

# Optimal Policies for the Sizing and Timing of Software Maintenance Projects

February 10, 2004

## **Abstract**

We present a model to determine the optimal point for maintaining a software application. We also address the question: given that a maintenance project has been initiated, should maintenance effort continue till the project is completed? Most previous literature has implicitly assumed that it is optimal to complete a maintenance project once it has been initiated. We analyze two policies: a work-based policy and a time-based policy. In the work-based policy, a fixed amount of work needs to be completed, and the time taken to accomplish the work is random. In the time-based policy, a fixed amount of time is allocated to maintenance, but a random amount of work is completed. We examine similarities and differences between the above two policies and provide insights into the management of software maintenance projects. A key insight of this study is that under a variety of situations, partial maintenance is sub-optimal.

**Keywords:** Software maintenance, dynamic programming, threshold policy

# 1 Introduction

The importance of the software engineering industry has grown with the widespread deployment of information technology in major industry sectors. Software quality issues have begun to exert a profound influence on the daily operations of organizations. To maintain quality, software applications need to be maintained on a regular basis. Thus, software maintenance is a resource-intensive activity and constitutes a large portion of the total costs spent on the entire product life (Lientz and Swanson 1981, Arthur 1988, Nosek and Palvia 1999). Software maintenance refers to activities associated with modifying a software system or component after delivery to correct faults, improve performance or other attributes, or adapt to a changed environment (Schneidewind 1987). Two major reasons accounts for huge software maintenance cost: a volatile user environment and deteriorating software maintainability (Chan, Chung and Teck 1996).

Empirical studies show a close relationship between software maintenance and the activities undertaken during software development. A field study by Dekleva (1992) examines the influence of the selected system development methodology on the maintenance effort. The work of Banker et al. (1998) also indicates that software maintenance is greatly influenced by design and development practices, and software complexity is a key intermediate factor. Both economies and diseconomies of scale in software maintenance were observed by Banker et al. (1997) in their study on the relationship between project size and software maintenance effort.

Most software maintenance occurs as a response to user requests for changes or enhancements to the system. According to Swanson (1976), maintenance requests can be classified into three types: adaptive, perfective and corrective. In most cases, the first two categories account for more than 75% of the total maintenance effort on a system (Lientz and Swanson 1980). Software maintenance decisions are usually based on the structure of maintenance cost. Barua and Mukhopadhyay (1989)

investigate the case of continuous maintenance with the possibility of replacing the system once during its useful life. However, the cost of delaying maintenance is not considered. Chan et al. (1996) develop a normative model of software maintenance and replacement. Tan and Mookerjee (2003) also study the trade-off between software maintenance and replacement, together with the impact of software reuse. Kulkarni and Sethi (2003) model software maintenance using a variation of  $M/G/1$  queues and obtain an optimal resource allocation policy.

In this paper, we consider software maintenance projects under a stable IS environment. That is, the software is operating in the organization, users are familiar with the system and the maintenance team has enough experience with the system. Thus, the focus here is on adaptive and perfective maintenance, rather than corrective activities. The total costs of maintenance includes the costs incurred by the maintenance team and the costs incurred by the users who wait for maintenance to be completed.

We present a dynamic programming formulation of the software maintenance problem. Requests for maintenance are assumed to arrive in accordance with a Poisson process. The costs incurred by the maintenance team constitute a fixed and a linear cost that depends on the amount of maintenance work done. Delaying maintenance results in a per unit time user waiting cost. Two major questions are addressed: (1) when to start maintenance? (2) how long should the maintenance activity go on, or how much work should be completed in one round of maintenance? To answer the first question, we start with a simplified model under the assumption of instantaneous maintenance. We show that there is an optimal system state, or a threshold level, at which maintenance should be initiated. This result can be further extended to the case when the maintenance duration is not negligible. Furthermore, we propose two policies to answer the second question. In the work-based policy, the target system state after maintenance is fixed, and the maintenance duration is random; in the time-based policy, a planned maintenance duration is specified, and the system state

after maintenance is random. The work-based policy aims at controlling the quality of the system, whereas the time-based policy provides greater administrative control of the maintenance project. We provide comparison of the two policies and show that they are basically equivalent under the optimal threshold.

In the next section, we develop the basic model in which maintenance is completed instantaneously and prove the optimality of a threshold policy. We extend the model in Section 3 with the inclusion of non-zero maintenance time. Two classes of policies are discussed and optimal cost functions are derived. Additional numerical results of the models are presented in Section 4. We conclude the paper in Section 5.

## 2 Basic Model

Following previous empirical research, we model change requests from software users according to a Poisson process with a constant rate  $\lambda$  (Chan et al. 1996). A *waiting cost* of  $b$  per unit time occurs for each outstanding request. The state of system, i.e., the number of pending requests that have accumulated, is monitored, and depending on this number, a decision is taken whether or not to initiate maintenance. If maintenance is performed, the *cost of maintenance* consists of a fixed cost  $K$  and a correction cost  $c$  per request. We start with a simple model where we assume that maintenance can be instantaneously performed. When the rate of request correction is much faster than the request arrival rate, the time required to perform maintenance can be ignored. A more general model with nonzero maintenance time is discussed in the next section.

The goal of the model is to optimize the long-run total discounted cost (waiting cost plus maintenance cost). A constant discount rate of  $r$ ,  $0 < r < 1$ , is assumed. The problem is to choose the state of the system at which maintenance should be performed.

Let  $V(B)$ ,  $B \geq 0$ , be the smallest possible cost (in steady state) of state  $B$ . Let  $V^c(B)$  denote the cost of carrying out maintenance, and  $V^n(B)$  be the cost of not maintaining the system at state  $B$ . Let  $T$  be the random variable denoting the inter-request time. Then the dynamic programming (DP) equations for the problem can be written as

$$V^c(B) = K + cB + V(0), \quad (1)$$

$$V^n(B) = Ee^{-rT}V(B+1) + E \int_0^T e^{-rt}bBdt = qV(B+1) + q\frac{bB}{\lambda}, \quad (2)$$

$$V(B) = \min\{V^c(B), V^n(B)\}, \quad (3)$$

where  $q = \lambda/(\lambda + r)$  and  $p = 1 - q$ .

Next we propose a threshold policy: *maintain the system when it reaches a fixed state  $F$* . The state of the system follows a stochastic process with an infinite sequence of cycles of length  $F$ . Let  $V(B|F)$ ,  $0 \leq B \leq F$ , denote the associated cost function under the threshold policy. If  $B = F$ , then maintenance is initiated. We have

$$V(F|F) = K + cF + V(0|F), \quad (4)$$

If  $B < F$ , then from (2) we have

$$V(B|F) = qV(B+1|F) + q\frac{bB}{\lambda}. \quad (5)$$

Applying (2) recursively gives

$$\begin{aligned} V(B|F) &= q^2V(B+2|F) + q^2\frac{b(B+1)}{\lambda} + q\frac{bB}{\lambda} \\ &= \dots \\ &= q^{F-B}V(F|F) + \frac{b}{r} \frac{1}{p} [q + pB - (q + pF)q^{F-B}], \quad \text{for } 0 \leq B < F. \end{aligned} \quad (6)$$

Setting  $B = 0$  in (6) gives  $V(0|F)$ . Substituting this expression in (4) and solving for  $V(F|F)$ , we get

$$V(F|F) = \frac{1}{1 - q^F} \left[ K + \left( c - \frac{b}{r} \right) F \right] + \frac{b}{r} F + \frac{bq}{rp}, \quad \text{for } 0 \leq B < F. \quad (7)$$

Putting (7) in (6) yields

$$V(B|F) = \frac{q^{F-B}}{1-q^F} \left[ K + \left( c - \frac{b}{r} \right) F \right] + \frac{b}{r} \frac{1}{p} (q + pB) \quad (8)$$

Equation (8) is the cost function under the threshold policy  $F$ . It is easily seen that if  $V(B|F)$  is minimized at  $F = F^*$  for some  $B$ , then it is also minimized at  $F = F^*$  for all  $B \leq F^*$ . This means that the optimal threshold  $F^*$  (if it exists) is independent of the state  $B$ . In Theorem 1, we establish the optimality of the threshold policy and characterize the optimal policy under different situations. We refer the reader to the proof of this theorem in Appendix A

