The resulting process can be interpreted as the workload process of a certain single-server queue in which customers arrive with a service request at a single server. Service requests of successive customers are independent and identically distributed random variables having a general distribution (corresponding to the general failure time distribution of the ON time). Upon arrival, the service request is registered. If the service request is less than \( a = \alpha \tilde{a} \), then, the next interarrival time corresponds to an OFF period of type 1; otherwise, the service time becomes exactly equal to \( a \) (is cut off at \( a \)), and the next interarrival time corresponds to an OFF time of type 2.

At first glance it is not easy to see that the fluid/inventory model and the queueing model are equivalent (in the sense that they have the same conditional steady-state laws). Kella and Whitt [13] were the first to prove it for a generalized version of the GI/G/1 queue (in which the interarrival times and service times are not necessarily independent); however, they did not focus on the equivalence between the queueing system and the production/inventory model.

The fluid production/inventory model is perhaps better motivated from the operational modeling point of view. However, the queueing interpretation, which is of independent interest due to the correlation feature, is better motivated from the stochastic analysis point of view. This interdependence feature will lead to the development of a new and original adaptation of level-crossing theory [8], which will augment the applications already presented by Perry and Posner [16–18]. For related works in production and manufacturing using fluid models see [14,15,7,21,19]. A detailed survey on replacement policies and preventive maintenance models can be found in [1,20].

We are unaware of any studies that explore queueing systems with correlation between interarrival times and preceding service times, as characterized in the present paper. However, there has been some work in which there is correlation between an interarrival time and the following service time. For example, Conolly [4] and Conolly and Hadidi [5] considered an M/M/1 queue in which the service time and the preceding interarrival time are linearly related; Conolly and Choo [6] extended this relationship to encompass a bivariate exponential distribution, and expressed the waiting time density in a series of partial fraction terms. For the same model, Hadidi [9] showed that the waiting times are hyperexponentially distributed, and later [10] he examined the sensitivity of the waiting time distribution to the value of the correlation coefficient. Jacobs [11] obtained heavy traffic results for the waiting time in queues with sequences of ARMA correlated, negative-exponentially distributed interarrival and service times. Boxma and Combe [3] and Borst et al. [2], respectively considered two variants of a single-server queue in which the service time of each customer
depends on the length of the preceding interarrival interval; here the dependence structure corresponds to a situation where work arrives at a gate which is opened at exponentially distributed intervals.

The paper is organized as follows. In Section 2 we describe the dynamics of the production/inventory model and develop the structure of the associated queueing model. In Section 3, we introduce the relationship between the buffer content process and the queueing process. In Section 4, we analyze the stationary law of the queueing system. In Section 5 we present the cost function and provide an optimization scheme, leading to some numerical results.

2. Model dynamics

Let \( \{Z(t): t \geq 0\} \) be the buffer content level process. Then, the sample paths of \( Z(t) \) are continuous and piecewise linear with \( Z(t) \geq 0 \). In order to derive the dynamics governing the evolution of \( Z(t) \) we first define an artificial net input contents process \( Y(t) \). The process \( Y(t) \) has the same properties as \( Z(t) \) except that negative inventory is allowed; that is, during ON times the inflow accumulates at rate \( \alpha > 0 \) and during OFF times the outflow is released at rate \( \beta > 0 \).

We now introduce the following notation: \( T_{n+1} = T_n + U_{n+1} + D_{n+1}, \ n \geq 0, \ T_0 = 0 \), where \( U_n \) and \( D_n \) are ON times and OFF times, respectively. Here \( T_n \) is the onset time of the \( n \)th ON period. The process \( \{Y(t): t \geq 0\} \) is then defined by \( Y(0) = 0 \), and

\[
Y(t) = \begin{cases} 
Y(T_n) + \alpha(t - T_n), & T_n \leq t < T_n + U_{n+1}, \\
Y(T_n) + \alpha \cdot U_{n+1} - \beta(t - T_n - U_{n+1}), & T_n + U_{n+1} \leq t < T_{n+1}.
\end{cases}
\]

Finally, the buffer content process \( \{Z(t): t \geq 0\} \) is obtained by applying a reflection map to the process \( Y(t) \), so that \( Z(t) = Y(t) - \min[0, \inf_{0 \leq s \leq t} Y(s)] \) (refer to Fig. 1).