**THEOREM 1** *Let  $F^*$  minimize (8). The cost function  $V(B) = V(B|F^*)$ ,  $0 \leq B \leq F^*$ , satisfies the dynamic programming equations (1), (2) and (3). Furthermore,*

1. *When  $c \geq b/r$ , the optimal cost function is strictly increasing and convex in the threshold, and the optimal policy is never to fix;*
2. *When  $c < b/r$  and  $K < p(b/r - c)$ , the optimal cost function is strictly decreasing and concave in the threshold, and the optimal policy is always to fix;*
3. *Otherwise, there exists a finite threshold policy  $F^*$  with  $F^* > 1$  such that the optimal cost function increases when  $1 \leq F < F^*$  and decreases when  $F > F^*$ . Moreover, there exists a unique switching point  $F^0$  such that the cost is convex for  $1 \leq F \leq F^0$  and concave otherwise.*

The above results are quite intuitive. Note that  $b/r$  stands for the discounted value of the total cost for leaving one request in the system forever, and  $c$  is the cost to fix a request now. If it is more expensive to fix a request than to leave it in the system forever, then one would never bother to maintain the system. Hence, when  $c \geq b/r$  the optimal policy is never to fix (see Fig.1.a). On the other hand, when the cost of fixing one request is less than the cost of postponing this request forever then

we would choose to fix the request at some time. Suppose one request is submitted now, and we choose one of the following: to respond to this request now or to respond it when the next request arrives. Because of discounting, a dollar at the time the next request arrives is worth  $\lambda/(\lambda+r)$  dollars now. Thus, we would spend  $c - \frac{\lambda}{\lambda+r}c (= pc)$  dollars more if we were to respond to this request now than at the next arrival. On the other hand, responding to this request now would reduce user waiting cost to the extent of  $\frac{1}{\lambda+r}b(= pb/r)$  dollars during the inter-arrival period. Hence, the net saving of responding to the request now is  $p(b/r - c)$ . If  $K < p(b/r - c)$  (i.e., the net saving exceeds the fixed cost), we would prefer to maintain the system whenever a request is submitted (see Fig.1.b). In other situations, where the net saving per request is positive, but not enough to cover the fixed cost, there is a best trade-off point  $F^*$  (see Fig.1.c).

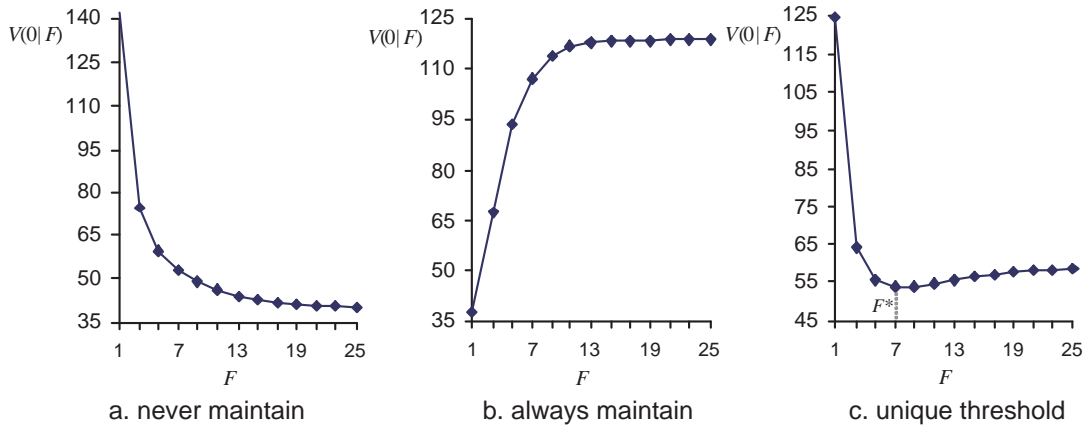


Figure 1: The Optimal Threshold Policy.

REMARK 1 *Under the threshold policy, we are only interested in the state set  $\{B|B \leq F\}$ . Any state above  $F$  is transient. In other words, in steady state,  $F$  is the highest number of requests that can accumulate in the system.*

### 3 Extended Model

In this section, we extend our discussion to consider nonzero time required for maintenance and study the behavior of two classes of policies in Sections 3.1 and 3.2. In addition to the fixed cost  $K$ , we also pay a correction cost of  $C$  per unit time spent in maintenance. We also assume that the time needed to fix each request follows an exponential distribution with a constant rate  $\mu$ .

#### 3.1 Work-Based Policy

##### 3.1.1 The Model

In this section, we investigate a work-based policy. If maintenance is initiated when the system state is  $B$ , a fraction  $\delta$  of the outstanding requests will be fixed in this round of maintenance. Note that  $\delta B$  should be an integer. So the policy is equivalent to one which specifies the amount of work  $m = \delta B$  in this round of maintenance. In other words,  $B - m$  is the target maintenance level. The maintenance cost function in (1) becomes <sup>1</sup>

$$V^c(B) = \min_{0 \leq m \leq B} \left\{ K + Ee^{-rT}V(B - m) + E \left[ \int_0^T e^{-r\theta}(C + bB)d\theta \right] \right\}. \quad (9)$$

In this case, the maintenance time  $T$  follows Erlang distribution with parameter  $(m, \mu)$ . The density of  $T$  is

$$f_m^\mu(T) = \frac{\mu e^{-\mu T} (\mu T)^{m-1}}{(m-1)!}.$$

Define  $s = \mu/(\mu + r)$ , then

$$Ee^{-rT} = s^m,$$

---

<sup>1</sup>In equation (9), we have assumed that new requests do not arrive during maintenance. As discussed in the Appendix C, the problem becomes quite complex if this assumption is relaxed. The implications of this assumption are also discussed there.



$$E \int_0^T e^{-r\theta} d\theta = \frac{1}{r}(1 - s^m).$$

Hence (9) can be simplified to

$$V^c(B) = \min_{0 \leq m \leq B} \left\{ K + s^m V(B - m) + \frac{C + bB}{r}(1 - s^m) \right\}. \quad (10)$$

### 3.1.2 Cost Functions under a Threshold Type Policy

We consider a work-based threshold policy of the form  $(F, m)$ . That is, we maintain the system when the system request level reaches  $F$  and fix  $m$  requests. Let  $W(B|F, m)$  be the optimal cost function under this policy. Then

$$W(F|F, m) = K + s^m W(F - m|F, m) + \frac{C + bF}{r}(1 - s^m). \quad (11)$$

Similar to the procedure in Section 2, we set  $V(B|F) = W(B|F, m)$  in (6), then substitute the expression  $W(F - m|F, m)$  in (11) and solve for  $V(F|F, m)$ . We get

$$W(F|F, m) = \frac{1}{1 - s^m q^m} \left[ K + \frac{C}{r}(1 - s^m) + \frac{b}{r}F + \frac{b}{r} \frac{1}{p} s^m [q - pm - (q + pF)q^m] \right].$$

Putting the above in (6) we obtain the cost function for  $0 \leq B \leq F$  as

$$\begin{aligned} W(B|F, m) &= \frac{q^{F-B}}{1 - s^m q^m} \left[ K + \frac{C}{r}(1 - s^m) + \frac{b}{r}F + \frac{b}{r} \frac{1}{p} s^m [q - pm - (q + pF)q^m] \right] \\ &\quad + \frac{b}{r} \frac{1}{p} [q + pB - (q + pF)q^{F-B}]. \end{aligned} \quad (12)$$

### 3.1.3 Suboptimality of Fractional Maintenance

Taking the first order difference of  $W(B|F, m)$  with respect to  $F$  in (12), we have

$$\Delta_F W(B|F, m) = \frac{q^{F-B} [-p(K + \frac{C}{r}(1 - s^m)) + \frac{b}{r}[q + (pm - q)s^m]]}{1 - s^m q^m}. \quad (13)$$

We make two observations from (13). First, for fixed finite  $m$ , if  $F = F^m$  minimizes  $W(B|F, m)$  for some  $B$ , then it minimizes  $W(B|F, m)$  for any  $B \leq F^m$ . Second, the

sign of (13) does not depend on the value of  $F$ . The second observation leads to the following result.

**PROPOSITION 1** *For any fixed finite  $m$ ,  $W(B|F, m)$  is minimized at  $F^m = m$  or  $F^m = \infty$ .*

Proposition 1 suggests that fractional maintenance is never optimal. Fig. 2 gives an intuitive explanation of this result, where we show the sample paths of two policies  $(F, m)$  and  $(m, m)$ . The shape of the two sample paths is the same, which means that the maintenance team incurs the same costs (the fixed cost and the correction cost) within a cycle under both policies. However, the system state level under the policy  $(F, m)$  is always  $F - m$  higher than that under the policy  $(m, m)$ . Thus, the user waiting cost is higher under fractional maintenance policy.

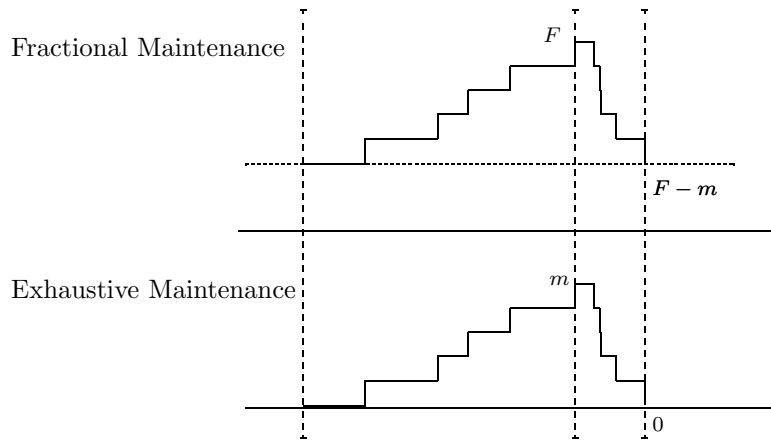


Figure 2: Fractional maintenance vs. exhaustive maintenance

The above argument assumes that the correction cost is linear in the maintenance time and the time to correct a request is stationary. One might argue that when the software system is very complex, it is possible that the effort to correct a request

increases with the number of requests. If we repeat the above sample path argument, we can easily see that by maintain partially, we not only pay a higher user waiting cost, but also a higher correction cost. Thus, the result in Proposition 1 applies to more general situation than the one studied here.

### 3.1.4 Optimality of the Threshold Type Policy

Using Proposition 1, the cost function in (12) reduces to

$$\begin{aligned}
W(B|F, F) = & \frac{q^{F-B}}{1-s^F q^F} \left[ K + \frac{C}{r}(1-s^F) + \frac{b}{r}F + \frac{b}{r} \frac{1}{p} s^F [q - pF - (q + pF)q^F] \right] \\
& + \frac{b}{r} \frac{1}{p} [q + pB - (q + pF)q^{F-B}]. \tag{14}
\end{aligned}$$

Notice that if we treat  $F$  as a continuous variable, then  $W(B|F, F)$  is clearly continuous in  $F$ . Hence, there exists an optimum  $F^*$  over the set of extended real numbers. We next establish the optimality of the threshold policy. The detailed proof is relegated to Appendix B.

**THEOREM 2** *Let  $F^*$  minimize (14). Then  $V(B) = W(B|F^*, F^*)$  satisfies the dynamic programming equations (2), (9) and (3).*

### 3.1.5 Computational Issues

The cost function under the optimal work-based threshold policy exhibits a nice property that helps reduce computation complexity. This property is stated in Proposition 2 below.

**PROPOSITION 2**  *$F = \infty$  minimizes  $W(B|F, m)$  for any value of  $m$ , if and only if either*

1.  $\bar{m} = \frac{b(qt+ps)-ptC}{bpt} \leq 0$ , or

$$2. pK + \frac{C}{r}p(1 - s^{\bar{m}}) \geq \frac{b}{r}[q + (p\bar{m} - q)s^{\bar{m}}].$$

REMARK 2 *We can deduce a more intuitive condition from the above proposition. Notice that a sufficient condition for Proposition 2 to be true is  $\bar{m} \leq 0$  (or  $(\mu + \lambda)(b/r) < C$ ). In one unit of time, either  $\mu$  requests get fixed, or  $\lambda$  new requests arrive (on average). Then, for each unit of time we spend in maintenance, we save  $(\mu + \lambda)(b/r)$  in user waiting cost. At the same time, we pay a correction cost of  $C$ . Thus, if the correction cost is larger than the saving per unit time, we should never bother to maintain the system. This explanation is similar to the one provided for the first case in Theorem 1.*

REMARK 3 *Another outcome of Proposition 2 is that there is a pair  $(m_1, m_2)$  such that  $F = \infty$  minimizes  $W(B|F, m)$  whenever  $F < m_1$  or  $F > m_2$ . Thus, it is always easy to find the optimum threshold provided it exists.*

## 3.2 Time-Based Policy

### 3.2.1 Cost Functions

Consider the following time-based policy. When the  $B^{th}$  request arrives, we decide whether or not to maintain the system. If we decide to maintain the system, we also specify the maintenance time  $\tau$ . Three possible situations may arise during maintenance:

1. With probability  $D_{m,\tau}^\mu = e^{-\mu\tau}(\mu\tau)^m/m!$ ,  $0 < m < B$ ,  $m$  requests are fixed during time  $\tau$ , and the system is released to the user.
2. With probability  $D_{0,\tau}^\mu$ , no request is fixed during time  $\tau$ . Then we undergo another maintenance of duration  $\tau$  (this is reasonable, because the system state is still at  $B$  and we have already incurred the setup cost).

3. If all  $B$  requests are fixed at some time  $\psi$  before  $\tau$ , then the system is released to the users at time  $\psi$ .

Thus, the actual maintenance time  $T = (n - 1)\tau + \min\{\tau, \psi\}$  is a stopping time, where  $(n - 1)$  is the number of times that case 2 happens in one maintenance cycle. Note that  $n$  follows a geometric distribution with success probability  $(1 - D_{0,\tau}^\mu)$ . It is easily seen that  $\psi$  follows an Erlang distribution, i.e., the density of  $\psi$  is given by

$$f_B^\mu(\psi) = \frac{\mu e^{-\mu\psi} (\mu\psi)^{B-1}}{(B-1)!}.$$

Note that the following relation holds

$$\sum_{m=0}^{B-1} D_{m,\tau}^\mu + \int_0^\tau f_B^\mu(\psi) d\psi = 1.$$

This shows that the three cases described above cover all the possibilities. The expected maintenance time can be obtained as follows.

$$\begin{aligned} ET &= \sum_{n \geq 1} \left[ \int_\tau^\infty n\tau f_B^\mu(\psi) d\psi + \int_0^\tau [(n-1)\tau + \psi] f_B^\mu(\psi) d\psi \right] (1 - D_{0,\tau}^\mu) (D_{0,\tau}^\mu)^{n-1} \\ &= \frac{\tau}{1 - D_{0,\tau}^\mu} + \int_0^\tau (\psi - \tau) f_B^\mu(\psi) d\psi. \end{aligned} \quad (15)$$

To set up the dynamic programming formulation, we define the optimal cost function and the optimal continuation cost function identical to (3) and (2), respectively. The optimal maintenance cost is denoted by  $V^c(B, I)$ , where  $I$  is an indicator variable. When  $I = 0$ , maintenance starts immediately after a new request arrives. When  $I = 1$ , case 2 happens and no additional setup cost is incurred. That is,

$$\begin{aligned} V^c(B, 1) &= \min_{\tau \geq 0} \left\{ \sum_{m=1}^{B-1} \left[ e^{-r\tau} V(B-m) + (C + bB) \int_0^\tau e^{-r\theta} d\theta \right] D_{m,\tau}^\mu + e^{-r\tau} V^c(B, 1) D_{0,\tau}^\mu \right. \\ &\quad \left. + \int_0^\tau \left[ e^{-r\psi} V(0) + (C + bB) \int_0^\psi e^{-r\theta} d\theta \right] f_B^\mu(\psi) d\psi \right\}, \end{aligned} \quad (16)$$

$$V^c(B, 0) = V^c(B, 1) + K \quad (17)$$

The only difference between (16) and (17) is that a setup cost  $K$  is applied to (17) but not (16). Inside the minimum of (16), the first, second and third term correspond to the cases 1, 2 and 3 discussed before. We substitute (16) in (17) and solve for  $V^c(B, 0)$  as follows.