We now construct the process \( \{V(t): t \geq 0\} \) by deleting the ON times and then gluing together the OFF times of the \( Z(t) \) process. Formally, we define \( \hat{T}_{n+1} = \hat{T}_n + D_{n+1}, \ n = 0, 1, \ldots \), with \( \hat{T}_0 = 0 \), and \( \hat{Y}(t) = \hat{Y}(\hat{T}_n) + \alpha U_{n+1} - \beta(t - \hat{T}_n) \), \( \hat{T}_n \leq t < \hat{T}_{n+1} \), with \( \hat{Y}(0) = 0 \). Then, \( V(t) = \hat{Y}(t) - \min[0, \inf_{0 \leq s \leq t} \hat{Y}(s)] \).

Note that \( U_n \) and \( D_n \) are the same for both \( Z(t) \) and \( V(t) \). However, unlike in \( Z(t) \), the sample paths of \( V(t) \) are not continuous; \( V(t) \) is in fact a right-continuous jump process with left-hand limits. These are depicted in Fig. 1, where it is clear that \( V(t) \) is a queueing model analogue of \( Z(t) \) in which the jump sizes (services times) are bounded by \( a = \hat{a} \cdot \alpha \), and the subsequent exponential interarrival process (repair [with mean \( \lambda^{-1} \)] or preventive maintenance time [with mean \( \lambda^{-1} \)]) is governed by whether or not the preceding jump was less than \( a \). Clearly, then, optimization with respect to \( \hat{a} \) or \( a \) are equivalent.

3. The relationship between \( Z(t) \) and \( V(t) \)

Assume that \( Z(t) \) is a regenerative process (so that \( V(t) \) is also a regenerative process), and define \( B_Z = \inf \{t > 0: Z(t) = 0\} \) and \( B_Y = \inf \{t > 0: V(t) = 0\} \). Here, \( B_Z \) and \( B_Y \) are the busy periods (busy cycles minus idle periods) associated with the processes \( Z(t) \) and \( V(t) \), respectively. Also, let \( B_0 = B_Z - B_Y \). In words, \( B_Y \) is the total amount of OFF time in a busy cycle during which the inventory level (buffer content) is always positive, and \( B_0 \) is the total amount of ON time during that busy cycle. Let \( N \) be the number of ON times during a cycle.

By the definition of the model, we have

\[
B_Y = \int_{B_Z} \sum_{n=0}^{N-1} 1_{\{T_n + U_{n+1} \leq t < T_{n+1}\}} \ dt
\]
Fig. 1. A typical sample function of $Z(t)$ and corresponding realization of $V(t)$.

and

$$B_0 = \int_{B_Z} \sum_{n=0}^{N-1} 1_{\{T_n \leq t < T_{n+1}\}} \ dt.$$  

Let $F_Z(x) = \lim_{t \to \infty} P(Z(t) \leq x)$ and $F_V(x) = \lim_{t \to \infty} P(V(t) \leq x)$, with $\pi_Z = F_Z(0)$ and $\pi_V = F_V(0)$. Then, $\tilde{F}_Z(x) = [F_Z(x) - \pi_Z]/[1 - \pi_Z]$ and $\tilde{F}_V(x) = [F_V(x) - \pi_V]/[1 - \pi_V]$ are the steady-state conditional distributions of $Z(t)$ and $V(t)$ given that the respective systems are not empty.

We now want to show that $\tilde{F}_Z(x) = \tilde{F}_V(x)$ for all $x > 0$. By renewal theory,

$$\tilde{F}_Z(x) = \frac{E \int_{B_Z} 1_{\{Z(t) \leq x\}} \ dt}{EB_Z} = \frac{E \int_{B_Y} 1_{\{Z(t) \leq x\}} \ dt + E \int_{B_0} 1_{\{Z(t) \leq x\}} \ dt}{EB_Z}$$  \hspace{1cm} (3.1)