$$\begin{aligned}
V^c(B) &= V^c(B, 0) \\
&= \min_{\tau \geq 0} \left\{ K + \frac{1}{1 - D_{0,\tau}^{\mu+r}} \left[ \sum_{m=1}^{B-1} D_{m,\tau}^{\mu} \left( e^{-r\tau} V(B-m) + (C + bB) \int_0^{\tau} e^{-r\theta} d\theta \right) \right. \right. \\
&\quad \left. \left. + \int_0^{\tau} f_B^{\mu}(\psi) \left( e^{-\psi\tau} V(0) + (C + bB) \int_0^{\psi} e^{-r\theta} d\theta \right) d\psi \right] \right\}. \tag{18}
\end{aligned}$$

This gives  $V^c(B)$ , the actual optimal maintenance cost function at state  $B$ .

### 3.2.2 Threshold Policy

We now consider a class of policies in form of  $(F, \tau)$ . That is, we maintain the system for a duration  $\tau$  if the request level in the system is at or above the threshold  $F$ . Let  $W(B|F, \tau)$  be the optimal cost function under this policy.

Use similar approach as before, we take  $B = F$  in (18), put  $V(F - m|F) = W(F - m|F, \tau)$  in (6), and solve for  $W(F|F, \tau)$ .

$$W(F|F, \tau) = \frac{(1 - D_{0,\tau}^{\mu+r})K + A_{F,\tau} + \frac{C+bF}{r}(1 - D_{0,\tau}^{\mu} - G_{F,\tau})}{1 - D_{0,\tau}^{\mu+r} - H_{F,\tau}}, \tag{19}$$

where,

$$H_{F,\tau} = \sum_{m=1}^{F-1} s^m q^m D_{m,\tau}^{\mu+r} + s^F q^F \sum_{m=F}^{\infty} D_{m,\tau}^{\mu+r}, \tag{20}$$

$$G_{F,\tau} = \sum_{m=1}^{F-1} s^m D_{m,\tau}^{\mu+r} + s^F \sum_{m=F}^{\infty} D_{m,\tau}^{\mu+r}, \tag{21}$$

$$A_{F,\tau} = \frac{b}{r} \frac{1}{p} \left[ q G_{F,\tau} + p \sum_{m=1}^{F-1} s^m (F - m) D_{m,\tau}^{\mu+r} - (q + pF) H_{F,\tau} \right] \tag{22}$$

$$s = \mu/(\mu + r), \text{ and } t = 1 - s. \tag{23}$$

The above together with (6) gives

$$W(B|F, \tau) = \frac{q^{F-B}}{1 - D_{0,\tau}^{\mu+r} - H_{F,\tau}} \left[ (1 - D_{0,\tau}^{\mu+r})K + A_{F,\tau} + \frac{C + bF}{r}(1 - D_{0,\tau}^{\mu} - G_{F,\tau}) \right] + \frac{b}{r} \frac{1}{p} [q + pB - (q + pF)q^{F-B}]. \quad (24)$$

Next, we establish in Theorem 3 the optimality of the time-based threshold policy. The proof Theorem 3 is similar to that of Theorem 2 and is omitted.

**THEOREM 3** *Let  $(F^*, \tau^*)$  minimize (24). The cost function  $V(B) = W(B|F^*, \tau^*)$  satisfies the dynamic programming equations (2), (18) and (3).*

### 3.3 Optimal Maintenance Policy under Time Constraint

In reality, maintenance activities may be constrained by resource limitations or requirements imposed by end users. As a result, maintenance should be completed within a given time window. In what follows, we propose a model to study the optimal policy under a time constraint.

Suppose that we are following the work-based policy.<sup>2</sup> An upper bound  $L$  is imposed as the maintenance time. Then, the maintenance cost when the maintenance starts at state  $B$  and  $m$  requests are planned to be fixed is given by

$$V_L^c(B) = \min_{0 \leq m \leq B} \left\{ K + \int_0^L \left[ e^{-r\tau} V(B - m) + (C + bB) \int_0^\tau e^{-r\theta} d\theta \right] f_m^\mu(\tau) d\tau + \sum_{i=0}^{m-1} \left[ e^{-rL} V(B - i) + (C + bB) \int_0^L e^{-r\theta} d\theta \right] D_{i,L}^\mu \right\}. \quad (25)$$

The second term on the left-hand side of (25) corresponds to the case when all  $m$  requests are fixed before time  $L$ . This event happens with probability  $\int_0^L f_m^\mu(\tau) d\tau$

---

<sup>2</sup>Note that if we impose the time constraint to the time-based policy, the problem reduces to one with only one decision variable  $F$ .

(where  $f_m^\mu$  is the density of the time to fix  $m$  requests), and the cost incurred is the same as that under the work-based policy. The second term on the left-hand side of (25) corresponds to the case when  $i$  requests, with  $i < m$ , are fixed before  $L$ . In this case, the cost incurred is the cost to fix  $i$  requests, and the corresponding probability is  $D_L^{\mu 3}$ .

Using the approach developed in Section 3.1 or 3.2, the cost function under the work-based threshold policy  $(F, m)$  is given by

$$V_L(B|F, m) = q^{F-B} \frac{K + Z_{m,L} + \frac{C}{r}(1 - G_{m,L})}{1 - H_{m,L}} + \frac{b}{r} \frac{1}{p} [q + pB - q^{F-B+1}] \quad (26)$$

where

$$Z_{m,L} = \frac{b}{r} \frac{1}{p} \left[ qG_{m,L} - p \left( \sum_{i=0}^{m-1} s^i i D_{i,L}^{\mu+r} + s^m m \sum_{i=m}^{\infty} D_{i,L}^{\mu+r} \right) - qH_{m,L} \right],$$

and  $H_{B,L}$ ,  $G_{m,L}$  are defined in (20), (21), respectively.

Taking the first order difference with respect to  $F$  in (26) gives

$$\Delta_F V_L(B|F, m) = q^{F-B} p \left( q - \frac{K + Z_{m,L} + \frac{C}{r}(1 - G_{m,L})}{1 - H_{m,L}} \right) \quad (27)$$

Clearly, the sign of (27) does not depend on the choice of  $F$ . Hence, similar to the result in Proposition 1,  $F = m$  holds in this situation as well. That is, it is optimal to fix all the requests accumulated in the system. One can also use the approach in Theorem 2 to establish the optimality of the work-based threshold policy under a time constraint.

## 4 Numerical Analysis

In this section, we compare the policies discussed in previous sections to provide additional insights.

---

<sup>3</sup>It is necessary to assume that  $L$  is large enough such that the probability that no request is fixed during  $L$  is negligible. Otherwise, the time constraint is easily violated since the request level may not decrease after maintenance even if the threshold is set at 1. As a result, the optimal policy would call for no maintenance.



As an illustrative example, we consider an organization using a software system, where the system administrator receives in average  $2(= \lambda)$  change requests each day. If the request is not responded to, users incur  $\$25(= b)$  every day for lost in productivity in operations. The cost to organize the maintenance team and set up the maintenance environment is  $\$80(= K)$ . The maintenance team can fix an average of  $6(= \mu)$  requests each day. The cost for request correction is  $\$20(= C)$  per day. Finally, the discount rate is  $0.9(= r)$ . Table 1 shows the above parameter settings as well as the range of values used for each of these parameters.