and

$$\tilde{F}_V(x) = \frac{E \int_{B_Y} 1_{\{Z(t) \leq x\}} \ dt}{EB_Y}.$$  \hspace{1cm} (3.2)
since both $Z(t)$ and $V(t)$ decrease at the same rate $\beta$. It follows by a simple geometric argument that, with probability 1,
\[ x \int_{B_0} 1_{\{Z(t) \leq x\}} \, dt = \beta \int_{B_0} 1_{\{Z(t) \leq x\}} \, dt, \quad x > 0. \quad (3.3) \]
Setting $x = \infty$ in (3.3) we thereby obtain $B_0 = (\beta/x)B_F$, and, since $B_0 = B_Z - B_F$, it follows that
\[ EB_Z = \left(1 + \frac{\beta}{x}\right) EB_F. \quad (3.4) \]
Substituting (3.3) and (3.4) into (3.1) we see that the right-hand side of (3.1) is the same as the right-hand side of (3.2). Thus, $F_Z(x) = F_F(x)$ (even if the interarrival times and the service times associated with the virtual waiting time process, $V(t)$, are not independent). This equality relation was also obtained by Kella and Whitt [13] and later generalized by Kaspi et al. [12], but using a different approach.

4. The stationary law of $Z(t)$ and $V(t)$

Let $\hat{G}$ be the distribution of the ON time of a machine in the fluid production–inventory model, and let $G$ be the distribution of the service requirement in the associated queueing model. Then, it is straightforward to see that $\hat{G}(x/x) = G(x)$; and in particular, $1 - \hat{G}(\hat{a}) = 1 - G(a)$ is the probability that the machine ON time [service requirement] is truncated at $\hat{a}[a]$. The only situation for which explicit formulas can be obtained is the case where both repair times and preventive maintenance times are exponential. In this case, the associated queueing system is a special version of a single-server queue with Markovian arrivals. We therefore assume that the repair time is exponentially distributed ($\lambda_0$) and preventive maintenance is exponentially distributed ($\lambda_1$). Then, the associated queueing system has the following properties. The service requirement has distribution $G$, and each arriving customer specifies his service requirement at his moment of arrival. If this service requirement is less than $a$, then the next interarrival time is exponentially distributed ($\lambda_0$). Otherwise, if his service requirement is at least $a$, then it is truncated to $a$, and the next interarrival time is exponentially distributed ($\lambda_1$). The $\hat{T}_n$’s are the arrival times of customers in that queueing system. If for some $n$, $\hat{T}_{n+1} - \hat{T}_n$ is exponentially distributed ($\lambda_0$), then, for all $\hat{T}_n \leq t < \hat{T}_{n+1}$, we say that the status of the system at time $t$ is $\{0\}$. Alternatively, if $\hat{T}_{n+1} - \hat{T}_n$ is exponentially distributed ($\lambda_1$), then, for all $\hat{T}_n \leq t < \hat{T}_{n+1}$, we say that the status at time $t$ is $\{1\}$. We can then define the status process at time $t$ by
\[ J(t) = \begin{cases} 1 & \text{if the status is } \{1\} \text{ at time } t, \\ 0 & \text{if the status is } \{0\} \text{ at time } t. \end{cases} \]
Note that the interarrival times in that queueing system are independent but they are not identically distributed. Also, the service requirements are i.i.d.; however, an interarrival time depends on the preceding service requirement.

We now apply level-crossing theory to compute the stationary density of $V(t)$. We first observe that the two-dimensional process $\{V(t), J(t); t \geq 0\}$ is a Markov process in which the state space of $V(t)$ is continuous on $[0, \infty)$ and $J(t)$ is either 0 or 1. Furthermore, we assume that conditions for stationarity prevail, so that the limiting distributions of $V(t)$, $J(t)$ and $Z(t)$ are given, respectively, by those of $V$, $J$, and $Z$.

In order to compute the law of the Markov process $\{V(t), J(t)\}$ in steady state, we separate it according to the status process $J(t)$. Then, we construct the Volterra integrals according to its generator. In our previous works a theoretical perspective was emphasized; in this study we focus more on the intuitive aspect. Accordingly, we partition the Borel $a$-field of the state space for each $0 \leq x < \infty$ into four types of disjoint subsets: $E_0 = \{(0,x), \{0\}\}$, $E^+ = \{[x, \infty), \{0\}\}$, $E_1 = \{(0,x), \{1\}\}$, and $E^+ = \{[x, \infty), \{1\}\}$. The
generator (or the infinitesimal transition rate) of the Markov process \( \{V(t), J(t)\} \) asserts simply that \( E_x \{0 \} \) can be reached (in a small interval of time) only from \( E_x \{0 \} \cup E_x \{1 \} \). By level-crossing theory, the transition rate of \( E_x \{0 \} \rightarrow E_x \{0 \} \) is the improper density \( f_0(x) \) which is the long-run average number of downcrossings of level \( x \) per unit time by \( V(t) \), while the status of the system is \( \{0 \} \). Also, to analyze the transition rate of \( E_x \{1 \} \rightarrow E_x \{0 \} \), we distinguish between two cases:

(i) \( x \leq a \). As with the Pollaczek–Khinchine formula, the transition rate is \( \lambda_1 \int_0^x G(x - \omega) \, dF_1(\omega) \), where, by PASTA [22], \( F_1(\omega) \) is the improper steady-state distribution of \( V \) when the status of the system is \( \{1\} \).

(ii) \( x > a \). The transition rate is \( \lambda_1 \int_0^{x-a} G(a) \, dF_1(\omega) + \lambda_1 \int_x^{x-a} G(x - \omega) \, dF_1(\omega) \).

Similarly, to leave the set \( E_x \{0 \} \) we distinguish between the same two cases:

(i) \( x \leq a \). The transition rate is \( \lambda_0 \int_0^x [1 - G(x - \omega)] \, dF_0(\omega) \), where \( F_0(\omega) \) is the improper steady-state distribution of \( V \) when the status of the system is \( \{0\} \).

(ii) \( x > a \). The transition rate is \( \lambda_0 \int_0^{x-a} [1 - G(a)] \, dF_0(\omega) + \lambda_0 \int_{x-a}^{x} [1 - G(x - \omega)] \, dF_0(\omega) \).

In an analogous manner, if \( x \leq a \), \( E_x \{1\} \) can be reached from \( E_x \{1\} \) with transition rate \( f_1(x) \); as well, the transition rate of \( E_x \{1\} \rightarrow E_x \{1\} \) is \( \lambda_1 \int_0^x dF_1(\omega) \). If \( x > a \), \( E_x \{1\} \) can be reached from \( E_x \{1\} \cup E_x \{0\} \) with transition rate \( f_1(x) + \lambda_0 \int_{x-a}^{x} [1 - G(a)] \, dF_0(\omega) \). Finally, the transition rate out of \( E_x \{1\} \) is \( \lambda_1 \int_0^{x-a} G(a) \, dF_1(\omega) + \lambda_1 \int_{x-a}^{x} dF_1(\omega) \).

To summarize the above discussion, we now present the resulting Volterra integrals obtained by this level-crossing analysis:

\[
f_0(x) + \lambda_1 \int_0^x G(x - \omega) \, dF_1(\omega) = \lambda_0 \int_0^x [1 - G(x - \omega)] \, dF_0(\omega), \quad x \leq a, \tag{4.1}
\]

\[
f_0(x) + \lambda_1 \int_0^{x-a} G(a) \, dF_1(\omega) + \lambda_1 \int_{x-a}^{x} G(x - \omega) \, dF_1(\omega) = \lambda_0 \int_{x-a}^{x} [1 - G(a)] \, dF_0(\omega) + \lambda_0 \int_0^{x-a} [1 - G(x - \omega)] \, dF_0(\omega), \quad x > a. \tag{4.2}
\]

\[
f_1(x) = \lambda_1 \int_0^x dF_1(\omega), \quad x \leq a, \tag{4.3}
\]

\[
f_1(x) + \lambda_0 \int_{x-a}^{x} [1 - G(a)] \, dF_0(\omega) = \lambda_1 \int_0^{x-a} G(a) \, dF_1(\omega) + \lambda_1 \int_{x-a}^{x} dF_1(\omega), \quad x > a. \tag{4.4}
\]

Associated boundary equations are \( f_i(0) = \lambda_i f_i, \quad i = 0, 1 \), where \( f_i = F_i(0) \), and the normalizing condition is \( F_0 + F_1 = 1 \), where \( F_1 = F_1(\infty) \). Observe that with this terminology a necessary and sufficient condition for stationarity is \( \int_0^x [1 - G(x)] \, dx \geq 0 \).

The set of Eqs. (4.1)–(4.4) together with the boundary equations and normalizing condition provide the unique solution for the stationary law of \( V(t) \). However, for any choice of \( G(\cdot) \), even exponential, the structure of the solution is exceedingly complex, and consequently, of limited utility. Accordingly, we construct a numerical computational approach in the appendix which is applicable for all (feasible) choices of \( G(\cdot) \); (refer to Fig. 2 for a typical realization of \( f \) for exponential \( G(\cdot) \) the resulting solution will then be employed in the next section in the formulation of a cost function to be optimized with respect to the decision variable \( a \).
5. Cost function and optimization

From the numerical solution carried out in the appendix, various functionals can be determined which bear directly on the construction of a cost function to be investigated in this section. The main functionals required are $\pi_Z \equiv \Pr(Z = 0)$, $\pi_V \equiv \Pr(V = 0)$, $EV$ and $EZ$. To obtain these, we must compute the various busy and idle periods for $Z$ and $V$ as $B_Z$, $I_Z$, and $B_V$, $I_V$, respectively.

We first note that

$$\pi_Z = \frac{EIZ}{EBZ + EIZ}$$

and

$$\pi_V = \frac{EIV}{EBV + EIV},$$

in which it is clear that

$$EIZ = EIV.$$  \hfill (5.1)

By renewal theory, the expected length of an idle period in $V$ is determined by

$$EI_V = \frac{f_0(0)}{f_0(0) + f_1(0)} \frac{1}{\lambda_0} + \frac{f_1(0)}{f_0(0) + f_1(0)} \frac{1}{\lambda_1} = \frac{f_0 + f_1}{\lambda_0 f_0 + \lambda_1 f_1},$$

using $f_i(0) = \lambda_i f_i$, $i = 0, 1$, where, invoking level-crossing theory, the ratio $f_i(0)/[f_0(0) + f_1(0)]$ is the probability that an idle period of type $i$ is initiated, and $1/\lambda_i$ is the corresponding length of this idle period. Furthermore, since a mean cycle time is given by $1/[f_0(0) + f_1(0)] = 1/[\lambda_0 f_0 + \lambda_1 f_1]$, it follows that

$$EB_V = \frac{1 - f_0 - f_1}{\lambda_0 f_0 + \lambda_1 f_1}.$$  \hfill (5.2)
includes the following functional components:

\[ V(K) \]

while a breakdown incurs an additional cost of the required model values are obtained from the numerical procedure of the appendix. A number of sample

The optimization with respect to \( a \) will be carried out for a net expected revenue accumulation rate which includes the following functional components:

(i) **Production revenue**: Production yields revenue of \( R_{PR} \) per unit produced. The production rate is \( x + \beta \), and the portion of time that the system is ON is given by the ratio of ON time \( (EB_0) \) to cycle time \( (EB_Z + EL_Z) \). Since \( EB_0 = EB_Z - EB_Y \), then using (3.4) and (5.3), we have the net production revenue rate as \( R_{PR} (x + \beta) (\beta/z)(1 - f_0 - f_1)/(1 + (\beta/z)(1 - f_0 - f_1)) \).

(ii) **Linear holding cost**: The holding cost rate is \( hEZ \), where \( h \) is holding cost per unit time per item held in inventory.

(iii) **Shortage cost**: A cost of \( K_{SH} \) is imposed for each demand arising during stockout. The corresponding cost rate is \( K_{SH} \beta \pi_z = K_{SH} [f_0 + f_1] / [1 + (\beta/z)(1 - f_0 - f_1)] \).

(iv) **Repair and PM costs**: A down machine incurs an ongoing cost of \( K_{DT} \) per unit time for repair and maintenance, while a breakdown incurs an additional cost of \( K_F \). The number of failures per unit time is \( \lambda_0 F_0 \) and the portion of time the system is OFF is given by the ratio of OFF time \( (EB_Y + EL_Y) \) to cycle time \( (EB_Z + EL_Z) \). As in (i) above, the net repair and PM cost rate is given by \( K_F \mu_0 F_0 + K_{DT} / (1 + (\beta/z)(1 - f_0 - f_1)) \).