Table 1: Parameter setting

Parameter	Basic Setting	Range	$F^* = \infty$ when	Figure
$K$	80	$0 \sim 105$	–	Fig. 3(a)
$C$	10	$0 \sim 45$	$c > 45$	Fig. 3(b)
$b$	25	$20 \sim 120$	$b < 20$	–
$\lambda$	2	$1 \sim 10$	$\lambda > 33$	Fig. 4(a)
$\mu$	6	$5 \sim 40$	$\mu < 5$	Fig. 4(b)
$r$	0.9	$0.1 \sim 1$	$r > 1.01$	Fig. 5

## 4.1 Results

### Maintenance costs

Increasing the fixed cost  $K$  favors postponing maintenance activities. As shown in Fig. 3(a), the maintenance threshold, the expected maintenance time and total cost increase with  $K$ . A similar pattern exists for the correction cost  $C$  (see Fig. 3(b)). The total cost function is convex in  $K$ , i.e., it is more sensitive to changes of  $K$  for a lower value of  $K$ . Also, the total cost function is almost linear in  $C$ .

It can be seen from the cost expressions (9) and (18) that user waiting cost  $b$  behaves in a opposite manner to the effects of  $K$  and  $C$ . The optimal threshold decreases and becomes less sensitive to  $b$  as  $b$  increases.

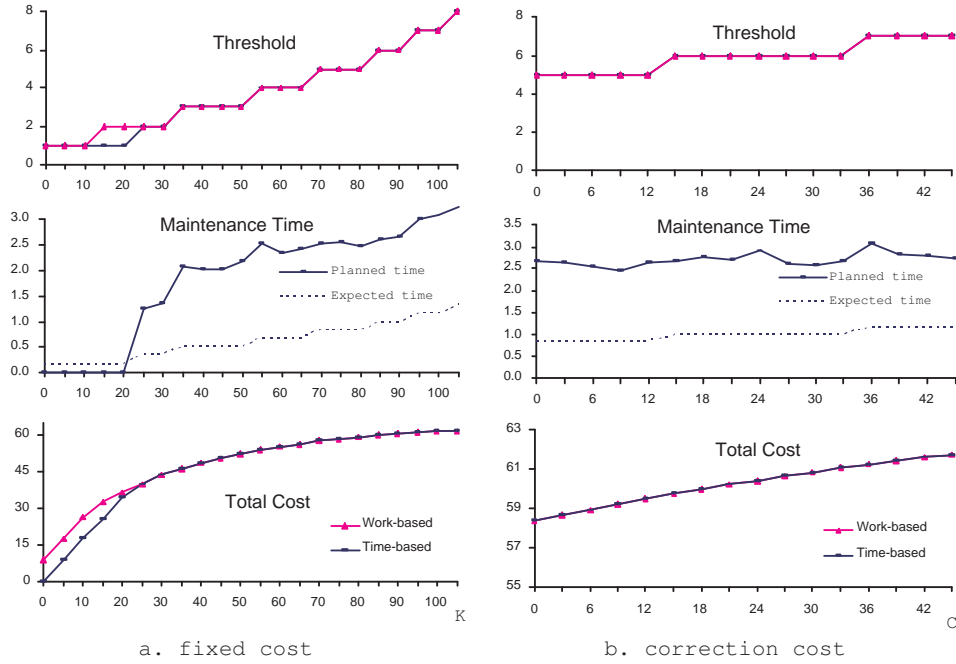


Figure 3: Change maintenance costs.

### Request arrival rate and maintenance rate

As shown in Fig. 4(a), the optimal total cost increases when requests arrive more frequently. At the same time, the optimal threshold decreases. When the request arrival rate  $\lambda \rightarrow 0$ , leaving a request in the system results in a high waiting cost. In this case, always responding to the incoming request is optimal. On the other hand, when  $\lambda$  becomes large, the requests accumulate in the system very fast. We can now respond to a large batch of requests to reduce fixed costs. Hence the threshold increases with the request arrival rate. We also observe that the optimal total cost decreases when the maintenance team is more efficient at responding to requests. However, the optimal threshold is not as sensitive to the maintenance rate as the request arrival rate.

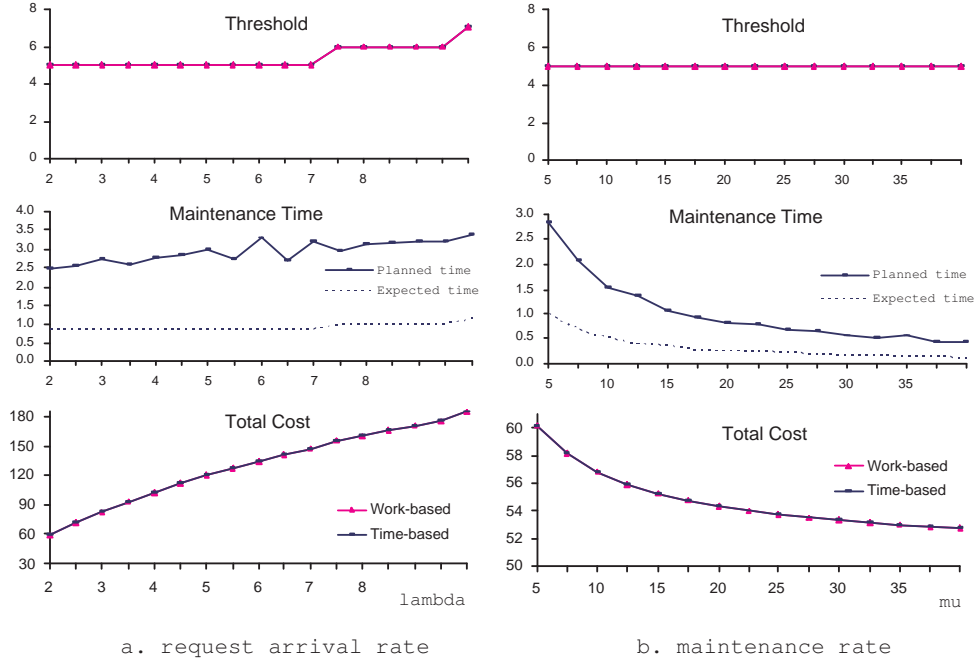


Figure 4: Change request incoming rate or maintenance rate.

## Discount rate

A higher discount rate  $r$  means a lower future value. Fig. 5 shows that the optimal policies tend to postpone maintenance when it is cheaper to pay in the future. As a result, the total cost decreases with the discount rate.

The difference between the total cost of work-based policy and that of the state-based policy is very small (less than 0.1%) in most of the parameter range. The difference is significant only in the extreme cases. For example, when the setup cost  $K$  is very small and the optimal threshold is set to be the boundary value of 1, the work-based policy reveals an advantage. Intuitively, the work-based policy is discrete, while the time-based policy is continuous. Thus, the time-based policy should always behave better than the work-based policy. Moreover, our results show, the difference of the two policy is only significant in the extreme cases, i.e. when the setup cost  $K$  or correction cost  $C$  is very low, or the waiting cost  $b$  is very high.

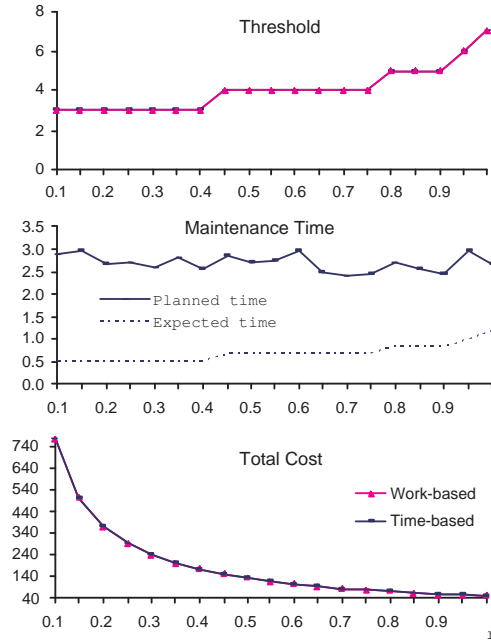


Figure 5: Change discount rate  $r$ .

Another interesting result is that, in the time-based policy, the expected actual maintenance duration  $ET$  in (15) is very close to the expected time to fix all  $F$  requests. As a result, the time-based policy also tends to clear all the existing maintenance requests once maintenance is initiated. This is consistent with our discussion concerning the work-based policy where it was found that fractional maintenance is not optimal. In addition, we observe that the planned maintenance time  $\tau$  is usually greater than the expected time to fix  $F$  requests. On average, the actual maintenance project ends earlier than planned.

## 4.2 Impact of Time Constraint

Next we discuss the impact of the time constraint on the optimal threshold policy and the optimal total cost. We compute the optimal policy for the model described in Section 3.3. As shown in Fig. 6, when the time schedule is extremely stringent,

it is optimal to never maintain the system. This is because there is a significantly large probability that the system state does not improve after maintenance. When the time constraint is relaxed, the total optimal cost decreases. Eventually, when the time constant becomes sufficiently loose, we can perform the maintenance following the optimal work-based threshold policy. As a result, the threshold value converges to the value of the case when no time constraint is imposed.

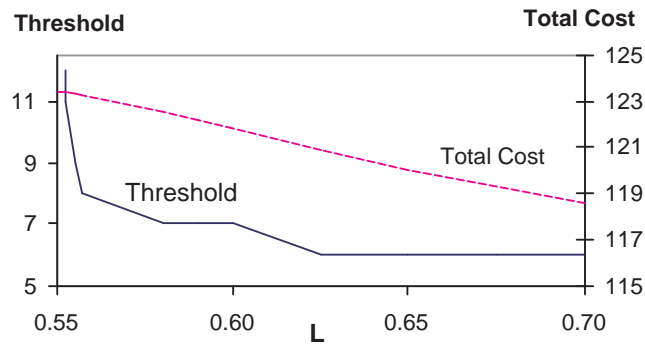


Figure 6: Impact of Time Constraint.

## 4.3 Approximation Methods

From our previous discussion, the models considering maintenance time are very complex. It would be helpful if we can use instantaneous maintenance model and get a quick approximation of the policy parameters. In this section, we examine how to use the results of instantaneous maintenance model for this purpose.

### 4.3.1 Using Instantaneous Model

One way to get a quick solution to the maintenance problem is to directly use the parameters of the instantaneous model. This gives a good approximation when the maintenance is costly but not time consuming. That is, the fixed cost  $K$  and the correction cost  $C$  are low relative to the waiting cost  $b$ , and the request arrival rate

$\lambda$  is low relative to the maintenance rate  $\mu$ . In such situations, the maintenance time becomes negligible. We remark that such a method can be used because we have shown that fractional maintenance is suboptimal.

### 4.3.2 Using Instantaneous Threshold Values

Another way to calculate the optimal solutions is to get the optimal threshold from the instantaneous model and estimate the parameters in the work-based and time-based model.

To get the approximate threshold value, we set  $c = C/\mu$  and find the optimum value of  $F$  in (6). We can then use this  $F$  value to compute the total cost for the work-based policy in (14).

To compute the time-based policy, we also need to estimate the planned maintenance time  $\tau$ . From our previous discussion, the optimal policy is to maintain the system completely, and the expected maintenance time  $ET$  is close to the expected time to correct all the  $F$  requests, i.e.,  $ET = F/\mu$ . Hence, we can compute  $\tau$  numerically from (15). Then, we can evaluate the policy in (24).

In general, such a method yields better approximations to the work-based policy than to the time-based policy. This is because an additional time parameter has to be estimated in the time-based model. For illustration purposes, we further discuss the approximation behavior of the time-based policy. We evaluate the approximation accuracy by computing the percentage deviation of the approximated total cost with respect to the optimal total cost.

Overall, the approximation accuracy is not sensitive to the fixed cost  $K$ , correction cost  $C$ , maintenance rate  $\mu$  and interest rate  $r$ . In our experiments, the approximation is generally very good (with differences less than 1%). The only exceptions are the case when the waiting cost is extremely high ( $b > 65$ ) as compared to  $K$  and  $C$ , and

Table 2: Total cost in response to the parameter values.

Total Cost	Basic Setting	$K = 40$	$C = 6$	$b = 40$	$\mu = 3$	$\lambda = 1$	$r = 0.5$
$(F^*, \tau^*)$	59.2761	48.2924	58.8978	83.0042	59.2763	59.3840	92.0484
$(F^* + 1, \tau^*)$	60.3763	51.0690	59.8817	92.3869	60.3764	59.2759	90.2708
$(F^* - 1, \tau^*)$	59.3844	49.9857	59.0990	83.3917	59.3849	59.8471	95.8719
$(F^*, 2\tau^*)$	59.2753	48.2917	58.8974	83.0033	59.2753	59.3825	92.0481
$(F^*, \tau^*/2)$	59.6448	48.5099	59.2061	83.3551	59.6853	59.8714	92.3054

the case when the request arrival rate  $\lambda$  is high. (see Fig. 7).

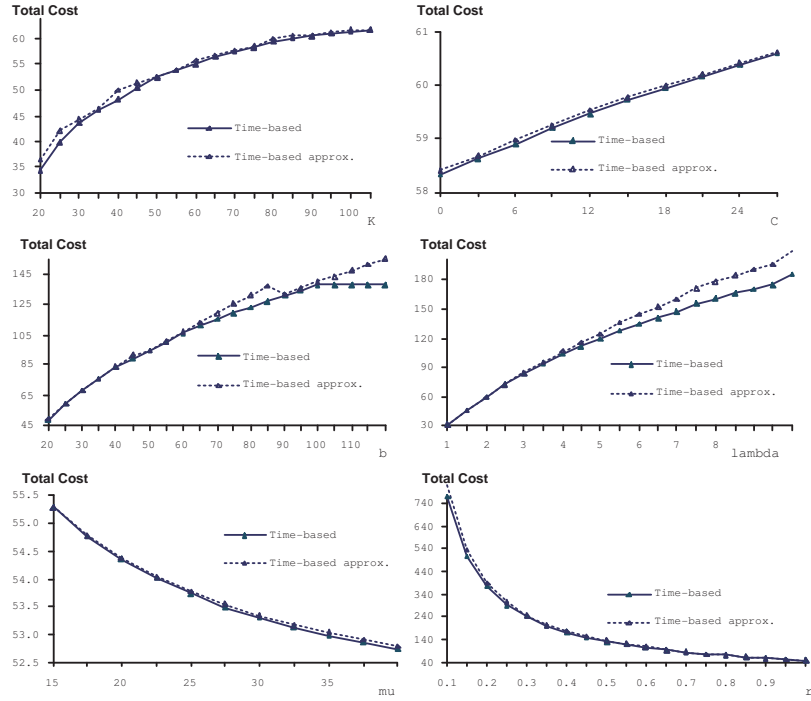


Figure 7: Approximation using instantaneous maintenance model.

We also find that the approximation accuracy is more sensitive to the choice of the threshold  $F$  than to the planned maintenance time  $\tau$ . Table 2 gives some examples.

## 5 Concluding Remarks

The major result of this paper is that we establish the optimality of the threshold policy in stable software maintenance projects. As we point out, in the presence of maintenance time, the work-based policy and the time-based policy are basically equivalent. Thus, an organization can choose either of the two policies according to the situation or preference. In practice, some companies do partial maintenance because of various reasons (e.g., budget constraints, time constraints, work-force constraints, etc.). We show, however, this is never optimal in terms of the total costs incurred.

There are a number of interesting extensions to the problem addressed here. In the models discussed in this study, incoming requests are blocked during maintenance. One could consider the case that requests continue to arrive during maintenance. However, this gives rise to mathematical difficulties. As the reader can see from Appendix C, the cost function becomes complicated and further analysis is extremely difficult.

Another interesting view of this problem is considering different classes of requests. In this case, service priority and service discipline will have to be specified and some results in queuing theory may be applicable to this problem.

One could also consider time-dependent arrival rate and maintenance rate. As maintenance is done, the system runs more smoothly, and the request arrival rate decreases. At the same time, the maintenance team becomes more familiar with the software system, and therefore more efficient.