The maximization of the net revenue is practically implemented via a grid search over values of \( a \), in which the required model values are obtained from the numerical procedure of the appendix. A number of sample

<table>
<thead>
<tr>
<th>Case 1: ( z_0^{-1} = 0.6, z_1^{-1} = 1 )</th>
<th>Case 2: ( z_0^{-1} = 1, z_1^{-1} = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu^{-1} )</td>
<td>( a^{*} )</td>
</tr>
<tr>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>0.50</td>
<td>0.39</td>
</tr>
<tr>
<td>0.75</td>
<td>0.51</td>
</tr>
<tr>
<td>1.00</td>
<td>0.59</td>
</tr>
<tr>
<td>1.50</td>
<td>0.67</td>
</tr>
<tr>
<td>2.00</td>
<td>0.70</td>
</tr>
<tr>
<td>2.50</td>
<td>0.72</td>
</tr>
<tr>
<td>3.00</td>
<td>0.72</td>
</tr>
<tr>
<td>3.50</td>
<td>0.73</td>
</tr>
</tbody>
</table>

using (5.4). Also,

\[
\pi_z = \frac{f_0 + f_1}{1 + \frac{\beta}{z}[1 - f_0 - f_1]} \tag{5.6}
\]

and \( \pi_f = f_0 + f_1 \) are easily confirmed.

Because of the demonstrated equivalence between \( \tilde{F}_Z \) and \( \tilde{F}_Y \), we have \( f_Z(x)/[1 - \pi_z] = f_Y(x)/[1 - \pi_f(x)] \), \( x > 0 \), whence \( f_Z(x)/f_Y(x) = [1 + \beta/z]/[1 + \beta/z(1 - f_0 - f_1)] \).
problems were considered, with the results, and optimal values \( a^* \) of \( a \), summarized in Table 1. In the table, the mean effective operating time of the machine is given by 
\[
\rho = \int_0^\infty \frac{1}{1 - G(t)} dt,
\]
for \( G(x) \) exponential (\( \mu \)), and 
\[
\rho = (\lambda_0 F_0 + \lambda_1 F_1) / \mu_0 \text{ is the analogue of the utilization factor for the associated queueing model.}
\]
Reflection upon the results presented illustrates sensitivity with respect to changes in mean repair time (Cases 1 and 4) or mean PM time (Cases 2 and 3) as the mean time to failure, \( \mu^{-1} \), varies. Of course, as \( \mu^{-1} \) increases beyond \( a^* \), the model stabilizes in the sense that well beyond \( a^* \) we expect PM almost exclusively.

Acknowledgements

We wish to thank O.J. Boxma for his careful reading of the manuscript and for providing useful comments and suggestions. Also, the financial support of the Natural Sciences and Engineering Research Council of Canada under grant OGP0004374 is gratefully appreciated.

Appendix Numerical solution

Define densities 
\[ f_i(x) = f_i(r) \]  
where \( f_i(rA) \), in which \( A \) is chosen so that \( a = NA \), with \( N \) arbitrary, but large. In addition, \( G_r \equiv G(rA) \) with \( G_r = 1 - G_r \) and \( f_i(i = 0, 1) \) are the zero-level probability masses (system idleness). Rewriting the main equations (4.1)–(4.4) gives

\[
\begin{align*}
&f_{0n} + \lambda_1A \sum_{r=0}^{n-1} G_{n-r} f_{1r} + \lambda_1G_n f_1 - \lambda_0 A \sum_{r=0}^{n} \tilde{G}_{n-r} f_{0r} - \lambda_0 \tilde{G}_n f_0 = 0, \quad n < N, \\
&f_{0n} + \lambda_1 \sum_{r=0}^{n-N} f_{1r} A \tilde{G}_N + \lambda_1 G_n f_1 + \lambda_1A \sum_{r=n-N}^{n-1} G_{n-r} f_{1r} - \lambda_0 A \tilde{G}_N \sum_{r=0}^{n-N} f_{0r} - \lambda_0 \tilde{G}_N f_0 = 0, \quad n \geq N, \\
&f_{1n} - \lambda_1 A \sum_{r=0}^{n} f_{1r} - \lambda_1 f_1 = 0, \quad n \leq N, \\
&f_{1n} + \lambda_0 A \tilde{G}_N \sum_{r=0}^{n-N} f_{0r} + \lambda_0 \tilde{G}_N f_0 - \lambda_1 G_N \sum_{r=0}^{n-N} f_{1r} - \lambda_1 G_N f_1 - \lambda_1 A \sum_{r=n-N}^{n} f_{1r} = 0, \quad n > N
\end{align*}
\]