## Appendix



## A Proof of Theorem 1

We establish theorem 1 through a series of lemmas. To examine the shape of  $V(0|F)$  as a function of  $F$ , we consider the first order difference and the second order difference of  $V(0|F)$ .

$$\begin{aligned}\Delta_F V(0|F) &= V(0|F+1) - V(0|F) \\ &= \frac{q^F}{(1-q^{F+1})(1-q^F)} \left[ -pK + (1-p(F+1) - q^{F+1}) \left( c - \frac{b}{r} \right) \right]\end{aligned}\quad (28)$$

$$\begin{aligned}\Delta_F^2 V(0|F) &= \Delta_F V(0|F+1) - \Delta_F V(0|F) \\ &= \frac{q^F [p^2(1+q^{F+1})K + [p^2(1+q^{F+1})(F+1) - p(1+q)(1-q^{F+1})] (c - \frac{b}{r})]}{(1-q^{F+2})(1-q^{F+1})(1-q^F)}\end{aligned}\quad (29)$$

We notice the properties of the following functions.

$$\text{Let} \quad S^1(x) = 1 - px - q^x, \quad S^1(1) = 0.$$

$$\text{Then} \quad \Delta_x S^1(x) = -p(1 - q^x) \leq 0.$$

$$\text{Thus} \quad S^1(x) \leq 0, \text{ for } x \geq 1.$$

$$\text{Let} \quad S^2(x) = p^2(1+q^x)x - p(1+q)(1-q^x), \quad S^2(1) = 0.$$

$$\text{Then} \quad \Delta_x S^2(x) = p^2[1 - q^x - pq^x x], \quad \Delta_x S^2(x)|_{x=1} = p^4 \geq 0,$$

$$\text{and} \quad \Delta_x^2 S^2(x) = p^4 q^x (x+1).$$

$$\text{Thus} \quad S^2(x) \geq 0, \text{ for } x \geq 1.$$

LEMMA 1  $V(0|F)$  is minimized at  $F = \infty$  if and only if  $c - \frac{b}{r} \geq 0$ , in this case  $V(0|F)$  is strictly decreasing convex in  $F$ .

PROOF. Clearly, when  $c - \frac{b}{r} \geq 0$ , (28) is strictly negative, i.e.  $V(0|F)$  is strictly decreasing in  $F$ . Thus,  $F = \infty$  is the only minimizer. Conversely, if  $c - \frac{b}{r} < 0$ , then (28) is always positive for large enough  $F$ . In this case  $F = \infty$  cannot minimize

$V(0|F)$ . In addition, when  $c - \frac{b}{r} \geq 0$ , (29) is clearly positive, which indicates that the cost function is strictly convex. (cf. the example in Fig.1.a).

□

Now we can restrict our attention to the case when  $c - \frac{b}{r} < 0$ . Let

$$S^3(x) = -pK + \left(c - \frac{b}{r}\right) (1 - q^{x+1} - p(x+1)),$$

$$\text{then } \Delta_x S^3(x) = -\left(c - \frac{b}{r}\right) (1 - q^{x+1})p > 0.$$

LEMMA 2 *When  $c - \frac{b}{r} < 0$  and  $K \leq p(b/r - c)$ ,  $V(0|F)$  is minimized when  $F = 1$ .*

PROOF. Note that  $S^3(x)$  is strictly increasing in  $x$ . Hence, if  $S^3(1) = -pK - (c - b/r)p^2 \geq 0$ , or the fixed cost  $K$  satisfies  $K \leq -p(c - b/r)$ , we will get  $\Delta_F V(0|F) \geq 0$ , for any  $0 < F < \infty$ . In this case,  $V(0|F)$  is minimized when  $F = 1$ . (cf. the example in Fig.1.b)

□

LEMMA 3 *When  $c - \frac{b}{r} < 0$  and  $K > -p(c - b/r)$ , there exists a unique minimizer  $F^*$ , with  $1 < F^* < \infty$ , of  $V(0|F)$ . Furthermore, there exists an  $F^0$  such that  $V(0|F)$  is convex for  $1 \leq F \leq F^0$  and concave for  $F > F^0$ .*

PROOF. To show the unique existence, we note that  $S^3(1) < 0$ ,  $\lim_{x \rightarrow \infty} S^3(x) > 0$  and  $S^3(x)$  is increasing in  $x$ . Hence, there exist a unique  $F^*$  such that  $V(0|F)$  is increasing for  $0 < F \leq F^*$  and decreasing for  $F > F^*$ .

Now let

$$S^4(x) = p^2(1 + q^{x+1})K + [p^2(1 + q^{x+1})(x+1) - p(1+q)(1 - q^{x+1})] \left(c - \frac{b}{r}\right),$$

$$\Delta_x S^4(x) = -p^3 q^{x+1} \left[ K + p \left(c - \frac{b}{r}\right) \right] + p^2 [1 + (p^2 - p(x+1) - 1)q^{x+1}] \left(c - \frac{b}{r}\right).$$

Clearly, when  $K > -p(c - b/r)$  the first term in  $\Delta_x S^4(x)$  is negative. We let

$$S^5(x) = 1 + (p^2 - p(x+1) - 1)q^{x+1}, \quad S^5(0) = p^2(1+q) > 0,$$

$$\Delta_x S^5(x) = p^2 q^{x+1}(q+x+1) > 0.$$

Since  $c - \frac{b}{r} < 0$  the second term in  $\Delta_x S^4(x)$  is strictly decreasing in  $x$  and thus negative. So we have  $\Delta_x S^4(x) < 0$ , that is,  $S^4(x)$  is strictly decreasing in  $x$ . In addition,  $S^4(0) = p^2(1+q)K \geq 0$  and  $\lim_{x \rightarrow \infty} S^4(x) \rightarrow -\infty$ . So there is a unique point where  $S^4(x) = 0$ . Put all these results in (29), we conclude that there exists an  $F^0$  such that the second order condition of  $V(0|F)$  in (29) is positive for  $1 \leq F \leq F^0$  and negative for  $F > F^0$  (cf. the example in Fig.1.c).

□

Finally, we establish the optimality of the threshold policy by showing that  $V(B) = V(B|F^*)$  satisfies the dynamic programming equation (3). For  $0 \leq B \leq F$  We substitute  $V(B) = V(B|F)$  in (1) and (2) and get

$$V^c(B) = K + cB + q^F V(F|F) + \frac{b}{r} \frac{1}{p} [q - (q + pF)q^F],$$

$$V^n(B) = V(B|F) = q^{F-B} V(F|F) + \frac{b}{r} \frac{1}{p} [q + pB - (q + pF)q^{F-B}].$$

Then

$$V^c(B) - V^n(B) = \frac{1 - q^B}{q^B} \left\{ \frac{q^B}{1 - q^B} \left[ K + \left( c - \frac{b}{r} \right) B \right] q^F V(F|F) \right\}$$

$$= \frac{1 - q^B}{q^B} [V(0|B) - V(0|F)].$$

Clearly, when  $F$  is the minimizer to  $V(0|B)$ , the above expression is negative. Hence, we established the optimality of the threshold policy.

## B Proofs in Section 3.2

PROOF OF PROPOSITION 2. Let

$$g_1(m) = p \left[ K + \frac{C}{r} (1 - s^m) \right] - \frac{b}{r} [q(1 - s^m) + pms^m].$$

Notice that  $g_1(0) = K > 0$ ,  $\lim_{m \rightarrow \infty} g_1(m) = p(K + C/r) - qb/r$  and

$$\Delta_m g_1(m) = \frac{s^m}{r} [ptc + (ptm - ps - qt)b].$$

So the solution to  $\Delta_m g_1(m) = 0$  is given by

$$\bar{m} = \frac{(qt + ps)b - ptC}{ptb}.$$

Now let

$$g_2(m) = ptc + (ptm - ps - qt)b.$$

$g_2(m)$  is clearly increasing in  $m$ . It follows that the first order difference of function  $g_1(m)$  is strictly increasing. Hence we conclude the proof. □

PROOF OF THEOREM 2. Let  $F$  be the minimizer to the cost function (14) for some  $B$ . It is easily seen that  $F$  also minimizes  $W(B|F, F)$  for any  $0 \leq B \leq F$ . Under policy  $(F, F)$ , we have

$$\begin{aligned} V^n(B) &= q^{F-B} W(F|F, F) + \frac{b}{r} \frac{1}{p} [q + pB - (q + pF)q^{F-B}], \\ V^c(B) &= K + \frac{C}{r}(1 - s^B) + \frac{b}{r} B(1 - s^B) + s^B V(0) \\ &= K + \frac{C}{r}(1 - s^B) + \frac{b}{r} B(1 - s^B) + \frac{b}{r} \frac{1}{p} [q - (q + pF)q^F] s^B + s^B q^F W(F|F, F). \end{aligned}$$

Then,

$$\begin{aligned} &\frac{q^B}{1 - s^B q^B} [V^c(B) - V^n(B)] \\ &= \frac{q^B}{1 - s^B q^B} \left[ K + \frac{C}{r}(1 - s^B) + \frac{b}{r} B(1 - s^B) + \frac{b}{r} \frac{1}{p} s^B [q - (q + pB)q^B] \right] \\ &\quad - q^F W(F|F, F) + \frac{b}{r} \frac{1}{p} [q - (q + pB)q^B] - \frac{b}{r} \frac{1}{p} [q - (q + pF)q^F] \\ &= W(0|B, B) - W(0|F, F). \end{aligned}$$

It follows by the minimality of  $F$  that the last expression is positive. Hence, we conclude the theorem. □

## C Allowing Incoming Requests During Maintenance

Next, we give the formulations to the problem where incoming requests are allowed during maintenance. In either cases, the problem becomes complicated.

### C.1 Work-Based Policy

In a work-based policy, we monitor the system state  $B$  and decide, whether or not go to maintenance. If we decide to maintain the system, then we specify an optimal target level  $B - m$  at which maintenance is complete. Let  $\{B_\theta, 0 \leq \theta \leq \tau\}$  is the process of the state during maintenance. Then the cost function of maintenance becomes

$$\begin{aligned} V^c(B) = & \min_{0 \leq m \leq B} \left\{ K + CE \int_0^{\tau_m} e^{-r\theta} d\theta + Ee^{-r\tau_m} V(B - m) \right. \\ & \left. + bE \left[ \int_0^{\tau_m} B_\theta e^{-r\theta} d\theta \middle| B_{\tau_m} = B - m, B_0 = B \right] \right\}. \end{aligned} \quad (30)$$

Given  $B_0 = B$ , the random time  $\tau$  is a *hitting time*, which is defined as

$$\tau_m = \inf\{t > 0, B_t = B - m\}.$$

By the memoryless property,  $\tau_m$  depends only on  $m$ , not on  $B$ . The  $\tau_m$  can be treated as  $m$  busy periods of an  $M/M/1$  queue. The distribution function of the latter can be computed in terms of Laplace transform (Kinatader and Lee 2000), by Ballot Theorem (Takacs 1962), or computed directly by conditioning on the first service time, similar approach can be found in Ross (2000). However, the calculation is messy.

### C.2 Time-Based Policy

Assuming  $\mu > \lambda$  (Otherwise the system is not stable and cost function goes to infinity as time goes to infinity). Suppose we maintain at level  $B$  and for a period of duration  $\tau$ . Let  $\{B_\theta, 0 \leq \theta \leq \tau\}$  be the stochastic process of the system state, with initial state

$B_0 = B$ . After maintenance the system reaches  $B_\tau$ . The maintenance cost function becomes

$$\begin{aligned}
V^c(B, 1) &= \min_{\tau \geq 0} \left\{ \int_0^\tau \left[ e^{-r\psi} P_{0,\psi}^B V(0) + C e^{-r\psi} + b E \left( \int_0^\psi e^{-r\theta} B_\theta d\theta \mid B_0 = B \right) \right] f(\psi) d\psi \right. \\
&\quad + \left[ e^{-r\tau} \left( \sum_{m=1}^{B-1} P_{m,\tau}^B V(m) + \sum_{m=B}^{\infty} P_{m,\tau}^B V^c(m, 1) + C \right) \right. \\
&\quad \left. \left. + b E \left( \int_0^\tau e^{-r\theta} B_\theta d\theta \mid B_0 = B \right) \right] \int_\tau^\infty f(\psi) d\psi \right\} \tag{31}
\end{aligned}$$

$$V^c(B, 0) = K + V^c(B, 1) \tag{32}$$

The distribution of  $B_\tau$  is given by

$$P_{m,\tau}^B = P(B_\tau = m \mid B_0 = B) = \begin{cases} \sum_{k=0}^{\infty} \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!} \frac{(\mu\tau)^k e^{-\mu\tau}}{k!}, & \text{if } m = B, \\ \sum_{k=0}^{\infty} \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!} \frac{(\mu\tau)^{B-m+k} e^{-\mu\tau}}{(B-m+k)!}, & \text{if } m < B, \\ \sum_{k=0}^{\infty} \frac{(\lambda\tau)^{m-B+k} e^{-\lambda\tau}}{(m-B+k)!} \frac{(\mu\tau)^k e^{-\mu\tau}}{k!}, & \text{if } m > B. \end{cases}$$

$B_\theta$  is a negative-drifted random walk with drift  $-(\mu - \lambda)$ . Thus,  $E(B_\theta \mid B_0 = B) = B - (\mu - \lambda)\theta$ . To compute the value functions we have to deal with Bessel functions, and the difference equation looks complicated.

## References

- [1] Arthur, L. J. (1988). *Software Evolution*, John Wiley & Sons, Inc., New York.
- [2] Banker, R. D., G. B. Davis and S. A. Slaughter (1998). Software Development Practices, Software Complexity, and Software Maintenance Performance: A Field Study, *Management Science*, **44**, 4, 433-450.
- [3] Banker, R. D. and S. A. Slaughter (1997). A Field Study of Scale Economies in Software Maintenance, *Management Science*, **43**, 12, 1709-1725.

- [4] Barua, A. and T. Mukhopadhyay (1989). A Cost Analysis of the Software Dilemma: To Maintain or to Replace, *Proceedings of the 22nd Hawaii International Conference on Systems Sciences*, 89-98.
- [5] Chan, T., S. Chung and T. Ho (1996). An Economic Model to Estimate Software Rewriting and Replacement Times, *IEEE Transactions on Software Engineering*, **22**, 8, 580-598.
- [6] Dekleva, Sasa M. (1992). The Influence of the Information Systems Development Approach on Maintenance, *MIS Quarterly*, **44**, 4, 355-372.
- [7] Kinatader, K. J. Kimberly and Lee, E. Y (2000). A New Approach to the Busy Period of The  $M/M/1$  Queue, *Queueing Systems*, **35**, 105-115.
- [8] Kulkarni, Vidyadhar G. and Sethi, S. P.(2003). Optimal Allocation of Effort to Software Maintenance: A Queuing Theory Approach, *working paper*.
- [9] Lientz, B.P. and Swanson, E.B. (1981). Software Maintenance Management, *Addison-Wesley*.
- [10] Nosek, J.T. and P. Palvia (1990). Software Maintenance Management: Changes in the Last Decade, *J. Software Maintenance*, **2**, 3, 157-174.
- [11] Ross, S. (2000). Introduction to Probability Models, 7<sup>th</sup> Edition, *Academic Press*, 320-323.
- [12] Schneidewind, N. F. (1987). The State of Software Maintenance, *IEEE Transactions on Software Engineering*, **SE-13(3)**, 303-310.
- [13] Swanson, E.B. (1976). The Dimensions of Maintenance, *Proceedings of the Second International Conference on Software Engineering*, 492-497.
- [14] Tan, Y. and V. S. Mookerjee (2003). Optimal Software Maintenance and Replacement Policies, *Working Paper*.

- [15] Takacs, L. (1962). Introduction to the Theory of Queues, *Greenwood Press*, 32-34, 58, 124.