and \( f_{i0} = \lambda_i f_i, \quad i = 0, 1 \). Finally, another relation between \( F_0 \) and \( F_1 \) can be found from \( F_0 + F_1 = 1 \). We observe, for example, that the rate of generating transitions out of preventive maintenance (PM) is balanced by the entry rate into PM. Thus, \( \lambda_1 G_1(a) = \lambda_0 F_0 \tilde{G}(a) \), leading to

\[
F_1 = \frac{\lambda_0 \tilde{G}(a)}{\lambda_1 G(a) + \lambda_0 \tilde{G}(a)}.
\]

and, using normality,

\[
F_0 = 1 - F_1 = \frac{\lambda_1 G(a)}{\lambda_1 G(a) + \lambda_0 \tilde{G}(a)}.
\]
From (A.1) and (A.2), for \( n = 1, 2, \ldots, N \), we can write the recursions

\[
\begin{align*}
    f_{0n} &= \frac{1}{1 - \lambda_0 A} \left\{ \lambda_0 A \sum_{r=0}^{n-1} \tilde{G}_{n-r} - \lambda_1 G_n f_1 - \lambda_1 A \sum_{r=0}^{n-1} G_{n-r} f_{1r} + \lambda_0 \tilde{G}_n f_0 \right\}, \\
    f_{1n} &= \frac{1}{1 - \lambda_1 A} \left\{ \lambda_1 A \sum_{r=0}^{n-1} f_{1r} + \lambda_1 f_1 \right\}.
\end{align*}
\]

(A.7) (A.8)

Define the column vectors \( g_n = [f_{0n} f_{1n}]' \) and \( \tilde{g} = [f_0 f_1]' \). Then, (A.7) and (A.8) can be restructured as the vector recursive equation,

\[
g_n = \sum_{r=0}^{n-1} A_{n-r} A g_r + A_n \tilde{g}, \quad n = 1, 2, \ldots, N
\]

(A.9)

with

\[
g_0 = \begin{bmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{bmatrix} \tilde{g}
\]

and

\[
A_k = \begin{bmatrix} \tilde{\lambda}_0 \tilde{G}_k & -\tilde{\lambda}_1 \tilde{G}_k \\ 1 - \lambda_0 A & 1 - \lambda_0 A \\ 0 & \lambda_1 \end{bmatrix}.
\]

The same procedure can also be applied to Eqs. (A.3) and (A.4), leading to the vector recursion,

\[
g_n = A^0 A \sum_{r=0}^{n-N} g_r + A \sum_{r=n-N}^{n-1} A_{n-r} g_r + A^0 \tilde{g}, \quad n = N + 1, N + 2, \ldots,
\]

(A.10)

in which

\[
A^0 = \begin{bmatrix} \tilde{\lambda}_0 \tilde{G}_N & -\tilde{\lambda}_1 \tilde{G}_N \\ 1 - \lambda_0 A & 1 - \lambda_0 A \\ -\lambda_0 \tilde{G}_N & \lambda_1 \tilde{G}_N \\ 1 - \lambda_1 A & 1 - \lambda_1 A \end{bmatrix}.
\]

Now, define matrices \( H_n \), \( n = 0, 1, 2, \ldots \), such that \( g_n = H_n \tilde{g} \). By making this substitution in (A.9) and (A.10), we can find these matrices from the resulting matrix recursions:

\[
H_n = \begin{cases} 
A \sum_{r=0}^{n-1} A_{n-r} H_r + A_n, & n = 1, 2, \ldots, N, \\
A^0 \sum_{r=0}^{n-N} H_r + A \sum_{r=n-N}^{n-1} A_{n-r} H_r + A^0, & n = N + 1, N + 2, \ldots,
\end{cases}
\]

in which

\[
H_0 = \begin{bmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{bmatrix}.
\]
By numerical integration, we have the total probabilities of the system being in status 0 and 1 given by
\[
\bar{g} + \Delta \sum_{r=0}^{\infty} g_r = \begin{bmatrix} F_0 \\ F_1 \end{bmatrix},
\]
where \(F_0\) and \(F_1\) are given in (A.5) and (A.6). Using \(g_r = H_r \bar{g}\), this converts into the unique solution for \(\bar{g}\) as
\[
\bar{g} = \left( I + \Delta \sum_{r=0}^{\infty} H_r \right)^{-1} \begin{bmatrix} F_0 \\ F_1 \end{bmatrix}.
\]
With \(\bar{g}\) known, all \(g_r\) are now determined, and hence the entire distribution is specified.

References